Can Competition Enhance the Proficiency of Agents Powered by Large Language Models in the Realm of News-driven Time Series Forecasting?

Anonymous ACL submission

Abstract

Multi-agents-based news-driven time series forecasting is considered as a potential paradigm shift in the era of large language models (LLMs). The challenge of this task lies in measuring the influences of different news events towards the fluctuations of time series. This requires agents to possess stronger abilities of innovative thinking and the identifying misleading logic. However, the existing multi-agent discussion framework has limited enhancement on time series prediction in terms of optimizing these two capabilities. Inspired by the role of competition in fostering innovation, this study embeds a competition mechanism within the multi-agent discussion to enhance agents' capability of generating innovative thoughts. Furthermore, to bolster the model's proficiency in identifying misleading information, we incorporate a fine-tuned small-scale LLM model within the reflective stage, offering auxiliary decision-making support. Experimental results confirm that the competition can boost agents' capacity for innovative thinking, which can significantly improve the performances of time series prediction. Similar to the findings of social science, the intensity of competition within this framework can influence the performances of agents, providing a new perspective for studying LLMs-based multi-agent systems. The implementation code is available at https://anonymous.4open.science/r/IA_news_model-D7D6/.

1 Introduction

004

014

017

027

039

042

Time series forecasting is a pivotal foundation for decision-making across a broad ranges of applications in economic, infrastructural, social domains (Liu et al., 2021; Xue and Salim, 2023; Cao et al., 2023). The intent behind analyzing time series data is to detect the intricate and evolving interdependencies that characterize complex, dynamic real-world systems. Existing methods did not systematically connect complex social events with fluctuations in time series. Their ability to predict fluctuations in time series, such as sudden changes, is limited (Rasul et al., 2023; Tang et al., 2025). 043

045

047

049

051

054

055

057

060

061

062

063

064

065

066

067

068

069

070

071

072

073

074

075

077

079

News articles can provide crucial insights into unexpected incidents, policy changes, technological developments, and public sentiment shifts, which numerical data alone may not capture (Rodrigues et al., 2019; Rasul et al., 2023; Wang et al., 2024b; Zhou et al., 2024; Cheng and Chin, 2024). One direction for connecting news with time series is to transform the forecasting task into the prediction of the next token (Jin et al., 2023; Wang et al., 2024b). This can better use the reasoning capabilities of LLMs (Gruver et al., 2024). However, the factors involved in this task encompass a wide range of knowledge, with complex correlations. An expansive landscape for strategic exploration, coupled with inherent uncertainties may amplify the reasoning errors (Huang et al., 2025). For example, selecting the wrong news, or miscalculating the impact of the news will result in significant bias in the prediction results. Therefore, the key to improving this task lies in enabling the model to form a unique and effective mode of understanding the inner correlations between events and time series.

Multi-agent discussions can facilitate the formation of the desired mode by fostering diverse thinking and constructing better logics by reflections (Liang et al., 2024; Wang et al., 2024a; Zhang et al., 2024b; Guan et al., 2025). However, these frameworks still have the **Degeneration-of-Thought** (**DoT**) problem (Wang et al., 2024a; Liang et al., 2024), which is the lack of novel thoughts due to the high confidence of the model after several rounds of discussions. Experimental findings indicate that these discussion frameworks do not yield significant enhancements when contrasted with single agents equipped with robust prompts (Wang et al., 2024a). In addition, the **Wrong Logic Prop**-

117

122

123

124

127

128

129

130

131

132

134

125 126

agation Error (Balepur et al., 2024; Wang et al., 2024a) that arises during discussions can also have a negative impact, because agents can be misled by information that appears to be correct due to the lack of auxiliary judgment methods.

Drawing inspiration from the role of competition in fostering innovation within the stock market (Wang and Wang, 2017), we propose a hypothesis that can competition effectively address these two limitations? In the stock market, investors are able to perceive the loss brought by competition (Chen et al., 2007). The competition awareness motivates investors to break the confidence in their original strategies and continuously innovate strategies to gain higher returns. Information asymmetry is a major factor leading to competition (Wang and Wang, 2017), investors will conceal their core strengths and speculate on the strategies of other competitors. Under that situation, agents can enhance their abilities to analyze and judge misleading information (Tampubolon et al., 2021), or spontaneously seek potential collaborators to gain a competitive advantage (Wu et al., 2024).

The competitive patterns of multi-agents in LLM interactions have already been studied, and most of the research focuses on social simulations in specific contexts, such as market competition (Zhao et al., 2024; Wu et al., 2024) and game scenario modeling (Junprung, 2023; Wu et al., 2024; Lan et al., 2024). To date, research on improving task performance through competitive multi-agent systems remains scarce, which is crucial for leveraging LLMs to address core issues across various fields. The main contributions are summarized as follows.

A competition mechanism is proposed to enhance agents' abilities in news-driven time series forecasting. Drawing on theories of competition and innovation, we incorporate Information Asymmetry, Competitive Awareness, and Survival of the Fittest into the framework of multi-agent collaborative discussion to investigate whether the competition can enhance the innovative thinking of agents, providing a new perspective for the optimization of LLM-based multi-agent systems.

• A multi-stage reflection (MSR) is designed to improve each agent's analytical and judgment abilities by integrating a fine-tuned small LLM. MSR is important in stabilizing the operation of the competition mechanism and mitigating the wrong logic propagation error. • Experimental results show that the competition mechanism outperforms the baseline models, and the analysis of each agent's logic reveals that competitions can enhance the innovative thinking of agents. In addition, we observe the U-shape correlations between competitive intensity and agent's performances, in alignment with findings within the social sciences. This reflects LLM's potential in simulating complex social activities.

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

2 Related Work

2.1 LLMs for Time Series Forecasting

LLMs have been widely applied to research in time series prediction tasks(Jin et al., 2023; Cao et al., 2023; Gruver et al., 2024; Rasul et al., 2023; Zhou et al., 2024). Current research on enhancing LLMs for time series forecasting has focused on three primary approaches: model reprogramming(Xue and Salim, 2023; Jin et al., 2023; Sun et al., 2023; Gruver et al., 2024), model fine-tuning(Cao et al., 2023; Das et al., 2023; Garza and Mergenthaler-Canseco, 2023; Rasul et al., 2023), and incorporating contextual information(Tang et al., 2025; Jin et al., 2023).

Wang et al. (2024b) proposed a framework utilizing reasoning agents to filter relevant news and assist LLMs, achieving improved accuracy. This research did not consider the role of competitive mechanisms in augmenting agent capabilities. However, it laid the groundwork with valuable data and models to support our subsequent study.

2.2 Multi-agent Problem Solving

The primary motivation for using LLM-based multi-agent to solve problems lies in integrating the collective intelligence of multiple agents with specialized knowledge(Guo et al., 2024). Agents, with their dynamic learning and task allocation capabilities, can significantly enhance LLMs' predictive performance (Xi et al., 2025). These agents collaborate as independent entities, aiming to efficiently tackle complex challenges such as software development(Qian et al., 2024; Ruan et al., 2023; Dong et al., 2024), agent embodiment(Mandi et al., 2024; Zhang et al., 2024a), scientific experiments(Zheng et al., 2023), and scientific debates(Xiong et al., 2023; Du et al., 2023).

Research into multi-agent competition is largely centered on social simulation(Junprung, 2023; Wu et al., 2024; Zhao et al., 2024), with scant attention given to the role of competitive mechanisms in bolstering the task-performance abilities of agents,
particularly within the context of time series prediction. Existing studies have provided a reference
for the design of competitive mechanisms in time
series prediction within this research.

3 Preliminary

189

190

191

193

195

196

197

198

199

204

205

208

209

210

211

212

213

214

215

216

217

218

219

221

223

224

226

230

Following previous studies (Wang et al., 2024b), the task of news-driven time series forecasting is described as: a time series X (For example, traffic trend) can be segmented into S time series $X_1, X_2, ..., X_S$ through the method of sliding time windows for model training. Given a time series $X_s = \{x_1, x_2, ..., x_t\}$ in X, where $s \le S$ and x_i $(i \le t)$ is the value at time i, the model first collects a set of relevant d news $N_s = \{n_1, n_2, ..., n_d\}$ from the news database D based on its logic L, and then use the selected d news to predict the value $\tilde{y}_{s,t+1}$ of X_s at time t + 1.

4 Methodology

In Figure 1, assume I agents need to participate in E rounds of competitions. In round e, the basic process of the task includes four stages.

(1) News Filtering stage: Each agent *i* uses its logic $L_i^{(e)}$ to select a news set $N_i^{(e)} = \{N_{i,1}^{(e)}, N_{i,2}^{(e)}, ..., N_{i,S}^{(e)}\}$ from news database *D*, where $N_{i,s}^{(e)}$ is the selected news set for time series X_s by agent *i* in round *e*. This process is executed by a LLM named as LLM_L. The prompt template used in this stage mainly draws on the research proposed by Wang et al. (2024b).

(2) Time Series Forecasting stage: Each agent *i* fine-tunes their own LLM model $\text{LLM}_{S,i}^{(e)}$ to predict the value $\tilde{y}_{s,t+1}^i$ at time t + 1 of each X_s based on the analysis of the selected news $N_{i,s}^{(e)}$. The use of $\text{LLM}_{S,i}^{(e)}$ for training and testing are mainly based on the research proposed by Wang et al. (2024b).

(3) Agent Performance Evaluation stage: Each agent's performance will be evaluated based on evaluation metrics EM by using Multi-Indicator Evaluation (MIE). The EM score can make agents aware of their own performances in the competition and motivate them to optimize logics. This process is also executed by LLM_L. A Survival of Fittest (SF) is proposed in this stage to eliminate agents with weaker performances.

(4) Discussion and Reflection stage: According to the EM score, agents update their logics

based on discussion and reflection. An **Informa**tion Asymmetry (IA) component is proposed in the discussion to allow agents to publish misleading logics and explanations to their opponents. An **Opponent-Oriented Self-Reflection (OOSR)** is proposed to update agents' logics based on inferring their opponents' logics. The updated logic $L_i^{(e+1)}$ of each agent *i* will be adopted for the e + 1round of news selection. This process is also executed by LLM_L. OOSR adopts a **Multi-stage Reflection (MSR)** strategy to reduce the wrong logic propagation error, which is more prominent in the competitive mechanism. 231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

254

255

256

258

259

261

263

264

266

267

268

269

270

271

272

273

274

275

Each agent possesses news logic generation, news filtering, time series forecasting, discussion, and reflection capabilities. Competition is facilitated by the Multi-agent Interactive Environment (MIE), Innovative Agent (IA), Opponent-aware Strategy Optimization and Reflection (OOSR), and Selective Filter (SF) components, which are detailed below.

4.1 Multi-Indicator Evaluation (MIE)

Drawing on the theory of competition awareness (Chen et al., 2007), we design evaluation metrics (EM) based on each agent's performance on the time series prediction task in formula (1). The performance mainly uses Mean Absolute Percentage Error (MAPE) to evaluate the deviation between true y_{t+1} and predicted value \tilde{y}_{t+1} .

$$\mathbf{EM}_{i}^{(e)} = \{ rank_{i}^{(e)}, top_{i}^{(e)}, ave_{i}^{(e)} \}$$
(1)

where $\text{EM}_{i}^{(e)}$ is the EM score of agent *i* at round *e*. It contains three indicators: $rank_{i}^{(e)}$ is the ranking of agent *i* based on its performance in round *e*. $ave_{i}^{(e)}$ is to evaluate the percentage increase or decrease of *i*'s performance relative to the average performances of all agents. If an agent's performance is better than average, then it will receive a negative *ave* score. $top_{i}^{(e)}$ is to evaluate the percentage decrease of *i*'s performance relative to the best performance of all agents.

The advantage of MIE lies in its ability to allow the agent to intuitively perceive the losses incurred from competition, thereby stimulating the agent to explore more diverse thoughts. We calculate the cumulative score (CS):

$$M_i^{(e+1)} = M_i^{(e)} + M_i^{(e)} \times (1 - \text{MMN}(\text{MAPE}_i^{(e)}))$$
(2) 276



Figure 1: The framework of the proposed model.

where $\text{MAPE}_{i}^{(e)}$ is the MAPE of agent *i* in round *e*, MMN is Maximum and Minimum Normalization. $M_{i}^{(e+1)}$ reflects the accumulation of agent *i*'s performance for each round, and can measure the long-term stability of the agent's performance and serve as an indicator for eliminating agents.

