InvestmentHelper-AI

Progress Meeting 4

Emir Şahin Dilli
M. Onur Özdemir
Hakan Muluk

Self-RAG

ICLR 2024 Conference Paper

Algorithm 1 SELF-RAG Inference

Require: Generator LM $\mathcal{M}$, Retriever $\mathcal{R}$, Large-scale passage collections $\{d_1, \ldots, d_N\}$

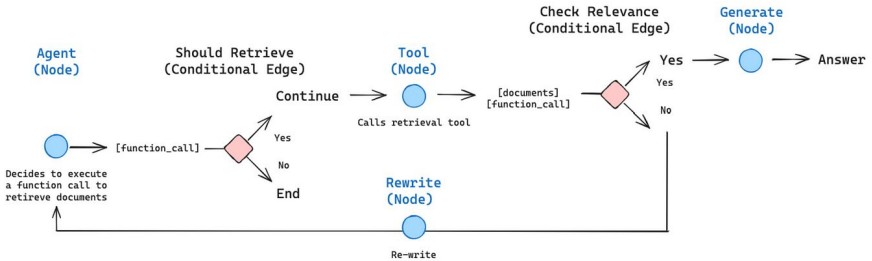
Input: input prompt x and preceding generation $y_{<t}$, **Output:** next output segment y_t
$\mathcal{M}$ predicts Retrieve given $(x, y_{<t})$
if Retrieve == Yes **then**
  Retrieve relevant text passages $\mathbf{D}$ using $\mathcal{R}$ given (x, y_{t-1}) ▷ Retrieve
  $\mathcal{M}$ predicts IsRel given x, d and y_t given $x, d, y_{<t}$ for each $d \in \mathbf{D}$ ▷ Generate
  $\mathcal{M}$ predicts IsSup and IsUse given x, y_t, d for each $d \in \mathbf{D}$ ▷ Critique
  Rank y_t based on IsRel, IsSup, IsUse ▷ Detailed in Section 3.3
else if Retrieve == No **then**
  $\mathcal{M}_{gen}$ predicts y_t given x ▷ Generate
  $\mathcal{M}_{gen}$ predicts IsUse given x, y_t ▷ Critique

Type	Input	Output	Definitions
Retrieve	x / x, y	{yes, no, continue}	Decides when to retrieve with $\mathcal{R}$
IsRel	x, d	{**relevant**, irrelevant}	d provides useful information to solve x.
IsSup	x, d, y	{**fully supported**, partially supported, no support}	All of the verification-worthy statement in y is supported by d.
IsUse	x, y	{**5**, 4, 3, 2, 1}	y is a useful response to x.

- Reduce hallucination, which might be crucial for a financial chatbot
- Might be a solution for “Lost in the Middle” problem
- Trade-off between robustness and time

Adaptation of SelfRAG for InvestmentHelper-AI

Example LangGraph Flow:

