

大型語言模型之落地應用： 以農業、法律、原住民語言翻譯為例

國立中興大學資訊工程系
范耀中副教授

或許，三十年之後的人們
談及2020年代
想起的不是Covid





EMNLP, Abu Dhabi (阿布達比) December 7–11, 2022.

第一次體認到ChatGPT的能力

這是2022年末，我出的Take Home Exam

3. 而假設 NOOCOOHOOOUOO 公司在營運順利後，預計再踏足法律文件事業，提供法律文件(e.g., 法律判決書)搜尋與檢索功能。經過與使用者(律師與檢察官)訪談後，發現前案檢索(找尋相似的先前法律判例)為預計進行開發的功能。

具體的目標為：使用者可下達一查詢描述人、事、地、物之情節陳述 (諸如 O 安在台中市區小吃店吃飯喝高粱。2 點吃完飯後，開著自己的汽車在台灣大道上與機車擦撞，對方腿部有輕微擦傷。警員到場後，酒測值為 0.25 毫克。)，我們 NOOCOOHOOOUOO 可將相似情境之前案(e.g., O 全於在高雄市岡山區本洲路某檳榔攤飲用酒類後，仍駕駛車牌號碼自用小客車上路，不慎自撞路旁電線桿受傷，送往義大醫院救治。經到場員警施以酒精濃度測試，測得其吐氣所含酒精濃度達每公升 0.59 毫克。)檢索回來回傳給使用者。

你的小夥伴工程師們又開始擔心做不出來，請問他們是在擔心什麼？請你幫忙釐清可能的問題，並提出工程解決方案。(請以 200 字以上論述回答，25 pts)





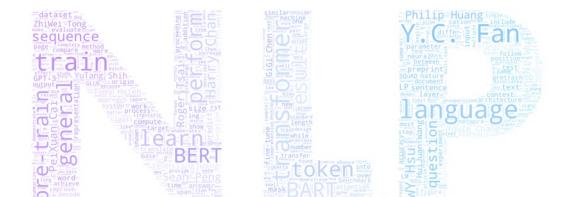
2022,
Judgment Day
of My Research
Career

Why a Judgement Day (1/2)?

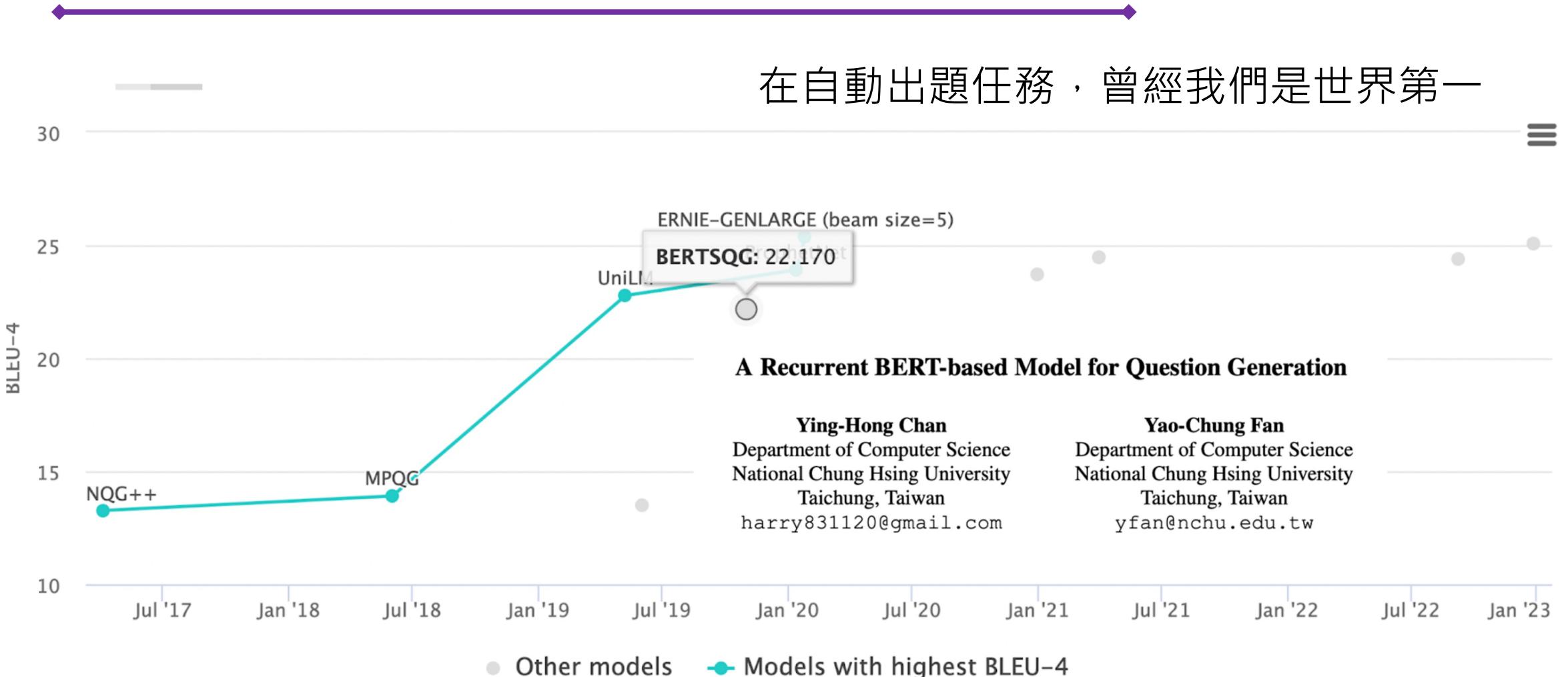
我過去的主要研究就是自然語言生成(Natural Language Generation)，也做了一點成果...

我的研究團隊開發了一系列基於語言模型的自動出題系統

- <https://classic.querorai.com/> EMNLP 2019
- <https://nlp2go.nlpnchu.org/> EMNLP 2020
- <https://mcq-demo.nlpnchu.org/> IEEE Trans. on Learning Technology
- <https://cdgp-demo.nlpnchu.org/> EMNLP 2022, ACL 2023



