

# Reconstruction of the Economic Policy Uncertainty Using Large Language Models

Jacky Yeh

This version: November 2023

## 0.0.1 Result

What utilize OpenAI's ChatGPT-3.5-turbo API to test the predictability and feasibility of the NLP foundation model in constructing textual indices, with our main focus on the EPU index. In our experiments, we employed, but were not limited to, three approaches: zero-shot prompting, few-shot prompting (3, 6, 8 shots), and fine-tuning. Essentially, one-shot prompting involves purely asking questions to ChatGPT without providing any examples. On the other hand, few-shot prompting does involve providing some examples, which can often result in more accurate and thorough answers from ChatGPT. Below are the news articles that were used in few-shot prompting.

Table 1: news for few-shot prompt

should be excluded (label 1)	shouldn't be excluded (label 0)	should be excluded (label 1)
澳洲央行昨日意外調高基準利率，隔夜拆款利率從原本 3%，調高為 3.25%，澳洲儲備銀行行長史蒂 (Glenn Stevens) 表示，澳洲經濟大幅緊縮的風險已明顯降低，為了讓經濟穩健成長，澳洲央行未來幾年將持續進行微調。巴克萊投資表示，澳洲與巴西、印度、印尼、台灣、韓國，是這一波金融風暴中，經濟提早反彈的國家，領先指標 3 月分，就出現反彈訊號，由於中國大陸原物料需求強勁，巴克萊預估，澳洲今年明年經濟成長率，可望達 0.9% 與 2.5%。巴克萊投資預估，英國、印度、韓國三國，可望在明年第 1 季升息，台灣升息時間，可望落在第 2 季；美國、中國與歐盟則要等到明年第 3 季才會升息。	千興為國內老牌不鏽鋼冷軋廠，廠區在台南麻豆，1972 年成立，主力產品為 300 系冷軋不鏽鋼。資本額 32.28 億元，董事長及總經理都是葉碩堂，目前員工約 100 多人。受到不鏽鋼多年不景氣的影響，千興過去 3 年呈現虧損，今年上半年每股稅後虧損 0.5 元。(記者羅倩宜)	儘管國際經濟情勢擾攘不安，MAZDA TAIWAN 執行長朱梅君在接受記者訪問時仍樂觀預期今年國內車市仍可望比去年成長 2 成左右，但對明年第 1、2 季則持保守態度，認為仍須取決總統大選後的經濟面和國際經濟局勢能否轉穩。她表示，隨著全新 Mazda5 的即將上市，未來除將強化 Zoom-Zoom 駕馭至上精神，也將持續營造 MAZDA 的日式高質感、高品味形象，同時亦將更側重服務品質的提升。由於市場預測未來小型房車和小型 SUV 都有成長空間，因此她透露明年將考慮引進全新 Mazda CX-5，也不排除國產化可能性。也因高競爭力新車的陸續加入，她誓言 2015 年後 MAZDA 總市占率將提升至 6%。
shouldn't be excluded (label 0)	should be excluded (label 1)	shouldn't be excluded (label 0)
【丁威 台北報導】國內消費市場不景氣，國內 KTV 龍頭之一的錢櫃 (8359) 今年前 4 月業績也衰退近 2 成，為了力抗景氣「寒流」，錢櫃推出 20 周年慶祝活動，自 4 月以來推出歡唱 8 折，啤酒降價的優惠活動後，業績成長近 20 30%，而隨著暑假旺季到來，錢櫃單月營收可望恢復 3 億元的水準。另外，錢櫃轉投資的錢櫃茶餐廳，從 1 月份開幕後業績穩定成長，4 月營收約 800 萬元，5、6 月應有單月千萬的實力。錢櫃協理呂嘉正表示，近 1、2 年來民眾荷包縮水，對於 KTV 歡唱需求降低，今年錢櫃 1 4 月營收較去年同期年減 18.24%。為了刺激民眾上門消費唱歌，今年滿 20 周年的錢櫃推出歡唱 8 折優惠活動。呂嘉正說，以往錢櫃最讓消費者介意的「開瓶費」和「酒錢」，4 月份開始降價促銷以來，人潮逐漸回籠，5 月份開始啤酒全面降價，業績已經明顯成長 2 成。	泰國前總理戴克辛流亡海外 17 個月後終於回國，他雖表示絕對不會重返政壇，但分析家與過去反對他的人士均對此存疑，也認為他返國將為泰國政局投下不穩定的變數。新加坡國立大學學者蒙提桑諾 (Michael Montesano) 指出，泰國軍方前年發動的政變並未達成將戴克辛徹底逐出政壇的目的。戴克辛抱著應可獲判無罪的期待返國，勢必會重新開展政治生涯。泰國一家風險顧問公司總經理道荷堤也指出，戴克辛聲稱不再參政，「在我聽來不是實話」。2005 至 06 年間發動長期示威，終於導致戴克辛下台的抗議人士，本周一再度走上街頭，矢言將針對戴克辛及現任總理薩瑪展開反對運動。當年領導反對運動的媒體人林明達形容，現在的局勢就像「四處已灑遍燃料，只待有人放火」。台大政治系楊永明教授則分析，戴克辛流亡海外期間，軍政府無法提振泰國經濟，執政面臨極大困境，終致人民力量黨在選舉中大勝。如今戴克辛能夠回國，應是與泰皇及曼谷政界人士達成某種交換協議，對政局應有穩定作用。他認為，戴克辛日後仍將在泰國政壇扮演關鍵性角色。編譯李寧怡。	總統馬英九競選期間提出的募兵支票恐怕面臨跳票？國防部長陳肇敏昨在立法院外交國防委員會中表示，推動募兵制的確有困難，且要推動到全面募兵制，「是國防部嚴峻的考驗」。陳肇敏指出，國防預算佔 GDP (Gross Domestic Product，國內生產毛額) 百分之三，但以現有國軍員額來看，根本不可能達成全面募兵制，因此國軍要再精進是必然方向。而我國走的是一個海島守勢作戰，各軍種人員配比都要進行檢討，並將繼續推動「再精進案」，將目前二十七萬大軍減至二十到二十五萬人以內。陳肇敏並指出，目前募得許多素質相當不錯的國軍弟兄，主要與過去八年經濟不景氣，工作難找有關，未來經濟好轉後，是否還有誘因？另外，台灣人口有限，國防部會努力推動募兵制，但不能因滿足「量」的需求而減低「質」的要求。至於重大軍購案，陳肇敏強調向美國購買 F-16 C/D 戰機有其必要，但其他軍購案如潛艦，因目前我方連採購潛艇的型號都沒有，這部分會再研究，會就作戰需求再作評估。記者王 <U+70F1> 華