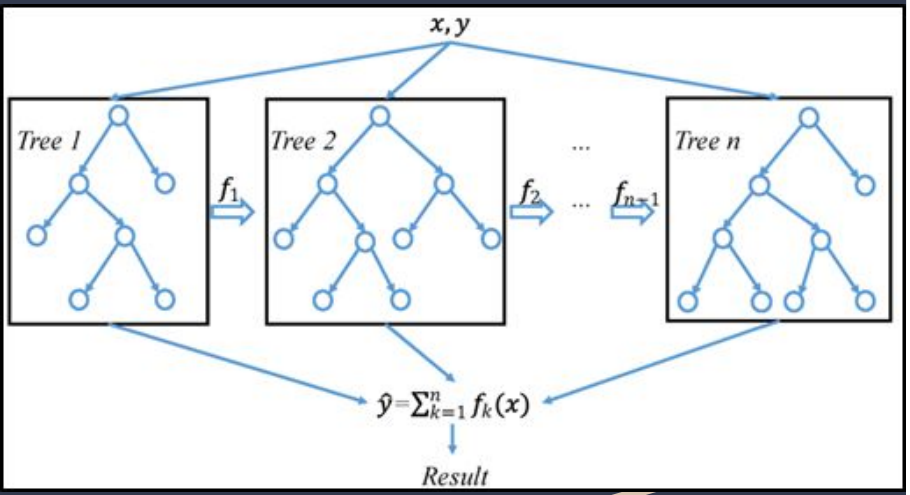


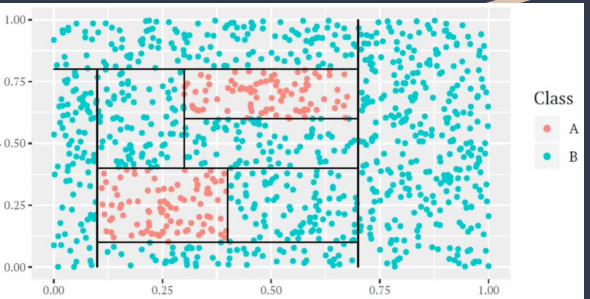
- Relevancy check can be done in parallel
- Instead of fine-tuning an LLM, we can use another LLM for these checks to improve the performance
- Easily implementable using LangGraph framework.
- LangGraph is a framework for building graph-based workflows for LLMs, where nodes represent tasks (e.g., retrieval, generation, validation) and edges define the flow of execution.
- Might be integrated to our chatbot as an option that user selects to get more accurate answers but answer generation will take longer

Stock Price Prediction

- Several different time-series prediction methods have been investigated for predicting NVIDIA stock prices(NVIDIA is selected as a case study to explore different methods).
- Relatively small dataset(6245 data instances)
- 90%-10% training and test sets
- Also aggregated Nasdaq data
- New features are added such as Relative Strength Index, Williams %R indicator)

Adaptive XGBOOST





- Modification to original XGBOOST algorithm for handling time series data
- Adaptive XGBoost uses a queue of trees that are added using gradient boosting
- Old trees(that are created with old data) are removed and new trees are added to the queue
- It is used for handling concept drifts-distribution shifts that occur in data
- Reached %56 accuracy and 0.64 F1 score
- Can be improved further by Dataset Augmentation

Confusion Matrix of Adaptive XGBOOST

	Response = 1	Response = 0
Prediction = 1	244	161
Prediction = 0	115	105

Notification System

User 1: “Notify me when A(1) occurs.”

User 2: “Notify me when A(2) occurs.”

User 3 “Notify me when A(3) occurs.”

.

.

.

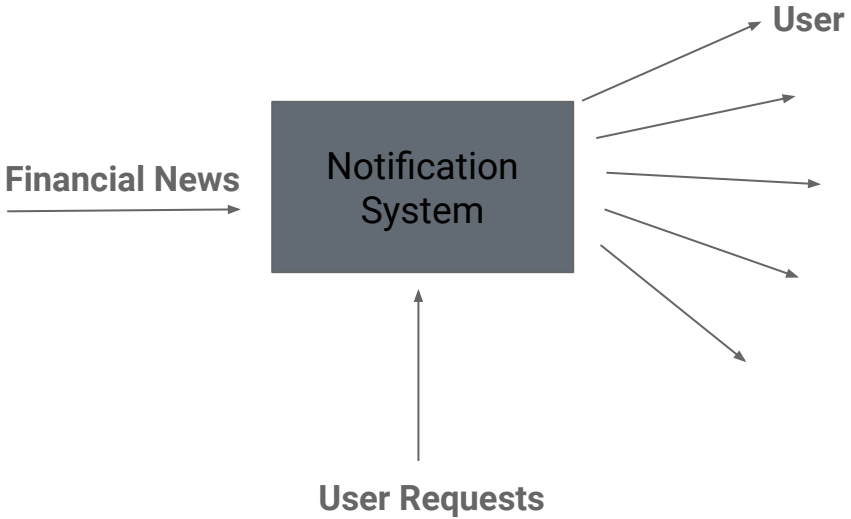
.

User N: “Notify me when A(n) occurs.”

Each user can one or more notification requests in the system.

