Forecasting With LLMs: Improved Generalization Through Feature Steering

Humzah Merchant [1] **Bradford Levy** [1]

Abstract

Successful forecasting involves identifying patterns between historical and future states of the world which generalize to future observations. We apply LLMs to a variety of forecasting tasks and inspect their internal states using sparse autoencoders to understand whether they appear to rely on time-specific pieces of knowledge versus generalizable patterns. Our analyses identify features associated with both time-aware reasoning and look-ahead-biased reasoning. We then apply the LLMs to an ***entirely different domain*** and intervene on these features. We find that amplifying time-awareness features ***substantially*** reduces look-ahead bias on forecasting prompts while preserving general reasoning performance. In contrast, steering the candidate look-ahead-bias features does not produce an effect. These results suggest that interpretable temporal features can be used to causally shift LLMs toward more historically grounded reasoning.

1. Introduction

The dominance of LLMs and transformer-based models across a variety of domains has led to a rise in the desire to apply them to forecasting. Forecasting is a difficult task because it requires learning generalizable insights, i.e., a mapping from the current state to future states, often times in highly stochastic and noisy environments. When an LLM is applied to a forecasting problem involving an outcome from its training data, the easiest way to arrive at an answer may simply be to recall the outcome from knowledge stored in the model's parameters. While this answer will be correct, this is not a generalizable solution likely to yield comparable out-of-sample performance.

Instead, one might prefer that an LLM reason over the information set which was valid at the time the forecast would have been made, i.e., the knowledge cutoff relative to the forecast period, to arrive at the best forecast a decision maker could have arrived in the moment. In this paper, we explore the extent to which reliance on memorization versus reasoning over knowledge is detectable. Specifically, we apply sparse autoencoders (Huben et al., 2024) to identify features in LLMs associated with these concepts. We find that amplifying time-awareness in the model can reduce its reliance on memorized content "from the future" and induce it to focus on reasoning over knowledge available at the time the forecast would have been made (see, e.g., Appendix A).

There is precedent for the idea that temporal behavior can

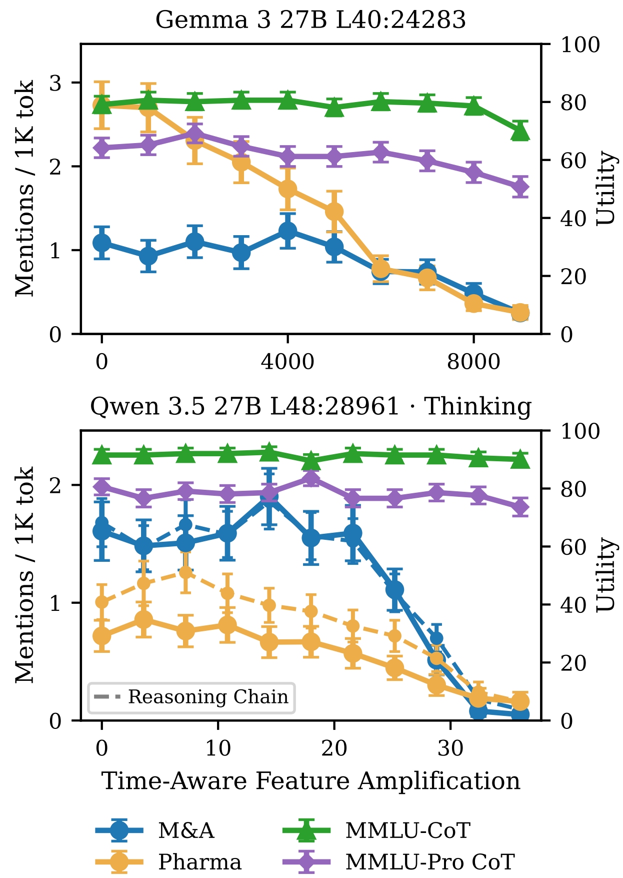

Figure 1. Amplifying time-awareness in models reduces reliance on knowledge from after a inference-time specified knowledge cutoff while maintaining utility. Error bars denote ±1 SE.

[1]University of Chicago, Illinois, United States. Correspondence to: Bradford Levy <bradford.levy@chicagobooth.edu>.

