

PIXIU: MÔ HÌNH NGÔN NGỮ LỚN, DỮ LIỆU HƯỚNG DẪN VÀ TIÊU CHUẨN ĐÁNH GIÁ TÀI CHÍNH

Nhóm 11



Thành viên

Đặng Đức Duy

Trịnh Hoàng Đức

Hoàng Mạnh Hùng

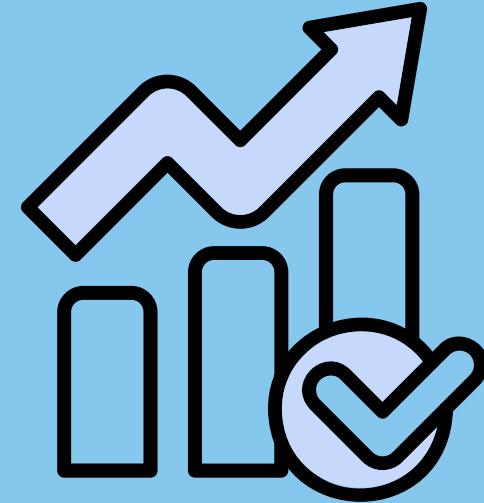
Phạm Trung Hiếu

Giới thiệu chung



Nhận xét chung

- Mặc dù vậy các mô hình ngôn ngữ lớn (LLMs) đã thể hiện hiệu suất vượt trội trong lĩnh vực tài chính
- Chưa có các mô hình LLM tài chính được công khai, bộ dữ liệu hướng dẫn và bộ tiêu chuẩn đánh giá chuyên biệt cho tài chính



Thúc đẩy sự phát triển mã nguồn mở của trí tuệ nhân tạo (AI) trong tài chính

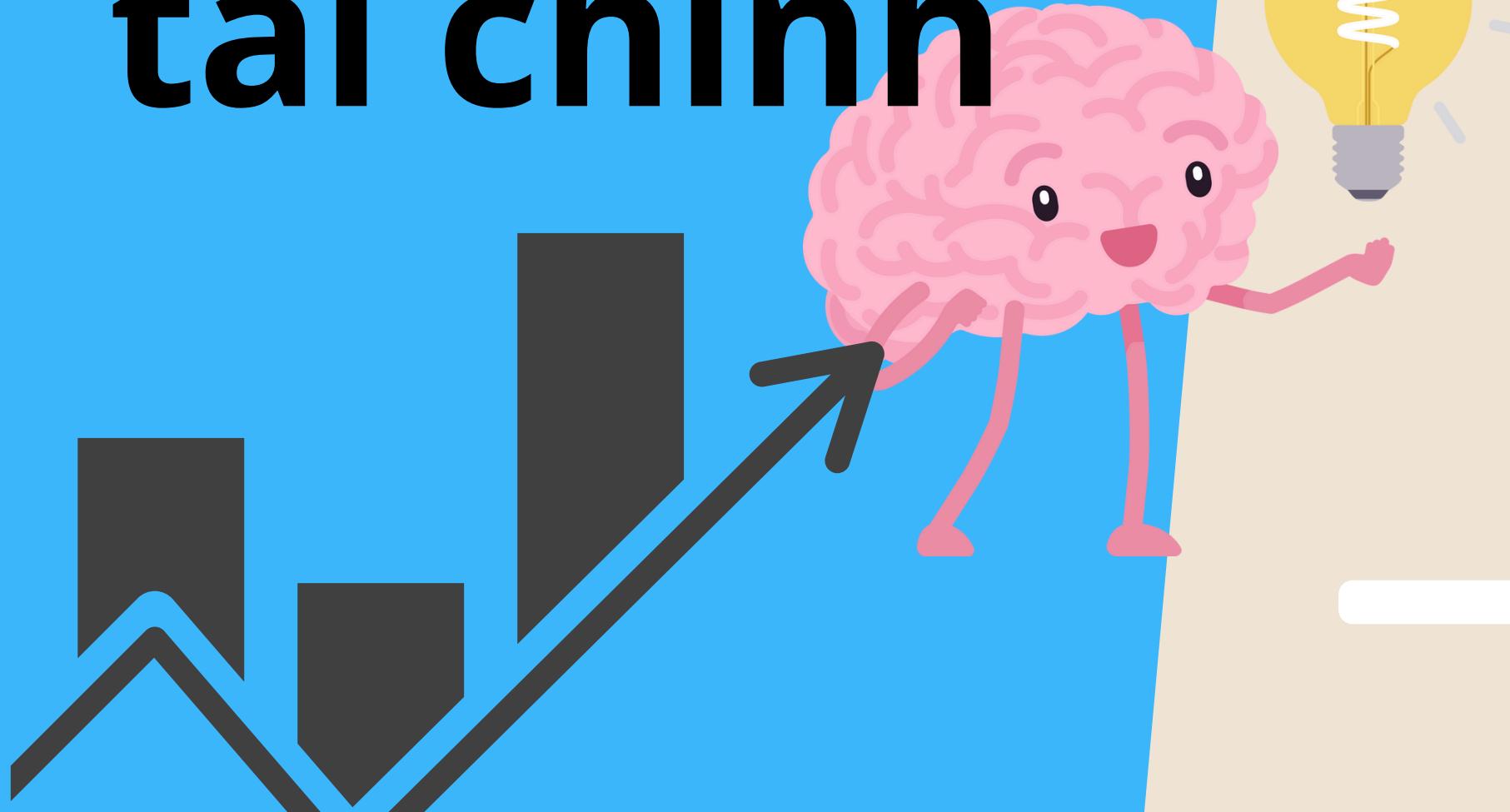
Vấn đề hiện tại với LLM tài chính



Bảng so sánh các mô hình ngôn ngữ được tiền huấn luyện (PLMs) và các mô hình ngôn ngữ lớn (LLMs) trong lĩnh vực tài chính

Model	Backbone	Size	Open Source Model	Open Source Data	Instruct	Language	Evaluation NLP	Evaluation Fin	Release Date
finBERT (Araci, 2019)	BERT	110M	✓	✓	✗	English	✓	✗	08/27/19
FinBERT (Yang et al., 2020)	BERT	110M	✓	✗	✗	English	✓	✗	06/15/20
Mengzi-fin (Zhang et al., 2021)	RoBERTa	103M	✓	✗	✗	Chinese	✓	✗	10/13/21
FLANG (Shah et al., 2022)	ELECTRA	110M	✓	✓	✗	English	✓	✗	10/31/22
BBT-FinT5 (Lu et al., 2023)	T5	220M	✓	✓	✗	Chinese	✓	✗	02/18/23
BloombergGPT (Wu et al., 2023)	BLOOM	50B	✗	✗	✗	English	✓	✗	03/30/23
FinMA	LLaMA	7/30B	✓	✓	✓	English	✓	✓	06/01/23

Vấn đề hiện tại với LLM tài chính



1. BloombergGPT không mở mã và không cung cấp dữ liệu huấn luyện.
2. Các mô hình hiện tại chưa hỗ trợ tinh chỉnh.
3. Thiếu dữ liệu hướng dẫn tài chính và bộ đánh giá toàn diện cho LLMs trong ngành



Thúc đẩy sự phát triển mã nguồn mở của trí tuệ nhân tạo (AI) trong tài chính

PIXIU

Giải pháp PIXIU - Là mô hình LLM tài chính đầu tiên được tinh chỉnh từ LLaMA với dữ liệu hướng dẫn và bộ đánh giá



