

Self-Consistency Guided Path Selection for Complex Reasoning in Large Language Models

Aislinn Elwood

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

Jorg Trommler

Department of Computer Science, University of Houston, Houston, TX, USA.
jorgtrommler84@uh.edu

Abstract

Large language models have demonstrated substantial capacity for complex reasoning when prompted to produce intermediate rationales, yet their outputs remain sensitive to sampling, prompt formulation, and opaque internal state. Self-consistency methods improve reliability by generating multiple reasoning paths and selecting or aggregating among them, but naive deployment becomes computationally expensive and operationally fragile as task complexity and model scale increase. This paper examines self-consistency guided path selection as a socio-technical systems problem rather than a narrow decoding heuristic. It argues that path selection must be understood through layered architecture, resource governance, robustness, fairness, and sustainability. The paper first reviews foundational work on prompting, reasoning, and alignment, then develops a conceptual architecture that separates path generation, scoring, consensus, retrieval, and observability. It analyzes trade-offs involving accuracy, latency, cost, interpretability, and environmental burden. It further considers how path selection interacts with adversarial inputs, biased training distributions, human feedback, auditability, and policy frameworks. A central claim is that guided selection can improve complex reasoning only when paired with transparent evaluation, calibrated uncertainty, and infrastructure-aware deployment. The discussion draws on research in chain-of-thought reasoning, self-consistency, retrieval augmentation, reinforcement learning from human feedback, model reporting, and green computing. The paper concludes that robust path selection is not merely a model-level optimization but a governance and infrastructure commitment that shapes the legitimacy, equity, and long-term viability of reasoning systems in high-stakes settings.

Keywords

large language models; self-consistency; reasoning path selection; socio-technical systems; governance; robustness; sustainability; fairness.

1. Introduction

Large language models have become central to contemporary artificial intelligence because they can perform many tasks through natural language interaction, including summarization, code generation, question answering, and multi-step reasoning. Foundational work on prompting demonstrated that few-shot examples can elicit capable behavior without parameter updates, and chain-of-thought prompting further showed that intermediate rationales can improve performance on arithmetic, commonsense, and symbolic tasks [1], [3]. The transformer architecture provided the computational substrate for these capabilities, while

scaling studies described predictable improvements as model size, data, and compute increase [4], [5], [6]. More recent systems have extended these gains through instruction tuning, preference optimization, and large-scale multimodal training [7], [8]. However, improved average performance does not guarantee reliable reasoning. Models may produce plausible but incorrect rationales, change answers under minor prompt variation, or fail when tasks require compositional inference over several steps.

Self-consistency was introduced as a response to this fragility. Instead of accepting a single chain of thought, the method samples multiple reasoning paths and selects the most consistent answer through aggregation or voting [2]. This approach improves accuracy on several reasoning benchmarks because correct reasoning paths tend to converge more often than erroneous ones. Yet self-consistency also introduces new system-level questions. How many paths should be generated? Who or what evaluates path quality? When should a path be discarded, merged, or escalated to human review? How can path selection remain efficient when models are deployed at scale, often under latency and cost constraints? These questions move the discussion beyond isolated prompting techniques and into architecture, governance, and infrastructure.

This paper treats self-consistency guided path selection as a socio-technical capability. The term guided path selection refers to the structured generation, filtering, scoring, and aggregation of reasoning trajectories so that final outputs are more reliable, auditable, and appropriate for the deployment context. The paper argues that such selection cannot be evaluated only by benchmark accuracy. It must also be assessed through robustness under distribution shift, fairness across user groups, interpretability for auditors, environmental cost, and compatibility with policy obligations. The analysis is organized around several interconnected concerns: conceptual architecture, resource trade-offs, robustness and fairness, deployment sustainability, evaluation metrics, and policy implications.

2. Background and Related Work

The current landscape of large language model reasoning rests on several converging research traditions. Transformer architectures enabled parallel sequence modeling and attention over long contexts, which made large-scale pretraining practical [4]. Scaling laws clarified that loss decreases predictably with model size, dataset size, and compute, encouraging the development of increasingly large systems [5]. Subsequent models such as PaLM and LLaMA demonstrated that scale, data curation, and architectural refinement could yield broad capabilities across languages and tasks [6], [7]. Systems such as GPT-4 further illustrated that multimodal input, extended context, and reinforcement learning from human feedback can produce behavior that appears more general and instruction-following [8]. These advances are important because path selection operates over the outputs of models whose behavior is shaped by pretraining, instruction tuning, and alignment.

