DLP Final Project Prposal: Market Guided Stock Transformer

Group 7
110705009 陳重光、313551047 陳以瑄、313554043 戴明貴

Outline

- Intoduction - Market Guided
- MASTER - AAAI'24
- Method
- Dataset
- Expected Result

Introduction - Market Guided

Stock prediction features can be divided into two types:

1. Individual Stock Features:
    - Open price, close price, etc.
    - Trading volume
2. Shared Market Features:
    - Market index
    - Macroeconomic indicators, e.g. interest rate

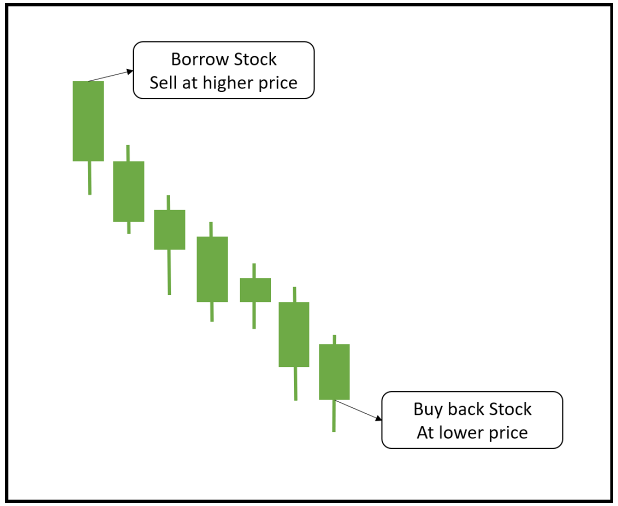
Introduction - Market Guided

The market feature impacts the effectiveness of other features.

Example: Short Selling

When investors believe a stock is overvalued.

1. Borrow stock, sell at high price.
2. Buy back at lower price when it falls.
3. Return to owner.



Short selling interest: the amount of stocks being short.

Introduction - Market Guided

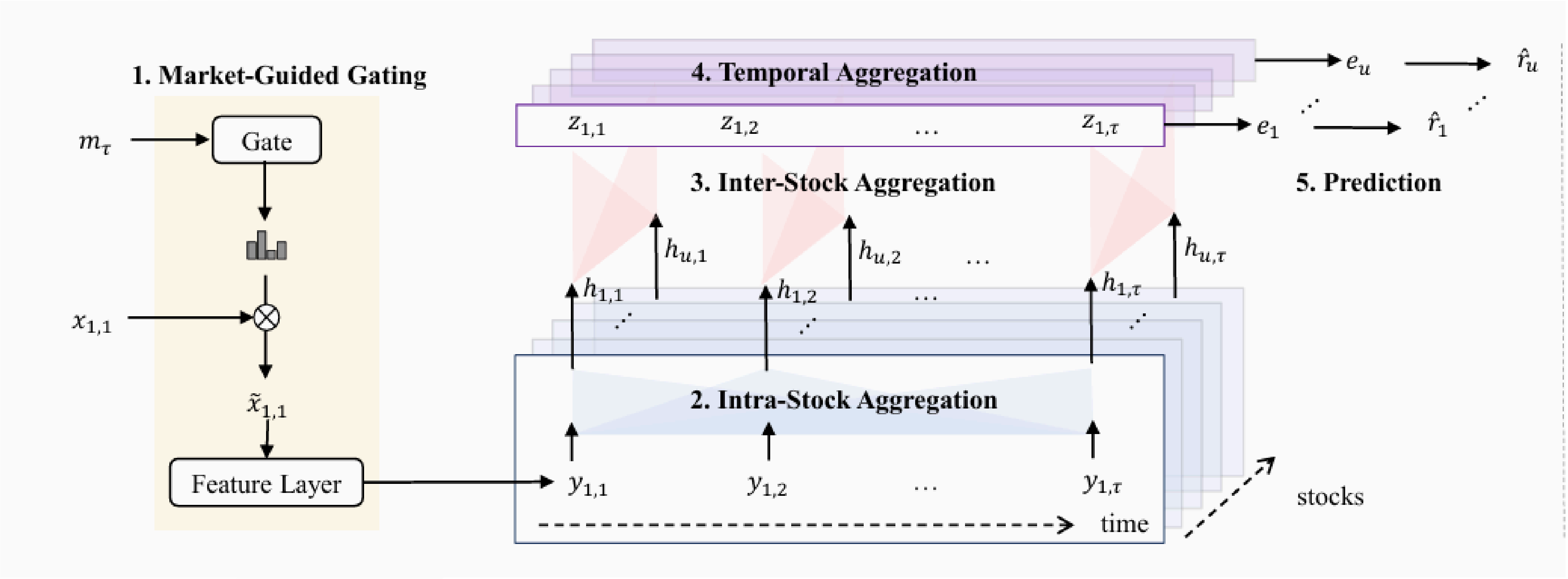
The market feature impacts the effectiveness of other features.