PIXIU

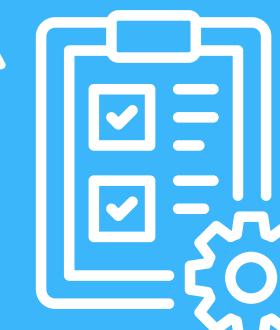


FinMA



Mô hình LLM tài chính

128.000 mẫu dữ liệu



FLARE

8 tác vụ tài chính

Đặc điểm PIXIU



Open Source

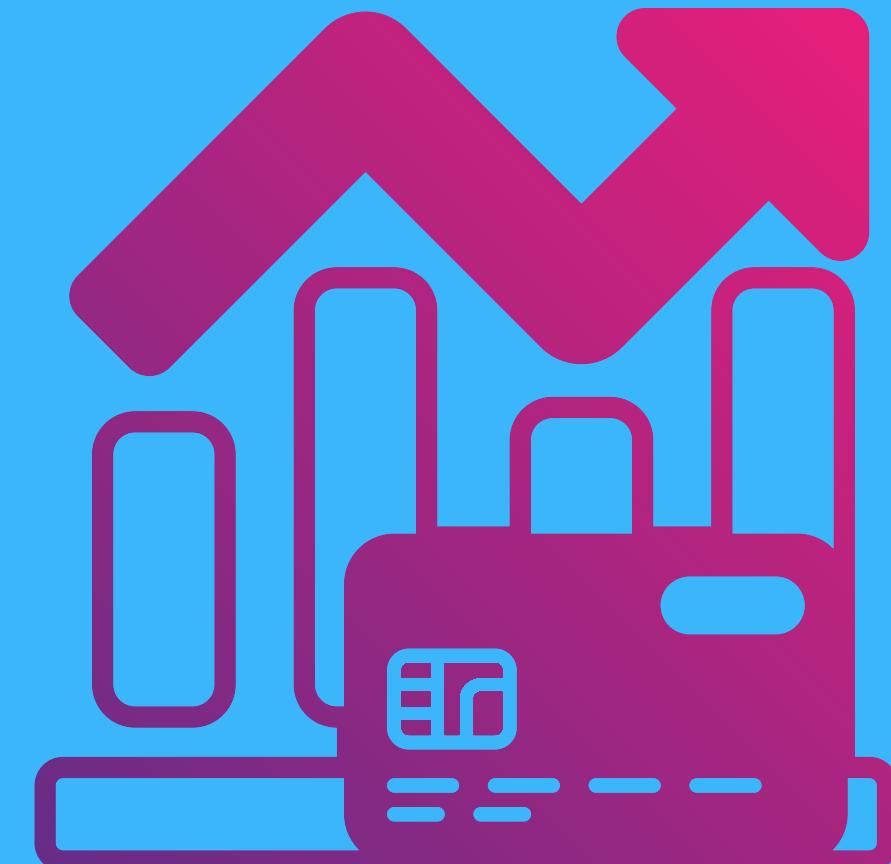


Mã nguồn mở



Đa phương thức

Đa tác vụ



Đa dạng

Các tác vụ

 Phân tích cảm xúc tài chính

 Phân loại Hawkish – Dovish

 Phân loại tiêu đề tin tức

 Tóm tắt văn bản tài chính

 Nhận diện thực thể có tên

 Dự đoán biến động chứng khoán

? Hỏi – đáp tài chính (QA)

 Xếp hạng tín dụng

Đóng góp chính của nghiên cứu

1. FIT – Bộ dữ liệu chỉ dẫn tài chính đầu tiên
2. FLARE – Bộ đánh giá toàn diện đầu tiên cho LLM tài chính
3. FinMA – Mô hình LLM tài chính đầu tiên mã nguồn mở có khả năng làm theo chỉ dẫn
4. Đánh giá toàn diện các mô hình LLM hiện có trong lĩnh vực tài chính

Dữ liệu và mô hình



Các nghiên cứu liên quan



01

Mô hình ngôn ngữ tài chính
(Financial Language Models)

02

Bộ đánh giá dữ liệu trong tài chính
(Financial Evaluation Benchmark)

03

Các mô hình ngôn ngữ lớn mã nguồn mở
(Open Sourced Large Language Models)

Mô hình ngôn ngữ tài chính (Financial Language Models)

2019

Araci đề xuất mô hình
PLM tài chính đầu tiên
FinBERT

Huấn luyện mô hình BERT
với các tập dữ liệu tài
chính được công bố công
khai như TRC2-financial và
Financial Phrase Bank

2020

Yang et al đề xuất mô
hình **FinBERT** khác

Huấn luyện BERT với tập
dữ liệu giao tiếp tài chính
gồm 4,9 tỷ từ

2022

Shah et al. đã đề xuất
FLANG

Một mô hình PLM tài chính
dựa trên BERT và ELECTRA
(Clark et al., 2020).

2023

Wu et al. đã giới thiệu
BloombergGPT

Mô hình ngôn ngữ lớn đầu
tiên cho lĩnh vực tài chính
với 50 tỷ tham số

Bộ đánh giá dữ liệu trong tài chính (Financial Evaluation Benchmark)

2022

Đề xuất bộ đánh giá hiệu suất không đồng nhất đầu tiên tên là
FLUE

- 01 Phân tích cảm xúc tài chính
- 02 Phân loại tiêu đề tin tức
- 03 Nhận diện thực thể có tên
- 04 Phát hiện ranh giới cấu trúc
- 05 Hỏi đáp

Phân tích cảm xúc tài chính

Nghiên cứu
tổng quát
về cảm xúc
của nhà
đầu tư và
cảm xúc
trong văn
bản tài
chính



Mục tiêu chính:

Phân loại cảm xúc theo các nhãn: Positive (Tích cực), Negative (Tiêu cực). Neutral (Trung tính)



Chi tiết về bộ dữ liệu:

- Gồm hơn 4,800 câu
- Gán nhãn thủ công bởi các chuyên gia
- Mỗi câu có thể được gán nhãn bởi nhiều người
- Dựa trên mức độ đồng thuận khác nhau (ví dụ: 75%, 90%...).



Ứng dụng:

- Huấn luyện mô hình phân tích cảm xúc tài chính (FinBERT, FinMA, v.v.)
- Giao dịch chứng khoán tự động
- Phân tích tác động lên thị trường
- Nghiên cứu tâm lý nhà đầu tư



Ưu điểm:

- Độ chính xác cao do chuyên gia gán nhãn
- Tập trung vào ngữ cảnh cảm xúc tài chính
- Công khai – rất hữu ích cho nghiên cứu

Phân loại tiêu đề tin tức

Tập trung vào
việc phân tích
nội dung tiêu
đề tin tức tài
chính để xác
định thông tin
quan trọng
liên quan đến
chuyển động
giá tài sản

Thông tin dữ liệu:

- Nguồn dữ liệu: Bao gồm hơn 11.000 tiêu đề tin tức
- Mỗi tiêu đề được chú thích thủ công bởi chuyên gia, mỗi tiêu đề có thể có nhiều nhãn.

Các nhãn phân loại:

-  1. Giá có được đề cập không (price or not)
- 2. Giá tăng (price up)
- 3. Giá giảm (price down)
- 4. Giá ổn định (price stable)
- 5. Giá trong quá khứ (past price)
- 6. Giá trong tương lai (future price)
- 7. Thông tin chung trong quá khứ (past general)
- 8. Thông tin chung trong tương lai (future general)
- 9. So sánh tài sản (asset comparison)

Ứng dụng:

- Huấn luyện mô hình phân tích cảm xúc tài chính (FinBERT, FinMA, v.v.)
- Giao dịch chứng khoán tự động
- Phân tích tác động của tin tức lên thị trường
- Nghiên cứu tâm lý nhà đầu tư

Nhận diện thực thể có tên

Có nhiệm vụ
trích xuất các
thực thể có
tên từ văn
bản, đặc biệt
là các văn bản
tài chính như
người, tổ chức,
địa điểm



Nguồn dữ liệu:

- Được thu thập từ các tài liệu pháp lý tài chính, công khai
- Loại văn bản: Văn bản hợp đồng pháp lý
- Thực thể chú thích: Tên người (PER), tổ chức (ORG) và địa điểm (LOC) được chuyên gia gán nhãn thủ công.



Ý nghĩa và ứng dụng:

- Xây dựng đồ thị tri thức tài chính
- Tìm kiếm chính xác hơn
- Tiền xử lý cho các tác vụ tài chính khác
- Tăng độ chính xác khi phân tích tài liệu hợp đồng

Phát hiện ranh giới cấu trúc

Xác định các ranh giới ngữ nghĩa hoặc định dạng trong văn bản, giúp tách nhãn

Phục vụ cho các tác vụ khác như tóm tắt, trích xuất thông tin, truy vấn hỏi-đáp, phân tích cảm xúc



Cách xử lý:

- Dựa vào mẫu định dạng (các tiêu đề in đậm, viết hoa, hoặc theo số thứ tự)
- Phù hợp với dữ liệu chuẩn hóa như PDF chuyển sang XML
- Mỗi đoạn văn được coi là một "token lớn"
- Gán nhãn cho từng đoạn như: HEADER, BODY, TABLE, FOOTNOTE, APPENDIX, v.v.
- Kết hợp ngữ nghĩa và thông tin định dạng/thị giác (font, layout, box...)
- Phù hợp cho văn bản như PDF, ảnh scan



Ứng dụng thực tế trong tài chính:

- Tự động phân tích báo cáo tài chính PDF
- Tóm tắt từng phần
- Tìm kiếm nội dung có cấu trúc(tóm tắt,báo cáo
- Tiền xử lý cho các mô hình như FinBERT, GPT(tránh sai ngữ cảnh)

Hỏi đáp

Nhận đầu vào là một câu hỏi và một đoạn văn bản ngũ cảnh, sau đó trả lời câu hỏi dựa trên nội dung của đoạn văn đó



Đặc điểm chính:

- Bộ dữ liệu bao gồm các đoạn văn bản từ tài liệu tài chính
- Mỗi đoạn có kèm câu hỏi và câu trả lời đúng
- Các câu hỏi liên quan đến: Doanh thu, mức tăng trưởng, chi tiết về công ty (tên, trụ sở, ngành,...), chiến lược tài chính, rủi ro, dự báo
- Các loại câu hỏi: Factual (thực tế), reasoning(suy luận), yes/no (đúng /sai), list-based (liệt kê)



Mục tiêu để đánh giá khả năng:

- Truy xuất thông tin đúng (Information Retrieval)
- Hiểu và suy luận tài chính (Financial Reasoning)
- Tóm tắt ý nghĩa từ câu hỏi yêu cầu (Answer Generation)



Ứng dụng thực tế trong tài chính:

- Chatbot tư vấn tài chính: trả lời câu hỏi nhà đầu tư từ báo cáo.
- Tìm kiếm có ngũ nghĩa: "tìm đoạn nào nói về rủi ro lạm phát?"
- Phân tích tài chính tự động: kết hợp QA để trích xuất insight từ báo cáo doanh nghiệp.