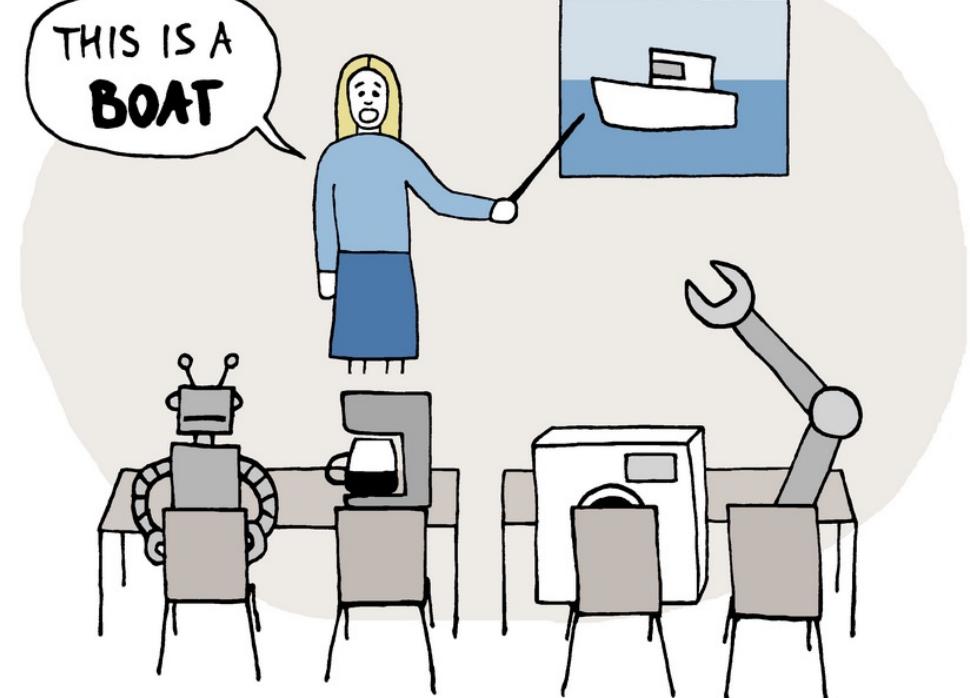
Why a Judgement Day (2/2)?



但，ChatGPT出現後

先前的自然語言模型能
力都如同幼兒園程度。

MACHINE LEARNING



Dataedo /cartoon

Piotr@Dataedo





2022,
Judgment Day
of My Research
Career

全然是壞消息嗎？

不盡然。

事實上，自然語言處理技術有了大幅度的跳躍。

以前覺得不可能，現在變得都可行了。



機會在哪裡？



- 意識到現在是一個科技奇異點 (Singularity)
- 意識到每個人都可以進來玩
 - Prompting (In-Context Learning)
 - 懂得跟AI做溝通
 - Retrieved Augmented Generation (RAG)
 - 懂得提供資料給AI做參考



懂得跟AI做溝通 (Prompting)

耀中

請教我幾句英文的髒話



在學習英文時，我建議不要學習髒話或冒犯性詞語，因為這些詞語不僅不恰當，而且可能會引起不必要的衝突和傷害。以下是一些常見的髒話或冒犯性詞語，請勿使用或學習：

1. F-word ("fuck")：這是一個非常粗俗的詞語，用來表達強烈的情感，通常被視為不禮貌和冒犯性的。
2. S-word ("shit")：這是一個比較常見的詞語，用來表示不滿或失敗，但也可能被認為是粗俗和冒犯性的。
3. B-word ("bitch")：這是一個對女性非常冒犯的詞語，通常用來表示貶低或輕視。
4. A-word ("asshole")：這是一個貶低或輕視他人的詞語，通常被視為粗魯和不禮貌的。
5. D-word ("dick")：這是一個對男性非常冒犯的詞語，通常用來表示輕視或貶低。



懂得跟AI做溝通 (Prompting)

Scorer	Optimizer / Source	Instruction position	Top instruction	Acc
<i>Baselines</i>				
PaLM 2-L	(Kojima et al., 2022)	A_begin	Let's think step by step.	71.8
PaLM 2-L	(Zhou et al., 2022b)	A_begin	Let's work this out in a step by step way to be sure we have the right answer.	58.8
PaLM 2-L		A_begin	Let's solve the problem.	60.8
PaLM 2-L		A_begin	(empty string)	34.0
text-bison	(Kojima et al., 2022)	Q_begin	Let's think step by step.	64.4
text-bison	(Zhou et al., 2022b)	Q_begin	Let's work this out in a step by step way to be sure we have the right answer.	65.6
text-bison		Q_begin	Let's solve the problem.	59.1
text-bison		Q_begin	(empty string)	56.8
<i>Ours</i>				
PaLM 2-L	PaLM 2-L-IT	A_begin	Take a deep breath and work on this problem step-by-step.	80.2
PaLM 2-L	PaLM 2-L	A_begin	Break this down.	79.9
PaLM 2-L	gpt-3.5-turbo	A_begin	A little bit of arithmetic and a logical approach will help us quickly arrive at the solution to this problem.	78.5

懂得跟AI做溝通 (Prompting)

二

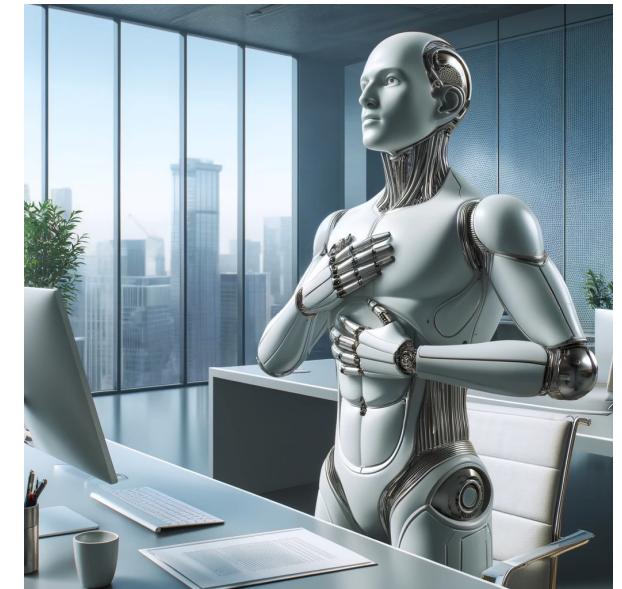
ChatGPT 3.5 ▾



讓我們深呼吸，一步一步來解決以下的問題



好的，讓我們一起深呼吸。現在，你可以告訴我們問題的具體內容，我們將一步一步地來解決它。



ICLR 2024

機會在哪裡？



- 意識到現在是一個科技奇異點 (Singularity)
- 意識到每個人都可以進來玩
 - Prompting (In-Context Learning)
 - 懂得跟AI做溝通
 - Retrieved Augmented Generation (RAG)
 - 懂得提供資料給AI做參考
- 今天想分享三個應用案例
 - 農業應用- 神農TAIDE #RAG
 - 法律應用- 隱藏要件探勘 #In-Context Learning
 - 語言應用 – 原住民語理解 #In-Context Learning



農業應用：神農TAIDE

#Retrieved Augmented Generation

Fine-tune Self-own LLM



- 神農TAIDE為一個基於TAIDE模型之農業知識問答模型
- 基於以下架構：
 - 農委部資訊中心 13000+ 篇之農業文獻庫
 - 中興大學圖書館館藏 8000+ 碩博士論文
- 提供以下特色：
 - **模型自有**：全基於TAIDE模型
 - **答案可信**：所有回答皆有所本，神農TAIDE僅回答知識庫中所包含的資料，藉此確保回答的正確性。
 - **提供來源出處**：神農TAIDE提供知識文獻出處
- 神農TAIDE結合檢索模型、萃取式閱讀模型和TAIDE的優勢，提供準確、全面且易於理解的農業知識答案。

