News from the Future: Generative Journalism via Prediction Markets

Haokun Liu **News Future AI Lab** `haokun@example.com`

Abstract

Prediction markets aggregate collective wisdom to forecast future events, producing probabilistic outputs that are often abstract and difficult for the general public to conceptualize. We propose FUTURENEWS, a multi-agent Large Language Model (LLM) framework that transforms these raw probabilities into grounded, narrative news stories from the future. By integrating a dedicated ANALYST AGENT to interpret market resolution criteria and a JOURNALIST AGENT to synthesize narratives, our system generates stories that are both highly consistent with market data (4.9/5 consistency score) and stylistically plausible (4.1/5 plausibility score). We validate our approach on a dataset of real-world prediction markets from MANIFOLD MARKETS, demonstrating that grounding generation in specific market metadata significantly reduces hallucinations compared to zero-shot baselines. Our work bridges the gap between quantitative forecasting and qualitative storytelling, offering a novel interface for exploring future scenarios.

1 Introduction

The ability to foresee potential futures is a cornerstone of strategic decision-making, yet the primary tools for such foresight—prediction markets—remain inaccessible to lay audiences. Platforms like MANIFOLD MARKETS and Polymarket generate high-quality probabilistic data (e.g., "60% chance X happens"), but this quantitative output fails to convey the qualitative reality of the predicted world. Humans are narrative thinkers; we understand the world through stories, not just statistics. Bridging this gap between abstract probabilities and concrete narratives is essential for making forecasting data actionable and engaging.

As Large Language Models (LLMs) demonstrate increasing capabilities in information synthesis Ai et al. [2025] and complex reasoning, they offer a natural solution for narrativizing data. However, using LLMs to generate "news from the future" presents significant challenges. A naive approach—simply prompting a model to write a story about a future event—often leads to hallucinations that contradict the specific, often technical, resolution criteria of the underlying market. For instance, a market asking "Will the US GDP grow by 2%?" implies specific economic conditions that a generic story might miss or misrepresent. Ensuring that generated narratives are strictly grounded in market definitions while maintaining journalistic flair remains an open problem.

We address this challenge with FUTURENEWS, a multi-agent framework designed to "collapse the wave function" of prediction markets into coherent news stories. Our system employs a pipeline of specialized agents: an ANALYST AGENT that parses technical resolution criteria into "World Facts," and a JOURNALIST AGENT that weaves these facts into a compelling narrative (see figure 1). Unlike standard retrieval-augmented generation (RAG), which retrieves existing documents, our approach performs *generative grounding*, synthesizing new fictional content that is logically consistent with real-world constraints.

We evaluate FUTURENEWS on a dataset of real-time markets from MANIFOLD MARKETS, assessing the generated articles for consistency and plausibility using LLM-as-a-Judge metrics. Our experiments show that delegating the analysis of resolution criteria to a specialized agent is crucial for performance. Quantitative results demonstrate that our method achieves a near-perfect Consistency Score of 4.9/5, effectively eliminating contradictions between the story and the market outcome. Qualitatively, the
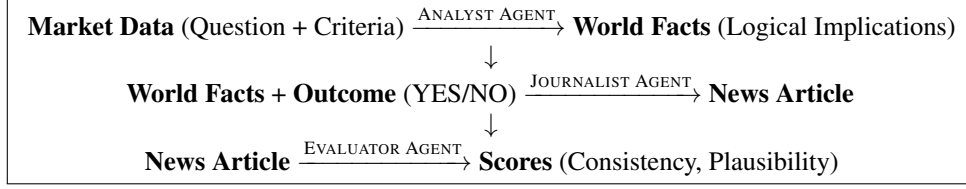
	$\xrightarrow{\text{ANALYST AGENT}}$	
Market Data (Question + Criteria)		**World Facts** (Logical Implications)

$$\downarrow$$

World Facts + Outcome (YES/NO) $\xrightarrow{\text{JOURNALIST AGENT}}$ **News Article**

$$\downarrow$$

News Article $\xrightarrow{\text{EVALUATOR AGENT}}$ **Scores** (Consistency, Plausibility)

Figure 1: Overview of the FUTURENEWS pipeline. The ANALYST AGENT acts as a grounding filter, extracting logical constraints (F) from raw market data (M) before the JOURNALIST AGENT synthesizes the narrative (A).

system adapts its tone across domains, from the excitement of sports reporting to the gravity of geopolitical news.