Bộ đánh giá dữ liệu trong tài chính (Financial Evaluation Benchmark)

2023

BBT-CFLEB được giới thiệu

- 01 Phân loại tin tức tài chính
- 02 Tóm tắt
- 03 Trích xuất quan hệ
- 04 Hỏi đáp
- 05 Xác định tin tức tiêu cực
- 06 Phân loại cảm xúc từ các văn bản mạng xã hội tài chính

Mô hình ngôn ngữ lớn mã nguồn mở (Open Sourced Large Language Models)

LLaMA một mô hình ngôn ngữ lớn mã nguồn mở (LLM) với số lượng tham số từ 7 tỷ, 13 tỷ đến 65 tỷ.



LLaMA 13B



GPT-3

FIT: Tập dữ liệu
điều chỉnh theo chỉ
dẫn trong lĩnh vực
tài chính



Dữ liệu gốc

1. Được gán nhãn bởi chuyên gia trong ngành, đảm bảo chất lượng cao.
2. Chi phí thấp và không bị hạn chế quyền sử dụng thương mại
3. Bao phủ nhiều loại văn bản khác nhau như tin tức, báo cáo, bài đăng mạng xã hội, và cả các dữ liệu đa phương thức như chuỗi thời gian, bảng số liệu và văn bản.

Data	Task	Raw	Instruction	Data Types	Modalities	License
FPB	sentiment analysis	4,845	48,450	news	text	CC BY-SA 3.0
FiQA-SA	sentiment analysis	1,173	11,730	news headlines,tweets	text	Public
Headlines	news headline classification	11,412	11,412	news headlines	text	CC BY-SA 3.0
FOMC	hawkish-dovish classification	496	496	FOMC transcripts	text	CC BY-NC 4.0
NER	named entity recognition	609	6,090	financial agreements	text	CC BY-SA 3.0
FiNER-ORD	named entity recognition	1,080	1,080	news articles	text	CC BY-SA 3.0
FinQA	question answering	8,281	8,281	earnings reports	text,table	MIT License
ConvFinQA	question answering	3,458	3,458	earnings reports	text,table	MIT License
ECTSum	text summarization	495	495	earning call transcripts	text	Public
EDTSum	text summarization	2,000	2,000	news articles	text	Public
BigData22	stock movement prediction	7,168	7,168	tweets,historical prices	text,time series	Public
ACL18	stock movement prediction	27,080	27,080	tweets,historical prices	text,time series	MIT License
CIKM18	stock movement prediction	4,971	4,971	tweets,historical prices	text,time series	Public
German	credit scoring	1,000	1,000	credit records	table	CC BY 4.0
Australia	credit scoring	690	690	credit records	table	CC BY 4.0

Các tác vụ

Phân tích cảm xúc tài chính

Nhằm phân tích thái độ tích cực/tiêu cực/trung lập từ các văn bản tài chính. Sử dụng 2 tập:

- FPB: câu tiếng Anh từ tin tức tài chính (nhãn: positive, negative, neutral).
- FiQA-SA: dữ liệu từ tin tức và microblog, cảm xúc trên thang [-1, 1].

Phân loại tiêu đề tin tức

Dự đoán các thẻ như: tăng giá, giảm giá, ổn định, giá quá khứ, giá tương lai, so sánh tài sản,... từ các tiêu đề tin tức về vàng.

Nhận diện thực thể có tên

Nhận dạng các thực thể như tên người (PER), tổ chức (ORG), địa điểm (LOC)... từ hợp đồng tài chính và bài báo.

Các tác vụ

Phân loại Hawkish – Dovish

Phân tích thái độ trong văn bản chính sách tiền tệ.

- Hawkish: xu hướng thắt chặt tiền tệ (chống lạm phát).
- Dovish: xu hướng nới lỏng (hỗ trợ tăng trưởng).
- Sử dụng tập từ biên bản họp FOMC.

? Hỏi – đáp tài chính (QA)

Từ báo cáo tài chính doanh nghiệp (S&P 500), tạo cặp hỏi – đáp để huấn luyện mô hình trả lời câu hỏi tài chính.

Tóm tắt văn bản tài chính

- ECTSum: tóm tắt dạng gạch đầu dòng từ bản ghi cuộc họp tài chính.
- EDTSum: tóm tắt tiêu đề cho bài báo tài chính (dạng rút gọn tự động – abstractive)

Các tác vụ

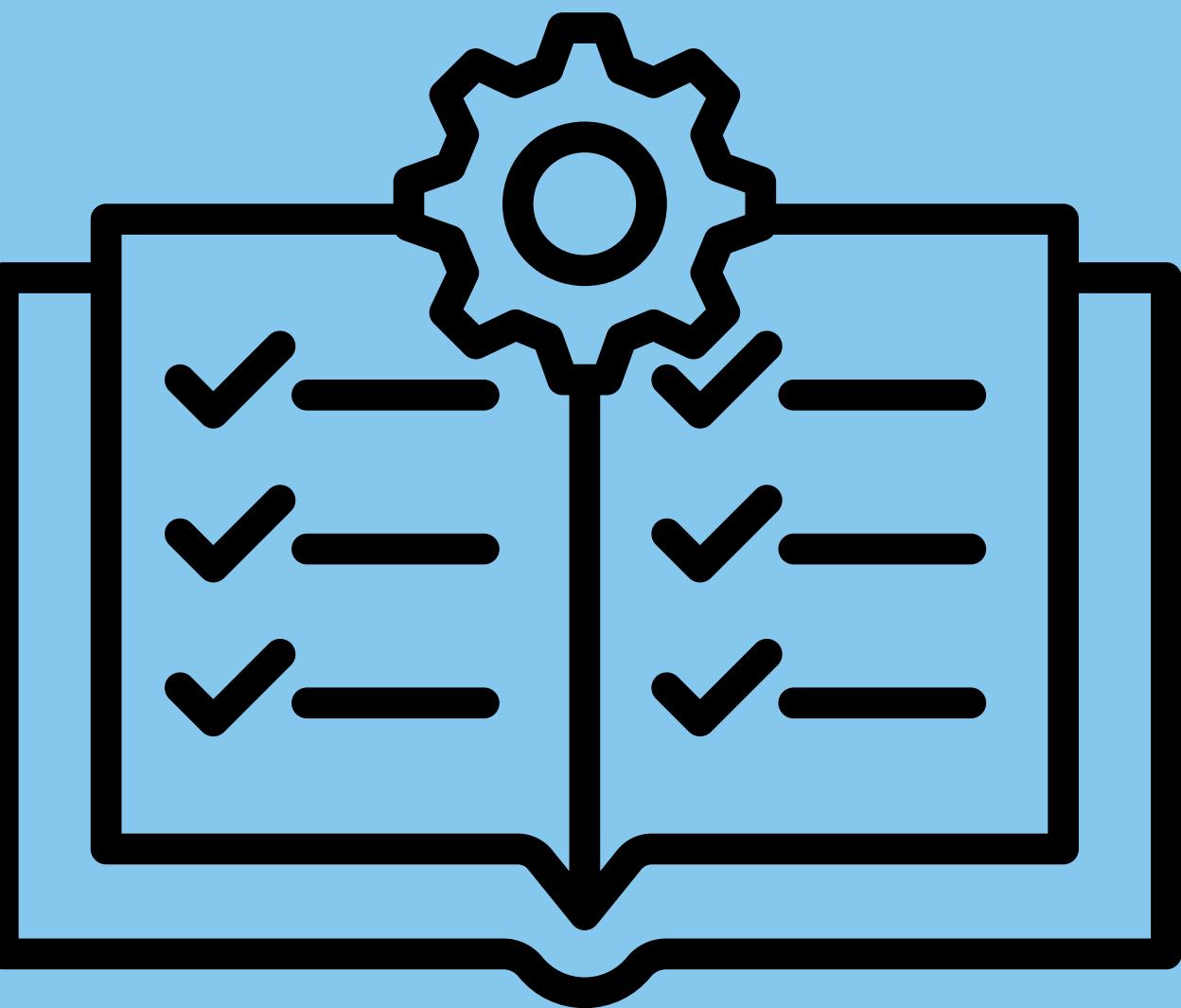
Dự đoán biến động chứng khoán

Kết hợp thông tin từ tweet + chuỗi thời gian giá cổ phiếu lịch sử để dự đoán chiều hướng biến động (tăng, giảm).