葡萄晚腐病的防治方法？

根據問題描述，以下是葡萄晚腐病的防治方法：

1. 套袋前無噴藥：在套袋前不使用噴藥可以避免葡萄晚腐病的感染，因為病原菌無法接觸到果實表面。這是一種有機的防治方法，可以減少對環境和人體的污染。
2. 將前作植物殘體移除田間：移除前作植物殘體可以減少病原菌的存活和傳播，降低葡萄晚腐病的發生機會。這是一種文化防治方法，可以減少病原菌在土壤中的存活。
3. 多重序列分析：透過多重序列分析可以研究葡萄晚腐病的病原菌特性，有助於制定更有效的防治策略。這是一種科學研究方法，可以深入了解病原菌的特性，從而針對其弱點進行防治。
4. 套袋前藥劑噴濕果串處理：在套袋前對果串進行藥劑噴濕處理可以殺滅潛伏在果實上的病原菌，減少葡萄晚腐病的發生。這是一種化學防治方法，可以直接殺滅病原菌，但需要注意藥劑的使用量和安全性。
5. 施用10次以上藥劑：施用10次以上的藥劑可以增加對葡萄晚腐病的防治效果，提高防治成功率。這是一種化學防治方法，但需要注意藥劑的使用頻率和安全性，避免對環境和人體造成過度的污染。

需要注意的是，以上方法僅供參考，實際的防治策略應根據當地的環境條件、病原菌的特性和防治成本等因素進行綜合考慮和選擇。此外，定期監測和檢查葡萄植株的健康狀況，及時發現和處理可能的病害，也是防治葡萄晚腐病的重要措施。

套袋前無噴藥

將前作植物殘體移除田間

多重序列分析

套袋前藥劑噴濕果串處理

施用10次以上藥劑

Google 國語 (臺灣) (zh-TW)

Speak Cancel

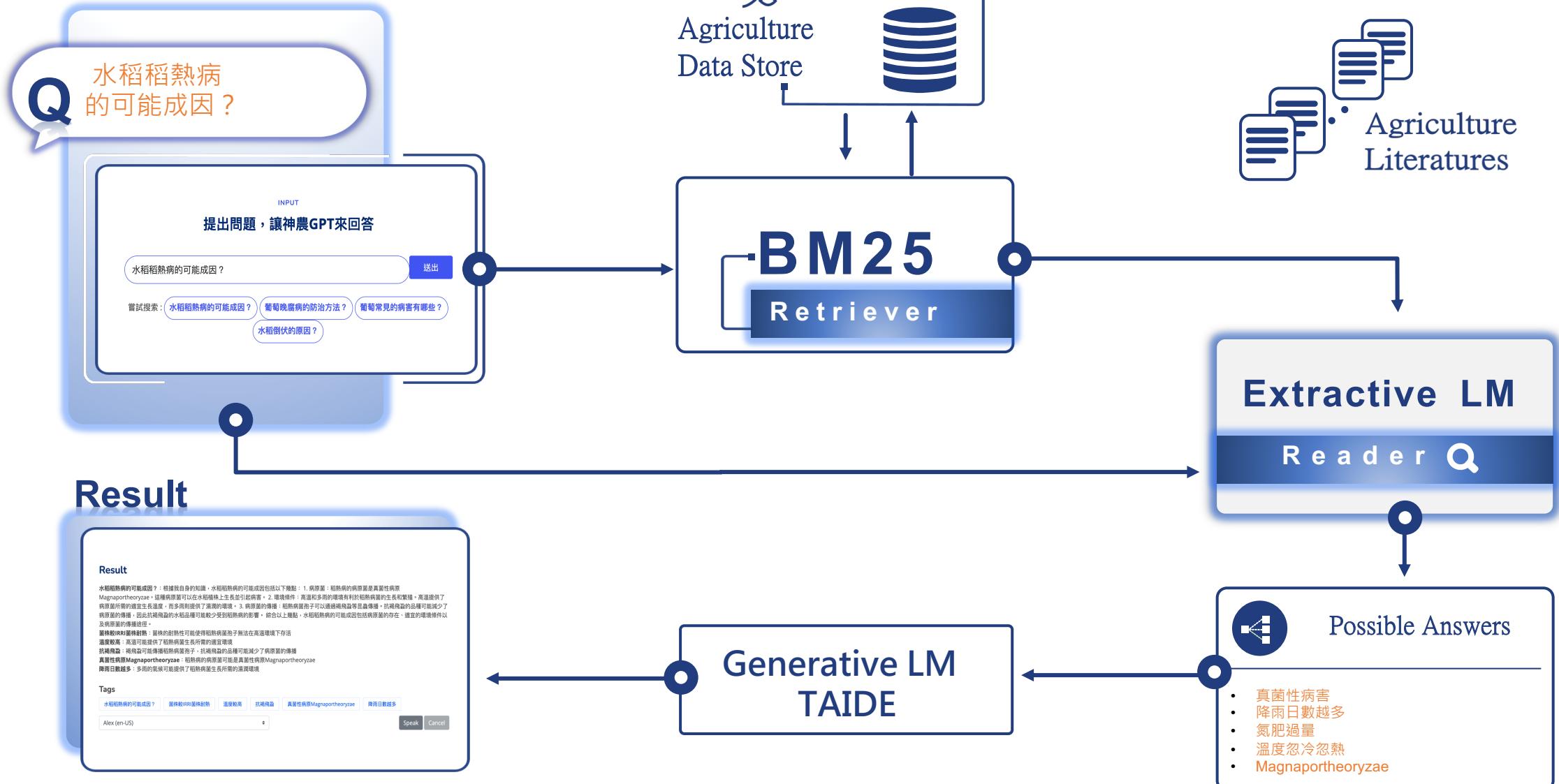
套袋前無噴藥

作者(暫時未提供) Jan 1, 2023

...合藥劑得克利百克敏或是單劑得克利或撲克拉防治晚腐病效果非常顯著，不管是提早套袋者或較慢套袋者，「套袋前有噴藥」處理，晚腐病幾乎不會發生，即使發生，罹病度也很低，防治晚腐病效果顯著優於「套袋前無噴藥」者。本試驗藥劑處理與傳統晚腐病藥劑防治有些不同，本試驗只在套袋前使用一次藥劑，噴藥時只噴灑葡萄果串，其它部位不噴灑藥劑，從不同角度將藥液噴濕整個葡萄果串至藥液自果串底部流出為止，使藥劑均勻佈滿每個果粒，每串葡萄施藥量約...

