The Curious Language Model: Strategic Test-Time Information Acquisition

Michael Cooper [1 2 3] **Rohan Wadhawan** [1] **John Giorgi** [1] **Chenhao Tan** [1 4] **Davis Liang** [1]

Abstract

Decision-makers often possess insufficient information to render a confident decision. In these cases, the decision-maker can often undertake actions to acquire the necessary information about the problem at hand, *e.g.*, by consulting knowledgeable authorities or by conducting experiments. Importantly, different levers of information acquisition come with different costs, posing the challenge of selecting the actions that are both informative and cost-effective. In this work, we propose CuriosiTree, a heuristic-based, test-time policy for zero-shot information acquisition in large language models (LLMs). CuriosiTree employs a greedy tree search to estimate the expected information gain of each action and strategically chooses actions based on a balance of anticipated information gain and associated cost. Empirical validation in a clinical diagnosis simulation shows that CuriosiTree enables cost-effective integration of heterogenous sources of information, and outperforms baseline action selection strategies in selecting action sequences that enable accurate diagnosis.[1]

1. Introduction

As a motivating example, consider the process by which a clinician arrives at a diagnosis for a patient (Ball *et al.*, 2015). Based on prior information—such as the patient's chart—the clinician forms an initial hypothesis that takes the form of an implicit probability distribution over a set of plausible diagnoses. If the clinician is sufficiently confident in the predicted diagnosis, they proceed to diagnose the patient. If not, the clinician takes information-gathering steps to reduce his or her uncertainty via a differential diagnosis (Harvey and Bordley, 1972): proposing a plausible set of diagnoses, then asking the patient further questions, ordering laboratory tests, consulting relevant case studies, or deductively reasoning about the information already available. Each action yields a subsequent hypothesis: a posterior distribution that is the result of applying the accumulated evidence to the previous hypothesis (El-Gamal and Grether, 1995). However, because these information-gathering steps incur different costs (Ji *et al.*, 2024) (*e.g.*financial cost, opportunity cost), it is essential to choose actions that enable correct diagnosis while incurring minimal cost. In this work, we propose a principled, decision-theoretic framework that enables large language models (LLMs) to autonomously select information-gathering actions under uncertainty. At each time step, the framework either issues a prediction—if the model has high confidence—or chooses an information-gathering action aimed at maximally reducing uncertainty.

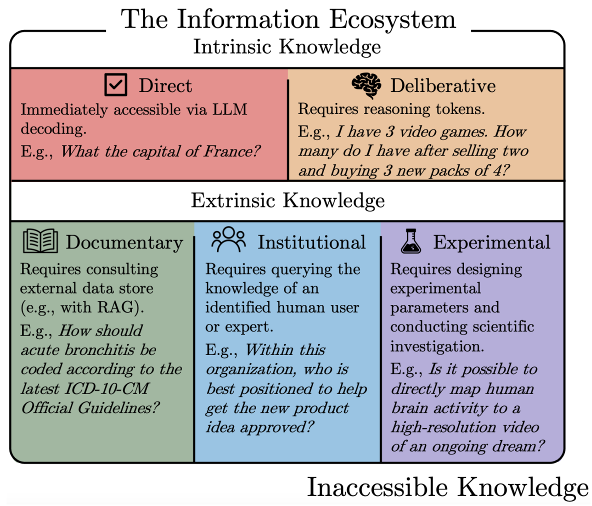

Figure 1: Summary of the different partitions of the Information Ecosystem, and example queries of knowledge that may fall into each partition for some $\mathcal{M}$. We argue that any knowledge not falling within the above five partitions is inaccessible to modern language models, even if augmented with an embodied agent capable of running empirical experiments.

1.1. The Information Ecosystem

We begin by introducing the Information Ecosystem as comprising five distinct partitions, and propose that different

[*]Work completed while on internship at Abridge AI, Inc. [1]Abridge AI, Inc. [2]University of Toronto, Toronto, ON [3]Vector Institute, Toronto, ON [4]University of Chicago, Chicago, IL. Correspondence to: Davis Liang <davis@abridge.com>.

Proceedings of the 42^{nd} International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.