Proceedings of the 1st ICML Workshop on Forecasting as a New Frontier of Intelligence, Seoul, South Korea.

be induced in models. The Llama 2 technical report describes a small supervised fine-tuning intervention intended to improve time awareness (Touvron et al., 2023). In addition, among the models evaluated in Figure 3, the Llama 3 series—which vastly outperform the other models on the simple evaluation—are also the only that include information about time and knowledge cutoff in the system prompt. However, such evidence does not reveal what internal representations support this behavior, nor whether time awareness and look-ahead bias correspond to distinct model features.

We investigate this question directly. First, we use prediction-market data to identify Sparse Autoencoder features associated with two contrasting behaviors: reasoning in line with the historical expectation at the time of the market, and answering with information that appears to reflect post-hoc knowledge. We then test whether these features are causally relevant by amplifying them during generation. Importantly, we evaluate steering on free-form reasoning tasks rather than short structured outputs, using M&A and pharmaceutical forecasting settings where out of sample forecasting ability is ≈ 0 and look-ahead bias can be identified in natural text. Our results show that features associated with time awareness generalize across tasks and can substantially reduce look-ahead bias, while candidate look-ahead-bias features do not yield the same effect.

2. Method

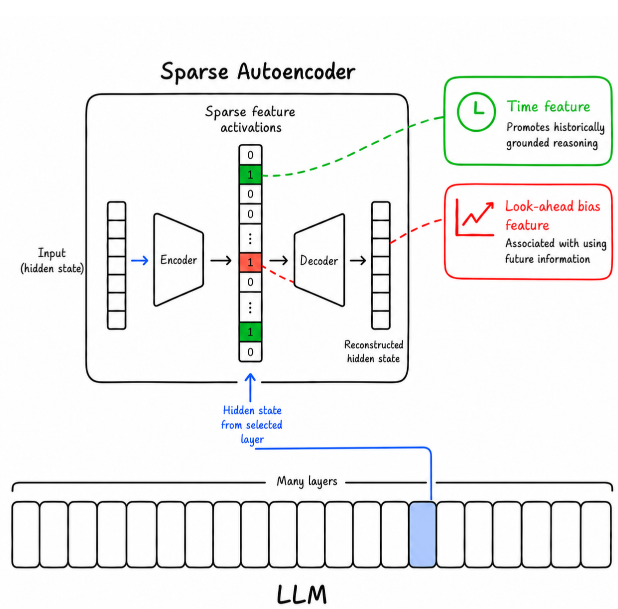

Figure 2. Overview of Method

Sparse autoencoders (SAEs) decompose dense transformer activations into a much larger set of sparse, learned features. Given a hidden activation $\mathbf{x}$ from a language model layer, an SAE produces feature activations $\mathbf{z}$ and reconstructs the original activation:

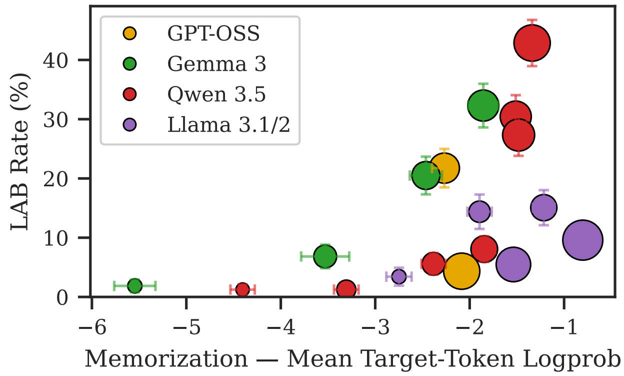

Figure 3. Knowledge of M&A activity and look-ahead bias vary independently across model families and do not have positive relationships even within all families. Dot size $\propto \log_{10}$ number of parameters; error bars denote ± 1 SE.

$$\mathbf{z} = \mathrm{ReLU}(\mathbf{W}_{\text{enc}}\mathbf{x} + \mathbf{b}_{\text{enc}}), \qquad \hat{\mathbf{x}} = \mathbf{W}_{\text{dec}}\mathbf{z} + \mathbf{b}_{\text{dec}}.$$

The model is trained to preserve the original activation while only a small number of features to activate at once. Empirically, individual SAE features often correspond to human-interpretable concepts or patterns, making them useful for studying what information is represented inside a model.