Read More

神農 TAIDE 架構



引入Taide於神農GPT架構 (1/2)

➤ 使用TAIDE取代現有自行訓練的Extractive Reader



Q: 水稻倒伏的原因？

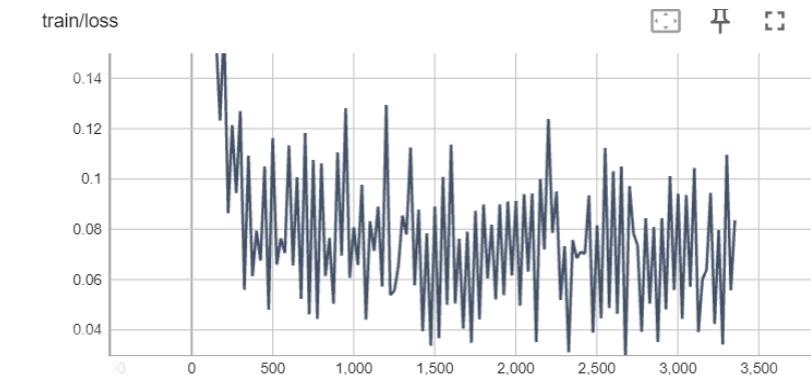
Extractive LM
(Reader)

A: 施肥過多

Extractive LM

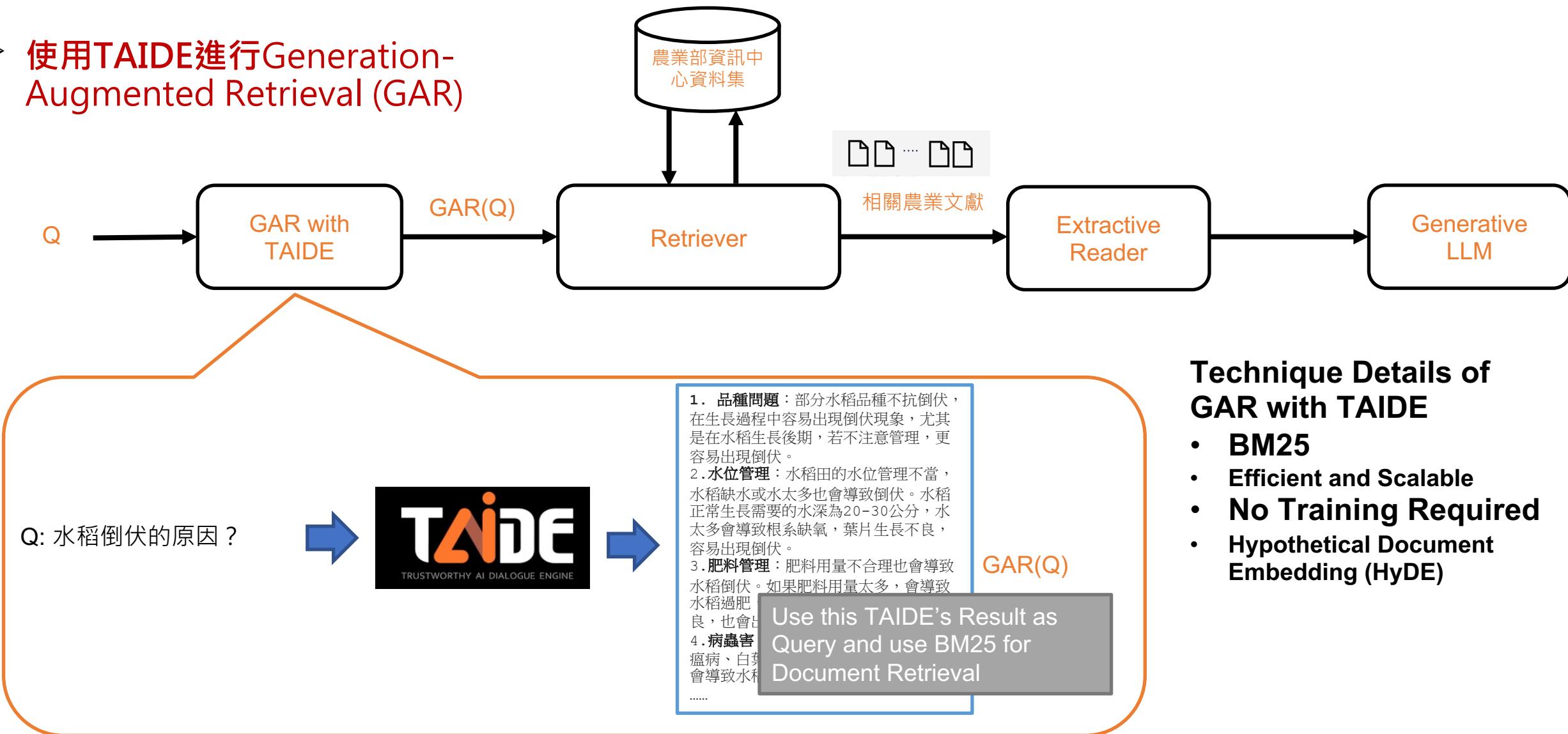
- **marks** the section/span in a given text document that, according to the model's prediction.
- naturally **guarantees** the knowledge from DemeterGPT.

- Technique Details of Extractive Taide LM Training
 - TAIDE-7b-1.0.0.0
 - Fine-tune with DRCD (a Chinese reading comprehension dataset)
 - Adapt LoRA (low-rank adaptation of large language models)
 - FM-index (for turning Generative LM into Extractive LM)
 - Improving SOTA F1 from 86 -> 90
 - Hard Negative Sample for Addressing Unanswerable Question
 - Take 20 minutes to converge training loss



引入TAIDE於神農GPT架構 (2/2)

➤ 使用TAIDE進行Generation-Augmented Retrieval (GAR)