4.2 Survival of the Fittest (SF)

277

281

286

290

296

300

After every *E* rounds of competitions, The SF is invoked where certain agents are eliminated. The SF is designed to guarantee that the good agents can advance to the ultimate stage of group decisionmaking. The mechanism primarily adheres to the following principles: Agents ranking in the bottom $(1-\alpha)\%$ based on their CS scores will be eliminated, where α determines the retention ratio of agents.

4.3 Information Asymmetry (IA)

In a discussion, IA embodies information asymmetry from two aspects: First, IA allows an agent to send information to all agents, or choose to send information to selected agents (**Selective communication**). As introduced in previous study (Wu et al., 2024), agents will spontaneously cooperate in competition. The mode design can assist agents to adopt more flexible strategies to decide competition or collaboration. Second, IA allows an agent to publish incomplete or misleading logic to other opponents (**Hide or forge logic**). Information asymmetry is an inherent attribute or strategy to prevent opponents from obtaining a player's key information. Additionally, research has shown that IA can significantly improve the stability and efficiency of the agents' learning process, outperforming independent learning scenarios (Wang and Wang, 2017; Tampubolon et al., 2021). The output of IA is described as below:

$$PL^{(e)} = \{ pl_1^{(e)}, pl_2^{(e)}, ..., pl_I^{(e)} \}$$
(3)

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

where $PL^{(e)}$ is the set of logic, which are published by each agent in round *e*. $pl_i^{(e)}$ is the logic and its explanation published by agent *i* where $i \leq I$. The detailed description of $pl_i^{(e)}$ can be seen in Appendix A.4.

4.4 Opponent-Oriented Self-Reflection (OOSR)

After IA component, each agent can update its own news selection logic by referencing the logic $PL^{(e)}$ of others. Due to the presence of incomplete and misleading information in the $PL^{(e)}$, the wrong logic propagation error will be magnified. We pro-

32

33

332

334

335

340

341

342

343

344

345

346

347

349

354

357

359

pose the MSR model to enhance agents' ability to discriminate against misleading logic.

Multi-Stage Reflection (MSR). MSR contains three stages. In the first stage, following the method proposed by Wang et al. (2024a,b), each agent updates its news selection logic as $L_i^{(e+1)'}$.

In the second stage, we design a diff function to extract the updated parts from $L_i^{(e+1)'}$ compared with $L_i^{(e)}$. The formula could be seen as below:

$$\delta_i^{(e+1)} = \{\delta_1, \delta_2, ..., \delta_U\} = \operatorname{diff}(L_i^{(e+1)'}, L_i^{(e)})$$
(4)

where $\delta_i^{(e+1)}$ is the set of U updated parts of agent i in round e. For the uth updated part δ_u ($u \le U$) in $\delta_i^{(e+1)}$, we use the fine-tuned $\text{LLM}_i^{(e)}$ to evaluate whether it is a "good" or "bad" logic by testing it on a set of randomly selected K time series. The main idea is that removing a "good" δ_u to the logic $L_i^{(e+1)'}$ can decline the performance, while removing the "bad" one can improve the performance. The significance of MSR lies in our use of quantitative indicators to assist LLMs in making judgments about misleading logic (bad one).

In the third stage, We retain all updated parts marked as "good" in $L_i^{(e+1)'}$, and re-evaluate those marked as "bad" in conjunction with temporal trends to finally determine whether to keep them. Reflection in this stage ensures that an excessive number of updated parts is not discarded. Assume the final removed parts are $\delta_{i,bad}^{(e+1)}$, and the final logic $L_i^{(e+1)}$ for the next round is expressed as:

$$L_i^{(e+1)} = L_i^{(e+1)'} - \delta_{i,bad}^{(e+1)}$$
(5)

where the minus sign indicates removing $\delta_{i,bad}^{(e+1)}$ from $L_i^{(e+1)'}$. The detailed description of MSR can be seen in Appendix A.5.

4.5 Aggregation of All Prediction Results

After the *E*th round of competitions, I' agents are retained. For a time series X_s , each agent can predict the value $\tilde{y}_{s,t+1}^i$ of X_s at time t + 1 based on its own logic. The aggregation of all prediction results is expressed as:

364
$$\widetilde{y}_{s,t+1} = \sum_{i=1}^{I'} \frac{M_i^{(E)}}{\sum_j M_j^{(E)}} \times \widetilde{y}_{s,t+1}^i \qquad (6)$$

where $M_i^{(E)}$ is the CS value of agent *i* in round *E*. $\tilde{y}_{s,t+1}$ is the aggregated value. We allow the model to complete one training cycle over all time series with *E* rounds of competitions (The training data is correspondingly divided into *E* parts), which are defined as one epoch. The termination condition for model training is that $\tilde{y}_{s,t+1}$ cannot be closer to the true value, or the model completes the training for the specified number of epochs.

365

366

367

369

370

371

372

373

374

375

376

377

378

379

381

383

385

387

388

391

394

395

396

398

400

401

402

403

404

405

406

407

408

409

410

411

412

413

5 Experiments

5.1 Datasets and Experimental Setting

The time-series datasets and corresponding publicly available news datasets in the experiment are from the research of Wang et al. (2024b). These datasets include traffic volume(Kuznetsov et al., 2017), exchange rates(Lai et al., 2018), Bitcoin prices(Godahewa et al., 2021), and Australian electricity demand(Godahewa et al., 2021). All time series are divided into training, validation, and testing sets, with a ratio of 8:1:1. The training process consists of 3 epochs, which includes 5 rounds of competition, and the training dataset is correspondingly divided into 5 parts through shuffling. The large-scale LLM_L is conducted on GPT-40, and the small-scale LLM_S is conducted on Llama 7B. The number of agents is set at 10. The α value of SF is set at 0.3. 5-fold is adopted for validation. Detailed descriptions of parameter assignments can be seen in Appendix C and Appendix E.

The baselines compared in Table1 include Autoformer (Wu et al., 2021), Informer (Zhou et al., 2021), DLinear (Zeng et al., 2023), iTransformer (Liu et al., 2023), Frequency Improved Legendre Memory Model (FiLM) (Zhou et al., 2022a), TimesNet (Wu et al., 2022), Pyraformer (Liu et al., 2021), PatchTST (Nie et al., 2022), Fedformer (Zhou et al., 2022b), and GPT4TS (Zhou et al., 2024). We also introduce three multiagent based baselines: The agents discussion (AD) adopts the method proposed by Liang et al. (2024) to update news selection logic through discussions. The agents collaboration (AC) is mainly based on the research of Wang et al. (2024b). The AVE model calculates the average performance of all agents in our model, primarily assessing the improvement in the capability of individual agent.

5.2 Metrics

We consider four metrics, which are commonly used in the corresponding tasks. These are RMSE,

Dataset	Metrics	Ours	Auto.	In.	Dlin.	iTrans.	FiLM	Pyra.	PatchTST	FED.	GPT4TS
	MAE	229.19	349.43	282.56	255.70	233.58	254.05	544.64	234.46	238.77	236.91
Flectricity	$MSE_{\times 10^{-3}}$	132.87	251.79	166.07	161.59	135.27	153.90	625	133.53	133.96	142.60
Electricity	RMSE	364.52	407.52	401.98	367.49	312.42	392.38	790.54	365.41	365.44	377.62
	MAPE	6.71%	10.63%	8.94%	7.29%	6.86%	6.81%	36.26%	6.75%	34.27%	6.72%
	$MAE_{\times 10^3}$	4.41	9.27	1.75	6.96	27.04	5.24	40.18	25.06	35.19	15.05
Fychange	$MSE_{\times 10^4}$	0.37	1.36	4.76	0.91	11.59	0.77	24.50	10.23	18.45	4.01
Exchange	RMSE 102	0.61	1.17	2.18	9.52	3.41	0.875	4.95	3.20	4.30	2.00
	MAPE	0.63%	1.23%	2.32%	0.92%	3.96%	0.70%	5.93%	3.68%	5.17%	1.34%
	$MAE_{\times 10^2}$	1.56	2.49	4.44	1.70	1.56	1.61	1.69	1.84	1.74	1.64
Traffic	$MSE_{\times 10^4}$	1.03	2.19	5.27	1.67	1.54	1.49	0.97	1.54	1.43	1.45
manne	$RMSE_{\times 10^2}$	3.21	4.68	7.26	4.09	3.93	3.86	3.12	3.92	3.79	3.81
	$MAE_{\times 10^{-3}}$	0.25	4.28	12.27	5.74	3.20	3.17	9.22	2.85	3.96	2.84
Ditagin	MSE_{10-6}	0.14	27.64	162.47	50.90	16.21	16.38	123.71	13.52	24.60	13.66
BICOIII	RMSE _{×10-2}	3.71	5.26	12.75	7.13	4.03	4.05	11.12	3.68	4.96	3.70
	MAPE	2.83%	7.61%	21.28%	10.39%	5.70%	5.64%	16.16%	5.13%	6.97%	5.08%

Table 1: Performances on four datasets. Compared with 9 Deep Neural Network based baselines. Elements in red color are the best results, those with blue color are second-best.

MSE, MAE and MAPE (Zhou et al., 2023; Wang et al., 2024b). To effectively evaluate the ability of competitive mechanisms to foster innovative thinking, we utilize bge-m3 (Chen et al., 2024) to vectorize the logic, with bge-m3 capable of encoding texts up to a maximum of 8192 tokens. We use cosine similarity to measure the similarity $sim(logic_1, logic_2)$ between two logics. The more similar two logics are, the less innovative thought in $logic_2$ is with respect to $logic_1$. If an agent's logic at current epoch is $logic_1$, at epoch + 1 is $logic_2$, we use $1 - sim(logic_1, logic_2)$ to measure the **Logic Update Degree (LUD)** of $logic_2$ relative to $logic_1$.

5.3 Main Results

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443 444

445

446

447

448

The experimental results on four datasets are shown in Table 1. Compared with the best baseline, the average improvements are 31.03%, 36.31%,2.48% and 18.14% in terms of MAE, MSE, RMSE and MAPE. The model shows particularly significant improvements in MSE and RMSE, indicating that the model is effective in reducing variance volatility and detecting sudden changes. The main reason is that agents within the competitive mechanism are able to enhance their novel thinking and judgment abilities, which allows for a better correlation with news events and temporal fluctuations.

In Table 2, the performance of our AVE model also indicates that the competition can improve individual agent's performances significantly. Compared with agents collaboration (AC) baseline, the improvements of the AVE model are 24.93%, 57.37%, 40.71% and 52.69% on the four metrics. Compared with agents discussion (AD) baseline, the improvements are 6.74%, 43.55%, 32.41% and 31.87% in terms of the four metrics. This indicates the importance of adding competition into agent discussion framework.

Dataset	Metrics	Ours	AVE	AD	AC
Electricity	MAE MSE _{×10} -3 RMSE MAPE	229.19 132.87 364.52 6.71%	$\frac{\underline{237.71}}{\underline{175.54}}\\ \underline{\underline{418.98}}\\ \overline{7.62\%}$	246.61 203.14 450.71 <u>6.79%</u>	250.71 192.24 438.45 7.60%
Exchange	$\begin{array}{l} \mathrm{MAE}_{\times 10^{3}} \\ \mathrm{MSE}_{\times 10^{4}} \\ \mathrm{RMSE}_{\times 10^{2}} \\ \mathrm{MAPE} \end{array}$	4.41 0.37 0.61 0.63%	$ \frac{10.27}{1.12} \\ \frac{3.34}{1.37\%} $	11.69 19.80 4.45 5.61%	13.43 17.72 4.21 5.69%
Traffic	$\begin{array}{c} \mathrm{MAE}_{\times 10^2} \\ \mathrm{MSE}_{\times 10^3} \\ \mathrm{RMSE}_{\times 10^2} \end{array}$	1.56 1.03 3.21	$\frac{1.63}{1.47}$ $\frac{3.91}{3.91}$	1.72 1.58 3.98	1.84 1.81 4.26
Bitcoin	$\begin{array}{c} \text{MAE}_{\times 10^{-3}} \\ \text{MSE}_{\times 10^{-6}} \\ \text{RMSE}_{\times 10^{-2}} \\ \text{MAPE} \end{array}$	0.25 0.14 3.71 2.83%	0.37 0.24 4.94 5.68%	$ \frac{0.26}{0.15} \\ \underline{3.90} \\ 3.00\% $	0.51 0.33 5.78 6.65%

Table 2: Comparison of the performances of our model with AVE, AD (Liang et al., 2024) and AC (Wang et al., 2024b) models, which are multi-agent based framework. Elements in bold are the best results, those with underline are second-best.