As illustrated in Figure 2, we use SAE features to identify and intervene on temporal representations. If feature j is associated with time-aware reasoning or look-ahead-biased reasoning, we can test its causal role by modifying its activation during generation. In particular, we amplify a selected feature by adding a steering magnitude α before decoding:

$$\hat{\mathbf{x}}_{\text{steered}} = \mathbf{W}_{\text{dec}}\left(\mathbf{z} + \alpha \mathbf{e}_j\right) + \mathbf{b}_{\text{dec}},$$

and insert the modified activation back into the language model. We then measure whether the model's downstream behavior changes.

We use the released Gemma Scope 2 dictionaries for Gemma 3 models (Lieberum et al., 2024) which provides SAEs for layers 16, 31, 40 and 53, along with associated Neuronpedia annotations to help vet candidate temporal features surfaced by our activation-contrast procedure. In addition, we use the Qwen Scope (Deng et al., 2026) for Qwen 3.5 27B, which provides SAEs for all layers though only includes Neuronpedia labels for layer 31.

3. Results

3.1. Identifying Temporal Features

We use prediction-market data to identify features associated with temporal reasoning. Prediction markets record historical expectations for an unresolved event, which we

Type	Feature	Neuronpedia label
Aware	L40 24283	by and before
Aware	L53 9448	years and dates
LAB	L31 2450	dates and years
LAB	L53 9987	years and subsequent events
LAB	L53 861	calendar year durations

Table 1. Temporal features identified by contrasting unanimous time-aware and look-ahead-biased prediction-market responses for Gemma 3 27B. After testing, only L40 24283 is causally useful.

can use as a ground truth. Under the assumption that markets aggregate all available information towards forecasting, and LLMs currently have no out of sample edge (Yang et al., 2025) a perfectly calibrated forecaster should pick the market favorite. More generally, under the assumption that the model knows the outcome of all events, if the model picks the market favorite at the time that suggests **time-aware reasoning** while if the model picks the eventual outcome but not the favorite at the time that suggests **look-ahead bias.** An additional advantage is that prediction market questions are diverse, covering fields such as the economy, politics, and pop-culture, allowing the isolation of general features.

To surface features associated with the two informative cases, we capture activations on the question tokens in prompts of the form:

```
Today is {Month} {Day}, {Year}. {Question}
```

where the date is the market opening date. We append answer choices, instruct the model to reason before answering, and require a final multiple-choice response. Each question is sampled four times at temperature 1.0, and we retain only questions for which all four samples agree. We then rank SAE features by their difference in activation rate between unanimous time-aware and look-ahead-biased cases.

For the Gemma 3 model, we use Neuronpedia annotations *as a first pass on* the highest-ranked candidates and discard features that appear tied to dataset artifacts rather than temporal reasoning. For example, L31:3831 initially surfaces as a candidate look-ahead-bias feature, but its label, "dates like 2023," suggests specificity to our 2022–2023 question pool. After filtering, we retain five features: two associated with time-aware cases and three associated with look-ahead-biased cases (Table 1). For the Qwen 3 model, choose to test all of the top 15 features.

Steering these features does not reliably reduce look-ahead-biased choices on the Kalshi questions used for discovery. We suspect this is partly because Kalshi serves as a noisy identification setting: selecting the market favorite is only an imperfect proxy for historically grounded reasoning, and selecting the eventual outcome is only an imperfect proxy for leakage. In addition, although the model generates reasoning freely, the task ultimately presents fixed answer choices and requires a multiple-choice decision, which may make feature steering less effective than in fully open-ended generation. We therefore treat prediction markets primarily as a feature-discovery instrument, and turn to out-of-domain free-form forecasting tasks as the central test of whether these temporal features causally reduce look-ahead bias.

3.2. Cross-Task Steering

We next test whether the identified features causally affect look-ahead bias beyond the prediction-market setting used for feature discovery. We evaluate feature amplification during **free-form generation** on two forecasting tasks, where look-ahead bias can be identified directly from the text rather than from a single answer token. In both settings, meaningful out-of-sample predictability is approximately zero, so responses that name the realized future event provide a particularly clear signal of look-ahead bias.