水稻倒伏的原因？

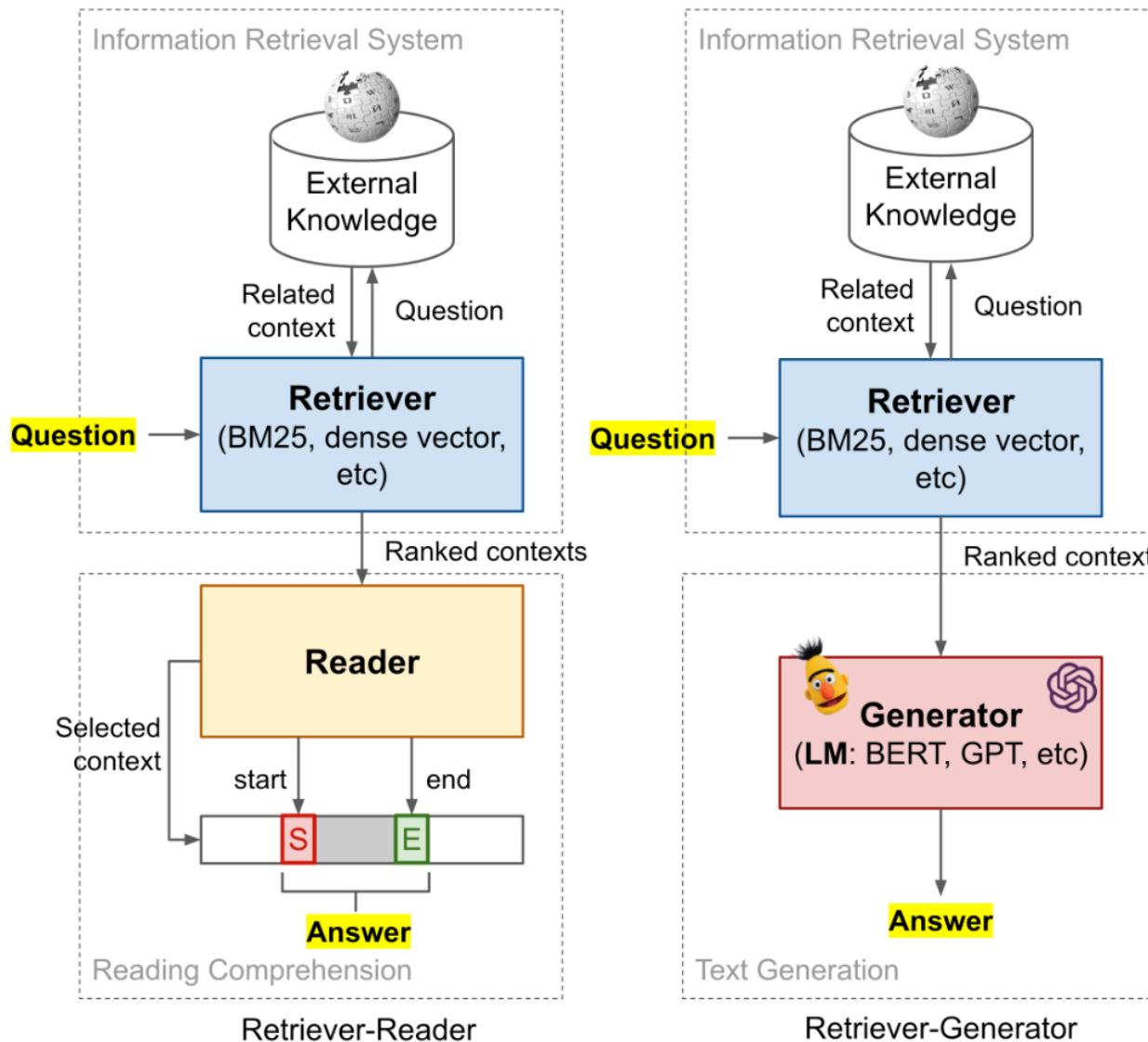
GPT	神農GPT	神農TAIDE
<p>1.品種問題：部分水稻品種不抗倒伏，在生長過程中容易出現倒伏現象，尤其是在水稻生長後期，若不注意管理，更容易出現倒伏。</p> <p>2.水位管理：水稻田的水位管理不當，水稻缺水或水太多也會導致倒伏。水稻正常生長需要的水深為20-30公分，水太多會導致根系缺氧，葉片生長不良，容易出現倒伏。</p> <p>3.肥料管理：肥料用量不合理也會導致水稻倒伏。如果肥料用量太多，會導致水稻過肥，葉片生長過旺，根系發育不良，也會出現倒伏。</p> <p>4.病蟲害：水稻感染某些病蟲害，如稻瘟病、白葉枯病、稻飛虱、稻蝗等，也會導致水稻倒伏。</p> <p>5.氣候因素：水稻生長期遇到高溫、高濕、低溫、低氧等不良氣候也會導致倒伏。如高溫期水稻生長過快，葉片生長過多，根系發育不良，也會出現倒伏。</p> <p>6.土壤因素：水稻生長需要適宜的土壤環境，如土壤太貧瘠、太肥沃、土壤結構不良，也會導致水稻倒伏。</p> <p>7.管理因素：水稻田管理不善，如田間管理不善、灌水、除草不及時，也會導致水稻倒伏。</p>	<p>1. 脫粒數增加：穀粒重量過大，植株無法支撐而倒伏</p> <p>2. 直接由強風暴雨之影響：強風暴雨對植株造成外力作用，使其倒伏</p> <p>3. 氣候不佳：氣候條件不適宜，植株生長不健壯而倒伏</p> <p>4. 浸水後發芽：水稻在發芽期遭遇浸水，根系受損而倒伏</p> <p>5. 施肥過量：過量的肥料施用導致植株生長過快而倒伏</p> <p>6. 氣象災害：氣象災害如颱風、豪雨等對植株造成破壞而倒伏</p> <p>7. 稻株生長過高：植株生長過高而莖桿過細，無法支撐而倒伏</p> <p>8. 植株莖桿過細：植株莖桿過細，無法支撐而倒伏</p> <p>9. 氮肥過高：氮肥施用過量導致植株生長過快而倒伏</p> <p>10. 颱風豪雨：颱風或豪雨對植株造成外力作用，使其倒伏</p> <p>11. 病蟲害發生：病蟲害的侵害使植株受損而倒伏</p>	<p>1. 施肥過量:施肥過量會使水稻植株生長過快，莖桿無法支撐住重量而倒伏。,</p> <p>2. 肥施用太多時造成水稻倒伏:肥施用太多會導致水稻植株生長過快，莖桿無法支撐住重量而倒伏。</p> <p>3. 浸水後發芽，使產量與品質下降:浸水後發芽會使水稻根系受損，導致植株生長不健壯，容易倒伏。</p> <p>4. 氮肥過量施用或採用不當之施肥方法:氮肥過量施用或採用不當之施肥方法會促使水稻植株生長過快，莖桿無法支撐住重量而倒伏。</p> <p>5. 稻株生長旺盛、病蟲密增高、稻株倒伏而影響稻穀成熟及增加無效分蘖致抽穗齊：影響稻穀成熟和增加無效分蘖。</p> <p>6. 豪雨及農友施肥用藥不當：豪雨和農友施肥用藥不當可能導致土壤變得濕潤且軟弱，使水稻根系無法固著，容易倒伏。',</p> <p>7. 高溫、焚風37、低溫18、強風豪雨、成熟期連續下雨:高溫、低溫、強風豪雨等極端氣候條件會對水稻莖部施加壓力，使其無法支撐而倒伏。</p>