Reasoning-specific research has focused on eliciting and improving intermediate computation. Chain-of-thought prompting showed that explicit rationales can improve multi-step tasks, particularly when examples include reasoned steps [1]. Self-consistency extended this by sampling diverse reasoning paths and selecting the answer with the highest marginal consistency [2]. Later approaches introduced tree-structured search, verbal reinforcement, and iterative self-refinement, allowing models to explore, critique, and revise intermediate reasoning [18], [19], [20]. These methods share a common premise: complex reasoning benefits from considering multiple candidate trajectories rather than committing to an initial generation. However, they also introduce computational overhead and create new evaluation

challenges, because a better final answer may not correspond to a faithful or interpretable reasoning process.

Alignment and feedback research provides additional context. Reinforcement learning from human preferences demonstrated that models can be trained to align with human judgments using comparison data [17]. Instruction tuning with human feedback scaled this idea to general-purpose assistants [16]. Constitutional AI and related approaches explored how explicit principles and AI feedback can reduce harmful outputs while preserving helpfulness [15]. These methods matter for path selection because the criteria used to score reasoning paths are not neutral. They reflect human preferences, institutional priorities, and safety constraints. A path that is logically coherent may still be harmful, biased, or policy-violating. Conversely, a safe answer may be under-informative. Guided selection therefore requires a multi-objective view that balances correctness, harm avoidance, transparency, and usefulness.

Retrieval augmentation and dense retrieval broaden the evidence base available to reasoning systems. Retrieval-augmented generation combines parametric memory with non-parametric documents, allowing models to ground answers in external sources [21]. Dense passage retrieval improves the efficiency and accuracy of document selection for open-domain question answering [22]. In path selection, retrieval can provide evidence for or against candidate rationales, but it also introduces dependencies on index freshness, source quality, and licensing. Systems research has long warned that machine learning deployments accumulate hidden technical debt through entanglement, configuration debt, data dependencies, and feedback loops [11]. AI safety research has similarly identified specification gaming, negative side effects, and distributional shift as persistent risks [12]. Interpretability research argues that rigorous evaluation of explanations requires attention to both model behavior and the social context of use [13]. These strands suggest that path selection is not only a decoding problem but a systems and governance problem.

3. Conceptual Architecture for Self-Consistency Guided Path Selection

A practical architecture for self-consistency guided path selection begins with a generation layer that produces multiple candidate reasoning paths for a given query. The generation layer should control diversity through sampling temperature, prompt variants, or role-conditioned instructions. The goal is not maximum randomness but principled coverage of plausible reasoning strategies. If all paths are nearly identical, consensus becomes trivial and uninformative. If paths are excessively diverse, aggregation may obscure systematic errors. The design must therefore tune diversity against task structure, model capability, and risk tolerance. Chain-of-thought prompting and self-consistency provide the conceptual basis for this layer, but production systems must also consider token budgets, concurrency, and model routing [1], [2].

The second layer is scoring and filtering. Candidate paths can be evaluated through learned verifiers, rule-based checks, retrieval evidence, or self-critique. Self-refinement and reflexion methods show that models can critique and revise their own outputs, although self-evaluation is not automatically reliable [18], [20]. Tree-structured search methods demonstrate that explicit branching and evaluation can improve deliberate problem solving [19]. In a guided selection architecture, scoring should be decomposed into multiple dimensions, including logical validity, evidential support, internal consistency, safety, and policy compliance. This decomposition allows different stakeholders to inspect which criteria drive selection. It also enables escalation when no candidate path satisfies minimum thresholds. A single scalar score may be operationally convenient but can hide contested value judgments.

The third layer is consensus and aggregation. In standard self-consistency, the final answer is often chosen by majority vote or marginalization over reasoning paths [2]. In complex domains, simple voting may be insufficient because paths can be semantically equivalent while differing in wording, or superficially similar while relying on different assumptions. Aggregation may therefore require semantic clustering, entailment checking, or weighted voting based on path quality. The aggregation layer must also distinguish between answer consensus and reasoning consensus. A system may find that many paths arrive at the same answer through flawed reasoning, or that a minority path reaches a correct answer through a more robust argument. Governance policies should specify when answer consensus is sufficient and when reasoning consensus is required.