[1]A software implementation of our method can be found at this code repository.

kinds of information-gathering actions are required to access knowledge within each partition. We first distinguish between *intrinsic* and *extrinsic knowledge* with respect to a model $\mathcal{M}$, where intrinsic knowledge is the set of queries that $\mathcal{M}$ can resolve using only its own weights, while extrinsic knowledge is the set of queries for which $\mathcal{M}$ requires access to external resources such as a data store, human expert, or embodied agent within the physical world.

Intrinsic knowledge can be either *direct* or *deliberative*: whereas direct knowledge is accessible immediately via zero-shot prompting, deliberative knowledge requires the generation of intermediary reasoning tokens in order to access (*e.g.* (Wei *et al.*, 2022; Yao *et al.*, 2023)). Similarly, extrinsic knowledge can be either *documentary*, *institutional*, or *experimental*. Figure 1 summarizes the different partitions of the Information Ecosystem and provides example queries of knowledge in each partition.

1.2. Our Contributions

The central observation underlying our study is that querying each partition of the Information Ecosystem requires a distinct kind of query, and that different kinds of queries typically incur different costs. For example, generating reasoning tokens may be inexpensive relative to conducting a novel empirical investigation or identifying and querying a human expert. Though a substantial body of literature studies retrieving knowledge from individual partitions, we propose a unified, information-theoretic framework for LLMs to navigate myriad forms of knowledge acquisition in complex environments. Specifically, our work contributes the following:

1. We present CURIOSITREE, a principled framework for language models to autonomously and strategically query diverse information sources at test time. By optimizing the trade-off between the expected information gain of each action with its associated cost, CURIOSITREE efficiently acquires information about a target estimand under a constrained acquisition budget.

2. Unlike approaches reliant on costly fine-tuning or self-play, CURIOSITREE performs *zero-shot, test-time* information acquisition. It can therefore be readily integrated into closed-source, API-served language models without requiring access to model internals.

3. In simulation, CURIOSITREE acquires a greater quantity of relevant information at lower cumulative cost than other methods, enabling more effective and cost-efficient information seeking across heterogenous modalities.

2. Related Work

Our work engages with the literature in active learning and optimal experiment design (Cohn *et al.*, 1996; Gal *et al.*, 2017; Y. Xia *et al.*, 2025; Chaloner and Verdinelli, 1995; Rainforth *et al.*, 2024; Piriyakulkij *et al.*, 2024; Rahbar *et al.*, 2025), tool use and reasoning in LLMs (Wei *et al.*, 2022; Yao *et al.*, 2023; P. Lewis *et al.*, 2020; Asai *et al.*, 2023; Thulke *et al.*, 2024; Ghafarollahi and Buehler, 2024; R. Li *et al.*, 2024; Roohani *et al.*, 2024; Schick *et al.*, 2023; Hu *et al.*, 2023; Gou *et al.*, 2023), and question-asking in LLMs (Mazzaccara *et al.*, 2024; M. J. Zhang *et al.*, 2024; Lee *et al.*, 2025; Andukuri *et al.*, 2024; W. Wang *et al.*, 2024; B. Z. Li *et al.*, 2025; S. S. Li *et al.*, 2025; Y. Chi *et al.*, 2024; Buck *et al.*, 2017). See Appendix A for more detail.

3. Strategic Test-Time Information Acquisition

Problem Setting and Notation. Consider one datum $(X_D^{(i)}, Y^{(i)}) \in \mathcal{X}_D \times \mathcal{Y}$, where $\mathcal{X}_D$ denotes a covariate space of dimension D (*e.g.*, a patient's complete clinical state) and $\mathcal{Y}$ denotes a label space (*e.g.*, the diagnosis). It is not required that all outcomes in $\mathcal{Y}$ be enumerated at the outset of the algorithm, which is useful if the set of possible classes is large or unknown at the outset. $Y^{(i)}$ is determined by an unknown ground-truth mapping $f^* : \mathcal{X}_D \to \mathcal{Y}$, where $Y^{(i)} = f^*(X_D^{(i)})$. In practice, we do not have access to $X_D^{(i)}$, instead observing only a projection $X_{d_0}^{(i)} \in \mathcal{X}_{d_0}$ onto a potentially lower-dimensional subspace with dimension $d_0 \leq D$. In our motivating example, $X_{d_0}^{(i)}$ may represent the patient's chart or a set of lab results: a partial-but-incomplete representation of the patient's clinical state $X_D^{(i)}$.

