William Brock, Josef Lakonishok, and Blake LeBaron. Simple technical trading rules and the stochastic properties of stock returns. *The Journal of Finance*, 47(5):1731–1764, 1992. doi: https://doi.org/10.1111/j.1540-6261.1992.tb04681.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1992.tb04681.x.

Chaojie Wang, Yanchen Deng, Zhiyi Lyu, Liang Zeng, Jujie He, Shuicheng Yan, and Bo An. Q*: Improving multi-step reasoning for llms with deliberative planning. *arXiv preprint arXiv:2406.14283*, 2024.

Xiang Zhang, Muhammad Abdul-Mageed, and Laks VS Lakshmanan. Autoregressive+ chain of thought= recurrent: Recurrence's role in language models' computability and a revisit of recurrent transformer. *arXiv preprint arXiv:2409.09239*, 2024a.

Puyuan Liu, Xiang Zhang, and Lili Mou. A character-level length-control algorithm for non-autoregressive sentence summarization. *Advances in Neural Information Processing Systems*, 35:29101–29112, 2022b.

Xiang Zhang, Senyu Li, Ning Shi, Bradley Hauer, Zijun Wu, Grzegorz Kondrak, Muhammad Abdul-Mageed, and Laks VS Lakshmanan. Cross-modal consistency in multimodal large language models. *arXiv preprint arXiv:2411.09273*, 2024b.

Pan Lu, Swaroop Mishra, Tanglin Xia, Liang Qiu, Kai-Wei Chang, Song-Chun Zhu, Oyvind Tafjord, Peter Clark, and Ashwin Kalyan. Learn to explain: Multimodal reasoning via thought chains for science question answering. *Advances in Neural Information Processing Systems*, 35:2507–2521, 2022.

Xiang Zhang, Ning Shi, Bradley Hauer, and Grzegorz Kondrak. Bridging the gap between babelnet and hownet: Unsupervised sense alignment and sememe prediction. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 2789–2798, 2023b.

Mingyang Gong, Jing Fan, Guohui Lin, Bing Su, Zihan Su, and Xiang Zhang. Multiprocessor scheduling with testing: improved online algorithms and numerical experiments. *Journal of Scheduling*, pages 1–15, 2025.

Jiaqi Wei, Hao Zhou, Xiang Zhang, Di Zhang, Zijie Qiu, Wei Wei, Jinzhe Li, Wanli Ouyang, and Siqi Sun. Alignrag: Leveraging critique learning for evidence-sensitive retrieval-augmented reasoning. *arXiv preprint arXiv:2504.14858*, 2025a.

Jiaqi Wei, Yuejin Yang, Xiang Zhang, Yuhan Chen, Xiang Zhuang, Zhangyang Gao, Dongzhan Zhou, Guangshuai Wang, Zhiqiang Gao, Juntai Cao, et al. From ai for science to agentic science: A survey on autonomous scientific discovery. *arXiv preprint arXiv:2508.14111*, 2025b.

Xiang Zhang, Juntai Cao, Jiaqi Wei, Chenyu You, and Dujian Ding. Why prompt design matters and works: A complexity analysis of prompt search space in llms. *arXiv preprint arXiv:2503.10084*, 2025.

Steven B. Achelis. *Technical Analysis from A to Z*. McGraw-Hill, 2nd edition, 2013. ISBN 9780071826297. Publication date: December 13, 2013.

A Why Technical Analysis Alone Can Suffice for Trading

QuantAgent is a multi-modal, multi-agent high-frequency trading LLM framework that provides market prediction based solely on price data, disregarding other information such as news, social media, etc. This strategy is referred to as technical analysis (Pring, 1991). Technical analysis is based on the premise that price alone is enough for capturing market movement and predicting future trends, and has been extensively studied by previous research (Murphy, 1999). In this section, we present in detail why technical analysis alone can suffice for trading.

The first principle of technical analysis is that all relevant information, whether economic, political, psychological, or otherwise, is already reflected in market prices (Fama, 1970). In other words, prices adjust quickly to new developments because people act on the information they receive by buying or selling (Kahneman and Tversky, 1979). These actions are recorded in price changes (Lo et al., 2000). Therefore, by observing how prices move, it is possible to indirectly understand how the market as a whole is reacting to both public and private information, without needing to process that information explicitly (Edwards et al., 2018). Our system therefore also follows this principle and has each of its agents perform analysis solely based on price data.

Technical analysis assumes that price movements are not entirely random (Lo et al., 2000). Instead, they tend to follow patterns over time. When prices begin to rise, they often continue to rise for some period, and similarly, downward trends can persist before reversing. These trends often reflect collective human behavior, such as fear during declines or optimism during rallies. By identifying such trends early, technical traders aim to make decisions that align with the general direction of the market (Jegadeesh and Titman, 1993). QuantAgent operates such that it captures market patterns and leverages this price movement assumption.