The fourth layer is retrieval and tool interaction. Reasoning paths are more trustworthy when they can be grounded in external evidence, computational tools, or databases [21], [22]. Retrieval can support path selection by confirming factual claims, identifying contradictions, or supplying missing context. Tool use can extend reasoning beyond the model's parametric knowledge, but it also creates new failure modes involving API errors, stale data, and unauthorized access. The architecture should treat retrieval and tools as governed resources rather than transparent add-ons. Each path may need provenance metadata indicating which sources were consulted and how they influenced the conclusion.

The fifth layer is observability and governance. Because path selection involves multiple candidates and scoring criteria, it produces rich operational data that can support auditing, debugging, and continuous improvement. Interpretability research emphasizes that explanations must be evaluated in relation to their intended audience and decision context [13]. Model cards and similar reporting instruments provide structured documentation of intended use, limitations, and performance across groups [25]. In path selection, observability should capture prompt versions, path counts, scoring thresholds, selection rationales, retrieval sources, and human overrides. This record enables accountability without requiring disclosure of proprietary model weights. The architecture is thus modular, but its modules are bound together by governance requirements.

4. System-Level Trade-Offs and Resource Considerations

Self-consistency guided path selection creates a direct trade-off between reasoning quality and computational cost. Generating many paths increases the chance of finding a correct or robust rationale, but it also multiplies inference cost, energy consumption, and latency. Scaling laws show that model performance improves with compute, yet they do not imply that unlimited test-time compute is economically or environmentally sustainable [5], [6]. Energy and policy analyses of deep learning have documented substantial carbon footprints from training and inference [23]. Green AI arguments further call for efficiency metrics that treat computation as a first-class outcome rather than an invisible input [24]. Path selection must therefore be evaluated not only by accuracy gains but by marginal benefit per unit of compute.

Latency is another constraint. In interactive applications, users may tolerate a few seconds of delay, while real-time decision support may require responses within strict limits. Generating many reasoning paths sequentially can violate such constraints. Systems can mitigate this through parallel sampling, speculative decoding, caching, and adaptive path counts. However, each mitigation introduces complexity. Caching may raise privacy concerns if reasoning traces contain sensitive data. Adaptive path counts require confidence estimation, which may itself be unreliable. Hidden technical debt in machine learning systems often arises from such

accumulated configuration choices and feedback loops [11]. A mature deployment must document these choices and monitor their operational consequences.

Resource-constrained environments face especially difficult trade-offs. Smaller parameter models may be cheaper to serve but less capable of complex reasoning. Recent work on scalable reasoning path filtering and alignment for small parameter large language models illustrates how filtering and alignment can make limited models more useful without requiring the largest available architectures [14]. This is significant for edge deployment, privacy-preserving local inference, and organizations with limited infrastructure. Yet smaller models may also be more sensitive to prompt framing and less robust under distribution shift. Guided path selection can help, but it cannot eliminate capability gaps. The system should therefore route tasks according to risk, allowing escalation from small to larger models when confidence is low or when policy requires higher assurance.

Cost allocation is a governance issue as well as an engineering issue. In multi-tenant platforms, one user's request for extensive reasoning may consume shared resources and increase costs for others. If path selection is unlimited, it can create incentives for overuse. If it is tightly capped, it may disadvantage users with complex legitimate needs. Fair resource allocation may require tiered service levels, transparent quotas, or priority rules for high-stakes domains such as healthcare, legal aid, and public benefits. These decisions are not purely technical. They reflect institutional values about who receives reliable reasoning support and under what conditions.

5. Robustness, Fairness, and Governance

Robustness is a central concern because path selection can amplify as well as reduce errors. Adversarial prompts may induce a model to generate many superficially consistent but incorrect paths. If the selection mechanism rewards consensus, it may reinforce a shared bias or hallucination. AI safety research has long identified specification gaming, negative side effects, and distributional shift as risks that emerge when systems optimize proxies in complex environments [12]. A robust path selection system must therefore include adversarial testing, out-of-distribution evaluation, and mechanisms for detecting anomalous reasoning patterns. It should also avoid assuming that internal consistency equals external correctness.