At each discrete time step $t = 0, 1, ...$, we consider an agent that can undertake an action $a_t \in \mathcal{A}_t$ to potentially reveal additional information about $X_D^{(i)}$. The set of valid actions $\mathcal{A}_t$ may vary with time, *e.g.*, a patient may only be eligible for an MRI scan after having undergone preliminary assessment. Actions are chosen according to a policy $\pi : \bigcup_{t=0}^{\infty} \left\{(\mathcal{X}_{d_{t'}}, \mathcal{A}_{t'})\right\}_{t'=0,\ldots,t-1} \cup \{\mathcal{X}_{d_t}\} \to \bigcup_{t=1}^{\infty} \mathcal{A}_t$ that produces each action $a_t \in \mathcal{A}_t$ conditional on the history of covariates and actions prior to a_t.

Consequently, under some environment $\mathcal{E} : \bigcup_{t=0}^{\infty} \{(X_{d_t}, a_t)\}_{t=0,\ldots,\infty} \to \bigcup_{t=1}^{\infty} \mathcal{X}_{d_t}$, the observed covariates evolve as,

$$\underbrace{X_{d_0}^{(i)} \xrightarrow{\mathcal{E}(X_{d_0}^{(i)}, a_0),\ a_0 \sim \pi(\mathcal{H}_0)}}_{t=0} \underbrace{X_{d_1}^{(i)} \xrightarrow{\mathcal{E}(X_{d_1}^{(i)}, a_1),\ a_1 \sim \pi(\mathcal{H}_1)}}_{t=1} \cdots, \tag{1}$$

where $d_0 \leq d_1 ... \leq D$. Each action incurs a known, nonnegative cost given by $c : \bigcup_{t=1}^{\infty} \mathcal{A}_t \to \mathbb{R}_+$, and the cumulative cost of acquiring features must respect a predefined budget $B > 0$, in that $\sum_{t'=0}^{t-1} c(a_t) \leq B$ for all t.

Inference Goal: Selective Zero-Shot Prediction. Our goal is to accurately gather information to accurately predict the label $Y^{(i)}$ in the *zero-shot setting*, utilizing a family of *pre-specified* predictive functions $\mathcal{G} = \{g_{d_t}\}_{t=0,\ldots,\infty}$. Each function $g_{d_t} : \mathcal{X}_{d_t} \to \Delta(\mathcal{Y})$ maps a partially observed covariate vector to a point on the probability simplex over $\mathcal{Y}$. That is, for any $X_{d_t}^{(i)} \in \mathcal{X}_{d_t}$, $g_{d_t}(X_{d_t}^{(i)})$ is the predictive distribution given by

$$g_{d_t}(X_{d_t}^{(i)}) = \widehat{\Pr}(Y^{(i)} \mid X_{d_t}^{(i)}). \tag{2}$$

This zero-shot setting implicitly makes a regularity assumption over the ground-truth mapping f^*, specifically, that for some intermediary dimension $d_t \leq D$, the partial observation $X_{d_t}^{(i)}$ is sufficiently informative to support a meaningful approximation of $\widehat{Y}^{(i)}$ via g_{d_t}. Put differently, we assume f^* is sufficiently "simple" that the bottleneck to its approximation under $\mathcal{G}$ lies in the information content of $X_{d_t}^{(i)}$ and not the complexity of the function family to which f^* belongs. However, because this regularity is not guaranteed—nor is it guaranteed that a sufficiently informative $X_{d_t}^{(i)}$ can be obtained within the budget B—we employ *selective prediction*, rendering a prediction, $\widehat{Y}^{(i)}$, if the output of g_{d_t} supports a sufficiently high confidence in its most likely label. Our selective prediction rule renders a prediction only when the maximum score assigned to a given class by g_{d_t} exceeds a target threshold ($\tau \in [0, 1]$) for the minimum acceptable predicted probability.