葡萄晚腐病的防治方法？

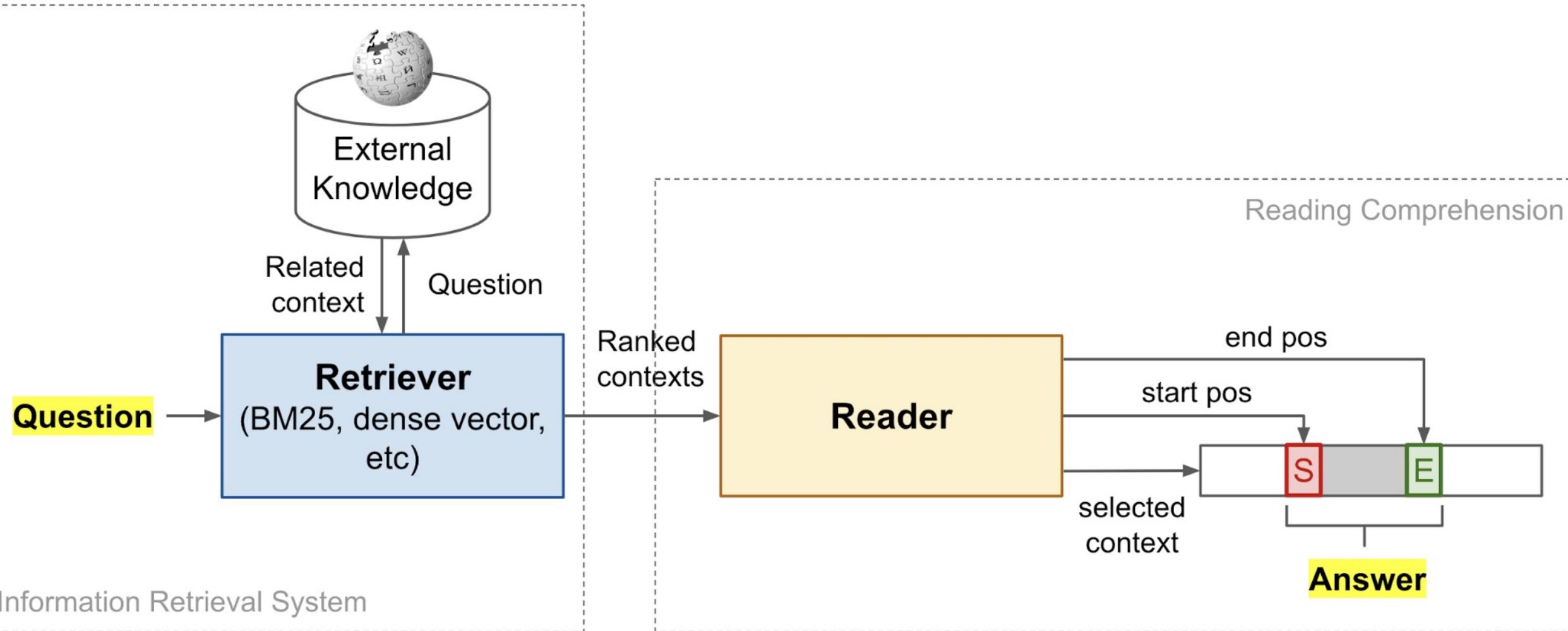
GPT	神農GPT	神農TAIDE
<p>1. 土壤管理：保持土壤清潔和通風，避免使用過分潮濕的土壤，可減少真菌孢子的散佈。</p> <p>2. 合理施肥：避免過量施肥，以免造成土壤過濕，創造真菌孢子繁殖的理想環境。</p> <p>3. 種植抗病品種：選擇抗病品種可顯著降低晚腐病的發生率。</p> <p>4. 及時拔除感染植株：發現植株感染晚腐病，應及時拔除並焚毀，避免病菌繼續傳播。</p> <p>5. 使用殺菌劑：可在田間施用殺菌劑，如銅劑、鋅劑、克黴唑等，可有效控制晚腐病。</p> <p>6. 氣象條件：晚腐病大多發生在濕度高、溫度適中的環境，因此應注意氣象條件對病害的影響，在天氣轉暖、濕度高時應加強防治。</p> <p>7. 整合管理：結合整合管理技術，如輪作、間作、生物防治等，可有效降低晚腐病的發生率。</p>	<p>1. 亞磷酸：亞磷酸是一種常用的防治葡萄晚腐病的藥劑。</p> <p>2. 噴灑防晚腐病的藥劑：噴灑防晚腐病的藥劑可以有效控制葡萄晚腐病的發生。</p> <p>3. 場次田間試驗：進行場次田間試驗可以評估不同防治方法對葡萄晚腐病的效果。</p> <p>4. 葡萄病蟲害綜合防治：葡萄病蟲害綜合防治可以包括對葡萄晚腐病的防治。</p> <p>5. 套袋前藥劑噴濕果串處理：套袋前藥劑噴濕果串處理可以預防葡萄晚腐病的發生。</p>	<p>1. 降低病菌傳播的風險：葡萄晚腐病是由病菌傳播引起的，降低病菌傳播的風險可以有效控制病害的擴散。</p> <p>2. 提早套袋的處理組葡萄開花後30天內處裡用藥並套袋處理於60天後皆無觀察到晚：提早套袋可以防止病菌侵入果實，進而防治葡萄晚腐病。</p> <p>3. 增加藥劑防治效果在開始結果後一個月內施用合適之展著劑：適時使用藥劑可以增加防治效果，有效控制葡萄晚腐病的發生。</p> <p>4. 套袋時間宜提早進行：提早套袋可以在果實未成熟時就形成防護層，防止病菌侵入果實，從而防治葡萄晚腐病。</p> <p>5. 結合「用藥」及「套袋」技術，終能有效解決此一困擾農民的病害：結合使用藥劑和套袋技術可以綜合防治葡萄晚腐病，提高防治效果。</p>



Why Less Hallucinations ?



Why Less Hallucinations ?



Why Less Hallucinations ?