Example: Short Selling
Effectiveness in different market status:

- Bull Market: short selling loses money, less concern.
- Bear Market: short selling signals pessimism, more significant.

→ **Using market status to select relevant features.**

MASTER:Market-Guided Stock Transformer for Stock Price Forecasting [1]



Limitation

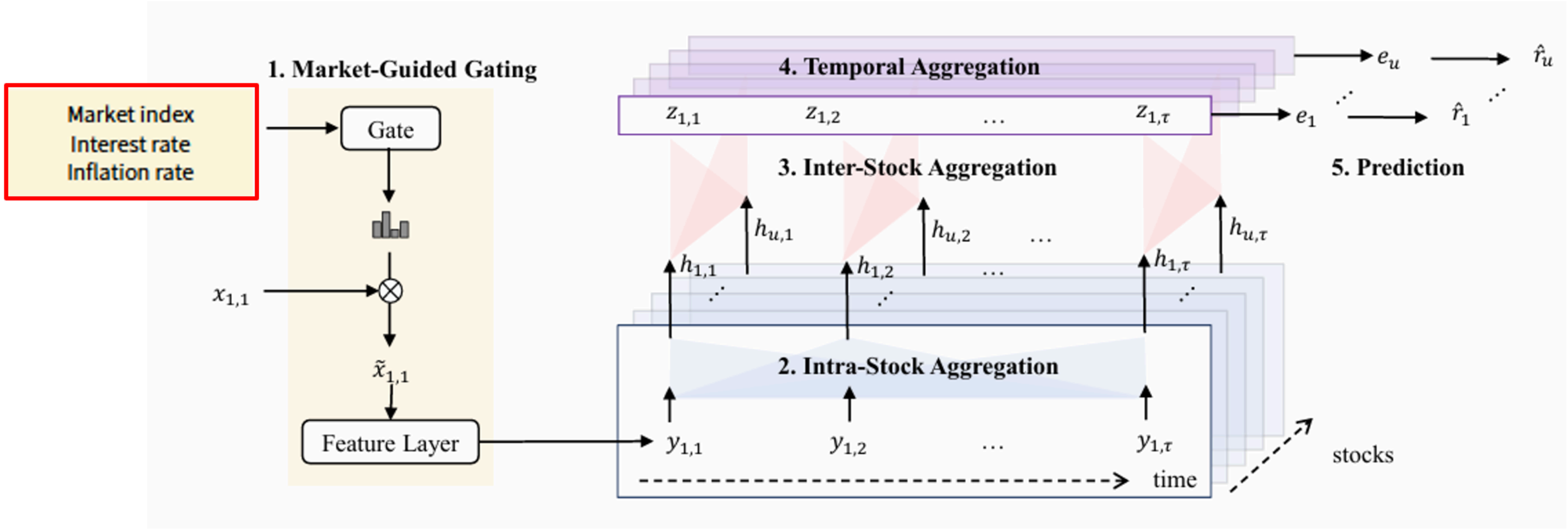
Simple Representation of Market Status:
Only market index prices and trading volumes are used as inputs.