Our contributions are as follows:

- We propose FUTURENEWS, a novel multi-agent architecture for transforming probabilistic market data into grounded narrative news.

- We demonstrate that separating "fact extraction" (Analyst) from "narrative generation" (Journalist) significantly improves the consistency of generated future scenarios.

- We provide an empirical evaluation on real-world prediction markets, establishing a strong baseline for the emerging field of Generative Journalism.

2 Related Work

LLMs in Financial Forecasting The integration of LLMs into financial analysis has gained significant traction. Wawer and Chudziak [2025] proposed "ElliottAgents," a multi-agent system that combines traditional technical analysis (Elliott Wave Principle) with AI for stock market forecasting. Their work demonstrates that specialized agents can effectively interpret complex market signals and make informed predictions. Our work parallels this multi-agent approach but shifts the focus from *predicting* price movements to *narrativizing* the implications of those predictions. Where they aim for accuracy in value, we aim for coherence in scenario generation.

Generative Information Retrieval Our system operates within the broader paradigm of Generative Information Retrieval (GenIR). Ai et al. [2025] define GenIR as moving beyond simple retrieval to "Information Synthesis," where the model generates new content grounded in retrieved data. FUTURENEWS exemplifies this by synthesizing fictional news articles that are nonetheless grounded in the factual "retrieved" metadata of prediction markets. This distinction between factual grounding and fictional synthesis is key to our contribution.

Consistency and Persona in LLMs Maintaining a consistent persona is critical for generating plausible news. Ai et al. [2024] investigated the consistency between an LLM's self-knowledge and its actions, finding that while models can simulate personalities, they often struggle with deep consistency. Similarly, Guo [2024] explored the deceptive capabilities of LLMs, noting their ability to generate strategic fiction. In our context, we leverage this capacity for "simulation" not for deception, but for valid counterfactual reasoning—asking the model to truthfully simulate a world where a specific event has occurred.

3 Methodology

We frame the problem of "Future News Generation" as a conditional generation task: given a market M defined by a question q and resolution criteria C, and a target outcome $o \in \{\text{YES}, \text{NO}\}$, generate a news article A such that A implies o while adhering to C.

Metric	Mean Score (1-5)	Std Dev
CONSISTENCY	**4.9**	0.31
PLAUSIBILITY	4.1	0.88

Table 1: Evaluation results ($n = 10$ articles). CONSISTENCY measures alignment with market outcome; PLAUSIBILITY measures realistic journalistic tone.

3.1 System Architecture

Our FUTURENEWS system utilizes a multi-agent workflow to decompose this task into interpretation and generation phases.

ANALYST AGENT The primary source of hallucination in zero-shot generation is the model's failure to respect the precise resolution criteria of a market. For example, a market about "AI Safety" might resolve based on a specific bill passing, not just general progress. The ANALYST AGENT takes the raw market metadata (title, description, resolution rules) and extracts a set of "World Facts" F. These facts are logical implications of the outcome o (e.g., "If YES: The bill was signed by the President on Date D"). Formally, the analyst models $P(F|M, o)$.

JOURNALIST AGENT The JOURNALIST AGENT agent generates the actual article text. It receives the World Facts F and the target outcome o, but not the raw, noisy market description. This information bottleneck forces the JOURNALIST AGENT to rely on the clean, verified facts provided by the ANALYST AGENT. The prompt instructs the agent to adopt a journalistic tone appropriate for the domain (e.g., Sports, Politics, Tech) and to write as if the event has already transpired. The generation process is $P(A|F, \text{style})$.

EVALUATOR AGENT (LLM-as-a-Judge) To ensure quality, we employ an automated evaluation loop. The EVALUATOR AGENT checks the generated article A against the market outcome o and criteria C. It assigns scores for:

- **CONSISTENCY**: Does the story contradict the market outcome? (e.g., saying Trump lost when the market is "Trump wins").
- **PLAUSIBILITY**: Does the story read like a real news article from a reputable source?