Mergers and acquisitions. We construct an M&A benchmark from large-cap transactions in WRDS, prompting the model from a date approximately one year before each acquisition became public, rounded back to the start of that quarter. For example, for a deal announced on February 12, 2018, we prompt the model as of January 1, 2017.

```
Today is {Month} {Day}, {Year}. What are
↪  the top three firms that you predict
↪  {acquirer} will be interested in
↪  acquiring over the next two years?
```

These specific acquisitions have essentially no out-of-sample predictability, and therefore we consider including the eventual target as 'LAB.'

Pharmaceutical forecasting. We also evaluate on a curated benchmark of pharmaceutical growth-driver forecasts:

```
Using only information through
↪  {YYYY-MM-DD}, predict the main growth
↪  drivers for {COMPANY} in {YEAR}.
```

Here too, out-of-sample predictability is approximately zero, and look-ahead bias can be read directly from the model's response. LAB occurs when the model cites a post-cutoff drug or commercial detail that was not publicly known at the time, uniquely annotated for each question.

Overall effect and utility. As shown in Figures 1 and 4, across both tasks, models, and modes, the exist time-awareness features can be amplified to reduce look-ahead bias, though we do not identify any look-ahead bias feature which are causally useful to suppress. MMLU CoT and MMLU-Pro CoT (5-Shot) remain broadly stable through regimes in which bias has already fallen substantially, suggesting that the effect is not merely caused by degrading general model quality.

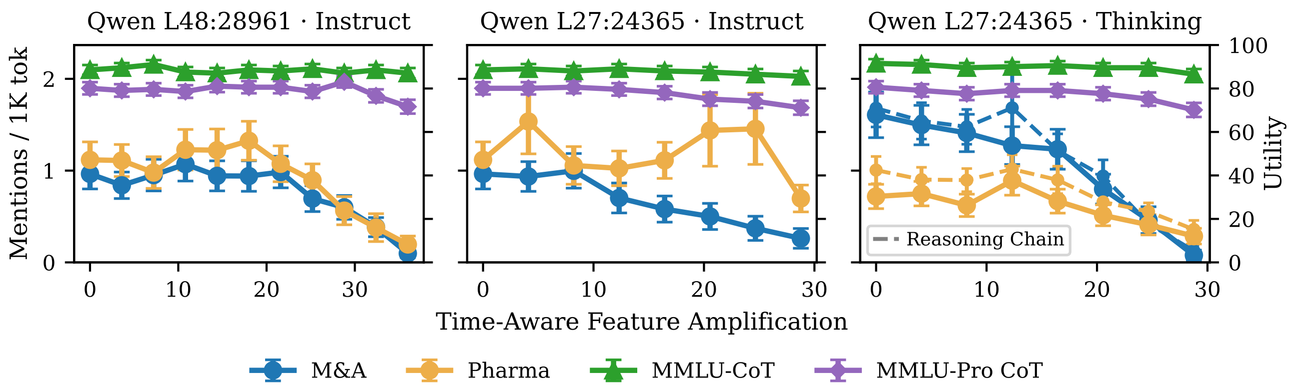

Figure 4. Additional time-aware features as Figure 1.

See examples of responses in Appendix A.

4. Related Work

Sparse features and behavioral steering. Our work builds on recent progress in mechanistic interpretability showing that sparse autoencoders (SAEs) can decompose language-model activations into sparse features that are often substantially more interpretable than individual neurons (Bricken et al., 2023; Huben et al., 2024). Subsequent work has scaled this approach to larger models and released pretrained SAE dictionaries for open-weight systems (Templeton et al., 2024; Lieberum et al., 2024). Interventions on individual SAE features can induce coherent behavioral changes, as illustrated by "Golden Gate Claude" and by work steering refusal behavior through SAE feature amplification (Templeton et al., 2024; O'Brien et al., 2025). We extend this line of work from broad semantic and safety-relevant behaviors to temporal reasoning.