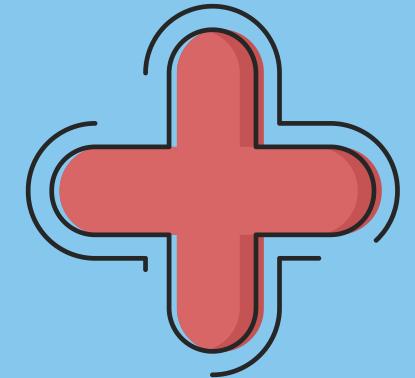
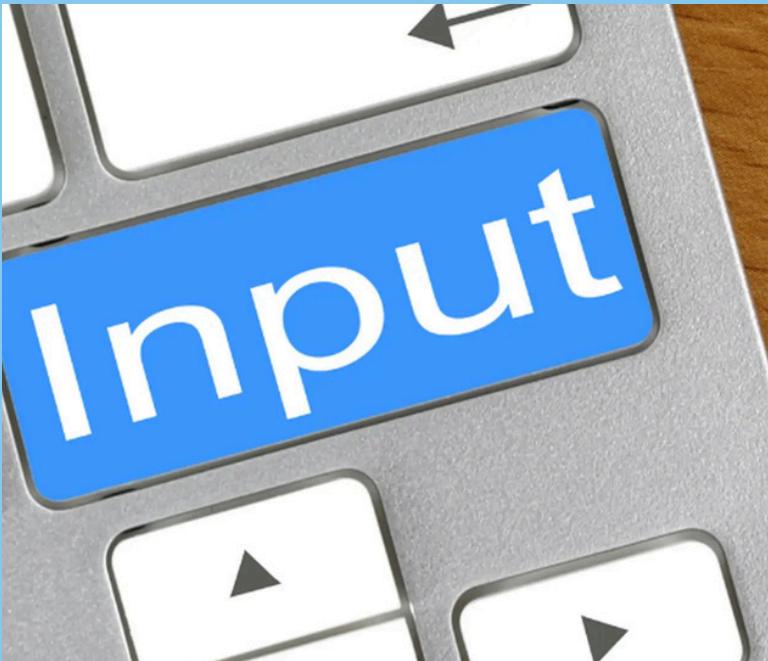
Xếp hạng tín dụng

Dự đoán mức độ tín dụng (tốt/xấu) dựa trên các thuộc tính tài chính như thu nhập, tuổi, lịch sử vay nợ,... từ bảng dữ liệu.

Xây dựng chỉ dẫn



Dữ liệu huấn luyện



FinMA: Mô hình Ngôn ngữ Lớn trong Tài chính

- FinMA được tinh chỉnh từ LLaMA bằng bộ dữ liệu FIT
- Có 4 phiên bản mô hình:
 - 7B (NLP)
 - 30B (NLP)
 - 7B-trade (dự báo),
 - 7B-full (tổng quát)
- Mô hình có input tối đa 2048 token, và cấu hình phần cứng rất mạnh:
 - FinMA-7B: 8 GPU A100
 - FinMA-30B: 128 GPU A100



NER

In the sentences extracted from financial agreements in U.S. SEC filings, identify the named entities that represent a person ("PER"), an organization ("ORG"), or a location ("LOC"). The required answer format is: "entity name, entity type". For instance, in "Elon Musk, CEO of SpaceX, announced the launch from Cape Canaveral.", the entities would be:

Elon Musk, PER

SpaceX, ORG

Cape Canaveral, LOC

Text: This LOAN AND SECURITY AGREEMENT dated January 27 , 1999 , between SILICON VALLEY BANK (" Bank "), a California - chartered bank with its principal place of business at 3003 Tasman Drive , Santa Clara , California 95054 with a loan production office located at 40 William St ., Ste .

Answer:

FPB

Analyze the sentiment of this statement extracted from a financial news article. Provide your answer as either negative, positive, or neutral. For instance, "The company's stocks plummeted following the scandal." would be classified as negative.

Shares of Standard Chartered (STAN) rose 1.2 % in the FTSE 100, while Royal Bank of Scotland (RBS) shares rose 2 % and Barclays shares (BARC)(BCS) were up 1.7 %.

• • •

BigData22

Contemplate the data and tweets to guess whether the closing price of \$ba will surge or decline at 2020-11-02. Please declare either Rise or Fall.

Context: date,open,high,low,close,adj-close,inc-5,inc-10,inc-15,inc-20,inc-25,inc-30
2020-10-19,1.2,2.2,-0.6,-0.1,-0.1,-1.4,-1.2,-0.8,-2.4,-2.4,-2.6

...

Answer:

NER

SILICON VALLEY BANK, ORG
Bank, ORG
California, LOC
bank, ORG
3003 Tasman Drive, LOC
Santa Clara, LOC
California, LOC
40 William St, LOC

FPB

Positive

• • •

BigData22

Rise

• • •

FinMA



Đánh giá và thực nghiệm



FLARE Benchmark – Tổng quan

Dựa trên FIT, thiết kế điểm đánh giá dự đoán và hiểu ngôn ngữ tự nhiên tài chính (FLARE).

- Mục tiêu: Đánh giá toàn diện LLM trên cả NLP tasks và Financial Prediction, khắc phục hạn chế của các benchmark trước (ví dụ: FLUE chỉ tập trung vào NLP).
- Cấu trúc: 6 NLP tasks: Sentiment Analysis, NER, QA... (10 datasets), 2 Prediction tasks: Stock Movement, Credit Scoring (5 datasets).
- Điểm đột phá: Đa phương thức (text + tables + time-series data). Tiêu chí đánh giá nghiêm ngặt (Accuracy, F1, MCC, ROUGE).

Data	Task	Valid	Test	Evaluation
FPB (Malo et al., 2014)	sentiment analysis	775	970	F1, Accuracy
FiQA-SA (Maia et al., 2018)	sentiment analysis	188	235	F1
Headlines (Sinha and Khandait, 2021)	news headline classification	1,141	2,283	Avg F1
NER (Alvarado et al., 2015)	named entity recognition	103	980	Entity F1
FiNER-ORD (Shah et al., 2023b)	named entity recognition	-	1080	Entity F1
FinQA (Chen et al., 2021)	question answering	883	1,147	EM Accuracy
ConvFinQA (Chen et al., 2022)	question answering	2,210	1,490	EM Accuracy
BigData22 (Soun et al., 2022)	stock movement prediction	798	1,470	Accuracy, MCC
ACL18 (Xu and Cohen, 2018)	stock movement prediction	2,560	3,720	Accuracy, MCC
CIKM18 (Wu et al., 2018)	stock movement prediction	431	1,140	Accuracy, MCC
ECTSum Mukherjee et al. (2022)	text summarization	-	495	ROUGE, BERTScore, BARTScore
EDTSum Zhou et al. (2021)	text summarization	-	2000	ROUGE, BERTScore, BARTScore
German Hofmann (1994)	credit scoring	-	1000	F1, MCC
Australian Quinlan (1987)	credit scoring	-	690	F1, MCC
FOMC Shah et al. (2023a)	hawkish-dovish classification	-	496	F1, Accuracy

FLARE: Financial NLP Tasks

Sentiment Analysis (FPB, FiQA-SA):

- **Phân loại cảm xúc tin tức tài chính (positive/neutral/negative).**
- **Named Entity Recognition (NER): Nhận diện thực thể (ORG, PER, LOC) trong báo cáo SEC**
- **Question Answering (FinQA, ConvFinQA): Yêu cầu lý luận số học từ báo cáo tài chính.**
- **Summarization (EDTSum, ECTSum): Tóm tắt báo cáo cuộc gọi thu nhập (earnings call).**

Data	Prompt
FPB	"Analyze the sentiment of this statement extracted from a financial news article. Provide your answer as either negative, positive or neutral. For instance, 'The company's stocks plummeted following the scandal.' would be classified as negative."
FiQA-SA	"What is the sentiment of the following financial {category} : Positive, Negative, or Neutral?"
Headlines	"Consider whether the headline mentions the price of gold. Is there a Price or Not in the gold commodity market indicated in the news headline? Please answer Yes or No."
NER	"In the sentences extracted from financial agreements in U.S. SEC filings, identify the named entities that represent a person ('PER'), an organization ('ORG'), or a location ('LOC'). The required answer format is: 'entity name, entity type'. For instance, in 'Elon Musk, CEO of SpaceX, announced the launch from Cape Canaveral.', the entities would be: 'Elon Musk, PER; SpaceX, ORG; Cape Canaveral, LOC'"
FiNER-ORD	"In the list of tokens, identify 'Person', 'Location', and 'Organisation' and label each accordingly. If the entity spans multiple tokens, use the prefix B-PER, B-LOC, or B-ORG for the first token, and I-PER, I-LOC, or I-ORG for the subsequent tokens of that entity. The beginning of each separate entity should always be labeled with a B-PER, B-LOC, or B-ORG prefix. If the token does not fit into any of the three named categories, or is not a named entity, label it as 'O'. Each line should contain one token and its corresponding label, separated by a colon. Do not combine tokens on your own. The format for each line should be: 'token:label'. Text: And all because you failed to prepare ! Answer:"
FinQA	"Given the financial data and expert analysis, please answer this question:"
ConvFinQA	"In the context of this series of interconnected finance-related queries and the additional information provided by the pretext, table data, and post text from a company's financial filings, please provide a response to the final question. This may require extracting information from the context and performing mathematical calculations. Please take into account the information provided in the preceding questions and their answers when formulating your response:"
BigData22	"Contemplate the data and tweets to guess whether the closing price of {tid} will surge or decline at {point} . Please declare with either Rise or Fall."
ECTSum	"Given the following article, please produce a list of 0 and 1, each separated by '' to indicate which sentences should be included in the final summary. The article's sentences have been split by ''. Please mark each sentence with 1 if it should be included in the summary and 0 if it should not."
EDTSum	"You are given a text that consists of multiple sentences. Your task is to perform abstractive summarization on this text. Use your understanding of the content to express the main ideas and crucial details in a shorter, coherent, and natural sounding text."
German	"Assess the creditworthiness of a customer using the following table attributes for financial status. Respond with either 'good' or 'bad'. And the table attributes including 13 categorical attributes and 7 numerical attributes are as follows:"
FOMC	"Examine the excerpt from a central bank's release below. Classify it as HAWKISH if it advocates for a tightening of monetary policy, DOVISH if it suggests an easing of monetary policy, or NEUTRAL if the stance is unbiased. Your response should return only HAWKISH, DOVISH, or NEUTRAL."