Fairness concerns arise because language models are trained on datasets that reflect historical inequalities and uneven representation [10]. Path selection can mitigate or exacerbate these patterns. If scoring criteria favor dominant linguistic styles, dialect varieties, or cultural assumptions, then minority users may receive lower-quality reasoning support. If retrieval sources are skewed, evidence-based path selection may reproduce those skews. Fairness evaluation should therefore examine outcomes across demographic groups, languages, task types, and geographic contexts. It should also consider allocative fairness, meaning who gets access to high-quality path selection, and representational fairness, meaning whose knowledge and reasoning styles are treated as legitimate.

Alignment methods provide tools for shaping path selection criteria, but they also raise governance questions. Reinforcement learning from human preferences and instruction tuning with human feedback depend on annotator judgments, which may be inconsistent, culturally situated, or influenced by platform incentives [16], [17]. Constitutional AI and principle-based approaches attempt to make value constraints more explicit, but the selection of principles remains a political and ethical matter [15]. In path selection, alignment signals could be used to score candidate paths for harmlessness or helpfulness. The system should

document which values are encoded, who participated in defining them, and how conflicts among values are resolved. Without such documentation, alignment can become a form of opaque policy enforcement.

Transparency and auditability are essential for legitimate governance. Interpretability research argues that explanation quality depends on the question being asked and the decision context [13]. Model cards and structured reporting can provide standard summaries of intended use, performance, and limitations [25]. For path selection, auditability requires more than a final answer. Auditors may need to inspect how many paths were generated, which scoring criteria were applied, whether retrieval sources were consulted, and whether human review occurred. However, full disclosure of reasoning traces can raise privacy, security, and intellectual property concerns. Governance frameworks must therefore define tiered access, redaction rules, and independent oversight mechanisms. The goal is not total transparency but accountable transparency.

Governance also extends to procurement and deployment. Organizations that purchase or adopt reasoning systems should require documentation of path selection policies, failure modes, and evaluation results. They should ask whether the system can be audited, whether it supports appeals, and whether it logs decisions sufficiently for regulatory review. Foundation model research emphasizes that risks and opportunities are distributed across the lifecycle, from data collection to deployment and decommissioning [9]. Path selection should be integrated into this lifecycle perspective rather than treated as a post-hoc wrapper. Without governance, self-consistency can create a false sense of reliability because multiple paths appear to confirm one another even when the underlying model is uncertain or biased.

6. Deployment, Infrastructure, and Sustainability

Deployment architecture shapes what path selection can realistically achieve. Centralized cloud inference can provide large models, high parallelism, and advanced retrieval, but it also concentrates control and raises data sovereignty concerns. Local or edge deployment can improve privacy and reduce latency, but it may limit model size and path diversity. Hybrid architectures can route simple tasks to local models and complex tasks to cloud services, with policy rules determining when data may leave the device. Retrieval-augmented generation and dense retrieval can support such routing by supplying domain-specific evidence without retraining the model [21], [22]. However, retrieval indexes must be maintained, access-controlled, and updated to avoid stale or unauthorized information.

Monitoring is a continuous requirement. Production systems should track answer accuracy, path agreement, selection confidence, refusal rates, latency, cost, and user feedback. These signals can feed into incident response and model improvement. Yet monitoring can also create feedback loops in which models are optimized for easily measured signals rather than genuine reasoning quality. Hidden technical debt in machine learning systems often emerges from such feedback loops and configuration dependencies [11]. Human feedback from instruction tuning and preference modeling can be valuable, but it must be collected with attention to annotator well-being, cultural diversity, and incentive design [16], [17]. Monitoring should therefore combine quantitative metrics with qualitative review and external audits.

Sustainability is both an environmental and an institutional concern. The energy and carbon costs of deep learning have been documented in natural language processing, with implications for research priorities and deployment practices [23]. Green AI argues for

reporting efficiency alongside accuracy, so that progress is not defined solely by larger models and greater compute [24]. Path selection multiplies inference calls, and therefore magnifies these concerns. A system that generates dozens of reasoning paths for every query may achieve marginal accuracy gains at substantial environmental cost. Sustainable design should include adaptive path counts, early stopping, model distillation, quantization, and caching where appropriate. It should also measure embodied emissions from hardware and data centers, not only operational electricity use.