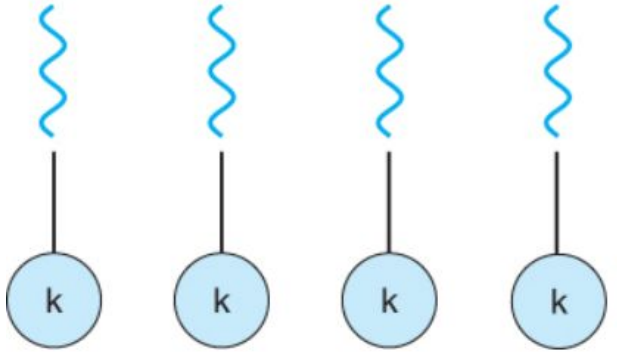
Complexity

The system processes financial news documents that have been collected, and, according to the each user's specific request, it redistributes the collected information.



Potential Problem

If each request is compared with the bulk financial news document separately, then, the complexity grows linearly with the number of requests, O(n).



One-to-one processing

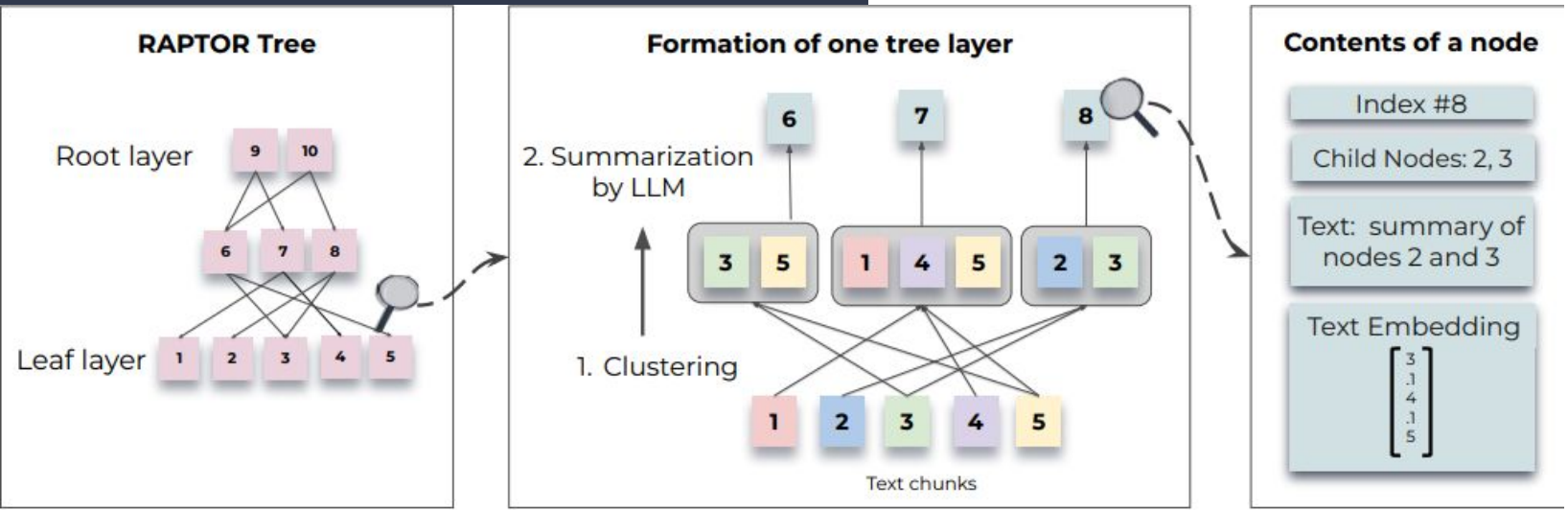
This is not ideal.

RAPTOR

(ICLR 2024 Conference Paper)

Tree construction process:

RAPTOR recursively clusters chunks of text based on their vector embeddings and generates text summaries of those clusters, constructing a tree from the bottom up. Nodes clustered together are siblings; a parent node contains the text summary of that cluster.



Coalitional Structures

Assumption:

Inherent Distribution of Requests: $\{C_1, \ldots, C_m\}$

A coalition is any nonempty subset $S \subseteq \mathcal{N}$. A coalition structure (or partition) $\mathcal{P}$ of $\mathcal{N}$ is a set of disjoint coalitions $\{S_1, S_2, \ldots, S_k\}$ such that $\bigcup_{\ell=1}^{k} S_\ell = \mathcal{N}$ and $S_i \cap S_j = \emptyset$ for all $i \neq j$.

We might assume incoming requests are coming from an inherent distribution, and hence try to assign each request to a coalition.