Bảng 5 trình bày tất cả các prompt cho từng tập dữ liệu trong tập dữ liệu hướng dẫn điểm chuẩn FLARE và FIT.

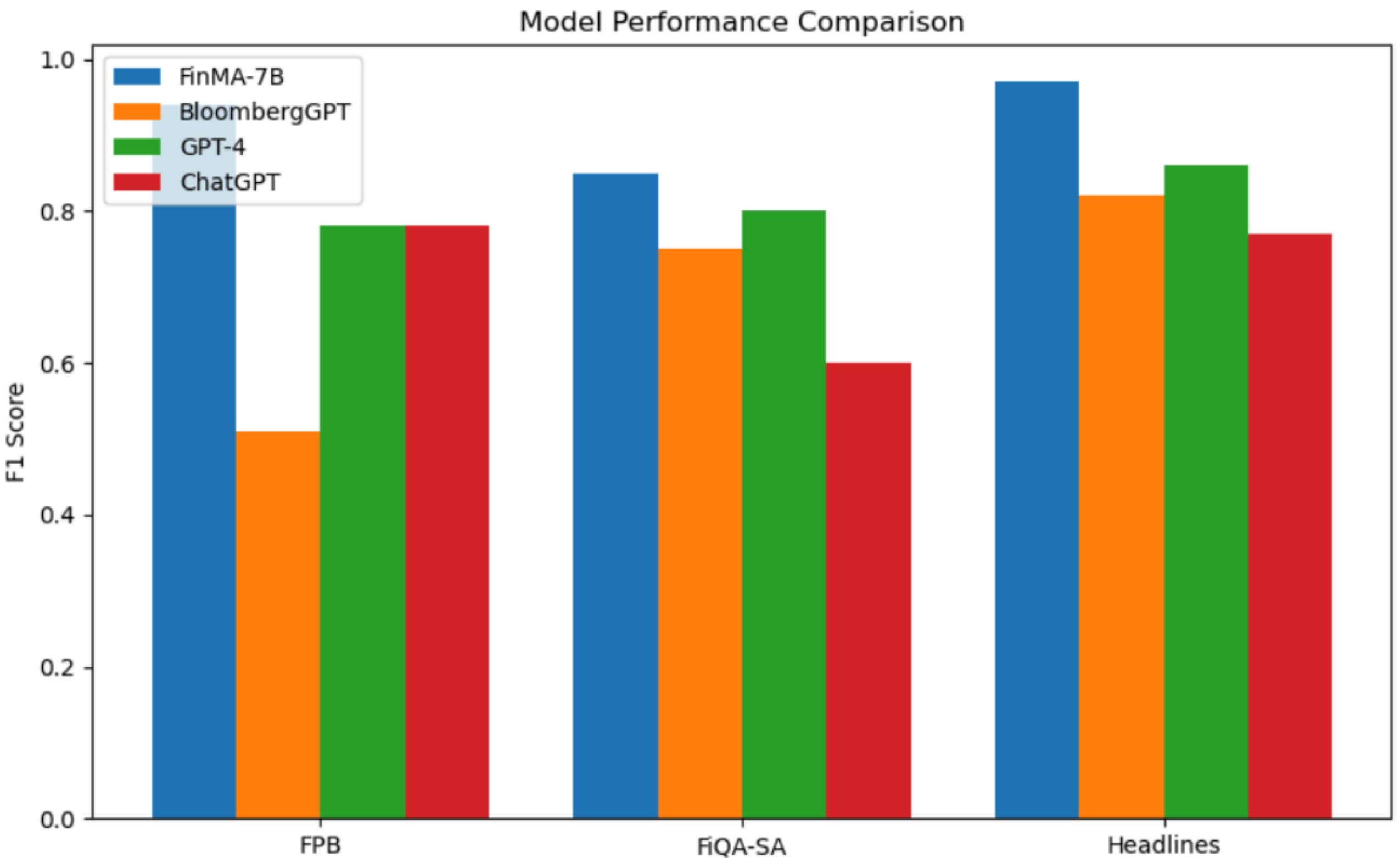
Chi tiết tác vụ Dự đoán Tài chính

- 1. Stock Movement Prediction:**
 Đầu vào: Tin tức + historical prices → Dự đoán "Rise/Fall".
 Thách thức: Độ chính xác thấp (~50-55% với LLM).
- 2. Credit Scoring:**
 Phân loại rủi ro tín dụng từ dữ liệu bảng (German/Australian dataset).
 GPT-4 đạt MCC cao nhất (0.47 trên Australian).

Dataset	Metrics	Chat GPT	GPT 4	Bloomberg GPT	Vicuna 7B	FinMA 7B	FinMA 7B-trade	FinMA 7B-full	FinMA 30B
FPB	F1	0.78*	0.78*	0.51*	0.29	0.94	0.03	0.94	0.88
	Acc	0.78*	0.76*	-	0.26	0.94	0.12	0.94	0.87
FiQA-SA	F1	0.60	0.80	0.75*	0.32	0.85	0.16	0.82	0.87
Headlines	AvgF1	0.77*	0.86*	0.82*	0.60	0.97	0.28	0.97	0.97
NER	EntityF1	0.77*	0.83*	0.61*	0.12	0.59	0.00	0.64	0.62
FINER-ORD	EntityF1	0.28	0.77	-	0.00	0.00	0.00	0.00	0.00
FinQA	EmAcc	0.58*	0.63*	-	0.00	0.06	0.00	0.04	0.11
ConvFinQA	EmAcc	0.60*	0.76*	0.43*	0.00	0.25	0.00	0.20	0.40
BigData22	Acc	0.53	0.54	-	0.44	0.45	0.45	0.51	0.47
	MCC	-0.025	0.03	-	-0.05	0.02	0.00	0.02	0.04
ACL18	Acc	0.50	0.52	-	0.50	0.49	0.49	0.51	0.49
	MCC	0.005	0.02	-	0.02	-0.01	0.03	0.03	0.00
CIKM18	Acc	0.55	0.57	-	0.44	0.43	0.43	0.50	0.43
	MCC	0.01	0.02	-	-0.03	-0.02	-0.003	0.08	-0.05
EDTsum	Rouge-1	0.17	0.2	-	0.22	0.09	0.05	0.13	0.17
	Rouge-2	0.08	0.09	-	0.10	0.04	0.02	0.06	0.08
	Rouge-N	0.13	0.15	-	0.17	0.08	0.05	0.10	0.14
	BertScore	0.66	0.67	-	0.61	0.56	0.51	0.38	0.54
	BartScore	-3.64	-3.62	-	-4.13	-6.12	-6.91	-5.71	-5.24
ECTsum	Rouge-1	0.00	0.00	-	0.00	0.00	0.00	0.00	0.00
	Rouge-2	0.00	0.00	-	0.00	0.00	0.00	0.00	0.00
	Rouge-N	0.00	0.00	-	0.00	0.00	0.00	0.00	0.00
	BertScore	0.00	0.00	-	0.00	0.00	0.00	0.00	0.00
	BartScore	-5.18	-5.18	-	-5.18	-5.18	-5.18	-5.18	-5.18
German	F1	0.20	0.55	-	0.52	0.17	0.52	0.17	0.53
	MCC	-0.10	-0.02	-	0.00	0.00	-0.07	0.00	-0.07
Australian	F1	0.41	0.74	-	0.26	0.41	0.26	0.41	0.46
	MCC	0.00	0.47	-	0.00	0.00	0.00	0.00	-0.01
FOMC	F1	0.64	0.71	-	0.19	0.49	0.10	0.49	0.43
	Acc	0.6	0.69	-	0.28	0.47	0.25	0.46	0.53

Kết quả Thí nghiệm – NLP Tasks

- **FinMA vượt trội trên phần lớn tác vụ:**
 - **FPB (F1: 0.94 vs BloombergGPT: 0.51).**
 - **Headlines (AvgF1: 0.97 vs GPT-4: 0.86).**
- **Hạn chế:**
 - **QA (FinQA EM Accuracy: 0.06 vs GPT-4: 0.63).**
 - **NER (Entity F1: 0.59 vs FLANG-BERT: 0.83).**