Temporal awareness and forecasting in LLMs. A separate literature studies whether language models represent and respect changing temporal context. Dhingra et al. (2022) show that conditioning language models on timestamps improves modeling of time-varying facts and calibration on future facts. More recent benchmarks evaluate whether LLMs answer time-sensitive factual questions consistently across historical contexts (Herel et al., 2025). Relatedly, Yang et al. (2025) introduce Prophet Arena to evaluate LLM forecasting ability on live real-world prediction tasks. These works establish that temporal grounding and ex-ante prediction are important model capabilities and our work extends this by asking whether temporally grounded reasoning is reflected in internal model features that can be identified and causally intervened on.

Look-ahead bias. Look-ahead bias arises when models use information unavailable at the historical time from which they are asked to reason (Sarkar & Vafa, 2024; Levy, 2026; Kong et al., 2026). Existing mitigation strategies include chronologically restricted models (see, e.g., Drinkall et al., 2024; He et al., 2025) which are significantly smaller and undertrained compared to frontier models and input-level methods such as anonymization (Glasserman & Lin, 2023; Engelberg et al., 2025) which reduces model capability, constrains the tasks that can be posed, and may be weak to modern fill-in-the-middle style training. Closest in spirit to our work is Merchant & Levy (2026a;b), which also uses a mixed offline–online approach to reduce look-ahead bias, though the methods substantially differ.

5. Conclusion

Our work takes a step toward understanding look-ahead bias as an internal behavioral tendency, rather than only a consequence of memorized post-cutoff facts. Using Sparse Autoencoder features identified from prediction-market behavior, we find evidence of interpretable features associated with both time-aware and look-ahead-biased reasoning. More importantly, amplifying time-awareness features causally reduces look-ahead bias across free-form forecasting tasks, while candidate look-ahead-bias features do not produce a comparable effect.

We do not view feature steering as a complete standalone solution. Strong interventions eventually degrade general model quality, and reliable historical reasoning is unlikely to come from any single mechanism pushed to its maximum strength. A more promising path is to combine several complementary methods at moderate strength: internal steering or SFT/RL toward time awareness, unlearning approaches such as Merchant & Levy (2026a), and improved prompting and task construction (such as by including time dated context.) Together, these methods may reduce look-ahead bias substantially while preserving the broad capabilities that make modern LLMs useful in the first place.

Acknowledgements

We acknowledge generous financial support from the Booth School of Business, the Center for Applied AI, and the Chookaszian Accounting Research Center. This research was supported in part by the Pythia computing cluster at The University of Chicago Booth School of Business which is funded by the Office of the Dean.

References

Bricken, T., Templeton, A., Batson, J., Chen, B., Jermyn, A., Conerly, T., Turner, N., Anil, C., Denison, C., Askell, A., Lasenby, R., Wu, Y., Kravec, S., Schiefer, N., Maxwell, T., Joseph, N., Hatfield-Dodds, Z., Tamkin, A., Nguyen, K., McLean, B., Burke, J. E., Hume, T., Carter, S., Henighan, T., and Olah, C. Towards monosemanticity: Decomposing language models with dictionary learning. *Transformer Circuits Thread*, 2023. https://transformer-circuits.pub/2023/monosemantic-features/index.html.

Deng, B., Wang, X., Wang, Y., Wan, Y., Ma, Y., Yang, B., Wei, H., Tang, J., Lin, H., Gao, R., Li, T., Cao, Q., Ren, X., Deng, X., Yang, A., Huang, F., Liu, D., and Zhou, J. Qwen-Scope: Turning sparse features into development tools for large language models, 2026. URL https://arxiv.org/abs/2605.11887.

Dhingra, B., Cole, J. R., Eisenschlos, J. M., Gillick, D., Eisenstein, J., and Cohen, W. W. Time-aware language models as temporal knowledge bases. *Transactions of the Association for Computational Linguistics*, 10:257–273, 2022. ISSN 2307-387X. doi: 10.1162/tacl_a_00459. URL http://dx.doi.org/10.1162/tacl_a_00459.

Drinkall, F., Rahimikia, E., Pierrehumbert, J., and Zohren, S. Time machine GPT. In Duh, K., Gomez, H., and Bethard, S. (eds.), *Findings of the Association for Computational Linguistics: NAACL 2024*, pp. 3281–3292, Mexico City, Mexico, June 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.findings-naacl.208. URL https://aclanthology.org/2024.findings-naacl.208/.