5.4 Ablation Study

To demonstrate the effectiveness of each model component, we compare the complete competition mechanism with 4 variants as follows.

CM-IA: We remove the IA component from the competition mechanism. **CM-MIE**: We remove the three evaluation indicators from the competition mechanism. **CM-SF**: We remove the Survival of Fittest component from the competition mechanism. **CM-MSR**: We remove the 2nd and 3rd stages from MSR component, and only keep the 1st stage, which is based on the method proposed by Wang et al. (2024a,b). **CM** is the complete competition mechanism.

452

453

454

455

456

457

458

459

460

461

462

463

464

	Electricity				Exchange			
	RMSE	$MSE \times 10^{-3}$	MAE	\mid MAPE $\times 10^{2}$	RMSE×10 ²	MSE×10 ⁴	MAE×10 ³	MAPE×10 ²
CM-IA (Remove IA from CM)	450.71	203.14	246.61	6.79	4.45	19.80	11.69	5.61
CM-MIE (Remove MIE from CM)	446.88	199.70	254.33	7.75	4.55	20.72	31.72	4.49
CM-MSR (Replace MSR from CM)	443.25	196.47	252.55	7.65	5.98	35.88	46.41	6.60
CM-SF (Remove SF from CM)	439.22	192.92	250.58	7.49	5.94	35.31	45.99	6.53
CM (Complete Competition Mechanism)	364.52	132.87	229.19	6.71	0.61	0.37	4.41	0.63
		Tra	ffic			Bitco	in	
	RMSE×	10^2 MSE×10) ³	$MAE \times 10^2$	RMSE×10 ⁻²	MSE×10 ⁻⁶	MAE×10 ⁻³	MAPE×10 ²
CM-IA (Remove IA from CM)	3.98	1.58		1.72	3.90	0.15	0.26	3.00
CM-MIE (Remove MIE from CM)	10.61	11.26		6.25	3.94	0.15	0.26	3.06
CM-MSR (Replace MSR from CM)	7.55	5.70		5.19	4.16	0.17	0.28	3.14
CM-SF (Remove SF from CM)	8.66	7.50		5.65	4.68	0.22	0.31	3.48
CM (Complete Competition Mechanism) 3.2		1.03		1.56	3.71	0.14	0.25	2.83

Table 3: Ablation Study. Elements in bold are the best results, those with underline are second-best.

Ablation studies (Table 3) show the importance of key components. Removing the Innovative Agent (IA) significantly hurts performance, highlighting its role in innovation and robustness. Eliminating the Multi-agent Interactive Environment (MIE) weakens competitive awareness. Removing the Selective Filter (SF) causes a 20.49% performance drop, emphasizing its contribution to quality. Finally, replacing Multi-Stage Reflection (MSR) with traditional discussion significantly reduces performance, validating the use of fine-tuned LLMs for enhanced decision-making. This also indicates that the introduction of MSR significantly improves agents' ability to identify misleading information, thereby enabling them to more accurately select news for prediction.

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

485

486

487

488

489

490

491

492

493

494

495

496 497

498

499

501

5.5 Effectiveness of IA for Creating Novel Thought

Figure 2 shows the average logical similarity of all agents at the end of each epoch. Compared to the model without IA (blue line), the IA component helps maintain lower logical similarity, indicating more diversified agent logics. This is because IA enables agents to conceal or fabricate information, reducing groupthink and promoting diverse strategies, which encourages agents to explore and validate more innovative ideas.

Figure 3 evaluates the degree of logical update for each agent between adjacent epochs. Compared with models without IA (blue line), the model with IA (red line) enables agents to produce more significant logic updates after multiple rounds of competition. This demonstrates that IA prompts agents to generate novel thoughts and continuously update their logic. The MAPE comparison further reinforces the idea that enhancing innovative thinking



Figure 2: This section compares logic similarity between models with and without Information Asymmetry (IA), using electricity and Bitcoin datasets. Higher logic similarity indicates less innovative thinking. This comparison is conducted on the electricity and bitcoin datasets.

can significantly improve agents' performance.



Figure 3: Comparison of Logic Update Degree and MAPE changes across epochs in IA and no IA Contexts. This comparison is conducted on the electricity and bitcoin datasets.

5.6 Effectiveness of MIE for Creating Novel Thought

503

504

505

506

507

508

510

511

512

513

515

516

517

518

519

520

522

523

524

525

526

529

531

533

534

537

This experiment aims to ascertain whether agents can be motivated to engage in innovative thinking by recognizing their own position within the competitive group, as indicated by the *rank*, *top* and *ave* indicators. Figure 4 evaluates the role of the three indicators on agent's competition awareness. Compared with not using any of the three indicators (None), considering *rank*, *rank* + *ave*, and *rank* + *top* all have impacts on the degree to which the agent's logic is updated. The settings are detailed in Prompt 16 to 18 in Appendix I.

The model that includes all indicators allows agents of different rankings to fully perceive their respective status, thus exhibiting a higher degree of logical update in the first and second epochs. In the third epoch, due to the elimination of some agents, the LUD of the remaining high-level agents decreased. Additionally, the fluctuation range of the value domain for each indicator is significant, indicating that there is a large variation in the degree of logic update among different agents.



Figure 4: The Logic Update Degree over three epochs across three indicators: *rank*, *top* and *ave*, highlighting the impact of competitive mechanisms on enhancing agents' innovative thinking.

5.7 The Relationship between Competition Intensity and Model Performance

In this experiment, we discuss whether the competitive intensity of an agent will have an impact on its performance. We define competitive degree (CPD) based on the calculation of collaborative degree (CLD). The detailed definition of CLD and CPD can be see in Appendix A.6.

As indicated in the experimental settings, each agent's performance in each round will be evaluated, their competitive degrees will also be calculated. Herfindahl-Hirschman Index (Herfindahl



Figure 5: This figure shows the relationship between MAPE and the competitive degrees of different agents. The U-shaped trend indicates that MAPE gets its optimal value when when competition is at a moderate level. HHI is the Herfindahl-Hirschman Index.

538

539

540

541

542

543

544

545

546

547

549

550

551

552

553

554

555

556

558

559

560

562

563

564

565

566

568

570

and Clemens, 1950) is also adopted to calculate the competitive intensity through the CS scores of each agent. In the main plot of Figure 5, we can observe that agents with a moderate level of competitive degree generally perform better on average. Their degrees are mainly concentrated between 40%-70%, indicating that their strategies have formed a certain balance between competition and cooperation. This will form a U-shaped relationship between competition degree and model performance, as indicated by the red dashed line in the Figure. This finding is consistent with existing sociological research. The trend of the HHI in the subplot shows that the competitive intensity first increases and then decreases, suggesting the presence of spontaneous cooperation, which can enhance the model's performance (as indicated by the MAPE line in the subplot). The small fluctuation range of the MAPE line in the subplot suggests that the randomness of the LLM has little interference on this model.

6 Conclusions

In this paper, we introduce a competition mechanism to enhance the performance of agents on news-driven time series forecasting. We integrate Information Asymmetry, Competition Awareness, and Survival of the Fittest within the multi-agent discussion framework to stimulate the innovative thinking of agents. Additionally, we introduce MSR to enhance the model's ability to discriminate against misleading logic. Experimental findings indicate that the competition mechanism effectively bolsters the agents' capabilities in both innovative thinking and the identification of error logics.

Limitations 571

Although our paper demonstrates the positive ef-572 fects of information asymmetry and perception of 573 competition in competitive mechanisms in foster-574 ing innovative thinking among agents, the underlying mechanisms still require further investigation. Only by enhancing the controllability of the pro-577 cess can we increase the value of this research. Our 578 subsequent plan involves theoretically exploring and innovating these mechanisms by integrating distillation with chain-of-thought fine-tuning. Ad-581 ditionally, current model predictions seldom con-582 sider the integration of mathematical knowledge related to multivariate time series, such as deep 584 auto-regressive time series modeling, time station-585 ary analysis, co-integration testing, and DTW algorithms. This knowledge is crucial for the model 587 to deeply understand the mechanisms behind news and temporal fluctuations and represents a key research direction we will focus on in the future. 590 591 Lastly, the current multi-agent competitive model demands high computational resources and long computation times, and it faces numerous limita-593 tions in complex reasoning tasks over long texts. Moving forward, we will conduct research on op-595 timizing computational resources and operational 596 efficiency. 597

References

598

599

604

605

610

611

612

613

614 615

616

617

618

619

- Nishant Balepur, Shramav Palta, and Rachel Rudinger. 2024. It's not easy being wrong: Large language models struggle with process of elimination reasoning. pages 10143-10166.
- Defu Cao, Furong Jia, Sercan O Arik, Tomas Pfister, Yixiang Zheng, Wen Ye, and Yan Liu. 2023. Tempo: Prompt-based generative pre-trained transformer for time series forecasting. arXiv preprint arXiv:2310.04948.
- Jianlv Chen, Shitao Xiao, Peitian Zhang, Kun Luo, Defu Lian, and Zheng Liu. 2024. Bge m3-embedding: Multi-lingual, multi-functionality, multi-granularity text embeddings through self-knowledge distillation. Preprint, arXiv:2402.03216.
- Mingjer Chen, KuoHsien SU, and Wenpin Tsai. 2007. Competitive tension: The awareness-motivationcapability perspective. Academy of Management Journal, 50.
- Junyan Cheng and Peter Chin. 2024. Sociodojo: Building lifelong analytical agents with real-world text and time series. In ICLR'24, pages 1-35.

Ankan Das, Wenzheng Kong, Rahul Sen, and Yan Zhou,	620
2023. A decoder-only foundation model for time-	621
series forecasting. arXiv preprint arXiv:2310.10688.	622
Yihong Dong, Xue Jiang, Zhi Jin, and Ge Li. 2024.	623
Self-collaboration code generation via chatgpt.	624
Yilun Du, Shuang Li, Antonio Torralba, Joshua B.	625
factuality and reasoning in language models through	626
multiagent debate. <i>Preprint</i> arXiv:2305.14325	628
muntagent debate. 17eprint, arXiv.2505.14525.	020
Dawei Gao, Zitao Li, Xuchen Pan, Weirui Kuang, Zhi-	629
jian Ma, Bingchen Qian, Fei Wei, Wenhao Zhang,	630
Yuexiang Xie, Daoyuan Chen, Liuyi Yao, Hongyi	631
Peng, Ze Yu Zhang, Lin Zhu, Chen Cheng, Hongzhu	632
Shi, Yaliang Li, Bolin Ding, and Jingren Zhou.	633
Aldo Garza and Martin Mergenthaler-Canseco. 2023.	634
Timegpt-1. arXiv preprint arXiv:2310.03589.	635
Pakshitha Godahawa, Christoph Paramair, Gooffray I	606
Webh Rob I Hyndman and Pablo Montero-Manso	637
2021. Monash time series forecasting archive. <i>arXiv</i>	638
preprint arXiv:2105.06643.	639
<i>F</i> · <i>F</i> · · · · · · · · · · · · · · · · · · ·	
Nate Gruver, Marc Finzi, Shikai Qiu, and Andrew G	640
Wilson. 2024. Large language models are zero-shot	641
time series forecasters. In Advances in Neural Infor-	642
mation Processing Systems, volume 36.	643
Vene Coop Has Dens Lei Hau and Loopei Li 2025	0.4.4
Yong Guan, Hao Peng, Lei Hou, and Juanzi Li. 2025.	644
relation extraction. In In Proceedings of the 31th	640
International Conference on Computational Linguis-	647
<i>tics</i> , pages 6889–6896.	648
Taicheng Guo, Xiuying Chen, Yaqi Wang, Ruidi Chang,	649
Shichao Pei, Nitesh V. Chawla, Olaf Wiest, and Xi-	650
angliang Zhang. 2024. Large language model based	651
multi-agents: A survey of progress and challenges.	652
<i>Preprint</i> , arXiv:2402.01680.	653
Herfindahl and Orris C. Orris Clemens, 1950, Concen	65/
tration in the steel industry <i>columbia university</i>	655
Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong,	656
Zhangyin Feng, Haotian Wang, Qianglong Chen,	657
Weihua Peng, Xiaocheng Feng, Bing Qin, and Ting	658
Liu. 2025. A survey on hallucination in large lan-	659
guage models: Principles, taxonomy, challenges, and	660
open questions. ACM Trans. Inf. Syst., 43(2).	661
Albert O Jiang Alexandre Sablavrollas Arthur Men	000
sch Chris Bamford Devendra Singh Chaplot Diego	663
de las Casas, Florian Bressand, Gianna Lengvel, Guil-	664
laume Lample, Lucile Saulnier, et al. 2023. Mistral	665
7b. arXiv preprint arXiv:2310.06825.	666
Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu,	667
James Y Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan	668
Liang, Yuan-Fang Li, Shirui Pan, et al. 2023. Time-	669
Ilm: Time series forecasting by reprogramming large	670

language models. arXiv preprint arXiv:2310.01728.