Kết quả Thí nghiệm – Prediction Tasks

- **Stock Movement:**

- **FinMA-7B-full đạt Accuracy cao nhất (ACL18: 0.56).**

- **MCC gần 0 → Khả năng dự đoán ngẫu nhiên.**

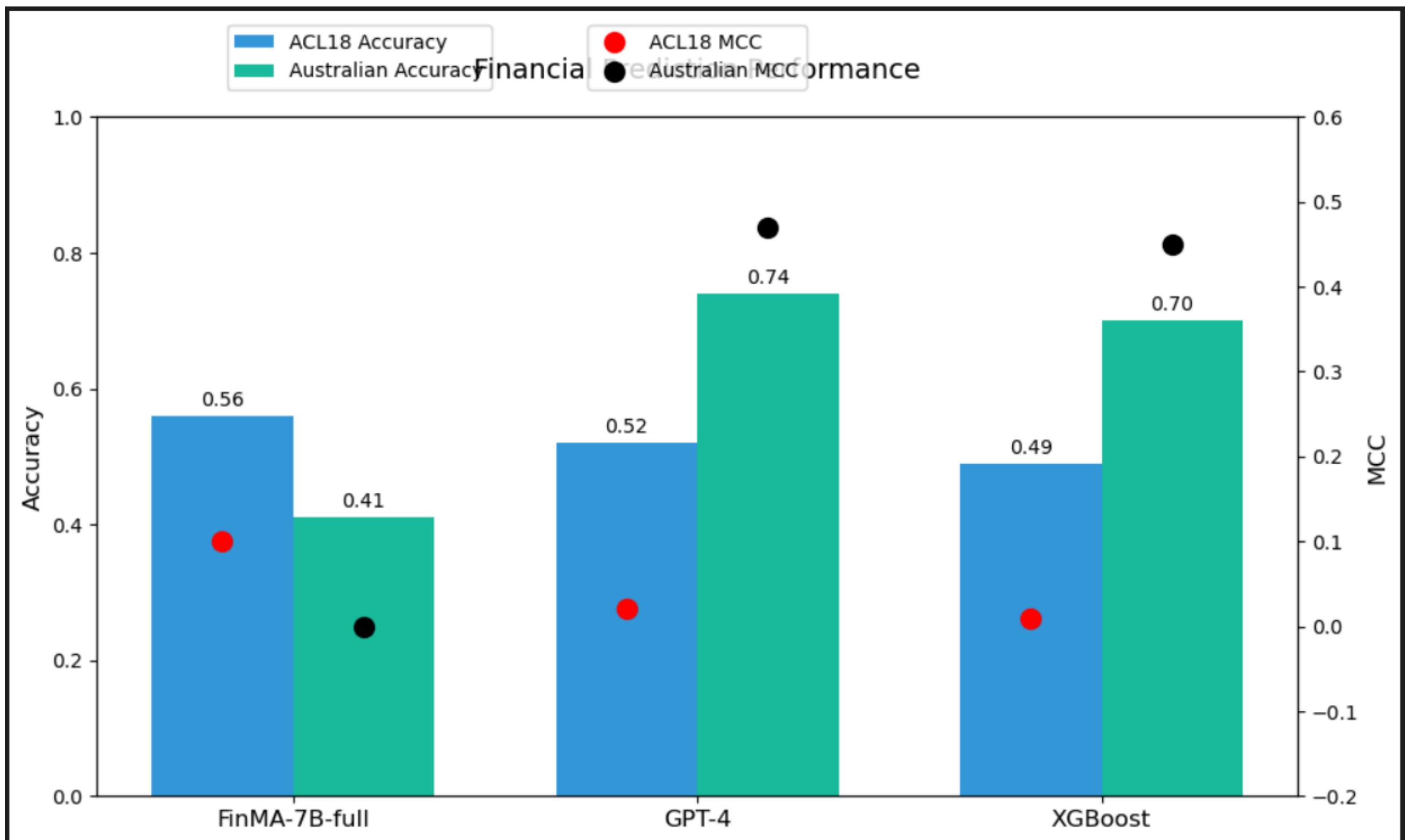
- **Credit Scoring:**

- **GPT-4 tốt nhất (Australian MCC: 0.47).**

- **FinMA không cải thiện so với baseline.**

- **Nhận định:**

- **Dự đoán tài chính cần tích hợp thêm dữ liệu structured.**



So sánh với BloombergGPT

Tiêu chí	FinMA	BloombergGPT
Mô hình	Open-source (LLaMA-based)	Proprietary (50B params)
Instruction Tuning	Có (FIT dataset)	Không
Hiệu suất	Tốt hơn trên sentiment/NER	Tốt hơn trên QA

-> FinMA phù hợp cho nghiên cứu mở, BloombergGPT tối ưu cho ứng dụng thương mại.

Key Contributions and Impact

- Đóng góp:

FinMA: LLM tài chính open-source đầu tiên.

FLARE: Benchmark đa nhiệm cho nghiên cứu AI tài chính.-

- Tác động:

Thúc đẩy phát triển LLM trong lĩnh vực tài chính.

Mở đường cho các nghiên cứu đa phương thức (văn bản + số liệu).

-> PIXIU là nền tảng quan trọng để xây dựng LLM tài chính trong tương lai.



Phân tích kết quả



Ảnh hưởng của Kích thước Mô hình

Dataset	Metrics	Chat GPT	GPT 4	Bloomberg GPT	Vicuna 7B	FinMA 7B	FinMA 7B-trade	FinMA 7B-full
FPB	F1	0.78*	0.78*	0.51*	0.29	0.94	0.03	0.94
	Acc	0.78*	0.76*	-	0.26	0.94	0.12	0.94
FiQA-SA Headlines	F1	0.60	0.80	0.75*	0.32	0.85	0.16	0.82
	AvgF1	0.77*	0.86*	0.82*	0.60	0.97	0.28	0.97
NER	EntityF1	0.77*	0.83*	0.61*	0.12	0.59	0.00	0.64
FINER-ORD	EntityF1	0.28	0.77	-	0.00	0.00	0.00	0.00
FinQA	EmAcc	0.58*	0.63*	-	0.00	0.06	0.00	0.04
ConvFinQA	EmAcc	0.60*	0.76*	0.43*	0.00	0.25	0.00	0.20
BigData22	Acc	0.53	0.54	-	0.44	0.45	0.45	0.51
	MCC	-0.025	0.03	-	-0.05	0.02	0.00	0.02
ACL18	Acc	0.50	0.52	-	0.50	0.49	0.49	0.51
	MCC	0.005	0.02	-	0.02	-0.01	0.03	0.03
CIKM18	Acc	0.55	0.57	-	0.44	0.43	0.43	0.50
	MCC	0.01	0.02	-	-0.03	-0.02	-0.003	0.08
EDTSUM	Rouge-1	0.17	0.2	-	0.22	0.09	0.05	0.13
	Rouge-2	0.08	0.09	-	0.10	0.04	0.02	0.06
	Rouge-N	0.13	0.15	-	0.17	0.08	0.05	0.10
	BertScore	0.66	0.67	-	0.61	0.56	0.51	0.38
	BartScore	-3.64	-3.62	-	-4.13	-6.12	-6.91	-5.71
ECTSUM	Rouge-1	0.00	0.00	-	0.00	0.00	0.00	0.00
	Rouge-2	0.00	0.00	-	0.00	0.00	0.00	0.00
	Rouge-N	0.00	0.00	-	0.00	0.00	0.00	0.00
	BertScore	0.00	0.00	-	0.00	0.00	0.00	0.00
	BartScore	-5.18	-5.18	-	-5.18	-5.18	-5.18	-5.18
German	F1	0.20	0.55	-	0.52	0.17	0.52	0.17
	MCC	-0.10	-0.02	-	0.00	0.00	-0.07	0.00
Australian	F1	0.41	0.74	-	0.26	0.41	0.26	0.41
	MCC	0.00	0.47	-	0.00	0.00	0.00	0.00
FOMC	F1	0.64	0.71	-	0.19	0.49	0.10	0.49
	Acc	0.6	0.69	-	0.28	0.47	0.25	0.46

-Nhận xét:

- FinMA-30B không vượt trội rõ rệt so với FinMA-7B trên các tác vụ NLP thông thường và dự đoán biến động giá cổ phiếu.
- Với các tác vụ phức tạp như FinQA và ConvFinQA, FinMA-30B thể hiện ưu thế rõ rệt → kích thước lớn giúp xử lý tác vụ toán học tốt hơn.

-Kết luận:

- Kích thước mô hình không quyết định tất cả → chất lượng dữ liệu instruction tuning mới là yếu tố then chốt.