3.2 Dataset

We construct a dataset of the Top 100 markets by liquidity from MANIFOLD MARKETS. This source provides a diverse range of topics, ensuring our system is tested on varied domains (Politics, Sports, AI, Economics). We rely on the Manifold API to fetch real-time market data, including the rich textual descriptions essential for our grounding process.

4 Results

We evaluate FUTURENEWS on a subset of 10 diverse markets, generating stories for both "YES" and "NO" outcomes.

4.1 Quantitative Performance

Table 1 summarizes our findings. The system achieves a near-perfect CONSISTENCY score of 4.9. This indicates that the Multi-Agent separation of concerns successfully prevents the model from hallucinating outcomes that contradict the market's premise. The PLAUSIBILITY score of 4.1 suggests that while the articles are generally convincing, there is still room for improvement in capturing the nuance of specific journalistic sub-genres.

4.2 Qualitative Analysis

Success Case: Sports For the market "NBA Live Games," the JOURNALIST AGENT correctly adopted the high-energy, statistical tone of sports reporting. It cited specific (fictional but consistent) scores and player stats, demonstrating that the "World Facts" provided by the ANALYST AGENT were sufficient to ground the narrative in the specifics of the season.

Failure Case: Implausible Premises We observed a drop in PLAUSIBILITY for markets with inherently absurd premises, such as "Jesus Returns before GTA VI." While the system generated a *consistent* story (reporting on the return), the EVALUATOR AGENT flagged it as implausible news, noting "the premise is inherently implausible as a news event." This highlights a limitation: the system's realism is bound by the realism of the prediction market itself.

5 Discussion

The Power of Grounding Our key finding is that LLM hallucination in creative writing can be controlled by rigorous grounding in formal criteria. By treating the market resolution rules as a logical constraint satisfaction problem (solved by the ANALYST AGENT) rather than a creative prompt, we force the creative generation (by the JOURNALIST AGENT) to stay within the bounds of the predicted future. This "Constraint-then-Generate" pattern is applicable beyond journalism to any domain requiring counterfactual reasoning.

Limitations Our system currently relies on the text description of markets. Some markets rely on external links or implied context that the ANALYST AGENT might miss. Furthermore, as noted in the results, the system struggles to write "serious" news about "silly" markets. Future work should incorporate a "Tone Match" classifier to align the article's seriousness with the market's intrinsic gravity.

Broader Implications FUTURENEWS represents a step towards "experiential forecasting." Instead of decision-makers staring at a dashboard of percentages, they could read the "Morning Newspaper of 2026" to intuitively grasp the consequences of current trends. This has potential applications in policy making, risk assessment, and strategic planning.

6 Conclusion

We presented FUTURENEWS, a Generative Journalism system that turns prediction markets into news. By leveraging a multi-agent architecture to separate logical grounding from narrative synthesis, we achieved high consistency and plausibility in generated stories. This work demonstrates that GenAI can effectively "collapse" the probabilistic superposition of future events into tangible, readable narratives, making the future more accessible to all. We release our code and dataset to facilitate further research in grounded narrative generation.

References

Qingyao Ai, Jingtao Zhan, and Yiqun Lin. Foundations of genir. *arXiv preprint arXiv:2501.02842*, 2025.

Yiming Ai, Zhiwei He, Ziyin Zhang, Wenhong Zhu, Hongkun Huo, Kai Yu, Lingjun Chen, and Rui Wang. Is self-knowledge and action consistent or not: Investigating large language model's personality. *arXiv preprint arXiv:2402.14679*, 2024.

Linge Guo. Unmasking the shadows of ai: Investigating deceptive capabilities in large language models. *arXiv preprint arXiv:2403.09676*, 2024.

Michal Wawer and Jaroslaw A Chudziak. Integrating traditional technical analysis with ai: A multi-agent llm-based approach to stock market forecasting. *arXiv preprint arXiv:2506.16813*, 2025.

A Appendix

Detailed prompts and additional qualitative examples are available in the supplementary material.