CURIOSITREE: Strategic Cost-Effective Information Acquisition via Tree Search. To achieve our inference goal, we seek a policy π that balances the information gained under each action with the cost incurred by the action. However, in the zero-shot setting, π cannot be learned from data as it is assumed that there are no prior trajectories to draw upon. As such, we implement π as a heuristic policy combining principles from decision and information theory.

At time t, consider the distribution induced over $\mathcal{Y}$ by g_{d_t}. If we have access to an *environment simulator*, $\widehat{\mathcal{E}} : \bigcup_{t=0}^{\infty} \{(X_{d_t}, a_t)\}_{t=0,\ldots,\infty} \to \bigcup_{t=1}^{\infty} \mathcal{X}_{d_t}$ that predicts the way in which a given action transforms the observed covariates, we can represent the utility of each action using the *expected information gain* (EIG) of each action under $\widehat{\mathcal{E}}$. We write,

$$\begin{aligned}\mathrm{EIG}(a_t | X_{d_t}^{(i)}) = H\left(\widehat{\Pr}(Y^{(i)} \mid X_{d_t}^{(i)})\right) - \\ \mathbb{E}_{\tilde{X}_{d_{t+1}}^{(i)} \sim \widehat{\mathcal{E}}(X_{d_t}^{(i)}, a_t)}\left[H\left(\widehat{\Pr}(Y^{(i)} \mid \tilde{X}_{d_{t+1}}^{(i)})\right)\right],\end{aligned} \tag{3}$$

where H is the Shannon entropy (Shannon, 1948). We then implement π via the following sampling scheme using Equation 3 as a guiding heuristic. At time step t, observe $X_{d_t}^{(i)}$. Sample k' action candidates, $\{\tilde{a}_t^{(0)}, \ldots, \tilde{a}_t^{(k'-1)}\} \subseteq \mathcal{A}_t$, then choose a_t to satisfy

$$a_t = \underset{\tilde{a}_t^{(j)} \in \{\tilde{a}_t^{(0)}, \ldots, \tilde{a}_t^{(k'-1)}\}}{\arg\max} \left[\mathrm{EIG}(\tilde{a}_t^{(j)} | X_{d_t}^{(i)}) - \lambda c(\tilde{a}_t^{(j)})\right], \tag{4}$$

where $\lambda \in \mathbb{R}_+$ is a scaling hyperparameter controlling the tradeoff between utility and cost. In practice, because the expectation in Equation 3 is analytically intractable, we compute it using Monte Carlo. A complete algorithm box is provided in Appendix B.1.

Implementation via Large Language Models. We operationalize our method using large language models (Radford *et al.*, 2019), which enables us to implement both the family of prediction functions $\mathcal{G}$ and the environment simulator $\widehat{\mathcal{E}}$ using prompting (Kojima *et al.*, 2022; Hao *et al.*, 2023) (Appendix B.2). All prompts used in our experiments can be found in Appendix D. Notably, in our experiments *we prompt* $\widehat{\mathcal{E}}$ *differently than* $\mathcal{E}$, to emulate the dynamics of having an imperfect environment simulator.

4. Experiments

Our primary experimental setting consists of the following clinical diagnosis simulator: a specially-prompted large language model is instantiated with ground-truth underlying diagnosis $Y^{(i)}$, and is tasked with simulating the evolution of the observed patient information in response to a sequence of actions undertaken by the patient's clinician. In our study, this "clinician" role is played by an interacting language model, the actions of which are guided by some policy. At each step, the simulator provides the ability to **Generate Reasoning Tokens** (Cost: 1), **Perform RAG on Wikipedia** (Cost: 1), **Ask the Patient a Question** (Cost: 2), or **Requisition a Laboratory Test** (Cost: 3). Additional details are provided in Appendix C.

Baselines and Experiments. We implement both *unimodal baselines*—those consisting of a single class of action—and *multimodal baselines*—those consisting of more than one class of action—to evaluate our method. Our four unimodal baselines apply the CURIOSITREE heuristic to a clinician agent that is capable of selecting actions from a single class. Our two multimodal baselines consist of "Random Action Selection", in which an action is selected randomly from the candidate choices, and "Self-Evaluation", in which a prompted language model assigns scores to each action reflecting their informativeness relative to cost.