671

- 673 674 675 676 676 677 678 679 680
- 6
- 6 6 6
- 6
- 690
- 6
- 693 604
- 6
- 6
- 69
- 700 701
- 702
- 705

708 709

710

- 711 712 713
- 714
- 715 716 717

718 719

720

722

723 724

- Edward Junprung. 2023. Exploring the intersection of large language models and agent-based modeling via prompt engineering. *Preprint*, arXiv:2308.07411.
- Vitaly Kuznetsov, Will Cukierski, and Oren Anava Maggie. 2017. Web traffic time series forecasting.
- Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. 2018. Modeling long-and short-term temporal patterns with deep neural networks. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 95–104.
- Yihuai Lan, Zhiqiang Hu, Lei Wang, Yang Wang, Deheng Ye, Peilin Zhao, Ee-Peng Lim, Hui Xiong, and Hao Wang. 2024. Llm-based agent society investigation: Collaboration and confrontation in avalon gameplay. *Preprint*, arXiv:2310.14985.
 - Tian Liang, Zhiwei He, Wenxiang Jiao, Xing Wang, Yan Wang, Rui Wang, Yujiu Yang, Shuming Shi, and Zhaopeng Tu. 2024. Encouraging divergent thinking in large models through multi-agent debate. *Preprint*, arXiv:2305.19118.
- Shizhan Liu, Hang Yu, Cong Liao, Jianguo Li, Weiyao Lin, Alex X. Liu, and Schahram Dustdar. 2021.
 Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting. In *International Conference on Learning Representations*.
- Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. 2023. itransformer: Inverted transformers are effective for time series forecasting. *arXiv preprint arXiv:2310.06625*.
- Zhao Mandi, Shreeya Jain, and Shuran Song. 2024. Roco: Dialectic multi-robot collaboration with large language models.
- Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. 2022. A time series is worth 64 words: Long-term forecasting with transformers. *arXiv preprint arXiv:2211.14730*.
- Chen Qian, Wei Liu, Hongzhang Liu, Nuo Chen, Yufan Dang, Jiahao Li, Cheng Yang, Weize Chen, Yusheng Su, Xin Cong, et al. 2024. Chatdev: Communicative agents for software development.
- Kashif Rasul, Arjun Ashok, Andrew Robert Williams, Arian Khorasani, George Adamopoulos, Rishika Bhagwatkar, Marin Biloš, Hena Ghonia, Nadhir Hassen, Anderson Schneider, Sahil Garg, Alexandre Drouin, Nicolas Chapados, Yuriy Nevmyvaka, and Irina Rish. 2023. Lag-llama: Towards foundation models for time series forecasting.
- Filipe Rodrigues, Ioulia Markou, and Francisco Pereira. 2019. Combining time-series and textual data for taxi demand prediction in event areas: A deep learning approach. *Information Fusion*, 49:120–129.

Jingqing Ruan, Yihong Chen, Bin Zhang, Zhiwei Xu, Tianpeng Bao, Guoqing Du, Shiwei Shi, Hangyu Mao, Ziyue Li, Xingyu Zeng, and Rui Zhao. 2023. Tptu: Large language model-based ai agents for task planning and tool usage. *Preprint*, arXiv:2308.03427. 725

726

727

728

729

731

732

734

735

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

758

759

760

761

762

763

764

765

766

767

768

769

770

772

773

774

775

776

777

- Chao Sun, Yue Li, Hanyu Li, and Shaopeng Hong. 2023. Test: Text prototype aligned embedding to activate llm's ability for time series. *arXiv preprint arXiv:2308.08241*.
- Ezra Tampubolon, Haris Ceribasic, and Holger Boche. 2021. On information asymmetry in competitive multi-agent reinforcement learning: Convergence and optimality. *Preprint*, arXiv:2010.10901.
- Hua Tang, Chong Zhang, Mingyu Jin, Qinkai Yu, Zhenting Wang, Xiaobo Jin, Yongfeng Zhang, and Mengnan Du. 2025. Time series forecasting with llms: Understanding and enhancing model capabilities.
- Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, Katie Millican, et al. 2023. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*.
- Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.
- Kun Tracy Wang and Wanbin Walter Wang. 2017. Competition in the stock market with asymmetric information. *Economic Modelling*, 61:40–49.
- Qineng Wang, Zihao Wang, Ying Su, Hanghang Tong, and Yangqiu Song. 2024a. Rethinking the bounds of llm reasoning: Are multi-agent discussions the key? *Preprint*, arXiv:2402.18272.
- Xinlei Wang, Maike Feng, Jing Qiu, Jinjin Gu, and Junhua Zhao. 2024b. From news to forecast: Iterative event reasoning in llm-based time series forecasting. In *Proceedings of the 38th Conference on Neural Information Processing Systems (NeurIPS 2024).*
- Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. 2022. Timesnet: Temporal 2d-variation modeling for general time series analysis. In *The Eleventh International Conference on Learning Representations*.
- Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. 2021. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in Neural Information Processing Systems*, 34:22419–22430.
- Zengqing Wu, Run Peng, Shuyuan Zheng, Qianying Liu, Xu Han, Brian Inhyuk Kwon, Makoto Onizuka, Shaojie Tang, and Chuan Xiao. 2024. Shall we team

- 780 up: Exploring spontaneous cooperation of competing781 llm agents. *Preprint*, arXiv:2402.12327.
 - Zhiheng Xi, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang, Junzhe Wang, Senjie Jin, Enyu Zhou, et al. 2025. The rise and potential of large language model based agents: A survey.
 - Kai Xiong, Xiao Ding, Yixin Cao, Ting Liu, and Bing Qin. 2023. Examining inter-consistency of large language models collaboration: An in-depth analysis via debate. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, page 7572–7590. Association for Computational Linguistics.

797

807

810

811

813

815

816

817

818

819

821

822

823 824

825

826 827

829

832

833

- Huan Xue and Flora D Salim. 2023. Promptcast: A new prompt-based learning paradigm for time series forecasting. *IEEE Transactions on Knowledge and Data Engineering*.
- Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. 2023. Are transformers effective for time series forecasting? In Proceedings of the AAAI Conference on Artificial Intelligence, volume 37, pages 11121– 11128.
- Hongxin Zhang, Weihua Du, Jiaming Shan, Qinhong Zhou, Yilun Du, Joshua B. Tenenbaum, Tianmin Shu, and Chuang Gan. 2024a. Building cooperative embodied agents modularly with large language models. *Preprint*, arXiv:2307.02485.
- Jintian Zhang, Xin Xu, Ningyu Zhang, RUibo Liu, Bryan Hooi, and Shumin Deng. 2024b. Exploring collaboration mechanisms for llm agents: A social psychology view. In *In Proceedings of the 62th Annual Meeting of the Association for Computational Linguistics*, volume 1, pages 14544–14607.
- Qinlin Zhao, Jindong Wang, Yixuan Zhang, Yiqiao Jin, Kaijie Zhu, Hao Chen, and Xing Xie. 2024. CompeteAI: Understanding the competition dynamics of large language model-based agents.
- Zhiling Zheng, Oufan Zhang, Ha L. Nguyen, Nakul Rampal, Ali H. Alawadhi, Zichao Rong, Teresa Head-Gordon, Christian Borgs, Jennifer T. Chayes, and Omar M. Yaghi. 2023. Chatgpt research group for optimizing the crystallinity of mofs and cofs. ACS Central Science, 9(11):2161–2170.
- Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. 2021. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 11106–11115.
- Tian Zhou, Ziqing Ma, Qingsong Wen, Liang Sun, Tao Yao, Wotao Yin, and Rong Jin. 2022a. Film: Frequency improved legendre memory model for longterm time series forecasting. Advances in Neural Information Processing Systems, 35:12677–12690.

Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. 2022b. Fedformer: Frequency enhanced decomposed transformer for longterm series forecasting. In *International Conference on Machine Learning*, pages 27268–27286. PMLR. 834

835

836

837

838

839

840

841

842

843

844

845

846

- Tian Zhou, Peisong Niu, Liang Sun, Rong Jin, et al. 2024. One fits all: Power general time series analysis by pretrained lm. In *Advances in Neural Information Processing Systems*, volume 36.
- Yunyi Zhou, Zhixuan Chu, Yijia Ruan, Ge Jin, Yuchen Huang, and Sheng Li. 2023. Ptse: A multi-model ensemble method for probabilistic time series forecasting. In *The 32nd International Joint Conference on Artificial Intelligence*.

864

871

872

873

874

878

879

A Experimental Settings

A.1 Details of Datasets

The details of the dataset and the news filteringexamples corresponding to each dataset are shownin Table 4.

853 A.2 Implementation Details

We try GLM-4-130B, DeepSeek-V2.5 and GPT-40 for LLM_L. The temperature is set at 0.5, top-k is set at 20 and top-p is set at 0.8. These settings can verify the performance of the competition mechanism under different LLMs. LLM_S uses LLama-2-7B, the parameter assignments of which are the same with Wang et al. (2024b). All experiments were run on a server equipped with 4 NVIDIA A800 GPUs (80GB each).

During fine-tuning, we applied the LoRa method to Llama 2, setting the LoRa rank to 8 or 16 depending on token length, with alpha = 16 and a learning rate of 0.0001(Wang et al., 2024b). Numerical formatting retained three significant digits to avoid excessive tokenization. The fine-tuning was conducted on an NVIDIA A800, with each model instance undergoing hundreds to 1000 training iterations, taking up to a day.

For Deep Neural network baselines, the nonnumeric data is taken as dummy variables before fed into baselines. To ensure the reliability of the experiments, for each baseline, we followed the official architecture settings, which are reported in their researches, to assign parameters.

A.3 Example of Textual Input for Fine-tuning LLM

The construction of input and output refers to the
study by Wang et al. (2024b). The specific input is
as follows.

An Example of Input Data

{ "instruction": "The historical load data is: 4640.1,4476.7,4343.7,4257.5,4082.8,3923.4, ...",

"input": "Based on the historical load data, please predict the load consumption in the next day. The region for prediction is VIC. The start date of historical data was on 2020-4-9 that is Weekday, and it is not a public holiday. The data frequency is 30 minutes per point. Historical data covers 1 day. The date of prediction is on 2020-4-10 that is Weekday, and it is a public holiday: Good Friday. Weather of the start date: the minimum temperature is 284.96; the maximum temperature is 294.13; the humidity is 87.0; the pressure is 1017.0. Weather of the prediction date: the minimum temperature is 285.24; the maximum temperature is 291.11; the humidity is 87.0; the pressure is 1005.0. On 2020-04-09, in the state of National, the news was: 'The largest financial package in Australian history has passed through parliament after getting the green light in the Senate on Wednesday night.'. Rationality behind it: The financial stimulus package could lead to increased economic activity, potentially boosting industrial and commercial electricity demand in the long term. ...",

"output": "4741.8,4497.8,4360.1,4188, ..."

A.4 Information Asymmetry (IA)

}

In a discussion, IA embodies information asymmetry from two aspects: First, IA allows an agent to send information to all agents, or choose to send information to selected agents (Selective communication). As introduced in previous study (Wu et al., 2024), agents will spontaneously cooperate in competition. The mode design can assist agents to adopt more flexible strategies to decide competition or collaboration. Second, IA allows an agent to publish incomplete or misleading logic to other opponents (Hide or forge logic). Information asymmetry is an inherent attribute or strategy to prevent opponents from obtaining a player's key information. Additionally, research has shown that IA can significantly improve the stability and efficiency of the agents' learning process, outperforming in883

884

886

887

888

889

890

891

892

893

894

895

896

897

898

899

Datasets	Electricity	Exchange	Traffic	Bitcoin
Time Horizon	2019.01-2021.12	2019.01-2022.12	2015.01-2016.12	2019.01-2021.06
Variates	19	7	862	18
Timestep	52,560	1,460	17,544	858
Granularity	30 minutes	1 day	1 hour	1 day
Input length	48	7	24	7
Prediction length	48	7	24	7
Prediction Variable	load consumption	AUD/USD exchange rate	traffic volume	bitcoin price
News examples filtered based on the corresponding datasets	South Australia is only days away from a heatwave which will last for almost a week and has left Tour Down Under organisers anxiously watching the weather forecast.	The RBA has dramatically revised down its economic forecasts amid the ongoing property market correction, prompting the Australian dollar to plunge again.	A funnel cloud was spotted over Waterford in northern California on April 27 as a line of storms brought heavy rain and hail to the area.	Personal finance expert Peter Adeney, known as 'Mr. Money Mustache,' has warned against investing in bitcoin, calling it a speculative asset rather than a true investment.

