**Labels**

* Label 1: Classifying Financial Events: Identify and categorize significant financial events within the articles.
* Label 2: Determining Relationships: Differentiate between causal and non-causal relationships within the context of these events.
* Label 3: Assessing Impact on Stock Returns: Analyze the impact of these financial events on the mentioned companies, particularly how they influence stock performance.

**Structure of GPT prompts**

1. Context and objectives: Instruct GPT to act as a financial analyst for the end goal of developing graph causal inferences
2. Concise event definitions: Provide GPT with one-line definitions of the five events to be classified. To help GPT distinguish between event types, we also try incorporating detailed descriptions and two-shot learning
   1. Detailed descriptions: Provide detailed descriptions obtained from five\_sample\_def.xlsx
   2. Two-shot learning: Provide examples of correctly labeled events
3. Event identification: Feed in the article and relevant information, and ask GPT to identify events based on previous definitions
   1. Article: title, descriptions and contents
   2. Stock returns: 30 days/ 5 days/ 3 days/ 1 day prior to news publication date, day of news publication, and 1 day/ 2 days/ 5 days/ 30 days after news publication
   3. Event keywords: keywords related to each event  
      E.g., "Industry Competitiveness" : {"competitive landscape", "market competition", "industry rivalry", "market share", "competitive position", "market leader", "industry dynamics", "market penetration", "competitive advantage", "pricing strategy", "market saturation", "competitive pressure", "market entrants", "brand competition", "competitive analysis", "market positioning", … }
4. Company classification: For each article, loop through the mentioned companies and events labeled, ask GPT to classify if a causal relationship exists; if so, label the impact to be positive, negative or neutral; if not, label the impact as not relevant

**Observations on event classification**

1. Industry Competitiveness

When an article mentions multiple companies, GPT labels as industry competitiveness

1. Revenue Discussions vs. Margin/ Profitability Discussions

GPT often label both these events together [stats]

* 1. Mentioning of net profit
  2. Mentioning of earnings

1. Others
   1. Industry-specific events
      1. Energy
         1. Energy - investment activities: GPT classifies as Others, but the event could potentially impact stock prices  
            E.g., “Chevron will invest more than $500 million to develop the Trapial block in western Neuquen province, home to the massive Vaca Muerta shale basin”
         2. Energy - greenhouse gas commitments, carbon capture projects: GPT classifies as Industry Competitiveness, as such news often mention plans of multiple companies, yet the impact is not straightforward
         3. Energy - discussions on gas futures or spot prices: GPT classifies as Others

E.g., “gas futures jumped 16% to a one-week high around $11 per mmBtu”

* + 1. Healthcare
       1. Trials of new drugs: GPT classifies as a combination
       2. FDA initiated approval process or approved new drugs
       3. New business or product lines
       4. Lawsuits
  1. Discussions on past stock performance

E.g., "This fund crushed it last year in part by heavily betting on one unique oil stock"

(xxx briefs)

* 1. Analysis of multiple stocks and investment recommendations

**Observations on impact analysis**

1. GPT is good at classifying the direction of impact
2. Mixture of positive and negative sentiments

E.g., Amazon.com Inc (NASDAQ:AMZN) beat expectations on revenue and profit for the first quarter but said it sees a slowdown.

**Adjustments on GPT prompts**

| **Issue** | **Treatment** |
| --- | --- |
| Industry Competitiveness | Feed in keywords identified related to “industry competitiveness” to GPT |
| Revenue Discussions vs. Margin/ Profitability Discussions | Feed in more detailed definitions of those two events based on Che’s event definitions |

**Statistics**

Accuracy of labeling (GPT vs. human) (event and impact labels)

Confusion matrix

Distribution of industries

…

**Other thoughts on data labeling**

1. Parts of an article utilized in manual labeling

* To classify event types, we first look at the title, description and summary of an article. If the event type is still unclear, we further read the first few sentences of the article.
* To identify whether a company is relevant to the event and the direction of impact, we look at the sentences in which the company ticker is mentioned.
* When designing GPT prompts, we can consider only providing such context to simplify prompts, but by doing so we might miss subtle information mentioned in the article.