Documentation and reporting are practical instruments for sustainable and accountable deployment. Model cards provide a structured way to report intended use, performance, limitations, and ethical considerations [25]. Similar reporting could be extended to path selection policies, including default path counts, scoring criteria, escalation thresholds, and known failure modes. Such documentation helps users calibrate trust and helps regulators identify gaps. It also supports procurement decisions by allowing organizations to compare systems on dimensions beyond benchmark accuracy. Sustainability reporting, fairness auditing, and robustness testing can be integrated into these documents, making governance part of the deployment artifact rather than an external afterthought.

7. Evaluation and Operational Metrics

Evaluation of self-consistency guided path selection must go beyond final-answer accuracy. Standard reasoning benchmarks are useful, but they often reward short answers and may not capture reasoning faithfulness. Chain-of-thought and self-consistency research has shown that accuracy can improve with intermediate rationales, yet a correct answer may still be produced through invalid reasoning [1], [2]. Evaluation should therefore include process-level measures such as logical validity, evidential support, and consistency under paraphrase. It should also test whether path selection improves calibration, meaning whether confidence scores correspond to actual correctness. A system that is accurate but overconfident is unsafe in high-stakes settings.

Calibration and uncertainty quantification are especially important because path selection often relies on agreement as a proxy for confidence. Agreement can be misleading when errors are correlated. If a model consistently overlooks a relevant document or misunderstands a constraint, many paths may share the same mistake. Evaluation should therefore include adversarial consistency tests, counterfactual prompting, and retrieval ablations. Interpretability methods can help identify whether selection criteria focus on superficial features or substantive reasoning [13]. However, interpretability itself requires validation. Explanations that satisfy human reviewers may still be incomplete or misleading.

Fairness metrics should be integrated into operational dashboards. These metrics may include performance disparities across user groups, languages, and task domains, as well as differences in refusal rates and escalation rates. Bias in training data and language technologies has been widely documented, and path selection can inherit or amplify these patterns [10]. Model cards and reporting frameworks can provide a template for documenting subgroup performance and known limitations [25]. Fairness evaluation should also consider procedural fairness. Users should know when a decision was made by an automated path selection process, what criteria were applied, and how to appeal. Without procedural safeguards, even statistically fair systems may lack legitimacy.

Cost and sustainability metrics complete the evaluation picture. Operational metrics should include latency percentiles, throughput, token consumption, energy per query, and carbon

intensity of the serving region. Energy and policy considerations for deep learning provide a foundation for such accounting [23]. Green AI argues for efficiency-aware evaluation that treats computation as a resource with environmental consequences [24]. A path selection system that achieves small accuracy gains at large cost may be inappropriate for broad deployment. Conversely, a modest accuracy gain may be justified in high-stakes domains if it reduces harmful errors. Evaluation should therefore be context-sensitive, with thresholds and weightings determined through governance rather than engineering convenience alone.

8. Policy Implications and Future Directions

Policy frameworks for reasoning systems should address path selection directly. Regulations that focus only on model weights or final outputs may miss the selection mechanisms that shape which reasoning is trusted. Standards bodies could define documentation requirements for path generation, scoring, and escalation. Procurement rules could require vendors to disclose default path counts, retrieval dependencies, and audit logs. Model cards and similar reporting instruments offer a starting point, but they may need extension for dynamic inference procedures [25]. Policymakers should also consider liability. If a system selects a harmful recommendation after generating multiple paths, responsibility may be distributed among model developers, deployers, and auditors.

Environmental policy is increasingly relevant. Energy and carbon reporting for deep learning has been proposed as a way to make computational costs visible [23], [24]. Mandatory or voluntary reporting could encourage efficiency improvements and discourage wasteful inference. However, reporting must be standardized to avoid greenwashing. Metrics should distinguish training from inference, operational from embodied emissions, and average from peak demand. Path selection systems that adaptively reduce computation when confidence is high could be rewarded. Systems that always generate maximum paths regardless of need could be flagged for review. Such policies would align technical design with sustainability goals.

Fairness and civil rights policy also intersect with path selection. If reasoning systems are used in housing, employment, credit, education, or public services, selection criteria may affect protected groups. Bias in language models and training data has been extensively critiqued [10]. Alignment approaches such as constitutional methods and human feedback can encode safety and helpfulness preferences, but they do not automatically ensure distributive justice [15], [16], [17]. Policymakers may need to require impact assessments, bias testing, and accessible appeal mechanisms. Independent auditors should have enough access to evaluate selection criteria without compromising privacy or security. The challenge is to build oversight that is technically informed but not captured by vendors.