This way, number of processes is constant with k.

Log in

Don't have an acount? Sign up

Continue with Google

Continue with Facebook

Continue with Apple

Or continue with email

Email address or user name

Password Hide

Forget your password

Remember me

Log in

Create an account

Already have an ccount? Log in

First name

Last name

Email address

Password

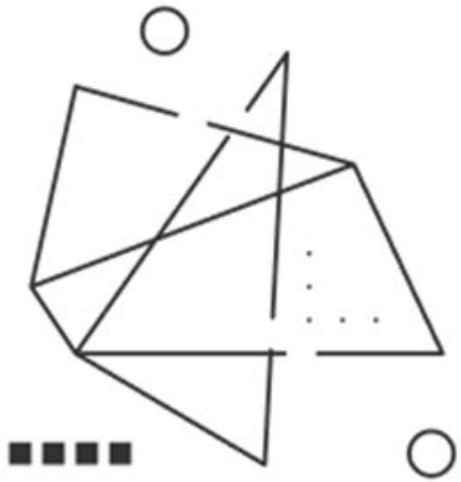
Confirm your password

Use 8 or more characters with a mix of letters, numbers & symbols

Show password

log in instead

Create an account

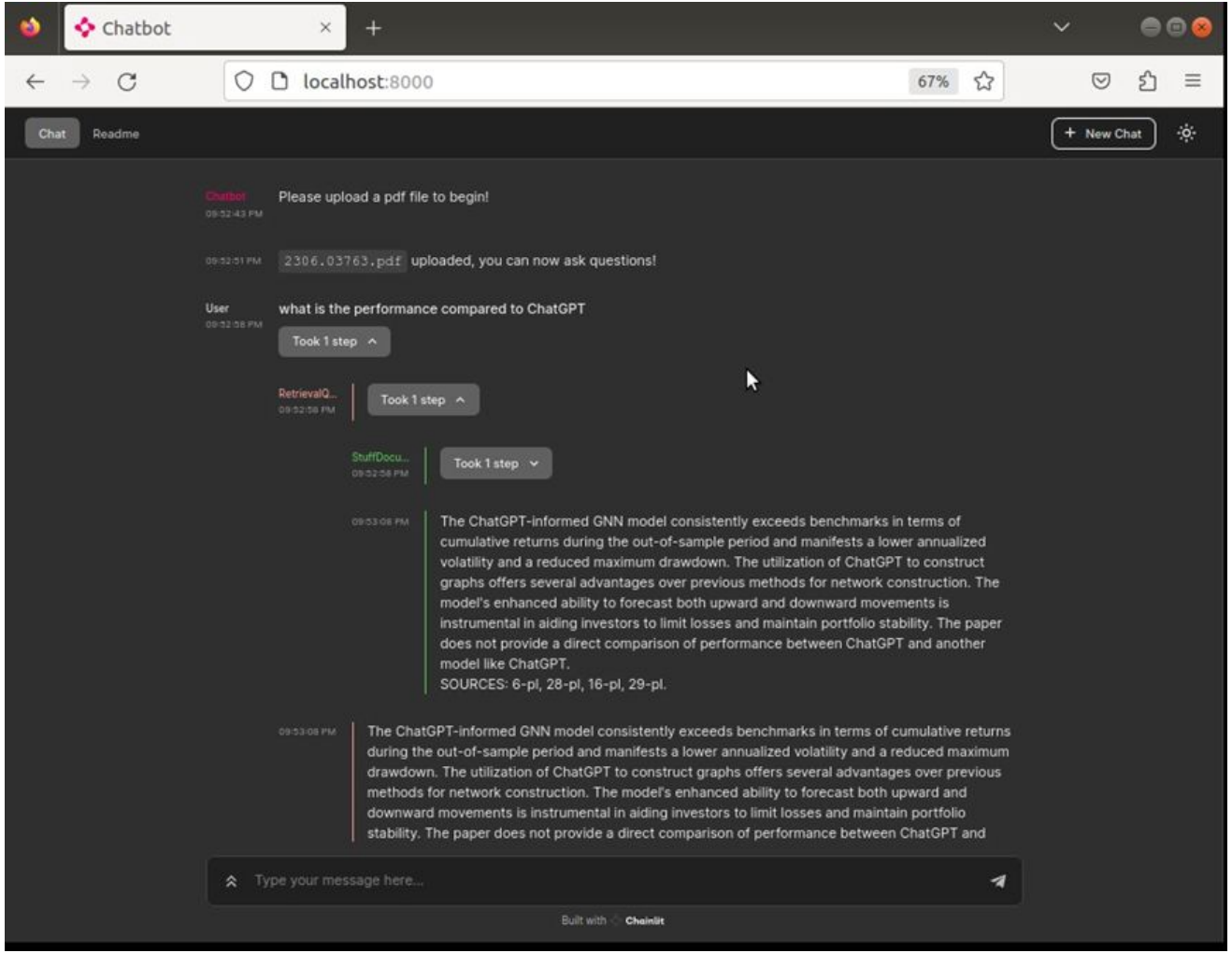


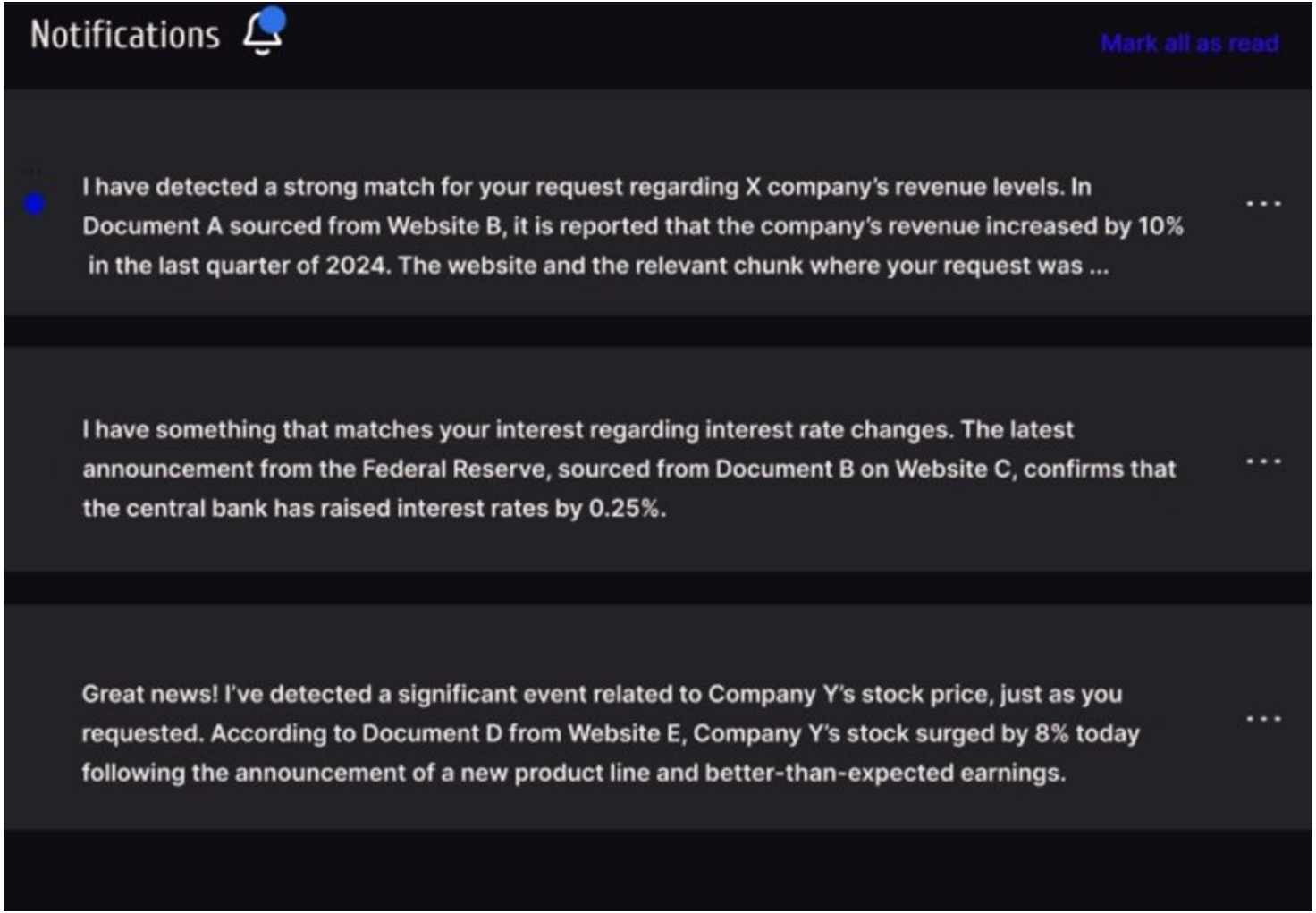
Companies | Sectors | Industries | Trending | Gainers & Losers | Most Visited