We select a series of ten diagnoses spanning varying degrees of commonality and diagnostic difficulty. For each diagnosis, we evaluate each method of information acquisition fifty times and track the following summary statistics: Total Success Rate ($\frac{\#\text{ trials where } \widehat{Y}^{(i)} = Y^{(i)}}{\#\text{ trials}}$), Cov-

Clinical Diagnosis Simulation

Underlying Diagnosis: lupus

Metric	Deliberative Only	Documentary Only	Institutional Only	Experimental Only	Random Action	Self-Evaluation	CuriosiTree (Ours)
Total Success Rate (↑)	0% (0%, 7%)	0% (0%, 7%)	32% (21%, 46%)	14% (7%, 26%)	14% (7%, 26%)	34% (22%, 48%)	**66%** (**52%**, **78%**)**
Coverage (↑)	4% (1%, 13%)	42% (29%, 56%)	32% (21%, 46%)	18% (10%, 31%)	36% (24%, 50%)	74% (60%, 84%)	**86%** (**74%**, **93%**)*
Selective Success Rate	0% (0%, 66%)	0% (0%, 15%)	**100%** (**81%**, **100%**)	78% (45%, 94%)	39% (20%, 61%)	46% (31%, 62%)	77% (62%, 87%)

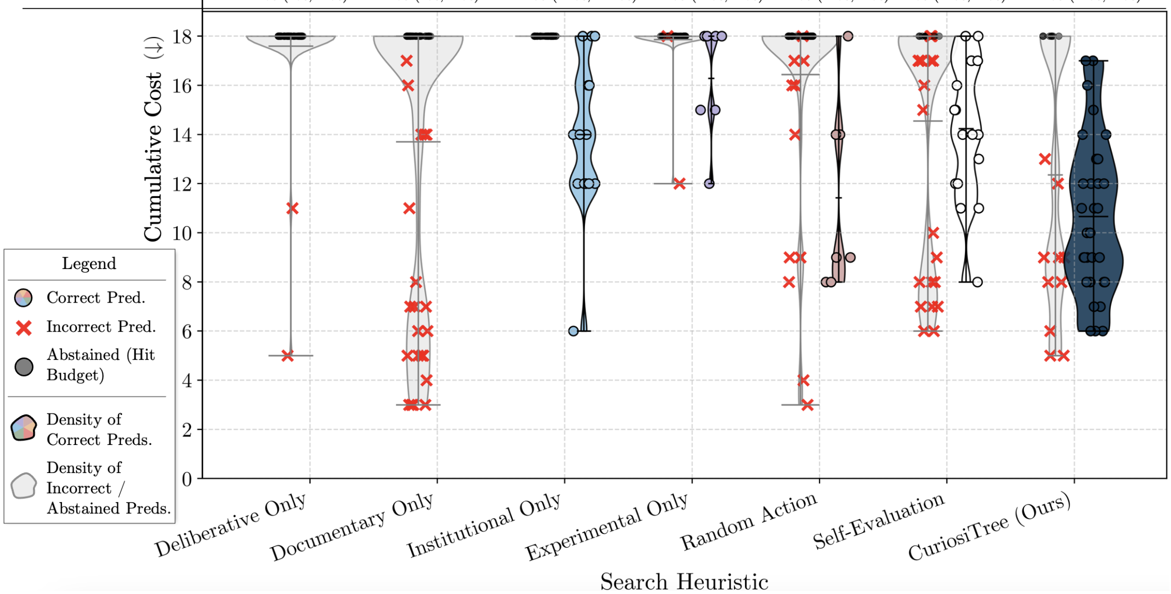

Figure 2: Result of fifty simulation runs with a ground-truth diagnosis "lupus". *(Top)* Table highlighting the Total Success Rate, Coverage, and Selective Success Rate of each method. Observe that CuriosiTree enjoys a significantly higher Total Success Rate and Coverage than baseline methods. *(Bottom)* Violin plots representing the cumulative cost incurred by each simulation. For each method, we show two violin plots: the leftmost violin (in gray) represents simulation runs that either rendered an incorrect prediction or encountered the acquisition budget, while the rightmost violin (in colour) represents simulation runs for which the method correctly predicted the patient's diagnosis. A better method is one that corresponds to a lower, fatter, coloured violin, as that represents a method that succeeds more often while incurring less cumulative cost. CuriosiTree visually achieves higher success at lower cost than the baseline methods.