Improvements: Expanding Shared Market Features

1. Macroeconomic features
2. Industry-level features
3. News-based features

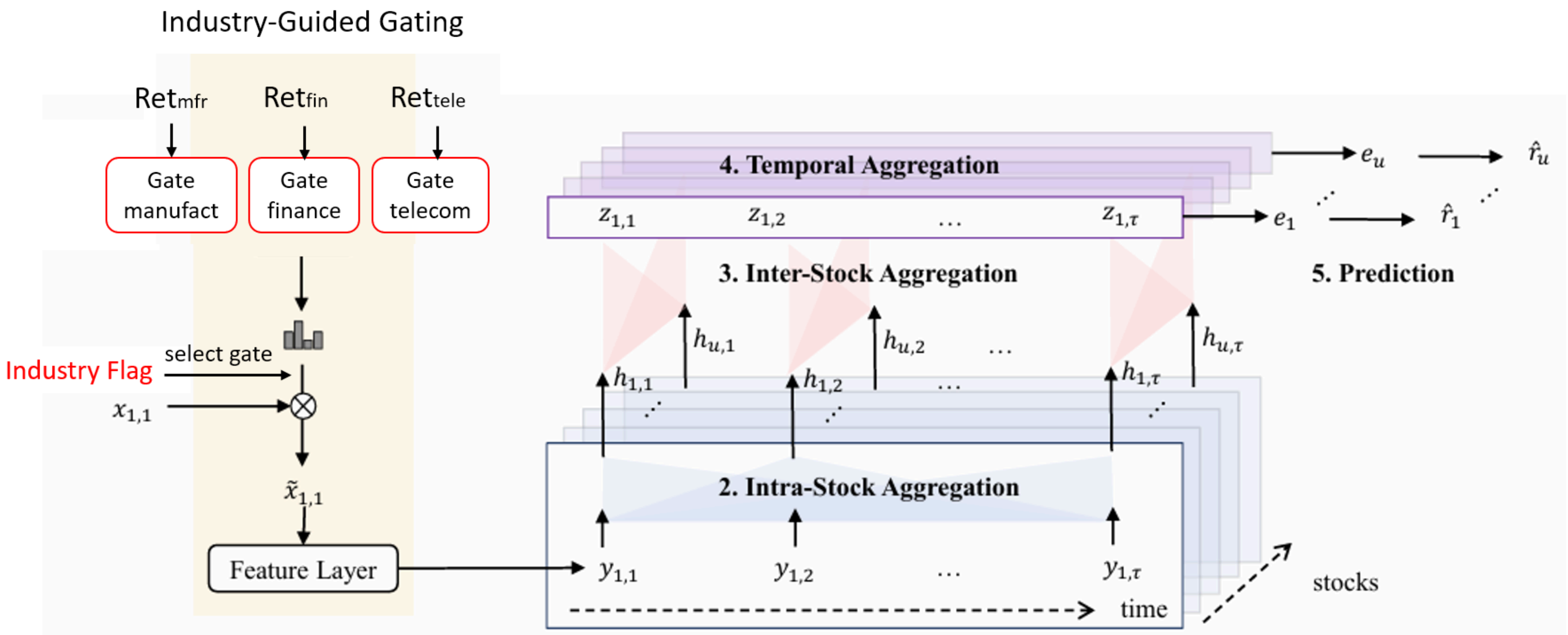
Idea 1 - Expand Market Features

Market-guided indicators expansion for richer market dynamics



Idea 2 - Industry-Level Features

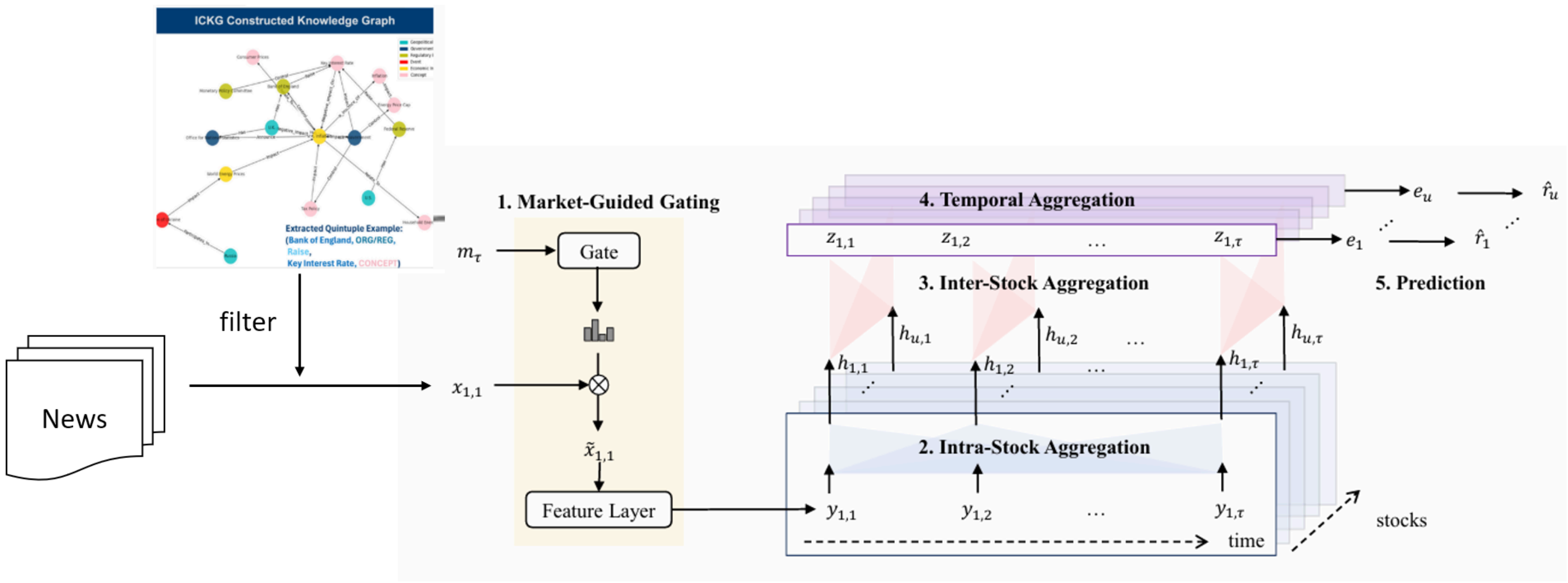
Each industry learns a different gate to capture its specific behavior. The industry flag is then used to select the relevant gate.



Idea 3 - News-Based Features

Use FinDKG[2] to identify business partners
Then use the firm correlation to select the relevant news.

Firm Correlation from FinDKG



Problem Definition

Given a set of stocks S with features $x_{u,t} \in \mathbb{R}^F$ collected at time steps $t \in [1, \tau]$:

For each stock, we consider:

- Individual stock features (price, volume)
- Shared market features(market index, macroeconomic indicators)
- Industry-Level feature (return)
- News-derived features (company and related party)

Output: The return ratio $r_u = \mathrm{Norm}_S((c_{u,\tau+d} - c_{u,\tau+1})/c_{u,\tau+1})$

Data Description

The dataset for input of this study consists of the following data:

- **Stock prices**
- **Industry**
- **Market index**
- **Economic indicator**
- **Sentimental scores**

Data Description (cont.)

- **Stock Price**:
  - **Base**: S&P 500 constituents
  - **Industry classification**: base on the Fama-French 12 industry [3].
  - **Number of stocks**: 8 firms * 12 industries = 96 firms
  - **Daily stock features**:
    1. price: open, high, low, close prices
    2. volume: trading volume
    3. others: short selling interest, etc.

Data Description (cont.)

- **Industry**:
  The Fama-French 12 industry returns.
- **Market Index**:
  S&P 500 market index
- **Economic Indicators**:
  VIX, interest rates, and other economic indicators
- **Sentimental Scores**:
  News sentiment score from RavenPack

Data Description (cont.)

NAME	NUMBER	TRAINING	TEST	SOURCE
Stock Price	96 * 5	2010 - 2022	2023	WRDS - CRSP
Industry	12	2010 - 2022	2023	Fama-French
Market Index	1	2010 - 2022	2023	CRSP
Economic Indicators	1	2010 - 2022	2023	VIX, FRED
Sentimental Scores	6	2010 - 2022	2023	Ravenpack