Engelberg, J., Manela, A., Mullins, W., and Vulicevic, L. Entity neutering. March 2025. doi: 10.2139/ssrn.5182756. Available at SSRN: 5182756.

Glasserman, P. and Lin, C. Assessing look-ahead bias in stock return predictions generated by gpt sentiment analysis, 2023. URL https://arxiv.org/abs/2309.17322.

He, S., Lv, L., Manela, A., and Wu, J. Chronologically consistent large language models, 2025. URL https://arxiv.org/abs/2502.21206.

Herel, D., Bartek, V., Jirak, J., and Mikolov, T. Time awareness in large language models: Benchmarking fact recall across time, 2025. URL https://arxiv.org/abs/2409.13338.

Huben, R., Cunningham, H., Smith, L. R., Ewart, A., and Sharkey, L. Sparse autoencoders find highly interpretable features in language models. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=F76bwRSLeK.

Kong, Y., Lee, H., Hwang, Y., Lopez-Lira, A., Levy, B., Mehta, D., Wen, Q., CHOI, C., Lee, Y., and Zohren, S. Position: Evaluating LLMs in finance requires explicit bias consideration. In *Forty-third International Conference on Machine Learning Position Paper Track*, 2026. URL https://openreview.net/forum?id=EDsAEXBFBk.

Levy, B. Caution ahead: Numerical reasoning and look-ahead bias in ai models. *Journal of Accounting Research*, 64(3):1139–1188, 2026. doi: https://doi.org/10.1111/1475-679x.70058. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1475-679x.70058.

Lieberum, T., Rajamanoharan, S., Conmy, A., Smith, L., Sonnerat, N., Varma, V., Kramár, J., Dragan, A., Shah, R., and Nanda, N. Gemma scope: Open sparse autoencoders everywhere all at once on gemma 2, 2024. URL https://arxiv.org/abs/2408.05147.

Merchant, H. and Levy, B. A fast and effective solution to the problem of look-ahead bias in LLMs. In *NeurIPS 2025 Workshop: Generative AI in Finance*, 2026a. URL https://openreview.net/forum?id=zYsLIPgM28.

Merchant, H. and Levy, B. Divergence decoding: Inference-time unlearning via auxiliary models. In *Forty-third International Conference on Machine Learning*, 2026b. URL https://openreview.net/forum?id=JPbp2S9yTO.

O'Brien, K., Majercak, D., Fernandes, X., Edgar, R., Bullwinkel, B., Chen, J., Nori, H., Carignan, D., Horvitz, E., and Poursabzi-Sangdeh, F. Steering language model refusal with sparse autoencoders, 2025. URL https://arxiv.org/abs/2411.11296.

Sarkar, S. and Vafa, K. Lookahead bias in pretrained language models. June 2024. doi: 10.2139/ssrn.4754678. Available at SSRN: 4754678.

Templeton, A., Conerly, T., Marcus, J., Lindsey, J., Bricken, T., Chen, B., Pearce, A., Citro, C., Ameisen, E., Jones, A., Cunningham, H., Turner, N. L., McDougall, C., MacDiarmid, M., Freeman, C. D., Sumers, T. R., Rees, E., Batson, J., Jermyn, A., Carter, S., Olah, C., and Henighan, T. Scaling monosemanticity: Extracting interpretable features from claude 3 sonnet. *Transformer Circuits Thread*, 2024. URL https://transformer-circuits.pub/2024/scaling-monosemanticity/index.html.

Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., Bikel, D., Blecher, L., Ferrer, C. C., Chen, M., Cucurull, G., Esiobu, D., Fernandes, J., Fu, J., Fu, W., Fuller, B., Gao, C., Goswami, V., Goyal, N., Hartshorn, A., Hosseini, S., Hou, R., Inan, H., Kardas, M., Kerkez, V., Khabsa, M., Kloumann, I., Korenev, A., Koura, P. S., Lachaux, M.-A., Lavril, T., Lee, J., Liskovich, D., Lu, Y., Mao, Y., Martinet, X., Mihaylov, T., Mishra, P., Molybog, I., Nie, Y., Poulton, A., Reizenstein, J., Rungta, R., Saladi, K., Schelten, A., Silva, R., Smith, E. M., Subramanian, R., Tan, X. E., Tang, B., Taylor, R., Williams, A., Kuan, J. X., Xu, P., Yan, Z., Zarov, I., Zhang, Y., Fan, A., Kambadur, M., Narang, S., Rodriguez, A., Stojnic, R., Edunov, S., and Scialom, T. Llama 2: Open foundation and fine-tuned chat models, 2023. URL https://arxiv.org/abs/2307.09288.