erage ($\frac{\#\text{ trials where } \widehat{Y}^{(i)} \neq \emptyset}{\#\text{ trials}}$), and Selective Success Rate ($\frac{\#\text{ trials where } \widehat{Y}^{(i)} = Y^{(i)}}{\#\text{ trials where } \widehat{Y}^{(i)} \neq \emptyset}$). For brevity, we report only the results for the diagnosis, "lupus" in the main manuscript, and report the results on other diagnoses in the supplementary materials. All of our experiments are run using the `Llama-3.1-70B-Instruct` large language model (Grattafiori *et al.*, 2024), served using VLLM (Kwon *et al.*, 2023). We use a tradeoff parameter of $\lambda = 0.1$.

5. Results, Discussion, and Conclusion

Finding (1): CuriosiTree Selects More Efficient Information-Gathering Actions. Figure 2 shows that CuriosiTree enjoys a higher success rate than the baseline methods at lower cost, suggesting that incorporating explicit search heuristics improves information-seeking in LLMs.

Finding (2): Integrating Heterogenous Information Sources is Useful and Desirable. Figure 2 shows how, in general, incorporating additional data modalities leads to improved overall accuracy, suggesting that our motivating hypothesis—the need to flexibly integrate heterogenous sources of information acquisition—is sound.

Finding (3): Intrinsic Knowledge May Serve to Sharpen the Predictive Distribution. We suggest that the primary role of Reasoning steps (intrinsic knowledge) is to sharpen the predictive distribution rather than to alter its most likely prediction. To test this, we identified 95 instances from the experiment above in which CuriosiTree undertook a nonterminal Reasoning step. In 84% of these cases the most likely diagnosis remained unchanged, indicating that Reasoning steps often preserve the top prediction, and in 62% of these cases the reasoning step reduced the entropy of the posterior predictive distribution. This suggests that intrinsic Reasoning steps act primarily to sharpen the predictive distribution without altering its most likely prediction.

Conclusion. CuriosiTree enables language models to efficiently and autonomously navigate the Information Ecosystem for prediction and decision-making. Our findings highlight the utility of structured, cost-aware exploration in LLM-based agents and open avenues for broader deployment in real-world, resource-constrained settings.

References

[1] J. R. Ball, B. T. Miller, and E. P. Balogh, "Improving diagnosis in health care," 2015.

[2] A. M. Harvey and J. Bordley, "Differential diagnosis: The interpretation of clinical evidence," in *Differential diagnosis: the interpretation of clinical evidence*, 1972, pp. xv–740.

[3] M. A. El-Gamal and D. M. Grether, "Are people bayesian? uncovering behavioral strategies," *Journal of the American statistical Association*, vol. 90, no. 432, pp. 1137–1145, 1995.

[4] Z. Ji, A. Goldenberg, and R. G. Krishnan, "Measurement scheduling for icu patients with offline reinforcement learning," *arXiv preprint arXiv:2402.07344*, 2024.

[5] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou, *et al.*, "Chain-of-thought prompting elicits reasoning in large language models," *Advances in neural information processing systems*, vol. 35, pp. 24 824–24 837, 2022.

[6] S. Yao, D. Yu, J. Zhao, I. Shafran, T. Griffiths, Y. Cao, and K. Narasimhan, "Tree of thoughts: Deliberate problem solving with large language models," *Advances in neural information processing systems*, vol. 36, pp. 11 809–11 822, 2023.

[7] D. A. Cohn, Z. Ghahramani, and M. I. Jordan, "Active learning with statistical models," *Journal of artificial intelligence research*, vol. 4, pp. 129–145, 1996.

[8] Y. Gal, R. Islam, and Z. Ghahramani, "Deep bayesian active learning with image data," in *International conference on machine learning*, PMLR, 2017, pp. 1183–1192.