Future research should explore hybrid architectures that combine neural reasoning with symbolic verification, tool use, and structured knowledge. Retrieval-augmented generation and dense retrieval provide mechanisms for grounding reasoning in external evidence [21], [22]. Reflexion, tree-structured search, and self-refinement show that iterative critique can improve reasoning, but their costs and failure modes require further study [18], [19], [20]. Research is also needed on adaptive path selection that varies effort according to task risk, user need, and environmental constraints. Such adaptation could draw on resource-aware methods for smaller models and filtered alignment [14]. The broader agenda is to make reasoning systems robust, fair, auditable, and sustainable without treating any one objective as an absolute.

9. Conclusion

Self-consistency guided path selection offers a promising route to more reliable complex reasoning in large language models. By generating multiple candidate reasoning paths and selecting among them through structured scoring, consensus, and evidence, systems can reduce dependence on a single stochastic generation. However, the approach is not a simple accuracy patch. It changes the computational profile of inference, introduces new governance questions, and creates opportunities for both robustness and hidden failure. This paper has argued that path selection should be designed as a layered socio-technical architecture that connects generation, evaluation, retrieval, observability, and policy compliance. Each layer involves trade-offs among accuracy, latency, cost, fairness, and environmental impact.

The central conclusion is that guided path selection is only as trustworthy as the institutions and infrastructures that surround it. Without calibration, adversarial testing, fairness auditing, and transparent reporting, consensus can become a false assurance. Without efficiency planning and sustainability metrics, reasoning quality may be purchased at unacceptable environmental cost. Without governance, selection criteria may encode opaque values and resist appeal. Future systems should therefore treat path selection as a governed capability rather than a hidden implementation detail. The research agenda must span model design, systems engineering, policy analysis, and social accountability so that complex reasoning can be improved without reproducing old inequities or creating new ones.

References

1. 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.
2. Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., Narang, S., Chowdhery, A., & Zhou, D. (2023). Self-consistency improves chain of thought reasoning in language models. In *International Conference on Learning Representations*.
3. Brown, T., 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.
4. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998-6008.
5. Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., & Amodei, D. (2020). Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*.
6. Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., Barham, P., Chung, H. W., Sutton, C., Gehrmann, S., Schuh, P., Shi, K., Tsvyashchenko, S., Maynez, J., Rao, A., Barnes, P., Tay, Y., Shazeer, N., Prabhakaran, V., ... Fiedel, N. (2023). PaLM: Scaling language modeling with pathways. *Journal of Machine Learning Research*, 24(240), 1-113.

7. Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G. (2023). LLaMA: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971.
8. Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. arXiv preprint arXiv:2303.12712.
9. Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., ... Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.
10. 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). ACM.
11. 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.
12. Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete problems in AI safety. arXiv preprint arXiv:1606.06565.
13. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.
14. 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.
15. Bai, Y., Kadavath, S., Kundu, S., Askell, 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.
16. 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. Advances in Neural Information Processing Systems, 35, 27730-27744.
17. Christiano, P. F., Leike, J., Brown, T., Martic, M., Legg, S., & Amodei, D. (2017). Deep reinforcement learning from human preferences. Advances in Neural Information Processing Systems, 30, 4299-4307.
18. Shinn, N., Cassano, F., Gopinath, A., Narasimhan, K., & Yao, S. (2023). Reflexion: Language agents with verbal reinforcement learning. Advances in Neural Information Processing Systems, 36.
19. 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.

20. Madaan, A., Tandon, N., Gupta, P., Hallinan, S., Gao, L., Wiegrefe, S., Alon, U., Dziri, N., Prabhume, S., Yang, Y., Gupta, S., Majumder, B. P., Hermann, K., Welleck, S., Yazdanbakhsh, A., & Clark, P. (2023). Self-refine: Iterative refinement with self-feedback. *Advances in Neural Information Processing Systems*, 36.
21. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W. T., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 33, 9459-9474.
22. Karpukhin, V., Oğuz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., & Yih, W. T. (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). ACL.
23. 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). ACL.
24. Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the ACM*, 63(12), 54-63.
25. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 220-229). ACM.