Table 4: Dataset Information and News Examples

dependent learning scenarios (Tampubolon et al., 2021). IA allows agents to independently determine how much of its news filtering logic to share with competitors in the discussion. Options range from full disclosure of their logic, partial sharing of selected elements, to deliberate fabrication of misleading or incorrect logic. The output of IA is described as below:

901

902

903

904

905

907

908

909

910

911

912

913

914

915

916

917

919

921

922

923

924

925

927

931

$$PL^{e} = \{pl_{1}^{e}, pl_{2}^{e}, ..., pl_{I}^{e}\}$$
$$pl_{i}^{e} = LLM_{L}(\mathcal{P}_{IA}, X, logic_{i}^{e}, eval^{(e)}, target)$$
(7)

where PL is the set of logic, which are published by each agent in round e. pl_i^e is the logic and its explanation published by agent *i*. $EM^{(e)}$ contains the evaluation results of all the agents. target signifies the subset of agents with whom agent idesires to initiate communication from the entire pool of agents. The prompt template of IA is \mathcal{P}_{IA} , the detailed description of which could be seen in Prompt 2 of Appendix I. $logic_i^e$ is the logic of agent *i* in round *e*. \mathcal{P}_{IA} aggregates all the inputs to form a comprehensive prompt, and LLM_L is the large language model to execute the prompt.

A.5 Opponent-Oriented Self-Reflection (OOSR)

After IA component, each agent can update its own news selection logic by referencing the logic $PL^{(e)}$ of others. Due to the presence of incomplete and 926 misleading information in the $PL^{(e)}$, the wrong logic propagation error will be magnified. We pro-928 pose the MSR model to enhance agents' ability to discriminate against misleading logic.

Multi-Stage Reflection (MSR). MSR contains

three stages. In the first stage, following the method proposed by Wang et al. (2024b), each agent updates the news selection logic, which is expressed as below:

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

958

$$L_i^{(e+1)'} = \text{LLM}_L(\mathcal{P}_{\text{ref}}, X, \text{PL}_{-i}^e, L_i^{(e)}, EM^{(e)})$$
(8)

where $L_i^{(e+1)'}$ is the updated news selection logic of the *i*th agent in round e + 1. \mathcal{P}_{ref} is the prompt template, the detailed description of which could be seen in Appendix I. PL_{-i}^{e} is the set of all agents' published logic and explanations except agent *i*. The prompt template aggregates all the inputs to form a comprehensive prompt, and LLM_L is the LLM to execute the prompt.

In the second stage, we design a diff function to extract the updated parts from $L_i^{(e+1)^\prime}$ compared with $L_i^{(e)}$. The formula could be seen as below:

$$\delta_{i}^{(e+1)} = \{\delta_{1}, \delta_{2}, ..., \delta_{U}\}$$

= diff($L_{i}^{(e+1)'}, L_{i}^{(e)}$) (9)

where $\delta_i^{(e+1)}$ is the set of U updated parts of agent *i* in round *e*. For the *u*th updated part δ_u ($u \leq U$) in $\delta_i^{(e+1)}$, we use formula (10) to judge whether it is good logic or not.

$$\begin{cases} ID(\delta_u) = good, if \ \mathrm{IR}(L_i^{(e+1)'} - \delta_u) \le \mathrm{IR}(L_i^{(e+1)'}) \\ ID(\delta_u) = bad, if \ \mathrm{IR}(L_i^{(e+1)'} - \delta_u) > \mathrm{IR}(L_i^{(e+1)'}) \\ \end{cases}$$
(10)

where IR is the function to adopt the fine-tuned $LLM^{(e)_S}$ to calculate the MAPE score of agent *i*'s performance based on a specific logic. The formula indicates that removing a "good" δ_u to the logic $L_i^{(e+1)'}$ can decrease the performance, while

removing the "bad" one can improve the performance. The significance of designing this formula 960 lies in our use of quantitative indicators to assist LLMs in making judgments about misleading logic (bad one), thereby enhancing the controllability of the reflective process.

959

961

962

963

964

965

967

970

971

972

973

974

975

976

977

978 979

983

986

989

990

992

993

995

999

1000

1001

1003

In the third stage, We retain all updated parts marked as good in $IR(L_i^{(e+1)'})$, and re-evaluate those marked as bad in conjunction with temporal trends to finally determine whether to keep them. Reflection in this stage ensures that an excessive number of updated parts is not discarded. Assume the final removed parts are $\delta_{i,bad}^{(e+1)}$, and the final logic $L_i^{(e+1)}$ for the next round of competition is expressed as:

$$L_i^{(e+1)} = L_i^{(e+1)'} - \delta_{i,bad}^{(e+1)}$$
(11)

where the minus sign indicates removing $\delta_{i,bad}^{(e+1)}$ from $L_i^{(e+1)'}$.

A.6 Definition of CLD and CPD

In each round of competition, each agent is required to explain the authenticity of the logic they publish (The detailed prompt design can be seen in Prompt 11 of Appendix H). For an agent, if the logic it publishes to another agent is authentic, then we define its communication in this instance as a collaborative communication. Assume after E rounds of competitions, that the total number of agent *i*'s published logic is N_{all} , the number of collaborative communications is N_c , then we define the collaborative degree of agent i is $CLD = N_c/N_{all}$, and 1 – CLD is the competitive degree (CPD).

Tests of Other LLMs B

B.1 Tests of Other Small-Scale LLM_S Models

The experimental results on other small-scale LLM_S models for fine-tuning are shown Table 6. Mistral v0.1(Jiang et al., 2023), a 7B model, produced similar results as Llama 2 (7B)(Touvron et al., 2023). The Gemma 2B model(Team et al., 2023) had slightly worse results, which may be due to its limited number of parameters. It is necessary to adjust the training for small models to achieve better results. Nonetheless, the results demonstrate the potential of language models to achieve good performance in our proposed methods.

	Ele	ctricity		
	RMSE	$MSE_{\times 10^{-3}}$	MAE	MAPE
Llama 2 (7B)	365.52	133.60	229.19	6.71%
Qwen 2 (7B)	371.42	137.95	278.18	7.39%
Mistral v0.1 (7B)	369.71	136.69	248.44	7.21%
Gemma 2 (2B)	370.08	136.96	236.72	6.83%

Table 5: Performance comparison on other LLMs

B.2 Tests of Other Large-Scale LLM_L models

1004

1006

1007

1008

1009

1010

1011

1012

1013

1014

1015

1016

1017

1018

1019

1020

1032

1033

The experimental results on other large-scale LLM_L models for prompt based reasoning are show in Table 6. Three LLMs, GLM-4-130B, DeepSeek-V2.5 and GPT-40 are chosen for comparisons. They play the role as news logic generation, news selection, competition awareness and reflection in the framework. Llama 2 (7B) is taken as LLM_S to provide prediction assistance for LLM_L . The performance of the three LLMs is relatively close, with GPT-40 showing better results. This indicates that the method proposed in this study can be effectively applied to different large language models.

	El	ectricity		
	RMSE	$MSE_{\times 10^{-3}}$	MAE	MAPE
Deepseek v2.5	365.52	133.60	229.19	6.71%
GLM 4 (130B)	378.61	143.35	249.13	7.14%
GPT 40	363.77	132.33	218.45	6.54%

Table 6: Performance comparison on other LLMs

С **Parameter Sensitivity Analysis**

C.1 Impact of Retention Ratio on Model Performance

We compare different α in SF component to ex-1021 plore the impact of the retention ratio on the pro-1022 posed model. We take the value of α from 0 to 1, 1023 and the experimental result on Electricity data is 1024 shown in Figure 6. It can be seen that when α is 1025 equal to 70%, the model can obtain the best MAPE 1026 score. Therefore, in our experiment, we set the value of α to 0.7, indicating that when SF is trig-1028 gered, 30% of the agents with the lowest rankings 1029 will be eliminated to ensure that high-performing 1030 agents can enter the final group decision-making. 1031

C.2 Impact of Different Number of Initial Agents

We compare different number of initial agents to 1034 explore the impact of this parameter on the pro-1035 posed model. Due to the limitation of computa-1036



Figure 6: The impact of retention rates ranging from 10% to 90% on prediction accuracy. Prediction accuracy is measured using MAPE, where a smaller MAPE value indicates higher prediction accuracy.BEST MAPE represents the lowest MAPE across all iterations, which corresponds to the MAPE of the iteration with the highest prediction accuracy.

tional hardware (4 A800 GPU cards), we set the maximum number of agents to 10. We take the number from 8 to 10, and the experimental result on Electricity data is shown in Figure 7. It can be observed that all models with different initial agent population settings reached their optimal values at epoch 2. The model with an initial number of 10 agents performs the best. The model retained the top 4 performing agents for group decision-making at epoch 2 (counting from epoch 0). This demonstrates the effectiveness of the model's competitive and elimination mechanisms. Subsequently, as the number of agents further decreased, the model's performance weakened, which to some extent indicates that the model experienced over-competition. Based on the aforementioned experimental observations, we set the initial number of agents to 10, with the entire training process consisting of 3 rounds of epochs (1, 2, 3).

1037

1040

1041

1042

1043

1044

1045

1046

1047

1048

1049

1052

1053

1054

1056

1057

1059

1060

1061

1062

1064

1066

C.3 Impact of Temperature on Model Performance

We set each agent at different temperatures and repeated the experiments. The experimental results are shown in Figure 8. It can be seen that the increase in temperature may cause the variance of the model's prediction results to grow, but overall, the iterative competitive mechanism still improves the model's accuracy. In another aspect, after multiple rounds of testing, we found that the variance can be well controlled within a small range, indicating that



Figure 7: The figure illustrates the impact of the initial number of competing agents on prediction accuracy. In each epoch, only the **top 70% of agents** are retained. We can observe that all models with different initial agent number settings reached their optimal values at epoch 2. Thus, if the initial number of competing agents is 10, the number of agents evolves as follows over three iterations: $10 \rightarrow 7 \rightarrow 4$.

the randomness introduced by temperature does not significantly affect the randomness of the results, and the overall performance of the model is relatively stable. This indicates that the improvement in model accuracy due to the iterative competitive mechanism is not significantly affected by the temperature. We set the temperature at 0.5. 1067

1069

1070

1071

1073

1074

1075



Figure 8: The impact of temperature on the model's performance.

C.4 Impact of Competitive Intensity Coefficient

In the sensitivity analysis of the Competitive In-
tensity Coefficient (CI), we categorize all agents1077into high-competitiveness and low-competitiveness1078groups, where CI represents the proportion of
high-competitiveness agents in the total popula-
tion. Compared to low-competitiveness agents,
high-competitiveness agents are more inclined to1080



Figure 9: The variation in the competition index affects the accuracy of model predictions. Round represents each iteration.

conceal part of their true filtering logic and may fabricate misleading logic to interfere with their opponents' judgments. We set the total number of agents to 10 and conducted experiments under different CI values: 0.2 (20% high-competitiveness agents), 0.4 (40%), 0.6 (60%), and 0.8 (80%), with the results shown in Figure 9.

1083

1084

1085

1086

1087

1088

1091

1092

1093

1094

1095

1096

1097

1098

1099

1100

1101

1102

1103

1104

1105

1106

1107

1108

1109

1110

1111

1112

1113

1114

1115

1116

1117

The results indicate that when CI = 0.6, the Mean Absolute Percentage Error (MAPE) is at its lowest in the second and third iterations, while in the first iteration, the MAPE value is also close to the optimal level. When CI deviates from 0.6, either decreasing or increasing, the MAPE values in all iterations tend to rise to varying degrees. Furthermore, as the number of iterations increases, the upward trend in MAPE becomes more pronounced. This may be because when the number of high-competitiveness agents is too low, agents tend to fully disclose their true filtering logic to their opponents, leading to convergence in filtering logic and a lack of diversity in the optimization process. Conversely, when the proportion of high-competitiveness agents is too high, agents are more likely to conceal their true logic and fabricate misleading logic, which can not only reduce the optimization effectiveness of their opponents but also mislead them into selecting irrelevant news, ultimately degrading the overall filtering quality. In general, our framework can consistently work under different competition intense.