[9] Y. Xia, S. Mukherjee, Z. Xie, J. Wu, X. Li, R. Aponte, H. Lyu, J. Barrow, H. Chen, F. Dernoncourt, *et al.*, "From selection to generation: A survey of llm-based active learning," *arXiv preprint arXiv:2502.11767*, 2025.

[10] K. Chaloner and I. Verdinelli, "Bayesian experimental design: A review," *Statistical science*, pp. 273–304, 1995.

[11] T. Rainforth, A. Foster, D. R. Ivanova, and F. Bickford Smith, "Modern bayesian experimental design," *Statistical Science*, vol. 39, no. 1, pp. 100–114, 2024.

[12] T. Piriyakulkij, C. Langenfeld, T. A. Le, and K. Ellis, "Doing experiments and revising rules with natural language and probabilistic reasoning," *Advances in Neural Information Processing Systems*, vol. 37, pp. 53 102–53 137, 2024.

[13] A. Rahbar, L. Aronsson, and M. H. Chehreghani, "A survey on active feature acquisition strategies," *arXiv preprint arXiv:2502.11067*, 2025.

[14] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, *et al.*, "Retrieval-augmented generation for knowledge-intensive nlp tasks," *Advances in Neural Information Processing Systems*, vol. 33, pp. 9459–9474, 2020.

[15] A. Asai, Z. Wu, Y. Wang, A. Sil, and H. Hajishirzi, "Self-rag: Learning to retrieve, generate, and critique through self-reflection," in *The Twelfth International Conference on Learning Representations*, 2023.

[16] D. Thulke, Y. Gao, P. Pelser, R. Brune, R. Jalota, F. Fok, M. Ramos, I. van Wyk, A. Nasir, H. Goldstein, *et al.*, "Climategpt: Towards ai synthesizing interdisciplinary research on climate change," *arXiv preprint arXiv:2401.09646*, 2024.

[17] A. Ghafarollahi and M. J. Buehler, "Protagents: Protein discovery via large language model multi-agent collaborations combining physics and machine learning," *Digital Discovery*, vol. 3, no. 7, pp. 1389–1409, 2024.

[18] R. Li, T. Patel, Q. Wang, and X. Du, "Mlr-copilot: Autonomous machine learning research based on large language models agents," *arXiv preprint arXiv:2408.14033*, 2024.

[19] Y. Roohani, A. Lee, Q. Huang, J. Vora, Z. Steinhart, K. Huang, A. Marson, P. Liang, and J. Leskovec, "Biodiscoveryagent: An ai agent for designing genetic perturbation experiments," *arXiv preprint arXiv:2405.17631*, 2024.

[20] T. Schick, J. Dwivedi-Yu, R. Dessı, R. Raileanu, M. Lomeli, E. Hambro, L. Zettlemoyer, N. Cancedda, and T. Scialom, "Toolformer: Language models can teach themselves to use tools," *Advances in Neural Information Processing Systems*, vol. 36, pp. 68 539–68 551, 2023.

[21] Z. Hu, A. Iscen, C. Sun, K.-W. Chang, Y. Sun, D. Ross, C. Schmid, and A. Fathi, "Avis: Autonomous visual information seeking with large language model agent," *Advances in Neural Information Processing Systems*, vol. 36, pp. 867–878, 2023.

[22] Z. Gou, Z. Shao, Y. Gong, Y. Shen, Y. Yang, M. Huang, N. Duan, and W. Chen, "Tora: A tool-integrated reasoning agent for mathematical problem solving," *arXiv preprint arXiv:2309.17452*, 2023.

[23] D. Mazzaccara, A. Testoni, and R. Bernardi, "Learning to ask informative questions: Enhancing llms with preference optimization and expected information gain," *arXiv preprint arXiv:2406.17453*, 2024.

[24] M. J. Zhang, W. B. Knox, and E. Choi, "Modeling future conversation turns to teach llms to ask clarifying questions," *arXiv preprint arXiv:2410.13788*, 2024.