Instruction Tuning – Vai trò quan trọng

Dataset	Metrics	Chat GPT	GPT 4	Bloomberg GPT	Vicuna 7B	FinMA 7B	FinMA 7B-trade	FinMA 7B-full
FPB	F1	0.78*	0.78*	0.51*	0.29	0.94	0.03	0.94
	Acc	0.78*	0.76*	-	0.26	0.94	0.12	0.94
FiQA-SA	F1	0.60	0.80	0.75*	0.32	0.85	0.16	0.82
Headlines	AvgF1	0.77*	0.86*	0.82*	0.60	0.97	0.28	0.97
NER	EntityF1	0.77*	0.83*	0.61*	0.12	0.59	0.00	0.64
FINER-ORD	EntityF1	0.28	0.77	-	0.00	0.00	0.00	0.00
FinQA	EmAcc	0.58*	0.63*	-	0.00	0.06	0.00	0.04
ConvFinQA	EmAcc	0.60*	0.76*	0.43*	0.00	0.25	0.00	0.20
BigData22	Acc	0.53	0.54	-	0.44	0.45	0.45	0.51
	MCC	-0.025	0.03	-	-0.05	0.02	0.00	0.02
ACL18	Acc	0.50	0.52	-	0.50	0.49	0.49	0.51
	MCC	0.005	0.02	-	0.02	-0.01	0.03	0.03
CIKM18	Acc	0.55	0.57	-	0.44	0.43	0.43	0.50
	MCC	0.01	0.02	-	-0.03	-0.02	-0.003	0.08
EDTSUM	Rouge-1	0.17	0.2	-	0.22	0.09	0.05	0.13
	Rouge-2	0.08	0.09	-	0.10	0.04	0.02	0.06
	Rouge-N	0.13	0.15	-	0.17	0.08	0.05	0.10
	BertScore	0.66	0.67	-	0.61	0.56	0.51	0.38
	BartScore	-3.64	-3.62	-	-4.13	-6.12	-6.91	-5.71
ECTSUM	Rouge-1	0.00	0.00	-	0.00	0.00	0.00	0.00
	Rouge-2	0.00	0.00	-	0.00	0.00	0.00	0.00
	Rouge-N	0.00	0.00	-	0.00	0.00	0.00	0.00
	BertScore	0.00	0.00	-	0.00	0.00	0.00	0.00
	BartScore	-5.18	-5.18	-	-5.18	-5.18	-5.18	-5.18
German	F1	0.20	0.55	-	0.52	0.17	0.52	0.17
	MCC	-0.10	-0.02	-	0.00	0.00	-0.07	0.00
Australian	F1	0.41	0.74	-	0.26	0.41	0.26	0.41
	MCC	0.00	0.47	-	0.00	0.00	0.00	0.00
FOMC	F1	0.64	0.71	-	0.19	0.49	0.10	0.49
	Acc	0.6	0.69	-	0.28	0.47	0.25	0.46

- FinMA-7B-full (huấn luyện đa nhiệm) tốt hơn rõ rệt so với bản đơn nhiệm.
- Đặc biệt hiệu quả trên ACL18 (biến động cổ phiếu).
- Nhưng gặp khó khăn với nhiệm vụ có đầu ra phức tạp (FINER-ORD, ECTSUM).

Khả năng sinh văn bản – Generation

Dataset	Metrics	Chat GPT	GPT 4	Bloomberg GPT	Vicuna 7B	FinMA 7B	FinMA 7B-trade	FinMA 7B-full
FPB	F1	0.78*	0.78*	0.51*	0.29	0.94	0.03	0.94
	Acc	0.78*	0.76*	-	0.26	0.94	0.12	0.94
FiQA-SA	F1	0.60	0.80	0.75*	0.32	0.85	0.16	0.82
Headlines	AvgF1	0.77*	0.86*	0.82*	0.60	0.97	0.28	0.97
NER	EntityF1	0.77*	0.83*	0.61*	0.12	0.59	0.00	0.64
FINER-ORD	EntityF1	0.28	0.77	-	0.00	0.00	0.00	0.00
FinQA	EmAcc	0.58*	0.63*	-	0.00	0.06	0.00	0.04
ConvFinQA	EmAcc	0.60*	0.76*	0.43*	0.00	0.25	0.00	0.20
BigData22	Acc	0.53	0.54	-	0.44	0.45	0.45	0.51
	MCC	-0.025	0.03	-	-0.05	0.02	0.00	0.02
ACL18	Acc	0.50	0.52	-	0.50	0.49	0.49	0.51
	MCC	0.005	0.02	-	0.02	-0.01	0.03	0.03
CIKM18	Acc	0.55	0.57	-	0.44	0.43	0.43	0.50
	MCC	0.01	0.02	-	-0.03	-0.02	-0.003	0.08
EDTSUM	Rouge-1	0.17	0.2	-	0.22	0.09	0.05	0.13
	Rouge-2	0.08	0.09	-	0.10	0.04	0.02	0.06
	Rouge-N	0.13	0.15	-	0.17	0.08	0.05	0.10
	BertScore	0.66	0.67	-	0.61	0.56	0.51	0.38
	BartScore	-3.64	-3.62	-	-4.13	-6.12	-6.91	-5.71
ECTSUM	Rouge-1	0.00	0.00	-	0.00	0.00	0.00	0.00
	Rouge-2	0.00	0.00	-	0.00	0.00	0.00	0.00
	Rouge-N	0.00	0.00	-	0.00	0.00	0.00	0.00
	BertScore	0.00	0.00	-	0.00	0.00	0.00	0.00
	BartScore	-5.18	-5.18	-	-5.18	-5.18	-5.18	-5.18
German	F1	0.20	0.55	-	0.52	0.17	0.52	0.17
	MCC	-0.10	-0.02	-	0.00	0.00	-0.07	0.00
Australian	F1	0.41	0.74	-	0.26	0.41	0.26	0.41
	MCC	0.00	0.47	-	0.00	0.00	0.00	0.00
FOMC	F1	0.64	0.71	-	0.19	0.49	0.10	0.49
	Acc	0.6	0.69	-	0.28	0.47	0.25	0.46

- FinMA sinh văn bản kém trong tóm tắt (EDTSum).
- Vicuna-7B dù không được fine-tune vẫn vượt trội theo ROUGE, BERTScore.

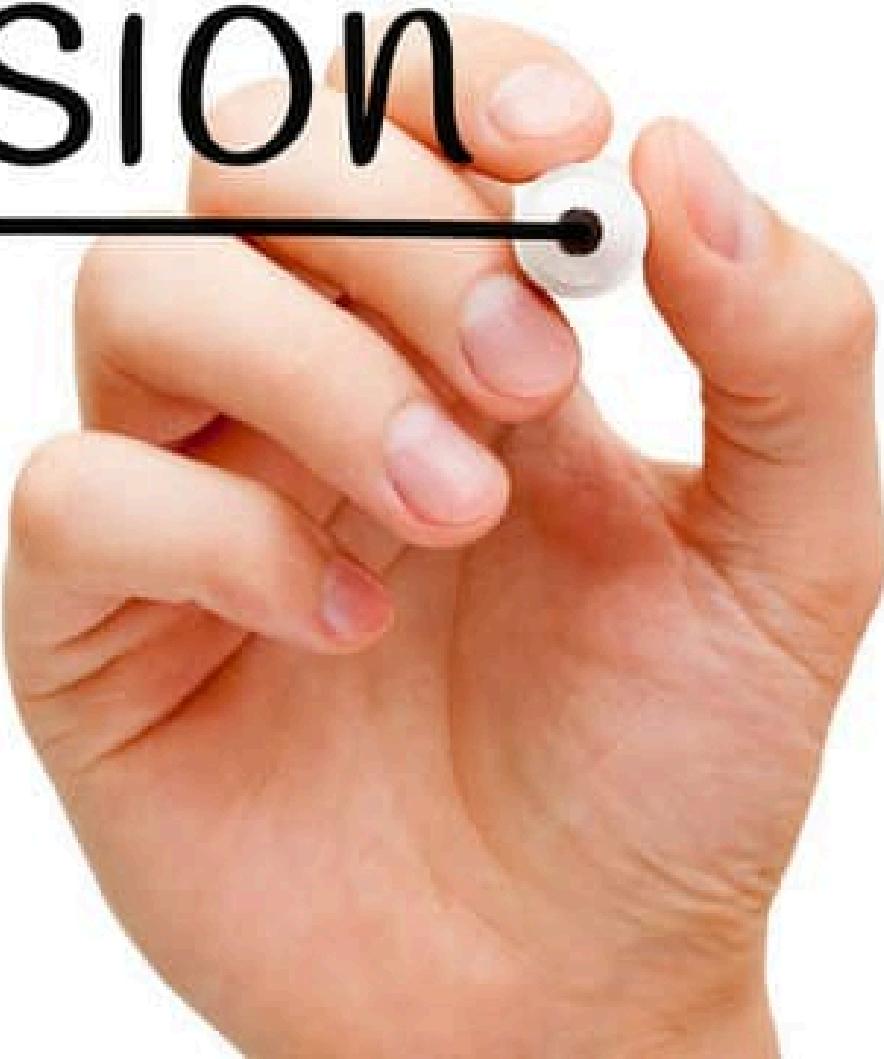
Hạn chế

- Mô hình chưa đủ lớn (tối đa 30B), chưa khai thác hết dữ liệu FIT.
- Xử lý toán và bảng biểu yếu do chưa pretrain dữ liệu toán.
- Instruction chưa đa dạng.
- Nguy cơ bị lợi dụng vào các mục đích tài chính phi đạo đức.