Notes: ADD NOTES HERE

**prompt**

**System:**

I am an economist working on constructing Taiwan's Economic Policy Uncertainty Index (EPU index). My primary goal is to classify whether a news should be excluded when constructing EPU index in Taiwan. There are two criteria I'm considering to exclude a news.

Criterion1:

The main idea of the news is either historical accounts or abstract subjective inferences, which won't impact Taiwan's economics for sure. Hence, this kind of news should be excluded.

Criterion2:

The main idea of the news is not related with Taiwan. For example, the people or companies mentioned in the news have nothing to do with Taiwan or the events in the news don't actually happen within Taiwan. I will exclude the news as well.

Notice that you can first justify whether there is a person, company or event in news related to Taiwan. If there isn't any, it should be excluded with high probability.

**{few shot examples}**

**Human:**

Help me complete the classification task identifying whether the given news should be excluded.

**News: {one new from test set}**

**Output Instructions:**

The output should be formatted as a JSON instance that conforms to the JSON schema below.

As an example, for the schema "properties": "foo": "title": "Foo", "description": "a list of strings", "type": "array", "items": "type": "string", "required": ["foo"] the object "foo": ["bar", "baz"] is a well-formatted instance of the schema. The object "properties": "foo": ["bar", "baz"] is not well-formatted.

Here is the output schema: **{json schema}**

Besides, don't forget to escape a single quote in the reason section and be aware of your reasoning's token length.

Furthermore, we explored additional prompting strategies, including the Chain of Thought (CoT) method proposed by the Google Research Brain Team. CoT has demonstrated a broad impact, and in prior studies, its effectiveness has been shown in tasks related to mathematical derivation and common-sense reasoning. In Table 3, we illustrate our endeavor to incorporate step-by-step reasoning into our approach.