Expected result

Dataset	Model	IC	ICIR	RankIC	RankICIR	AR	IR
CSI300	XGBoost	0.051 ± 0.001	0.37 ± 0.01	0.050 ± 0.001	0.36 ± 0.01	$\underline{0.23 \pm 0.03}$	1.9 ± 0.3
	LSTM	0.049 ± 0.001	$\underline{0.41 \pm 0.01}$	0.051 ± 0.002	0.41 ± 0.03	0.20 ± 0.04	$\underline{2.0 \pm 0.4}$
	GRU	0.052 ± 0.004	0.35 ± 0.04	0.052 ± 0.005	0.34 ± 0.04	0.19 ± 0.04	1.5 ± 0.3
	TCN	0.050 ± 0.002	0.33 ± 0.04	0.049 ± 0.002	0.31 ± 0.04	0.18 ± 0.05	1.4 ± 0.5
	Transformer	0.047 ± 0.007	0.39 ± 0.04	0.051 ± 0.002	$\underline{0.42 \pm 0.04}$	0.22 ± 0.06	2.0 ± 0.4
	GAT	$\underline{0.054 \pm 0.002}$	0.36 ± 0.02	0.041 ± 0.002	0.25 ± 0.02	0.19 ± 0.03	1.3 ± 0.3
	DTML	0.049 ± 0.006	0.33 ± 0.04	$\underline{0.052 \pm 0.005}$	0.33 ± 0.04	0.21 ± 0.03	1.7 ± 0.3
	MASTER	$\mathbf{0.064^* \pm 0.006}$	$\mathbf{0.42 \pm 0.04}$	$\mathbf{0.076^* \pm 0.005}$	$\mathbf{0.49 \pm 0.04}$	$\mathbf{0.27 \pm 0.05}$	$\mathbf{2.4 \pm 0.4}$
CSI800	XGBoost	0.040 ± 0.000	0.37 ± 0.01	0.047 ± 0.000	0.42 ± 0.01	0.08 ± 0.02	0.6 ± 0.2
	LSTM	0.028 ± 0.002	0.32 ± 0.02	0.039 ± 0.002	0.41 ± 0.03	0.09 ± 0.02	0.9 ± 0.2
	GRU	0.039 ± 0.002	0.36 ± 0.05	0.044 ± 0.003	0.39 ± 0.07	0.07 ± 0.04	0.6 ± 0.3
	TCN	0.038 ± 0.002	0.33 ± 0.04	0.045 ± 0.002	0.38 ± 0.05	0.05 ± 0.04	0.4 ± 0.3
	Transformer	0.040 ± 0.003	$\mathbf{0.43 \pm 0.03}$	0.048 ± 0.003	$\mathbf{0.51 \pm 0.05}$	0.13 ± 0.04	1.1 ± 0.3
	GAT	$\underline{0.043 \pm 0.002}$	0.39 ± 0.02	0.042 ± 0.002	0.35 ± 0.02	0.10 ± 0.04	0.7 ± 0.3
	DTML	0.039 ± 0.004	0.29 ± 0.03	$\underline{0.053 \pm 0.008}$	0.37 ± 0.06	$\underline{0.16 \pm 0.03}$	$\underline{1.3 \pm 0.2}$
	MASTER	$\mathbf{0.052^* \pm 0.006}$	$\underline{0.40 \pm 0.06}$	$\mathbf{0.066 \pm 0.007}$	$\underline{0.48 \pm 0.06}$	$\mathbf{0.28^* \pm 0.02}$	$\mathbf{2.3^* \pm 0.3}$

Table 1: Overall performance comparison. The best results are in bold and the second-best results are underlined. And * denotes statistically significant improvement (measured by t-test with p-value < 0.01) over all baselines.

Expected result (cont.)

The set of stocks has changed from Chinese stocks to U.S. stocks
→ Prior related works are no longer directly applicable.

We aim to compare performance between:

1. The original version of MASTER
2. MASTER with three improvements

References

[1] Li, T., Liu, Z., Shen, Y., Wang, X., Chen, H., & Huang, S. (2024). MASTER: Market-Guided Stock Transformer for Stock Price Forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, 38(1), 162-170.

[2] Li, X. V., & Sanna Passino, F. (2024). FinDKG: Dynamic Knowledge Graphs with Large Language Models for Detecting Global Trends in Financial Markets. In Proceedings of the 5th ACM International Conference on AI in Finance (ICAIF '24) (pp. 573–581).

References (cont.)

[3] Fama, E. F., & French, K. R. (1997). Industry costs of equity. Journal of financial economics, 43(2), 153-193.

Thank you for listening.