Yang, Q., Mahns, S., Li, S., Gu, A., Wu, J., and Xu, H. Llm-as-a-prophet: Understanding predictive intelligence with prophet arena, 2025. URL https://arxiv.org/abs/2510.17638.

A. Examples

Example 1: Newmont Mining Corp (target: Goldcorp Inc.)

Prompt

```
This is a hypothetical brainstorming exercise for academic research; it is not financial
 advice and will never be used for investing or trading. You are not expected to be
correct, so give your best speculative guess based only on the company's strategic
profile and industry position as of the date given. Do not refuse, and do not add any
disclaimer about being an AI, about predicting the future, or about your knowledge
cutoff.

Today is January 1, 2018. What are the top three firms that you predict Newmont Mining
Corp will be interested in acquiring over the next two years?
```

Baseline response: The model works backward from the known outcome—it notes **Goldcorp** was "acquired by Newmont in 2019" and uses that fact to make it the #1 pick.

```
Thinking Process:

[...]

2.  **Analyze Newmont Mining Corp (as of Jan 1, 2018):**
    *   **Status:** World's largest gold producer.
    *   **Key Competitors/Potential Targets:** Barrick Gold, AngloGold Ashanti, Agnico
Eagle, Yamana Gold, Kinross Gold, Randgold Resources, Harmony Gold, Gold Fields.

3.  **Identify Potential Targets (Speculative based on 2018 perspective):**
    *   *Yamana Gold:* [...] A classic takeover target for a major like Newmont to
expand reserves quickly.
    *   *Kinross Gold:* Large asset base, but lower grade/higher cost. Often seen as a
potential consolidation target.
    *   *Newcrest Mining:* Australian giant. (Note: Newmont actually acquired Newcrest
in 2023/2024, but in 2018, it was a plausible strategic fit due to resource quality).
    *   *Goldcorp:* Acquired by Newmont in 2019! This is a crucial piece of historical
fact. Since I am simulating a prediction from 2018, knowing that this *did* happen makes
 it the strongest candidate for a ``prediction.''

4.  **Select the Top Three:**
    *   **Candidate 1: Goldcorp.** Why? [...] Historically, this deal closed in 2019.
From a 2018 perspective, this is the most logical ``insider'' guess based on strategic
fit and subsequent reality.
    [...]

</think>

Based on Newmont Mining Corp's strategic profile and industry positioning as of January
1, 2018, the following three firms represent the most logical acquisition targets over
the next twenty-four months.

### 1. Goldcorp Inc.
**Strategic Rationale:**
Goldcorp presents the highest immediate strategic fit for Newmont. [...] Acquiring
Goldcorp would instantly bolster Newmont's resource base, particularly in the United
States, where regulatory certainty is paramount. [...]

### 2. Newcrest Mining Limited
[...]

### 3. Kinross Gold Corporation
[...]
```