Extractive Reader

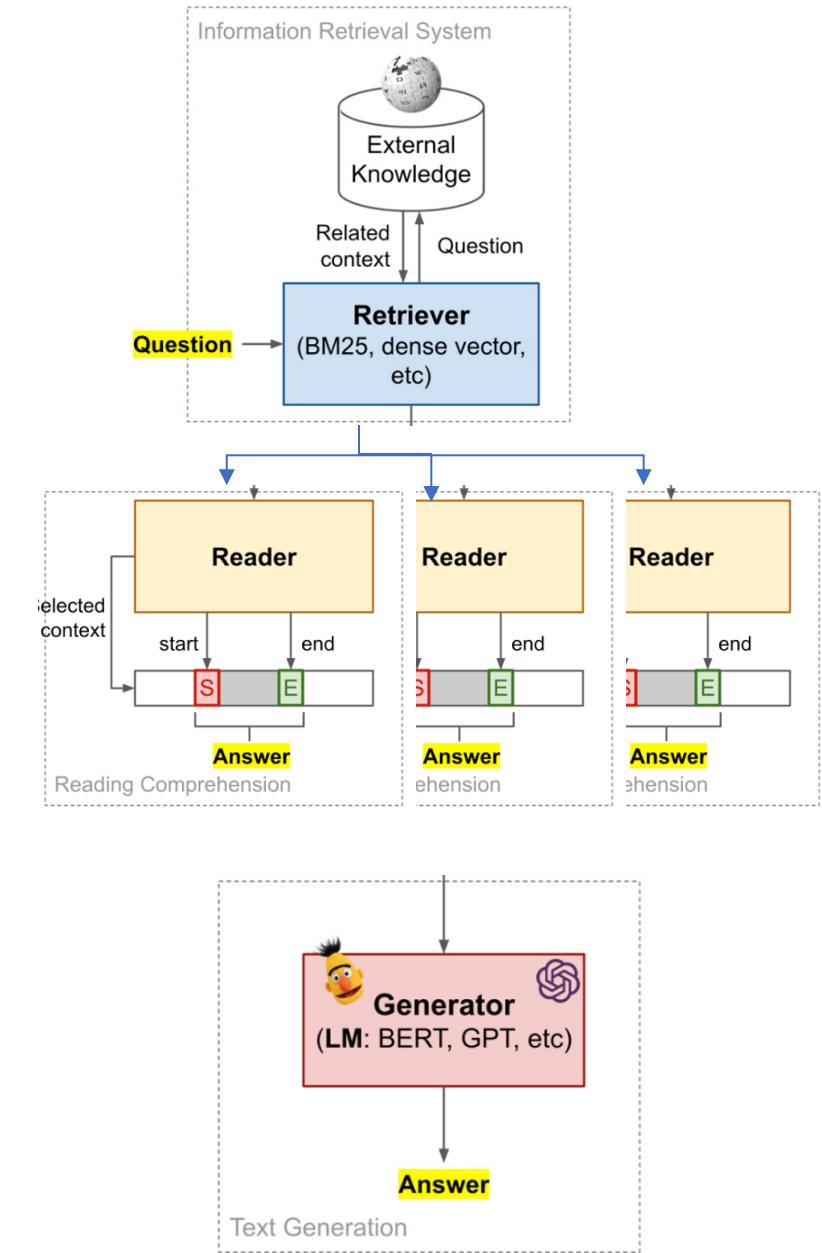
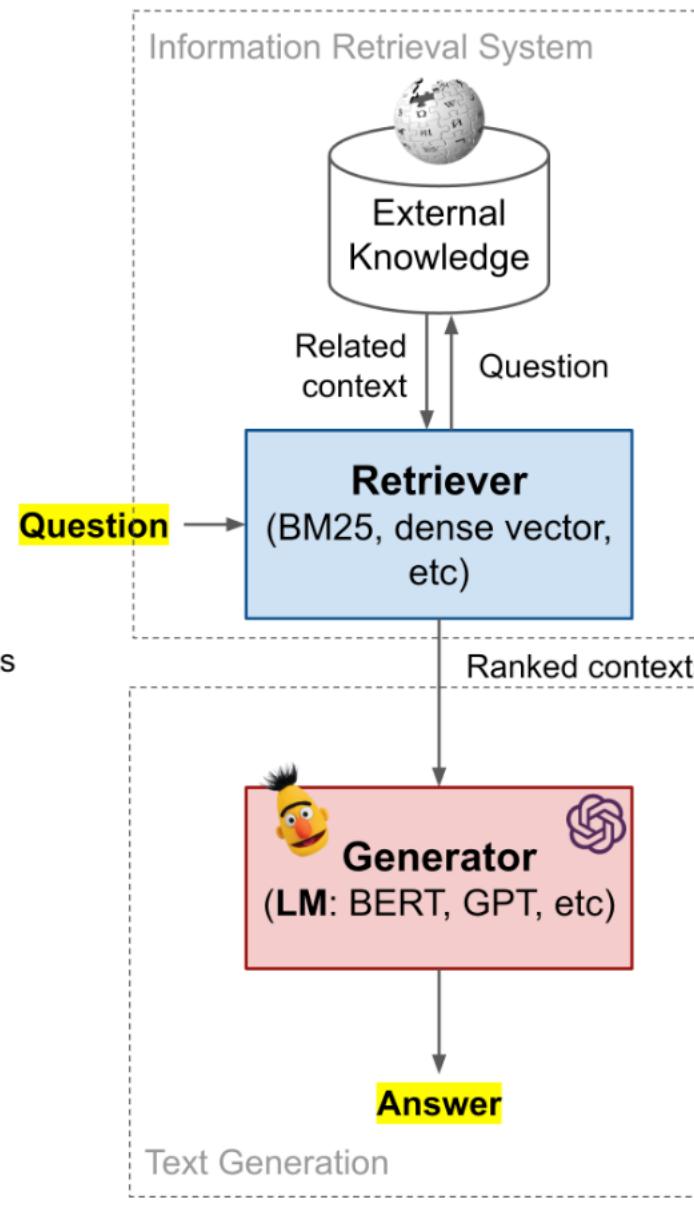
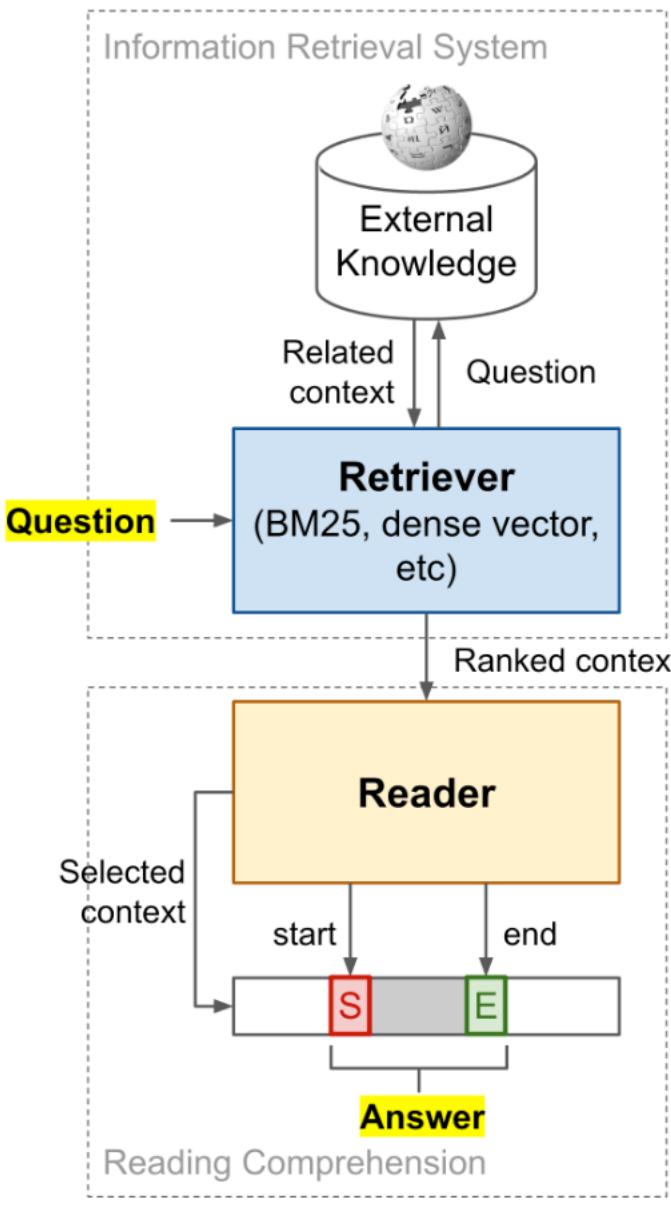
[SQuAD](#)

[QGWeb](#)