Table 2: CoT reasoning in few-shot prompt

should be excluded (label 1)	shouldn't be excluded (label 0)	should be excluded (label 1)
First, the news is related to Taiwan's economics as it mentions Taiwan is one of the countries which have experienced an early economic rebound. Second, the news is related to Taiwan's policy as it states the expectation timing of a interest rate hike. Finally, there isn't any further discussion about uncertainty. This news doesn't simultaneously mention economics, policy and uncertainty. Hence, it should be excluded when constructing Taiwan's EPU index.	First, the news is related to Taiwan's economics as the program discussed is related to economics and it's the Taiwan's Prime Minister who propose it. Second, the news is related to Taiwan's policy. Finally, the news is related to uncertainty as the writer of this news is doubt about the effectness of the policy. This news simultaneously mention economics, policy and uncertainty. Hence, it shouldn't be excluded when constructing Taiwan's EPU index.	First, the news is not directly related to Taiwan's economics as it discuss mainly about the global economics condition. Second, the news is not related to the policy. Finally, the news is related to the uncertainty as it discuss various downturns of the stock market in many countries. This news doesn't simultaneously mention economics, policy and uncertainty. Hence, it should be excluded when constructing Taiwan's EPU index.
shouldn't be excluded (label 0)	should be excluded (label 1)	shouldn't be excluded (label 0)
First, the news is related to Taiwan's economics as it points out the relation between the quality of voluntary military and the Taiwan's economics condtion. Second, the news is related to the policy regarding the voluntary military service. Finally, the news is related to the uncertainty as it discuss some possibilities and possible progress of the policy. This news simultaneously mention economics, policy and uncertainty. Hence, it shouldn't be excluded when constructing Taiwan's EPU index.	First, the news is related to Taiwan's economics as it's about the CEO of Mazda Taiwan expressing optimism about the domestic car market. Second, it's obvious the news isn't related to the policy. Finally, the news is related to the uncertainty as it states the existence of the uncertainty within the car market depending on the global economics condition and the results of the election. This news doesn't simultaneously mention economics, policy and uncertainty. Hence, it should be excluded when constructing Taiwan's EPU index.	First, the news is related to Taiwan's economics as the main discussion is about how Korean's political situaion will affect Taiwan'e financial market. Second, the news is related to the policy as the news metion the central bank will try to stabalize the market. Finally, it's obvious that the news is related to the uncertainty result from the Korean's political situation. This news simultaneously mention economics, policy and uncertainty. Hence, it shouldn't be excluded when constructing Taiwan's EPU index.

Notes: ADD NOTES HERE

The intuition behind CoT is analogous to solving a math problem. When faced with a mathematical challenge, our approach involves thinking step by step, linking these steps together to arrive at the final answer. Applying this idea of chaining, we break down our task into three components: Economy, Policy, and Uncertainty—the fundamental concepts of EPU.

Despite the disappointing results, the concept used to identify whether a news item is related to EPU appears to be more complex and abstract. Consequently, it's highly challenging to provide the reasoning part in few-shot examples, prompting the need for ChatGPT to break down the important concepts and make more logical statements or reasoning. We hope future literature can further unveil the hidden potential of this prompting technique and develop a more concrete strategy for the EPU construction task.

Table 3: ChatGPT answering

one of news in test set (label 1)	CoT answer	standard few-show answer
〔編譯許世函／綜合報導〕中國央行週四晚間出乎外界意料宣布降息一碼（．二五％）。美國聯準會（Fed）主席柏南克則表示，目前美國不會推出寬鬆政策；但副主席萊倫週三卻暗示，美國應推更多寬鬆政策。美、中兩國舉動，透露各國益發擔憂歐債危機惡化威脅到全球經濟成長。中國央行週四晚間宣布，自週五起下調金融機構一年期貸款基準利率一碼、至六．三一％，一年期存款基準利率下調一碼、至三．二五％，以支持經濟增長並推動金融系統改革，中國降息動作作為二八年金融風暴以來首見。...the following news are omitted	First, the news is not directly related to Taiwan's economics as it mainly discusses the actions and statements of the central banks in China and the US. Second, the news is not related to Taiwan's policy. Finally, the news is related to uncertainty as it mentions concerns about the European debt crisis and the global economic growth. This news doesn't simultaneously mention economics, policy and uncertainty. Hence, it should be excluded when constructing Taiwan's EPU index.	The news is about the actions of central banks in China and the United States, which are not directly related to Taiwan's economics.
one of news in test set (label 0)	CoT answer	standard few-show answer
【陳文蔚 台北報導】原以為大選紛擾將影響房地產市場，不過國有財產局發現，選後房市熱度不但沒降，標售成功案件還比去年同期多上 1 成，連決標價格也比去年高出 2 倍。國有財產局主任秘書蘇維成說：「政治干擾的短期影響慢慢解除。」總統大選過後，由於政治不確定因素升高，讓正值復甦期的房地產市場，蒙上陰影，連國產局也一度認為，政治因素可能對房地產出現短暫影響。...the following news are omitted	The news discusses the post-election impact on the real estate market, indicating that the market remains active with an increase in successful sales and higher bidding prices. This shows positive signs for the real estate market and contradicts the expectation of political interference. Hence, it should not be excluded when constructing Taiwan's EPU index.	The news provides information about the post-election real estate market in Taiwan, which is relevant to Taiwan's economy.

Notes: ADD NOTES HERE

So far, we’ve illustrated various prompt engineering concepts, and one might understand the benefits of providing some examples. However, can we input as many examples as we want without any restrictions? The answer is quite straightforward—no. There is a limit to restrict the input context length for ChatGPT. In our experiment, to utilize the research fund in the most efficient way, we used the base 3.5 model with 4k tokens as long as we could. For longer news, we had to switch to the 16k model, which comes with a much higher cost. It’s worth mentioning that the 3.5 model is not the most powerful model provided by OpenAI; there is a 4.0 version renowned for its better understanding not only of textual data but also images, known for its multimodal capability.

