**3. Methodology:**

* Dự báo lợi nhuận hàng tuần/tháng của cổ phiếu NASDAQ-100 dựa trên giá cổ phiếu, hồ sơ công ty và tin tức tài chính.
* Kết hợp mô hình ngôn ngữ lớn như GPT-4 và Open LLaMA.
* Use GPT-4 for zero-shot and few-shot.
* Use Open LLaMA for fine-tuning để cải thiện khả năng xử lý.
* Áp dụng Chain Of Thoughts để cải thiện hiệu suẩt mô hình.

**3.1. Data:**

**3.1.1. Stock Price Data:**

* Tải dữ liệu từ Yahoo Finance bằng pack yfinance, chuẩn hoá thành phần trăm thay đổi rồi phân loại thành các nhóm.

**3.1.2. Company Profile Data:**

* Use GPT-4 tạo hồ sơ công ty, mô tả chung về các doanh nghiệp và các yếu tố ảnh hưởng đến giá. (Fig 1)

**3.1.3. Finance/Economy News Data:**

* Use GG Custome Search API thu thập tin tức cho từng cổ phiếu.
* Use GPT-4 tóm tắt. (Fig 2)
* Tình hình tài chính cũng dùng phương pháp tương tự. (Fig 3 gần giống Fig 2 nhưng cô đọng hơn)

**3.2. Instruction-Based Zero-shot/Few-shot Inference with LLMs:**

* Few-shot Learning để cung cấp ví dụ từ các cổ phiếu tương tự tránh lặp lại -> CMR LLM có thể xem xét thông tin giữa các chuỗi từ nhiều cổ phiếu khác nhau.
* Dùng GPT-4 để xác định mối quan hệ giữa các cổ phiếu liên quan.
* Cấu trúc nhắc lệnh và hướng dẫn chia thành 2 phần để mô hình hiểu nhiệm vụ hơn.
* Dự đoán tóm tắt và từ khoá để giải thích dự đoán lợi nhuận cổ phiếu.
* Thử nghiệm Chain-Of-Thoughts (Tư duy từng bước) để cải thiện hiệu suất (sẽ nói ở section 4.2)
* GPT-4 nhận diện điểm quan trọng bị bỏ qua về “báo cáo thu nhập”

**3.3. Instruction-based Fine-tuning with Open LLaMA:**

* Tinh chỉnh Open LLaMA để xử lý các nhắc lệnh dựa trên hướng dẫn.
* Dữ liệu đào tạo gồm 30K dự báo hàng tuần, 7K dự báo hàng tháng.
* Do hạn chế phần cứng, phải nén xuống 1K token. (Ít hơn GPT-4 với 8K token)
* Use GPT-4 cô đọng dữ liệu thành các bản tóm tắt. (Thực tế có thể dùng Open LLaMA nhưng sẽ kém hơn đáng kể so với GPT-4).