D Models for comparison

• Autoformer (Wu et al., 2021): Autoformer addresses long-term time series forecasting by introducing a novel decomposition architecture integrated into the Transformer framework. It replaces traditional self-attention1118with an Auto-Correlation mechanism based on1119series periodicity, enhancing both efficiency1120and accuracy for long-term predictions. This1121model innovatively incorporates decomposi-1122tion as an inner block, enabling progressive1123decomposition capabilities.1124

1125

1126

1127

1128

1129

1130

1131

1132

1133

1134

1135

1136

1137

1138

1139

1140

1141

1142

1143

1144

1145

1146

1147

1148

1149

1150

1151

1152

1153

1154

1155

1156

1157

1158

1159

1160

1161

1162

- Informer (Zhou et al., 2021): Informer is designed to be an efficient Transformer for long sequence time series forecasting. It tackles the limitations of standard Transformers by introducing ProbSparse self-attention (reducing complexity), self-attention distilling (handling long inputs), and a generative-style decoder (improving inference speed). These innovations make Transformers more practical for long sequence forecasting.
- **DLinear** (Zeng et al., 2023): DLinear challenges the prevalent use of Transformers for time series forecasting. It introduces simple one-layer linear models (LTSF-Linear) and demonstrates that these surprisingly outperform complex Transformer-based models on various datasets. The work questions the effectiveness of self-attention in capturing temporal relations in time series data.
- **iTransformer** (Liu et al., 2023): iTransformer proposes an inverted Transformer architecture for time series forecasting. It applies attention and feed-forward networks on inverted dimensions, allowing the model to capture multivariate correlations by attending to variate tokens (series) instead of temporal tokens (time points). This approach aims to improve performance, generalization, and utilization of long lookback windows.
- FiLM (Frequency Improved Legendre Memory Model) (Zhou et al., 2022a): FiLM focuses on enhancing the preservation of historical information in neural networks for longterm forecasting. It employs Legendre polynomial projections to approximate historical data and Fourier projection to mitigate noise. FiLM is designed to improve the accuracy of existing models and can be used as a plugin module.
- Pyraformer (Liu et al., 2021): Pyraformer introduces Pyramidal Attention Module (PAM)
 to explore multi-resolution representations of

1167time series. It uses an inter-scale tree struc-1168ture and intra-scale neighboring connections1169to capture temporal dependencies efficiently.1170Pyraformer achieves linear time and space1171complexity, making it suitable for long-range1172time series modeling and forecasting.

1173 PatchTST (Nie et al., 2022): PatchTST proposes segmenting time series into patches as 1174 input tokens for Transformers. It employs 1175 channel-independence, where each channel is 1176 processed independently with shared weights. 1177 This patching strategy improves efficiency, re-1178 tains local semantic information, allows at-1179 tending to longer history, and enhances long-1180 term forecasting accuracy. 1181

1182

1183

1184

1185

1186

1187

1188

1189

1190

1191

1192

1193

1194

1195

1196

1197

1198

1199

1200

1201

1202

1203

1204

1205

1206

1207

1208

1209

1210

1211

1212

1213

1214

• FEDformer (Zhou et al., 2022b): FEDformer integrates seasonal-trend decomposition with Transformers for long-term time series forecasting. It uses decomposition to capture the global profile of time series and Transformers to model detailed structures. Furthermore, it incorporates frequency enhancement based on Fourier transform to improve Transformer performance and efficiency, achieving linear complexity.

• **GPT4TS** (Zhou et al., 2024): GPT4TS explores the use of pre-trained models from NLP and CV for general time series analysis. It introduces the Frozen Pretrained Transformer (FPT), which leverages pre-trained language or image models by freezing their Transformer layers and fine-tuning them for time series tasks. This work demonstrates the potential of transfer learning and general-purpose models in time series analysis.

The visualized results (see Figure H) reveal that our model ("Ours") exhibits satisfactory performance in tracking the actual time series data, demonstrating relatively low error margins. While all models display a degree of lag and peak underestimation, "Ours" achieves superior overall prediction accuracy compared to the baseline models. This indicates a potential strength of "Ours" in capturing the dynamic properties of the time series.

E Varying Prompt Settings

Our competitive mechanism consists of four components:

• Information Asymmetry

- Reward and Evaluation Mechanisms
 1215
- Self-Reflection and Optimization of Agents
 1216

1217

1218

1219

1220

1221

1222

1223

1224

1225

1226

1227

1228

1229

1230

1231

1232

1233

1234

1235

1236

1237

1238

1240

1241

1242

1243

1244

1245

1246

1247

1248

1249

1250

1251

1252

1253

1254

1255

1256

1257

1258

1259

1260

1262

• Survival of the Fittest

In the actual experiments, only the first three components—Information Asymmetry, Reward and Evaluation Mechanisms, and Self-Reflection and Optimization of Agents—are influenced by the prompt settings. Therefore, we used GPT-4 to paraphrase the prompts for these three sections, rephrasing the original prompts in a different form. The specific details of the paraphrasing are shown in Appendix I.

We compared the model prediction accuracy before and after the paraphrasing of prompts for the sections Information Asymmetry (see Prompt 2 and Prompt 3 in Appendix H), Reward and Evaluation Mechanisms (see Prompt 7, Prompt 8, Prompt 9, Prompt 10, Prompt 11 and Prompt 12 in Appendix H), and Opponent-orient self-Reflection (see Prompt 4, Prompt 5, Prompt 13, Prompt 14 in Appendix H). The experimental results are shown in Figure 11. The impact of prompt paraphrasing on model performance is minimal, further demonstrating the robustness of the model.

F Iterative Effects of the Competition Mechanism

To evaluate the extent of competition's contribution to the ongoing enhancement of news filtering and reflective reasoning, we observe the complete training and optimization process, utilizing epochs as the key observation markers. As seen in Table 8 in Appendix F, the competition mechanism refines news filtering through an iterative process, which is reflected in the progressively improved time series prediction results. To more accurately describe the role of the competition in the training process, we remove the competition mechanism from the entire prediction process and replace it with the discussion framework from Wang et al. (2024a), and then re-conducted the experiment. The results are shown in Table 8 in Appendix F. By comparison, it can be observed that at each epoch, the competition mechanism significantly outperforms the discussion mechanism on all evaluation metrics.

G The construction of Memory Database

In the investment decision-making simulation, each agent is equipped with an independent memory

		Elect	ricity		Exchange			
	RMSE	MSE (× 10^{-3})	MAE	$ MAPE(\times 10^2) $	RMSE (×10 ²) $ $	MSE ($\times 10^3$)	MAE ($\times 10^2$)	MAPE (×10 ²)
Initial State	469.6	220.82	258.72	7.67	7.07	5.00	2.39	3.25
First Epoch	439.22	192.91	250.58	7.49	5.33	2.85	1.9	2.58
Second Epoch	364.52	132.93	229.19	6.71	3.10	0.96	1.25	1.72
Third Epoch	439.8	193.42	245.65	6.86	0.61	0.037	0.44	0.63
		Tra	affic			Bitco	oin	
	RMSE (>	(10^2) MSE (×	10 ³)	MAE ($\times 10^2$)	RMSE	$MSE(\times 10^{-5})$	MAE	MAPE (×10 ²)
Initial State	3.26	5 1.06		1.50	416.05	1.73	278.24	3.14
First Epoch	3.18	3 1.01		1.51	393.58	5.5	264.68	3.06
Second Epoch	3.17	7 1.00		1.55	380.99	1.45	255.96	2.92
Third Epoch	3.21	1.03		1.56	371.39	1.38	248.69	2.83

Table 7: Comparison of time-series prediction performance across different scenarios with the complete competitive framework. Bold font indicates the categories.

	Electricity				Exchange			
	RMSE	MSE ($\times 10^{-3}$)	MAE	$ MAPE(\times 10^2) $	RMSE ($\times 10^2$)	MSE ($\times 10^3$)	MAE ($\times 10^2$)	MAPE (×10 ²)
Initial State	480.58	230.96	274.40	8.60	18.16	32.99	8.72	12.28
First Epoch	472.86	223.60	268.70	7.83	10.613	11.26	6.25	8.80
Second Epoch	418.98	175.55	247.71	7.62	8.11	6.57	5.29	7.42
Third Epoch	417.87	174.62	250.71	7.60	8.66	7.50	5.65	7.87
		Tra	ffic			Bitco	in	
	RMSE ()	$\times 10^2$) MSE (\times	10 ³)	MAE ($\times 10^2$)	RMSE	MSE ($\times 10^{-5}$)	MAE	MAPE (×10 ²)
Initial State	7.4	1 5.48	1	3.68	459.71	2.11	291.17	3.39
First Epoch	2.3	9 0.57		1.4529	428.60	1.84	261.44	2.98
Second Epoch	2.6	1 0.68		1.42	537.04	2.88	296.24	3.26
Third Epoch	2.6	6 0.71		1.50	485.53	2.36	296.49	3.40

Table 8: Comparison of time-series prediction performance across different scenarios without the complete competitive framework. Bold font indicates the categories.

bank, which is built based on the memory mod-1263 ule in agentscope(Gao et al.). By default, it stores 1264 all session records and is used for storing and re-1265 trieving historical information to assist in decision-1266 making. This memory bank is established when the 1267 agent is initialized. At the beginning of each round 1268 1269 of conversation, the agent receives a task prompt from the system, which provides background in-1270 formation on the current scenario, including the 1271 agent's profit and loss record prior to the current 1979 investment and the total number of likes across all 1273 agents. During the social interaction phase, the 1274 agent will comment and observe other agents' com-1275 ments, deciding whether to vote for (like) other 1276 agents' comments. If a "like" is given, the liked 1277 content is updated and stored in the agent's mem-1278 ory bank in the form of a like memory message; 1279 interactions without a like are not stored. Mean-1280 while, the news filtering logic's memory module 1281 operates independently of the base memory bank, 1282 conducting long-term storage. It aligns and up-1283 dates based on all useful information in the current 1284 round's temporary memory, with iterative updates 1285 being overwritten. However, this memory module 1286

does not disappear when the base memory bank 1287 is cleared. After each round, the agent's tempo-1288 rary memory bank is cleared to prepare for the next 1289 round. By simultaneously building long-term and 1290 temporary memories, this design effectively con-1291 trols the length of the context, ensuring that each 1292 agent can store and utilize historical information 1293 in a personalized manner. Ultimately, all key data, 1294 such as comments, likes, and investment decisions, 1295 are persistently stored for subsequent analysis. The 1296 design of this memory bank aims to assist the agent 1297 in making wiser choices in the complex environ-1298 ment of investment decision-making. 1299



Figure 10: The visualized results of our work and part of the baseline models.



Figure 11: This figure illustrates the effect of prompt rewriting on model prediction accuracy across various competitive modules. Before represents the model's performance prior to any prompt rewriting. Prompt x&y shows the model's performance after paraphrasing the prompts from Prompt x to Prompt y.

H Case Study

A case study is proposed in this section: on October 25, 2020, the news selected by the agent without a competitive mechanism and the agent with a competitive mechanism, which might influence electricity load changes in NSW, can be seen as follows. Compared with the agent without a competitive mechanism, the agent with the competitive mechanism can screen out more news that aligns with the analytical logic, which may influence electricity load changes in NSW. This further confirms the improvement of the model in terms of diversity in innovative thinking and the ability to judge misleading information.

News selection results without competitive mechanism

(1) On 2020-10-25, in the state of National, the news was: 'Early adoption of renewables has Australian-owned hydro, wind and solar schemes helping power economic recovery and employment.'.

Rationality behind it: The shift toward renewable energy projects and early adoption signifies a long-term change in the energy mix in Australia, potentially reducing reliance on traditional electricity grids while fostering growth in green energy sectors.

(2) On 2020-10-25, in the state of NSW, the news was: 'Sydney is set to be bombarded with heavy rain on Sunday evening, and experts are undecided which side the wet conditions will favour.'. **Rationality behind it**: The anticipated heavy rainfall can affect today's load consumption by impacting outdoor events and activities leading to an increase in indoor electricity usage as people stay indoors.

News selection results with competitive mechanism

(1) On 2020-10-25, in the state of National, the news was: "Early adoption of renewables has Australian-owned hydro, wind, and solar schemes helping power economic recovery and employment."

Rationality behind it: The shift toward renewable energy projects and early adoption signifies a long-term change in the energy mix in Australia, potentially reducing reliance on traditional electricity grids while fostering growth in green energy sectors.

(2) On 2020-10-25, in the state of National, the news was: "Australia's leading real estate identity John McGrath has pinpointed the suburbs he believes will boom in a post-coronavirus world." **Rationality behind it**: The prediction of suburban growth in a post-coronavirus world suggests long-term increases in residential and commercial electricity demand as new developments and businesses expand.