Kết luận và
định hướng
tương lai

Conclusion

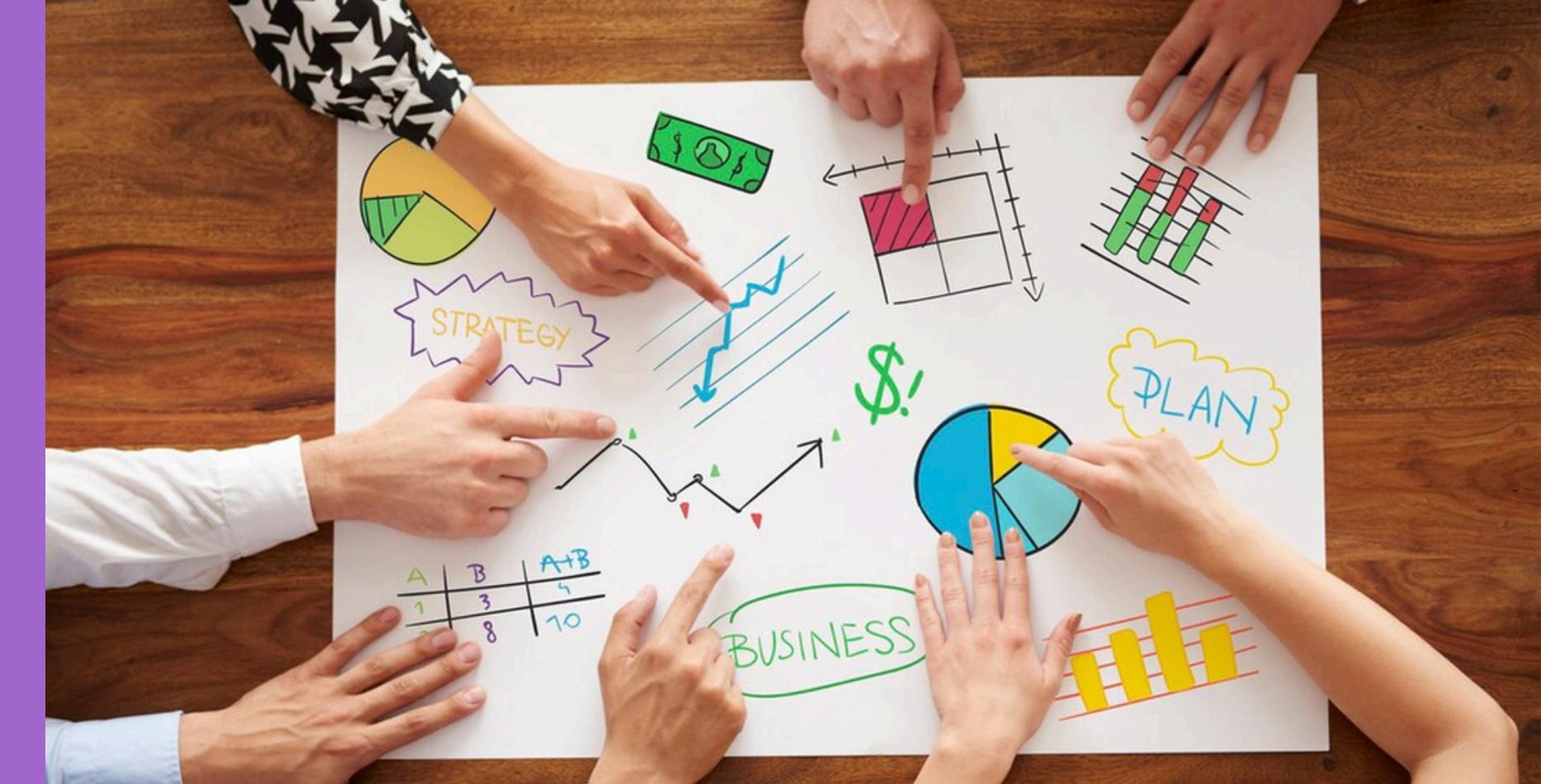


PIXIU đóng góp gì?

- Đề xuất một benchmark toàn diện cho đánh giá LLM tài chính.
- Ra mắt FinMA: mô hình ngôn ngữ tài chính đầu tiên được fine-tune theo instruction.
- Cung cấp bộ dữ liệu FIT và hệ thống đánh giá FLARE.

Hướng phát triển tương lai:

- Bổ sung dữ liệu toán học và bảng biểu trong giai đoạn tiền huấn luyện.
- Tăng tính đa dạng và tính thực tiễn của các instruction trong FIT.
- Mở rộng mô hình: quy mô lớn hơn, khả năng xử lý chuỗi dài và đầu ra phức tạp hơn.
- Xây dựng quy chuẩn đạo đức rõ ràng khi triển khai mô hình trong ứng dụng tài chính.



Xin cảm ơn!

Đừng ngần ngại liên hệ với chúng tôi nếu bạn có bất kỳ câu hỏi nào.

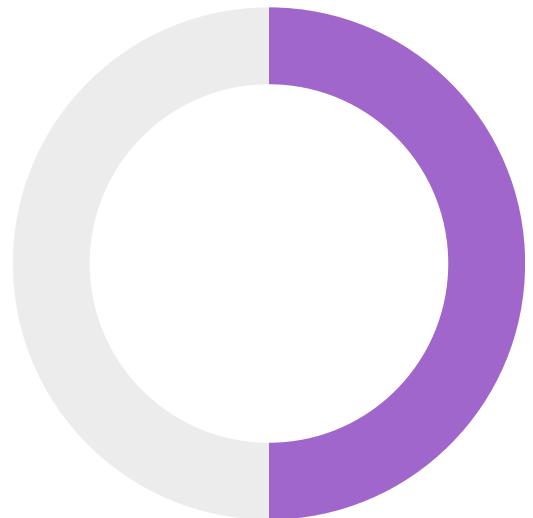
Cập nhật trạng thái dự án



Giai đoạn dự án 1

41% đã hoàn thành

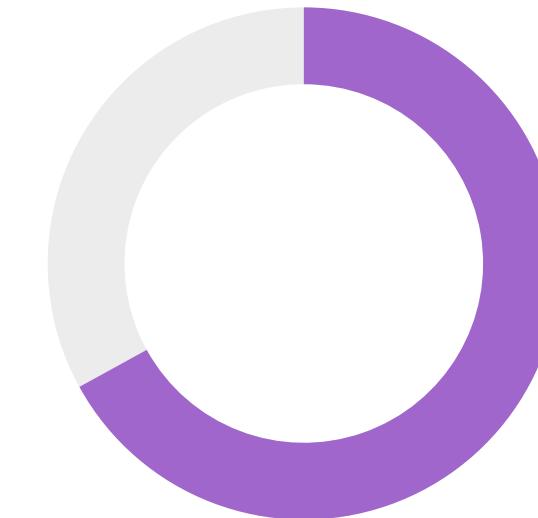
Thuyết trình dễ dàng và
khiến khán giả ngạc nhiên
với Bài thuyết trình Canva.



Giai đoạn dự án 2

50% đã hoàn thành

Thuyết trình dễ dàng và
khiến khán giả ngạc nhiên
với Bài thuyết trình Canva.



Giai đoạn dự án 3

67% đã hoàn thành

Thuyết trình dễ dàng và
khiến khán giả ngạc nhiên
với Bài thuyết trình Canva.

Sắp tới

Đang thực hiện

Đã hoàn thành

Tạm dừng

Viết một nhiệm vụ tại đây

TRUNG BÌNH

Viết một nhiệm vụ tại đây

THẤP

Viết một nhiệm vụ tại đây

CAO

Viết một nhiệm vụ tại đây

CAO

Viết một nhiệm vụ tại đây

CAO



Sắp xếp lại các nhiệm vụ theo trạng thái và mức độ ưu tiên tương ứng. Sao chép nếu cần thiết tùy thuộc vào nhu cầu trong dự án.

Ngày:

Trang tài nguyên

Thuyết trình trực tiếp không phải là thế mạnh của bạn? Đừng lo lắng! Ghi lại Bản thuyết trình Canva để mọi người có thể xem theo tốc độ của họ. Đừng quên xóa trang này trước khi thuyết trình.

Nhấp vào nút Chia sẻ ở góc trên cùng bên phải của màn hình và chọn 'Thuyết trình và Ghi hình'.

.....

Nhấp vào 'Đi đến phòng ghi hình' nơi bạn có thể chọn nguồn video và âm thanh cho bài thuyết trình video của mình.

.....

Thoải mái chọn 'Không máy ảnh' và chỉ ghi lại giọng nói của bạn.

.....

Bắt đầu ghi và nhấn tạm dừng giữa mỗi lần thu nếu cần.

.....

Sau khi hoàn tất, hãy tải về Bản thuyết trình Canva của bạn ở định dạng MP4 hoặc nhận đường liên kết đến Bản thuyết trình tự ghi hình và chia sẻ với người khác.

.....

Bạn có thể thu video bên trong trình biên tập! Di đến 'Tải lên' và nhấp vào 'Tự ghi hình'.