Notably, many existing technical analysis methods are based on the observation that certain price patterns tend to appear repeatedly. This repetition is attributed to stable behavioral tendencies in market participants. For example, traders often react similarly to price increases or decreases, leading to recurring patterns such as peaks, dips, and reversals. Recognizing and responding to these familiar structures allows technical systems to make predictions without needing to understand the specific causes of each movement (Edwards et al., 2018). Such observable repetition is a natural fit for an agentic framework, as LLM agents have shown strong capability in reasoning over patterns and trends, achieving human-like capabilities (Bommasani et al., 2021).

Occasionally, price changes occur well before any official information is made available to the public (Chakravarty et al., 1998). For example, a stock's price might begin rising days or even weeks before a company announces strong earnings. This can happen because certain investors—such as employees, suppliers, or professional analysts—may already have insights into the company's performance, such as increased sales activity or unusually high production volumes. Similarly, prices may fall before news of a scandal becomes public. If there are rumors of legal investigations or unusual management behavior, informed traders might start selling early, causing the price to decline in advance. In both cases, the price moves ahead of the news because the market collectively reacts to early signals, expectations, or private information (Chakravarty et al., 1998; Cao et al., 2025). Technical analysis captures these movements directly through price behavior, without requiring access to the underlying cause (Brock et al., 1992). This allows trading systems to respond to changes as soon as they appear in the market, rather than waiting for delayed or incomplete public disclosures (Lo et al., 2000).

In summary, our agent works under the principles of technical analysis, which offers a practical and self-contained approach to understanding market behavior. By assuming that all available information is already incorporated into price data, and that human reactions to price movements tend to follow consistent patterns, it becomes possible for our agent to forecast future price directions without relying on external inputs (Murphy, 1999).

B Why LLMs Are Well-Suited for Price-Based Technical Analysis

This appendix explains why large language models (LLMs) are not only capable of conducting technical analysis from price data but are *especially* well matched to it. The core claim is methodological: technical analysis is a structured, short-horizon reasoning problem over standardized inputs (OHLC bars, indicators, and chart geometry), and modern LLM capabilities—multi-step reasoning (Wang et al., 2024; Zhang et al., 2024a; Liu et al., 2022b), multimodal perception (Zhang et al., 2024b; Lu et al., 2022; Zhang et al., 2023b), tool use Gong et al. (2025), retrieval Wei et al. (2025a), and agentic Wei et al. (2025b) coordination—map directly onto these requirements.

Technical analysis converts recent OHLC sequences into a compact set of signals—momentum oscillators, moving-average relations, rate-of-change, and shape/level interactions. These signals compose into rules of the form "*if MACD crosses down, RSI exits overbought, and price rejects resistance, then short with risk r*." LLMs excel at synthesizing such heterogeneous but *symbolic* cues into consistent, short-horizon judgments, producing language-native rationales and machine-checkable action schemas. This yields decisions that are both human-auditable and executable (see our DecisionAgent design in the main text).

A large share of technical analysis is visual: trend channels, swing pivots, triangles, flags, double bottoms, and wedge compressions are geometric concepts. Modern multimodal LLMs can parse candlestick charts, identify pivots and boundaries, and align the detected structure with canonical pattern libraries, enabling the system to reason about *structure + context* rather than isolated indicators. Our PatternAgent and TrendAgent instantiate exactly this: tool-generated charts are analyzed for support, resistance, slope, and convergence before any prediction is issued.

Purely textual reasoning can drift; technical analysis benefits from *grounding* via tools. By binding indicator calculators (RSI, MACD, ROC, Stoch, Williams %R), trendline estimators, and execution simulators to the LLM, we constrain outputs to numerically verifiable quantities, reduce hallucination risk, and accelerate decisions—key in latency-sensitive settings. Tools also standardize feature extraction, so the model reasons over stable, low-dimensional summaries instead of raw, noisy prices.

Technical analysis is naturally modular. Splitting responsibilities across specialized agents—Indicator (numerical momentum/oscillators), Pattern (geometric formations), Trend (direction and slope), and Risk (position sizing and boundaries)—yields complementary views that can be *cross-validated*. An agentic LLM stack (e.g., via LangGraph) supports: (i) division of labor for lower latency, (ii) explicit debate/consensus protocols to down-weight conflicting signals, and (iii) clean hand-offs to an execution layer that emits structured orders with stops and take-profits. Our QuantAgent architecture operationalizes this workflow and demonstrates consistent gains over rule-based and ML baselines on 1h/4h horizons.

In summary, Price-based technical analysis poses a *structured, tool-grounded, multimodal, and modular* reasoning task. These properties align tightly with modern LLM strengths in stepwise reasoning, chart perception, retrieval, tool invocation, and agentic coordination—yielding decisions that are fast, interpretable, and risk-aware. Consequently, LLMs are natural engines for technical analysis on price data, especially on 30min–4h horizons where structure dominates noise.