When it comes to using a Large Language Model, there’s always a two-choice problem: whether to go for fine-tuning or simply conduct prompt engineering. Although, thanks to the pre-trained capability, ChatGPT is already ready-to-use in many tasks, we also gave fine-tuning a shot.

It’s crucial to emphasize that there are many aspects of fine-tuning techniques. Traditionally in the NLP domain, it refers to training the model on top of pre-trained model parameters. There might be concerns that the parameters could get polluted when there is a lot of noise in the personalized data of the downstream task. In our case, we used the fine-tuning API provided by OpenAI, and most importantly, we don’t know how it works. As our main goal is to explore the possibility of ChatGPT, this is the only way to perform fine-tuning.

One benefit of fine-tuning is that as long as we have the fine-tuned model, there is no need to provide few-shot examples in each request. In other words, we can use zero-shot prompting to replace few-shot prompting with ease. Nonetheless, the cost is expensive because the workflow is intensive, and it’s not as straightforward as asking a question and receiving a response. There is always a trade-off, and we are curious about which approach is more suitable for constructing the EPU index.

The result of fine-tuning can’t compete with the best-performing few-shot prompting, especially in the case where the provided training instances are only labels (without reason). We observed that the training process reaches a bottleneck when approaching 100 steps, and the loss also seems to have no room for improvement. Interestingly, when we add reasoning to our training instances, meaning we force ChatGPT to answer with a label plus a reason, the loss seems to improve, but the evaluation metrics don’t show much difference. This is because we have limited knowledge of the process, and only a few parameters can be tuned, such as the epoch and learning rate. We decided to conclude our exploration at this point.

In conclusion, the few-shot prompting with six examples performed the best. This outcome sheds

light on how powerful the foundation model is, revealing that even when provided with only six examples, it can yield substantial improvements. Although in the current state ChatGPT3.5 can't outperform deep learning models when constructing the EPU index, it provides an alternative that strikes a balance between training time and result performance.

Table 4: Evaluation Metrics for Test Set Contained in 7000 News Articles

	(1)	(2)	(3)	(4)
	micro_f1	macro_f1	weighted_f1	precision_0
8 shot with reason	0.707	0.705	0.799	0.707
6 shot with reason	0.672	0.674	0.762	0.677
3 shot with reason (fine-tuned 1000)	0.555	0.598	0.63	0.58
6 shot no reason	0.54	0.585	0.621	0.566
zero shot with reason (fine-tuned 1000)	0.417	0.579	0.59	0.472
zero shot no reason (fine-tuned 1000)	0.372	0.588	0.589	0.438
zero shot with reason	0.401	0.58	0.588	0.459
zero shot no reason	0.415	0.572	0.587	0.469

Notes: ADD NOTES HERE

Figure 1: confusion matrix of best performing model

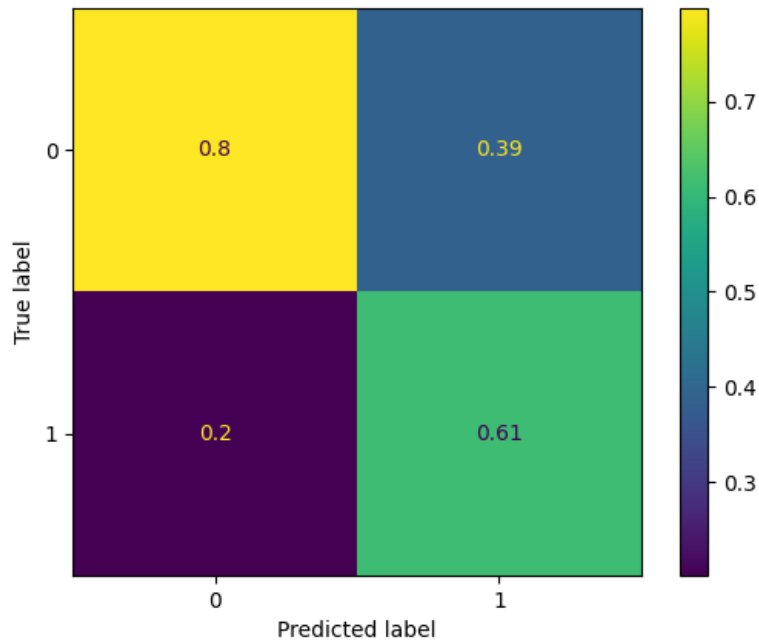


Figure 2: loss of fine tuning without reason



Figure 3: loss of fine tuning with reason