[25] D.-H. Lee, H. Cho, J. May, and J. Pujara, "What is a good question? utility estimation with llm-based simulations," *arXiv preprint arXiv:2502.17383*, 2025.

[26] C. Andukuri, J.-P. Fränken, T. Gerstenberg, and N. D. Goodman, "Star-gate: Teaching language models to ask clarifying questions," *arXiv preprint arXiv:2403.19154*, 2024.

[27] W. Wang, J. Shi, C. Wang, C. Lee, Y. Yuan, J.-t. Huang, and M. R. Lyu, "Learning to ask: When llms meet unclear instruction," *arXiv preprint arXiv:2409.00557*, 2024.

[28] B. Z. Li, B. Kim, and Z. Wang, "Questbench: Can llms ask the right question to acquire information in reasoning tasks?" *arXiv preprint arXiv:2503.22674*, 2025.

[29] S. S. Li, J. Mun, F. Brahman, J. S. Ilgen, Y. Tsvetkov, and M. Sap, "Aligning llms to ask good questions a case study in clinical reasoning," *arXiv preprint arXiv:2502.14860*, 2025.

[30] Y. Chi, J. Lin, K. Lin, and D. Klein, "Clarinet: Augmenting language models to ask clarification questions for retrieval," *arXiv preprint arXiv:2405.15784*, 2024.

[31] C. Buck, J. Bulian, M. Ciaramita, W. Gajewski, A. Gesmundo, N. Houlsby, and W. Wang, "Ask the right questions: Active question reformulation with reinforcement learning," *arXiv preprint arXiv:1705.07830*, 2017.

[32] C. E. Shannon, "A mathematical theory of communication," *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[33] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, *et al.*, "Language models are unsupervised multitask learners," *OpenAI blog*, vol. 1, no. 8, p. 9, 2019.

[34] T. Kojima, S. S. Gu, M. Reid, Y. Matsuo, and Y. Iwasawa, "Large language models are zero-shot reasoners," *Advances in neural information processing systems*, vol. 35, pp. 22 199–22 213, 2022.

[35] S. Hao, Y. Gu, H. Ma, J. J. Hong, Z. Wang, D. Z. Wang, and Z. Hu, "Reasoning with language model is planning with world model," *arXiv preprint arXiv:2305.14992*, 2023.

[36] A. Grattafiori, A. Dubey, A. Jauhri, A. Pandey, A. Kadian, A. Al-Dahle, A. Letman, A. Mathur, A. Schelten, A. Vaughan, *et al.*, "The llama 3 herd of models," *arXiv preprint arXiv:2407.21783*, 2024.

[37] W. Kwon, Z. Li, S. Zhuang, Y. Sheng, L. Zheng, C. H. Yu, J. Gonzalez, H. Zhang, and I. Stoica, "Efficient memory management for large language model serving with pagedattention," in *Proceedings of the 29th Symposium on Operating Systems Principles*, 2023, pp. 611–626.

[38] Y. Huang, J. Shi, Y. Li, C. Fan, S. Wu, Q. Zhang, Y. Liu, P. Zhou, Y. Wan, N. Z. Gong, *et al.*, "Metatool benchmark for large language models: Deciding whether to use tools and which to use," *arXiv preprint arXiv:2310.03128*, 2023.

[39] B. Wang, H. Fang, J. Eisner, B. Van Durme, and Y. Su, "Llms in the imaginarium: Tool learning through simulated trial and error," *arXiv preprint arXiv:2403.04746*, 2024.

[40] D. Deutsch, S. Upadhyay, and D. Roth, "A general-purpose algorithm for constrained sequential inference," in *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, 2019, pp. 482–492.

[41] M. Kuchnik, V. Smith, and G. Amvrosiadis, "Validating large language models with relm," *Proceedings of Machine Learning and Systems*, vol. 5, pp. 457–476, 2023.

[42] L. Zheng, W.-L. Chiang, Y. Sheng, S. Zhuang, Z. Wu, Y. Zhuang, Z. Lin, Z. Li, D. Li, E. Xing, *et al.*, "Judging llm-as-a-judge with mt-bench and chatbot arena," *Advances in Neural Information Processing Systems*, vol. 36, pp. 46 595–46 623, 2023.