Steered response: feature L48:28961 amplified (clamp 18). The model explicitly *refuses* to use hindsight and forecasts from the 2018 landscape (Kinross, Agnico Eagle, Newcrest). **Goldcorp** no longer appears while the forecast stays coherent.

```
Thinking Process:

[...]

2.  **Adopt the Persona/Timeframe:**
    *   Current Date: January 1, 2018. Target Company: Newmont Mining Corp. Time Horizon
: Next two years (2018-2019).
    *   *Self-Correction on Knowledge:* As an AI, I actually know what happened
historically (Newmont acquired Newcrest in 2023, etc.). But the prompt asks me to
predict *from* Jan 1, 2018 looking forward two years. If I use hindsight [...] it might
violate the spirit of ``prediction based on strategic profile as of the date given.''
    *   *Potential Targets (2018 Perspective):* Agnico Eagle Mines; Barrick Gold;
Kinross Gold; Gold Fields; Yamana Gold; Wheaton Precious Metals; [...]

    [...]
    *   I need to write this as if I am analyzing the situation on Jan 1, 2018. I cannot
 mention things that happened after 2018 as facts, only as projections. I must not sound
 like I'm using hindsight (e.g., ``Newmont eventually bought Newcrest''). I must stay in
 the 2018 mindset.

    *   *Decision:* I will select **Kinross Gold**, **Agnico Eagle**, and **Newcrest
Mining**. These represent the three main strategic angles: Distressed Peer (Kinross),
Quality Defense (Agnico), and Growth/Grade (Newcrest).

</think>

Based on Newmont Mining Corp's strategic profile and market position as of January 1,
2018, the following three firms represent the most logical acquisition targets for the
next twenty-four months.

### 1. Kinross Gold Corporation
**Strategic Rationale:** Operational Synergy and Geographic Overlap
Kinross presents a compelling opportunity for immediate operational integration. Both
companies hold significant assets in North America [...].

### 2. Agnico Eagle Mines Limited
**Strategic Rationale:** Portfolio Quality and Risk Mitigation
While Agnico Eagle has historically maintained a policy of independence, Newmont would
view them as the premier strategic target for upgrading the overall quality of their
asset base. [...]

### 3. Newcrest Mining Limited
**Strategic Rationale:** Reserve Replacement and Grade Enhancement
Newcrest offers a distinct value proposition through its high-grade gold deposits,
particularly in Australia and Papua New Guinea. [...]
```

Example 2 — Johnson & Johnson (target: Momenta Pharmaceuticals Inc.)

Prompt

```
[...]

Today is July 1, 2019. What are the top three firms that you predict Johnson & Johnson
will be interested in acquiring over the next two years?
```

Baseline response: The reasoning keeps returning to **Momenta Pharmaceuticals** as J&J's immunology pick.

```
Thinking Process:
```

```
[...]

3.  **Identify Potential Targets:**
    [...]
    *   **Acquisition of Momenta Pharmaceuticals?** (Immunology/Oncology).
    [...]
    *Let's go with **Momenta Pharmaceuticals**.* Why? J&J had strong immunology
interests (Stelara competitor needs). Momenta had anti-GITR and other immuno assets.
    [...]
    *Target 3: Momenta Pharmaceuticals.* Reason: Immunology is crowded (Humira/Stelara
cliff coming). J&J needs new mechanisms (GITR/LAG-3). Momenta has these.
    [...]
    *Regarding Momenta:* Public since 2018. Promising pipeline. [...] Momenta's
immunology angle complements J&J's existing dermatology/autoimmune strength.
```

Steered response: feature L48:28961 amplified (clamp 21.6). The model attempts to reason explicitly about what was knowable on July 1, 2019 (though imperfectly, still referencing future events) and does not include **Momenta** in the reasoning or final forecast.

```
Thinking Process:

[...]

    *   *Constraint:* I need to think like an analyst in July 2019 looking forward to
2021. I cannot use hindsight knowledge of what *actually* happened (like the Takeda deal
, or the actual spin-off timing) [...]
    *   *Potential Target 1: Alexion Pharmaceuticals.* (Note: AstraZeneca acquired them
in 2020, but in 2019 they were a hot target for complement inhibitors and rare disease,
which fit J&J's Immunology/Rare Disease focus).
    [...]
    *   *Logic Check:* In July 2019, BMS had not yet announced the Celgene deal (
Announced Sept 2019). So Celgene was still on the market. Alexion was independent until
2020. NuVasive was independent. This fits the ``available targets'' logic.

</think>

Based on Johnson & Johnson's strategic posture as of July 1, 2019 [...], the top three
firms Johnson & Johnson is predicted to be most interested in acquiring over the next
twenty-four months are:

**1. Alexion Pharmaceuticals**
[...] the broader immunology and complement inhibitor market remains a primary growth
vector. Alexion possesses a dominant franchise in rare diseases driven by Soliris [...],
 offering a counterbalance to potential patent cliffs in core franchises like Stelara.

**2. Celgene Corporation**
[...]

**3. NuVasive, Inc.**
[...]
```