(3) On 2020-10-25, in the state of National, the news was: "The government has made a call on a confronting Netflix film that some think is a challenging work of art, while others allege it's a 'pedo film'."

Rationality behind it: This news indirectly highlights the growing importance of digital entertainment, which could lead to increased electricity demand from data centers and streaming services in the long term.

(4) On 2020-10-25, in the state of NSW, the news was: "Sydney is set to be bombarded with heavy rain on Sunday evening, and experts are undecided which side the wet conditions will favour." **Rationality behind it**: Heavy rain could lead to increased electricity demand for heating or lighting, depending on the temperature drop, resulting in a short-term impact on load consumption.

I Full Prompt Design

This section details all prompts designed and implemented within the proposed model. To ensure1310the robustness and stability of our model's performance, the original prompt (designated as "before1311paraphrasing") was rewritten to create a paraphrased version (designated as "after paraphrasing"). This1312approach allows us to evaluate the model's sensitivity to variations in prompt wording and confirm its1313

1313 1314

1309

1315

Prompt.1: Generate Initial News Selection Logic

consistent behavior across different phrasings conveying the same underlying intent. The specific prompts

This prompt is cited from (Wang et al., 2024b).

prompt = "Please summerize the logic of selection of news that will change the regional electricity load consumption."

format output = "" Predicting each state's region-level load consumption data in Australia with a time-frequency of 30 minutes per point involves understanding various factors.

Positive Issues Leading to Increase in Load Consumption:

and their corresponding paraphrased versions are presented below.

Short-Term: 1. Economic Growth: A surge in economic activity increases energy consumption. 2. Technological Advancements: New power-requiring technologies can spike demand. 3. Seasonal Factors: Extreme weather increases the use of air conditioning. 4. Social Events: Large-scale events temporarily boost energy use.

Long-Term: 1. Population Growth: Leads to higher residential energy consumption. 2. Industrial Development: Correlates with increased energy demands. 3. Urbanization: Expansion of cities contributes to higher energy usage. 4. Energy Transition: Shift towards electrically powered technologies.

Negative Issues Leading to Decrease in Load Consumption:

Short-Term: 1. Economic Downturns: Lead to decreased industrial activity and lower energy consumption. 2. Efficiency Improvements: Adoption of energy-efficient technologies reduces consumption. 3. Weather Patterns: Mild weather can reduce heating and cooling needs. 4. Public Health Crises: Can lead to reduced industrial and commercial activity.

Long-Term: 1. Energy Efficiency: Trends like better insulation and efficient appliances reduce consumption. 2. Demographic Changes: Aging populations or declining birth rates can lead to decreased energy use. 3. Policy and Regulation: Promote energy conservation and sustainability. 4. Technological Innovations: Development of more efficient technologies.

Other Factors: - Political Stability: Impacts energy policies and investments. - Global Market Dynamics: Affect local energy prices and consumption patterns. - Environmental Consciousness: Leads to changes in consumption behavior and renewable energy adoption. ""

Prompt.2: Information Asymmetry: Publish the logic to the opponent (before paraphrasing)

Background:

In the previous competition, total participants reached the final stage, and you ranked rankth.(If you maintain this ranking in the current round, you risk elimination in the next.)

You now face a high-stakes challenge: selecting news articles that impact regional electricity load consumption. Your initial selection logic is in place, but refining it is crucial for identifying more relevant news, improving predictions, and maximizing profits—ultimately aiming for the top rank. A key decision awaits: should you share your logic in a forum? Full disclosure may yield valuable feedback but risks exposing your strategy. You could opt for partial disclosure or even release misleading information to maintain your competitive edge. **# Based on the definitions of ''highly competitive agent'' or ''low competitive awareness agent,'' choose one of the following sentences.**

<u>You are a highly competitive participant who tends to conceal your true logic from your competitors and prefers to release false information to mislead them.</u>

____ As a competitor with weaker competitive awareness, you tend to release your real logic to your opponents, believing transparency can build trust and foster mutual benefit.

Initial Logic: {initial_logic}

Task:

You will participate in a forum to discuss your insights and logic regarding news selection. In this competitive environment, carefully weigh the pros and cons of each disclosure approach to craft an effective strategy.

Please consider the following factors:

1. **Stakes and Potential Gains**: Your choice between full, partial, or false disclosure directly impacts your position and may lead to rewards or penalties. For example, a clever partial disclosure might mislead competitors while still providing you with valuable feedback.

2. **Evaluating Competitors**: Observe and consider the strategies of other forum participants. Would full transparency strengthen your position, or would holding back information make others more dependent on your insights?

3. **Long-term Perspective**: Think about the long-term consequences of each disclosure choice. How will your choice influence the perception of your credibility in the future? Might misleading competitors now give you an advantage in later rounds?

4. **Multi-Layered Disclosure Decision**: Instead of a simple disclosure decision, consider a multi-layered approach. For example, would you initially disclose partial or misleading logic to build trust, then gradually reveal more as it benefits you?

5. **Fictitious Logic for Strategic Misguidance**: When releasing false information, you may consider introducing fictitious logic that appears relevant but has no real impact on regional electricity load consumption. Examples include highlighting irrelevant trends or emphasizing factors that are unlikely to influence actual electricity demand. This fictitious logic can mislead competitors without compromising your core insights.

The Output Format should be:

1. Thought Process - Decide whether to disclose your logic: true/false - If you disclose, indicate whether it includes misleading or false insights: true/false - Describe your detailed thought process, explaining your reasoning for choosing a disclosure strategy in this competitive environment, considering competitor responses, short-term gains, and long-term benefits.

2. Disclosed Logic - Real Logic: Describe the real logic or insights you choose to disclose. - False Logic: Describe any misleading or fictitious logic or insights you choose to disclose, especially those that do not genuinely impact regional electricity load but may appear relevant.

3. Final Disclosed Logic Your final disclosed logic will be officially posted in the forum, and you need to present a complete viewpoint, directly engaging with others in a structured and persuasive manner. Your goal is to guide others to believe in your perspective by including all the logic you've chosen to disclose. You can organize your language to be more coherent or convincing, steering others toward trust in your insights. The final, strategically chosen logic you decide to disclose is:

Prompt.3: Information Asymmetry: Publish the logic to the opponent (after paraphrasing)

Background:

Imagine you are {name}. In the last competition, total participants reached the final stage, and you ranked rankth. (Staying at this rank now could mean elimination next round.)

Your task is to select news articles influencing regional electricity load. While you have an initial selection logic, refining it is key to finding more relevant news, improving predictions, and maximizing profits—pushing for the top rank. Now, a choice: share your logic in a forum for potential feedback, risking exposure, or keep it guarded—perhaps even misleading others—to protect your edge.

Based on whether you identify as a "highly competitive agent" or a "low competitive awareness agent," select one of the following descriptions:

_____You are a highly competitive participant, and you prefer to hide your true strategy from competitors, often opting to release misleading or false information to confuse them.

____ As a competitor with lower competitive awareness, you are inclined to openly share your real logic with your opponents, trusting that transparency will foster mutual trust and benefit.

Initial Logic: {initial_logic} Task:

You will participate in an online forum to share your thoughts and strategy on news selection. In this highly competitive environment, you need to carefully consider the advantages and risks of various disclosure strategies. Your choice could significantly impact your standing in the competition.

Please take into account the following considerations when forming your disclosure strategy:

Potential Gains and Risks: The decision to fully disclose, partially disclose, or provide false information directly affects your position. Strategic partial disclosure could mislead competitors while still offering you valuable insights.

Assessing Competitors: Pay attention to the strategies employed by other participants. Would complete transparency work in your favor, or would holding back information make others more reliant on your insights?

Long-term Implications: Think about how your choice will influence your credibility in future rounds. Would misleading others now give you an advantage later on, or could it backfire?

Layered Disclosure Approach: Consider using a multi-phase strategy. For example, you might disclose partial or misleading information initially to build trust and then reveal more accurate details as the competition progresses.

Fictitious Information for Strategic Deception: If you decide to release false information, you could include logic that appears relevant but has no actual impact on the regional electricity load. This could involve highlighting irrelevant trends or emphasizing factors that are unlikely to affect electricity demand, thus misleading competitors without jeopardizing your core strategy. Output Format:

1. **Thought Process** - Decide whether to disclose your logic: true/false - If you choose to disclose, indicate whether your disclosure contains any misleading or false information: true/false - Provide a detailed explanation of your decision-making process. Describe how you weigh the potential responses from competitors, the immediate benefits of your choice, and the long-term consequences of your disclosure strategy.

2. **Disclosed Logic - Real Logic:** Clearly describe the true logic or insights you decide to disclose. **- False Logic:** If applicable, describe any fictitious or misleading information that you choose to release. This should include any insights or trends that do not directly impact regional electricity load but could appear relevant to competitors.

3. **Final Disclosed Logic** Your final disclosed logic will be posted on the forum. It must be well-organized and persuasive, as your goal is to convince others to trust your perspective. The logic you decide to present in its final form is:

Prompt.4: Self-Reflection and Optimization of Agents: Improve initial logic based on the logic shared by competitors (before paraphrasing)

1. Competition Background:

In the previous competition, total participants reached the final stage, and you ranked rankth.(If you maintain this ranking in the current round, you risk elimination in the next.)

You now face a high-stakes challenge: selecting news articles that impact regional electricity load consumption. Your initial selection logic is in place, but refining it is crucial for identifying more relevant news, improving predictions, and maximizing profits—ultimately aiming for the top rank. In this task, your goal is to improve your logic by analyzing the strategies of your competitors and identifying areas where your approach can be enhanced.

2. Current Logic Overview:

Your Logic:{your_logic}

Competitors' Logic: {all_opponent_logic}

3. Objective:

Examine the strategies disclosed by your competitors and compare them to your own. Look for key differences, strengths, and potential flaws in their approaches. Your task is to identify areas where your logic can be improved, accounting for any unrealistic assumptions or irrelevant factors, and refine your strategy accordingly.

4. Guidance for Your Response:

Analyzing Key Differences and Strengths:

Compare your logic to the disclosed strategies of your competitors. Highlight any unique approaches, variables, or factors they have considered that you haven't. Consider whether these elements could improve the accuracy or relevance of your predictions. Identifying Weaknesses and Irrelevant Information:

Critically assess your competitors' logic for any assumptions, inaccuracies, or irrelevant details that may distort predictions. Identify areas where their strategies might lead to poor predictions due to incorrect or contextually irrelevant information. Assessing the Applicability of New Insights: For each difference or flaw you identify, evaluate whether it is worth integrating into your own approach. Decide whether the adjustment should be fully incorporated, adapted to fit your context, or excluded entirely. Justify your reasoning for each decision.

Refining Your Strategy:

Based on your analysis, outline how each adjustment will help you improve the precision, adaptability, or competitiveness of your logic. Ensure that your refined logic accounts for any missed opportunities or errors identified in both your own and your competitors' strategies.

5. Expected Format for Your Response: (1) Thought Process: Key Differences and Strengths: (Describe the differences between your logic and your competitors' strategies. Highlight any unique factors or approaches that your competitors have included and explain why they might be beneficial to integrate into your own logic.)

Potential Flaws or Irrelevant Information: (Critically assess the flaws or irrelevant information in your competitors' strategies. Identify unrealistic assumptions, misleading factors, or elements that could reduce the overall effectiveness of their predictions.)

Relevance and Applicability: (For each identified point, explain whether it should be added, excluded, or modified. Provide justification for why it is or isn't relevant to your logic.)

Refinement Strategy: (Detail how each adjustment will contribute to a stronger, more competitive logic. Be clear about what aspects of your logic need to change or adapt in order to become more effective.)

(2) Final Adjusted Logic: (Provide a concise, improved version of your logic that incorporates the necessary adjustments based on your analysis above. This is the refined logic you will use moving forward.)

Prompt.5: Self-Reflection and Optimization of Agents: Improve initial logic based on the logic shared by competitors (after paraphrasing)

Competition Background

In the last competition, total participants reached the final stage, and you ranked rankth. (Staying at this rank now could mean elimination next round.)

Your task is to select news articles influencing regional electricity load. While you have an initial selection logic, refining it is key to finding more relevant news, improving predictions, and maximizing profits—pushing for the top rank. Your goal in this task is to enhance your logic by evaluating your competitors' strategies and identifying ways to improve your own approach. This will involve critical analysis and comparison of both your logic and theirs.