☆	#	Company	Price	1D %	1M %	YTD %	M Cap
★	1	Apple AAPL	$207.23	+0.36%	-1.48%	+16.06%	$3.22 T
☆	2	Microsoft MSFT	$417.14	+0.18%	-1.11%	+9.42%	$3.00 T
☆	3	NVIDIA NVDA	$123.61	+0.18%	+12.00%	+183.00%	$2.928 T
☆	4	Alphabet GOOG	$104.25	+0.05%	+1.6%	+21.51%	$2.021 T
☆	5	Amazon AMZN	$161.93	+0.24%	+6.96%	+31.3%	$1.873 T
☆	6	Meta Platforms META	$494.09	+1.02%	-3.85%	+61.72%	$1.318 T
☆	7	Berkshire Hathaway BRK-B	$444.51	+0.24%	-3.66%	+24.79%	$1.025 T
☆	8	TSMC TSM	$174.54	+0.05%	+8.06%	+88.23%	$890.44 B
☆	9	Eli Lilly LLY	$165.72	-0.78%	-9.14%	+38.28	$960.02 B

Add to watchlist


FinanceHelper AI
Markets
Chatbot
Settings
Indicators
Tools
Watchlist
Alerts
120.00
118.00
116.00
114.00
112.00
110.00
108.00
106.00
104.00
102.00
100.00
98.00
96.56
96.00
94.00
92.00
90.00
88.00
86.00
2023
5
9
13
17
21
25
Jul
5
9
13
17


Chatbot
localhost:8000
67%
Chat
Readme
+ New Chat
Chatbot
09:52:43 PM
Please upload a pdf file to begin!
09:52:51 PM
2306.03763.pdf uploaded, you can now ask questions!
User
09:52:58 PM
what is the performance compared to ChatGPT
Took 1 step
RetrievalQ...
09:52:58 PM
Took 1 step
StuffDocu...
09:52:58 PM
Took 1 step
09:53:08 PM
The ChatGPT-informed GNN model consistently exceeds benchmarks in terms of cumulative returns during the out-of-sample period and manifests a lower annualized volatility and a reduced maximum drawdown. The utilization of ChatGPT to construct graphs offers several advantages over previous methods for network construction. The model's enhanced ability to forecast both upward and downward movements is instrumental in aiding investors to limit losses and maintain portfolio stability. The paper does not provide a direct comparison of performance between ChatGPT and another model like ChatGPT.
SOURCES: 6-pl, 28-pl, 16-pl, 29-pl.
09:53:08 PM
The ChatGPT-informed GNN model consistently exceeds benchmarks in terms of cumulative returns during the out-of-sample period and manifests a lower annualized volatility and a reduced maximum drawdown. The utilization of ChatGPT to construct graphs offers several advantages over previous methods for network construction. The model's enhanced ability to forecast both upward and downward movements is instrumental in aiding investors to limit losses and maintain portfolio stability. The paper does not provide a direct comparison of performance between ChatGPT and
Type your message here...
Built with Chainlit


Notifications
Mark all as read
I have detected a strong match for your request regarding X company's revenue levels. In Document A sourced from Website B, it is reported that the company's revenue increased by 10% in the last quarter of 2024. The website and the relevant chunk where your request was ...
I have something that matches your interest regarding interest rate changes. The latest announcement from the Federal Reserve, sourced from Document B on Website C, confirms that the central bank has raised interest rates by 0.25%.
Great news! I've detected a significant event related to Company Y's stock price, just as you requested. According to Document D from Website E, Company Y's stock surged by 8% today following the announcement of a new product line and better-than-expected earnings.