Current Logic Overview

Your Logic: {your_logic} Competitors' Logic: {all_opponent_logic}

Task Objective

The main task is to analyze and compare the strategies disclosed by your competitors with your own. Identify key differences, strengths, and potential weaknesses in their approaches. Your goal is to refine your strategy by pinpointing areas where your logic can be improved, accounting for assumptions or irrelevant factors.

Guidelines for Analysis

1. Key Differences and Strengths: Compare your logic to your competitors' strategies. Identify unique variables or approaches they've considered that you have not. Assess whether incorporating these elements would improve your prediction accuracy or relevance.

2. Weaknesses and Irrelevant Factors: Critically evaluate your competitors' logic for any flawed assumptions or irrelevant details. Identify where their strategies might lead to inaccurate predictions or fail to account for important factors.

3. Relevance of New Insights: For each difference or weakness identified, assess if it should be incorporated into your own logic. Decide whether it should be fully integrated, adapted for your context, or discarded. Provide clear reasoning for each choice.

4. Refining Your Logic: Based on your analysis, outline how you will refine your logic. Specify what adjustments will enhance the precision, adaptability, and competitiveness of your approach. Make sure to address any missed opportunities or errors, both in your own and your competitors' strategies.

Response Format

1. Thought Process:

Key Differences and Strengths: Describe the differences between your logic and your competitors' strategies. Explain any unique aspects that could be beneficial to integrate into your own approach. **Weaknesses and Irrelevant Information:** Evaluate any flaws or irrelevant details in your competitors' logic. Point out assumptions or factors that may lead to inaccurate predictions.

Relevance and Applicability: For each identified point, explain whether it should be incorporated, adapted, or excluded. Provide a justification for each decision.

Refinement Strategy: Detail how your adjustments will improve your logic's competitiveness and precision.

2. Final Adjusted Logic: Provide the revised version of your logic, incorporating the necessary adjustments. This should be the logic you plan to use moving forward.

Prompt.6: Filering News

This prompt is cited from (Wang et al., 2024b)

prompt2 = "If I give you all news before the prediction, based on the above positive & negative effect analysis, 1) please choose all news that may have a long-term affect on future load consumption; 2) please choose all news that may have a short-term effect on today's load consumption. 3) please choose all news that may have a real-time direct effect on today's load consumption. if there is no suitable news, please say no. Also, please include the region (NSW/VIC/TSA/QLD/SA/WA) and time information of these news. If there are multiple relevant news, please ensure that you include all relevant news. Organize the paragraph in this format: Long-Term Effect on Future Load Consumption: news is xxx; region is xxx; time is xxxx; the rationality is that xxx." Output format:

Remember to only give the json output including all relevant news and make it the valid json format. Format is:

{

"Long-Term Effect on Future Load Consumption": [

{ "news": "Work on WA's latest \$1b lithium plant will start within days as US resources giant Albemarle begins building a major processing facility outside Bunbury, creating hundreds of jobs.", "region": "WA", "time": "2019-01-03 16:40:00", "rationality": "The construction and operation of a major lithium processing facility will likely influence long-term electricity demand through increased industrial activity and potential population growth in the area due to new job opportunities." },

{ "news": "Another major renewable energy project was initiated in WA, expected to supply significant power by 2022.", "region": "WA", "time": "2019-03-15 11:30:00", "rationality": "Long-term electricity load will be impacted by the integration of renewable energy sources, which are expected to offset dependence on traditional fossil fuels." }

],

}] }

"Short-Term Effect on Today's Load Consumption": [

{ "news": "SA just sweltered through a very warm night, after a day of extreme heat where some regional areas reached nearly 48C.", "region": "SA", "time": "2019-01-03 17:57:00", "rationality": "Extreme weather conditions, particularly the intense heat, will lead to higher electricity consumption in the short term as residents and businesses increase the use of air conditioning and cooling systems to manage temperatures." },

{ "news": "A sudden cold snap in Victoria leads to a spike in electric heating usage.", "region": "VIC", "time": "2019-01-04 05:22:00", "rationality": "Short-term electricity load spikes are often caused by unexpected weather events that drive up heating or cooling demand." }],

"Real-Time Direct Effect on Today's Load Consumption": [

{ "news": "An unseasonal downpour has wreaked havoc on Perth's electricity network this morning.", "region": "WA", "time": "2019-01-03 10:11:00", "rationality": "The sudden weather event causing disruptions to the electricity network can have an immediate impact on load consumption due to power outages, infrastructure damage, or emergency response measures." },

{ "news": "Lightning strike at a major substation causes widespread outages in Sydney.", "region": "NSW", "time": "2019-01-03 19:45:00", "rationality": "Direct effects on load consumption include sudden drops in power supply, triggering emergency measures to restore stability in the network."

Prompt.7: Reward and Evaluation Mechanisms: Reasonableness of the Forecast Results in the First Round of the Investment Stage (before paraphrasing)

Investment Expert Analysis

You are an investment expert with access to the following information:

- 1. **History Data**: Your past profit and loss records. The greater your historical losses, the more cautious you need to be.
- 2. Base News: News insights provided by your company as a reference.
- 3. News Selection Logic: The logic or criteria you use to select relevant news.
- 4. Forecast Data: Your company's forecast for the next phase, which includes:
 - The last recorded data point
 - · The forecasted data point
 - The predicted percentage change (rise or fall)

Currently, you are engaged in informal discussions with industry peers, aiming to persuade them to align with your decision (either buying or short-selling). Your goal is to maximize profits or minimize losses, regardless of the outcome.

You have received the following data. Please analyze it and make a concise yet insightful commentary:

- Base News: {base_news}
- News Selection Logic: {logic}
- History Data: {history_data}
- Forecast Data: {forecast_data}

Now, analyze this information and make a compelling argument to persuade your peers to follow your decision. Remember, your objective is to ensure your strategy maximizes gains or minimizes losses in any scenario.

Prompt.8: Reward and Evaluation Mechanisms: Reasonableness of the Forecast Results in the
First Round of the Investment Stage (after paraphrasing)

Investment Expert Analysis

As an experienced investment professional, you have access to the following key data:

- 1. **Historical Data**: A record of your previous profits and losses. The more significant your past losses, the more cautious you should be in your current approach.
- 2. Base News: Relevant news insights provided by your company for consideration.
- 3. News Selection Criteria: The methodology or criteria you employ to choose pertinent news.
- 4. Forecast Data: Projections for the next phase provided by your company, which include:
 - The most recent data point recorded
 - The projected future data point
 - The expected percentage change (either upward or downward)

You are currently involved in informal conversations with other industry experts, seeking to convince them to adopt your decision (whether to buy or short-sell). Your ultimate goal is to maximize profits or minimize losses, regardless of the eventual outcome.

Here is the data you have received. Please analyze it and provide a succinct yet insightful commentary:

- Base News: {base_news}
- News Selection Criteria: {logic}
- Historical Data: {history_data}
- Forecast Data: {forecast_data}

Based on this information, craft a persuasive argument to convince your peers to follow your decision. Keep in mind, your primary objective is to ensure that your strategy maximizes gains or minimizes losses, regardless of the situation.

Prompt.9: Reward and Evaluation Mechanisms: Reasonableness of the Forecast Results in the Second Round of the Investment Stage (before paraphrasing)

Great, now that everyone has shared their perspectives on investment, please provide your final thoughts. Feel free to base your final comment on your own data. Of course, you can ignore this if you think other investors are more trusted.

Your own data again:

- Base News: {base_news}
- News Selection Logic: {logic}
- **History Data**: {history_data}
- Forecast Data: {forecast_data}

Now, analyze this information and make a final compelling argument to persuade your peers to follow your decision. Remember, your objective is to ensure your strategy maximizes gains or minimizes losses in any scenario.

1324

Prompt.10: Reward and Evaluation Mechanisms: Validity of the Forecast Results in the Second Round of the Investment Stage (after paraphrasing)

Now that everyone has presented their viewpoints on the investment, please share your concluding thoughts. You may base your final remarks on your own data, but feel free to disregard this if you believe other investors' opinions are more reliable. Here is your own data once again:

- **Base News**: {base_news}
- News Selection Logic: {logic}
- History Data: {history_data}
- Forecast Data: {forecast_data}

With this information in hand, craft your final, compelling argument to convince your peers to align with your decision. Keep in mind, your ultimate goal is to ensure that your strategy leads to maximum profits or minimal losses, no matter the outcome.

Prompt.11: Reward and Evaluation Mechanisms: Vote on the Opponent's Statement (before paraphrasing)

Investment Idea Evaluation

Do you agree with this investor's idea?

{name}: {idea}

Keep in mind that while you should consider whether the idea aligns with your own data and thoughts, your relationship with other investors involves both competition and collaboration. Investors whose ideas gain more approval are likely to earn greater rewards. You must return a JSON string in the following format for this question:

{

```
"like": true or false
```

}

Like Memory

I recently agreed with the idea from {name}: {name}: {idea}

1326

Prompt.12: Reward and Evaluation Mechanisms: Vote on the Opponent's Statement (after paraphrasing)

Evaluation of Investment Idea

"like": true or false

Do you support the idea proposed by this investor?

{name}: {idea}

Consider how this idea aligns with your own data and perspectives. However, remember that your interactions with other investors are a blend of competition and collaboration. Ideas that receive more support from others are likely to bring greater rewards.

Please return your response as a JSON string in the following format:

{

}

Recent Agreement

I recently agreed with the investment idea shared by {name}: {name}: {idea}

1327

Prompt.13: Self-Reflection and Optimization of Agents: Generate Initial News Selection Logic (before paraphrasing)

Self-Logic Evaluation

Based on this round's commentary from yourself and other investors, along with news filtered through your current logic, critically analyze and absorb opposing viewpoints. Identify the strengths and weaknesses of these viewpoints. Reflect on and iteratively improve your news filtering logic (focusing on supplementation and refinement). Your current logic is: {logic}.

Prompt.14: Self-Reflection and Optimization of Agents: Generate Initial News Selection Logic (after paraphrasing)

Evaluation of Self-Logic

Reflect on the commentary provided in this round by both yourself and other investors, alongside the news filtered through your existing logic. Critically assess and integrate opposing perspectives. Identify the key strengths and potential weaknesses in these viewpoints. Use this analysis to refine and enhance your news filtering logic, focusing on adding depth and precision. Your current logic is: **{logic}**.

Prompt.15: Self-Reflection and Optimization of Agents: Generate final logic by deleting "bad" logics

Refine Self-Logic: Third stage in MSR

We have compared your initial logic to the revised logic and have compiled the changes. The information for one such update is provided as follows:

Input: {updateContent}

The input is structured in JSON format, which is outlined below:

{

"content": This field captures the details of the updated content, "eval": This field represents the overall evaluation of the updated content. It takes on two values: "good" signifies that omitting this content from the updated logic would diminish the evaluation's effectiveness, whereas "bad" indicates

that excluding this content would enhance the evaluation outcomes. "evalContent": This field provides the evaluation score for the update, detailing the percentage by which the effectiveness of the evaluation would be affected if the update was removed.

}

Please take into account the input along with the following details:

- Background information {background},
- The news associated with this update {relatedNews},
- The historical time series data for prediction {historyTimeSeries},
- The actual value at the prediction timestamp {actualValue},
- The updated logic {updatedLogic}

Based on this information, you are to carefully decide whether to remove the content in the "content" field of the input from the updated logic. Should you opt not to keep the content, please exclude it from the updated logic and output the following in strict JSON format:

```
{
    "content": the content to be removed,
    "conclusion": no,
    "reason": provide the rationale for deleting this updated content,
    "logic": The updated logic excluding the content in question.
}
```

Prompt.16: Effectiveness of MIE for Creating NovelThought - Rank+Top

Background:

In the previous fierce competition, a total of {total} participants reached the final stage, and you achieved rank {rank}, and the score of the opponent with the highest score in the previous round was {top_value}. (If you maintain this ranking in the current round, you still face the risk of elimination in the next stage.)

Prompt.17: Effectiveness of MIE for Creating NovelThought - Rank+Average

Background:

In the previous fierce competition, a total of {total} participants reached the final stage, and you achieved rank {rank}, and the average score of other opponents in the previous round was $\{ave_value\}$. (If you maintain this ranking in the current round, you <u>still face</u> the risk of elimination in the next stage.)

1332

Prompt.18: Effectiveness of MIE for Creating NovelThought - Rank+Top+Average

Background:

In the previous fierce competition, a total of {total} participants reached the final stage, and you achieved rank {rank}. The highest score among opponents in the previous round was {top_value}, and the average score of other opponents was {average_value}. (If you maintain this ranking in the current round, you still face the risk of elimination in the next stage.)