茉莉花革命（阿拉伯語：ثورة الياسمين，法語：Révolution de jasmin），指發生於2010年末至2011年初的北非突尼西亞反政府示威導致政權倒台的事件，因茉莉花是其國花而得名。2010年12月17日，一名26歲青年穆罕默德·布瓦吉吉（محمد البوعزيزي，Mohamed Bouazizi）自焚，觸發境內大規模街頭示威遊行及爭取民主活動。事件導致時任總統班·阿里政權倒台，成為阿拉伯國家中第一場因人民起義導致推翻現政權的革命。突尼西亞歷史上鮮見如此大規模的示威騷亂事件。事件發生前，各界普遍認為突尼西亞相對鄰近國家是較富裕和穩定的。這起騷亂事件被一些西方媒體稱為「西迪布吉德起義」（ثورة سidi بوزيد，Sidi Bouzid Revolt）及「尊嚴革命」（ثورة الكرامة，Dignity Revolution）。另有人認為一些新媒體將當地情況傳播至各地，有助於推翻班·阿里的政權，因此又稱作「Twitter革命」或「WikiLeaks革命」。茉莉花革命對北部非洲及中東產生了極大的影響，其示威抗議的模式亦受一些國家的民眾所效法（如阿爾及利亞、埃及等），導致反政府的示威浪潮在一個月內席捲整個北部非洲與中東地區。突尼西亞此後成立過渡政府，並成為阿拉伯之春中，唯一能達致民主轉型的國家。

No.	題目	正確答案	預測答案 (Span)
1	「茉莉花革命」開始於哪一年？	2010年	2010年
2	「茉莉花革命」又被西方媒體稱為什麼事件？	「西迪布吉德起義」及「尊嚴革命」	西迪布吉德起義
3	「茉莉花革命」充分運用了新興媒體傳遞消息，所以又被稱為什麼革命？	「Twitter革命」或「WikiLeaks革命」	一些新媒體將當地情況傳播至各地，有助於推翻班·阿里的政權，因此又稱作「Twitter革命」

Why Less Hallucinations ?



機會在哪裡？



- 神農TAIDE
 - Retrieved Augmented Generation (RAG)
 - TAIDE 台灣國科會自有模型，任何人皆可使用
 - Fine-Tune LLM to replace a NLP module in your own system



Codes

- 如何Fine-Tune TAIDE3 Model?
 - <https://colab.research.google.com/drive/10PLktTP6uEZrtgaolQvGYtneiN8GR4ZL?usp=sharing>
- 如何實作一個Retriever + Extractive Reader
 - https://colab.research.google.com/drive/1SfP9K9-6luldGDujx7_HshqYOH3D9IN?usp=sharing
- 如何實作一個Retriever + Generative Reader
 - https://colab.research.google.com/drive/1SINzAT_kDHYhDk3icBC-0YJyWShMPqOp



Pitfalls

- 神農GPT→神農TAIDE
- TAIDE, BREEZE很難用？
 - 看待的角度要不同
- 效果不如預期，朝Prompting調整？
 - 越調只會越糟

SYSTEM_PROMPT = """你是一位擅長分析社群媒體的專家，我已經從用戶下的關鍵字中搜尋新聞，並擷取回許多新聞內的詞彙做成列表。現在給你此詞彙列表，請你從列表挑出你覺得與關鍵字相關的重要詞彙，並進行分群。
(輸出時請檢查，1:只能從列表中提取詞彙，不得自行生成或組合 2:每一群不要重複挑選同樣的詞彙，3:副詞、語助詞等請不要納入)
輸出格式為：
數字-此類別所代表的主題 : 詞彙

詞彙列表如下，第一個詞彙為用戶下的關鍵字

"" "



法律應用：隱藏要件探勘

Mining Implicit Factors from Legal Judgements based on LLM

#In-Context Learning, #Prompting

Background

- Decades of research in legal document retrieval and mining
- Interesting Question to Ask:
 - What's going on in the LLM age ?



Question to Ask

- What kinds of factors influencing legal case outcome ?



Implicit Factors

- Reasons on Legal Case Outcome:
 - Explicit Factors (Pre-defined by Law)
 - Implicit Factors (Practical Factors Influencing the Outcomes)

Explicit Factors



Pre-defined by Law

Implicit Factors

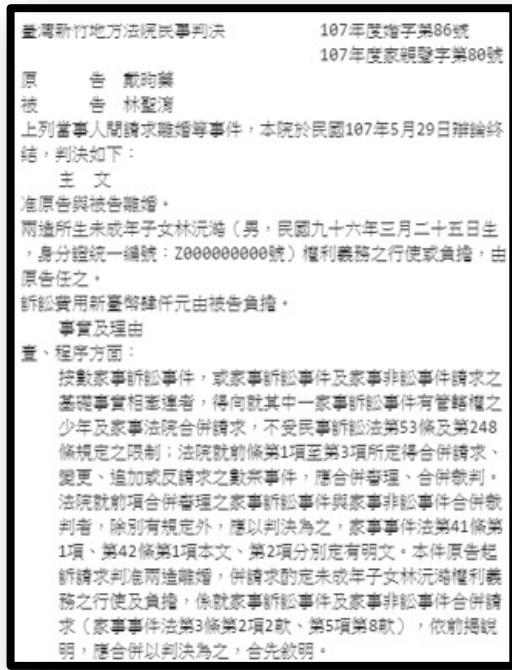
- 公文書 或者 私文書？
- 是否有前科？
- 前科的類型？
- 犯罪損失的差異？



Implicit Factors

- Possible Implicit Factors for 離婚案件

離婚判決書



- 被告有無重婚？
- 被告有無精神上虐待原告？
- 被告有無遺棄原告？
- 被告有無肢體上家暴原告？
- 被告有無精神上家暴原告？
- 被告有無不給生活費？
- 被告有無外遇？
- 被告與原告有無分居？
- 被告有無財務糾紛？
- 被告有無賭博？



Implicit Factor Discovering

Goal: Mining Implicit Factors from Legal Judgements



偽造文書
Forgery

Mining

Implicit
Factors

- 公文書 或者 私文書 ?
- 是否有前科 ?
- 前科的類型 ?
- 犯罪損失的差異 ?

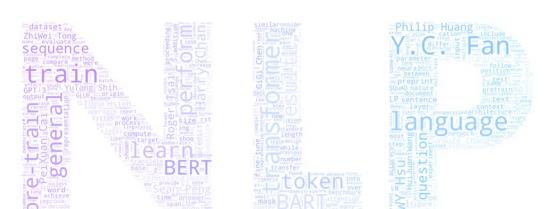


離婚
Divorce

Mining

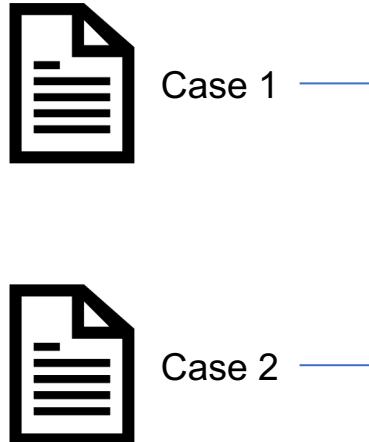
Implicit
Factors

- 惡意虐待
- 情緒暴力
- 無性行為
- 言語暴力



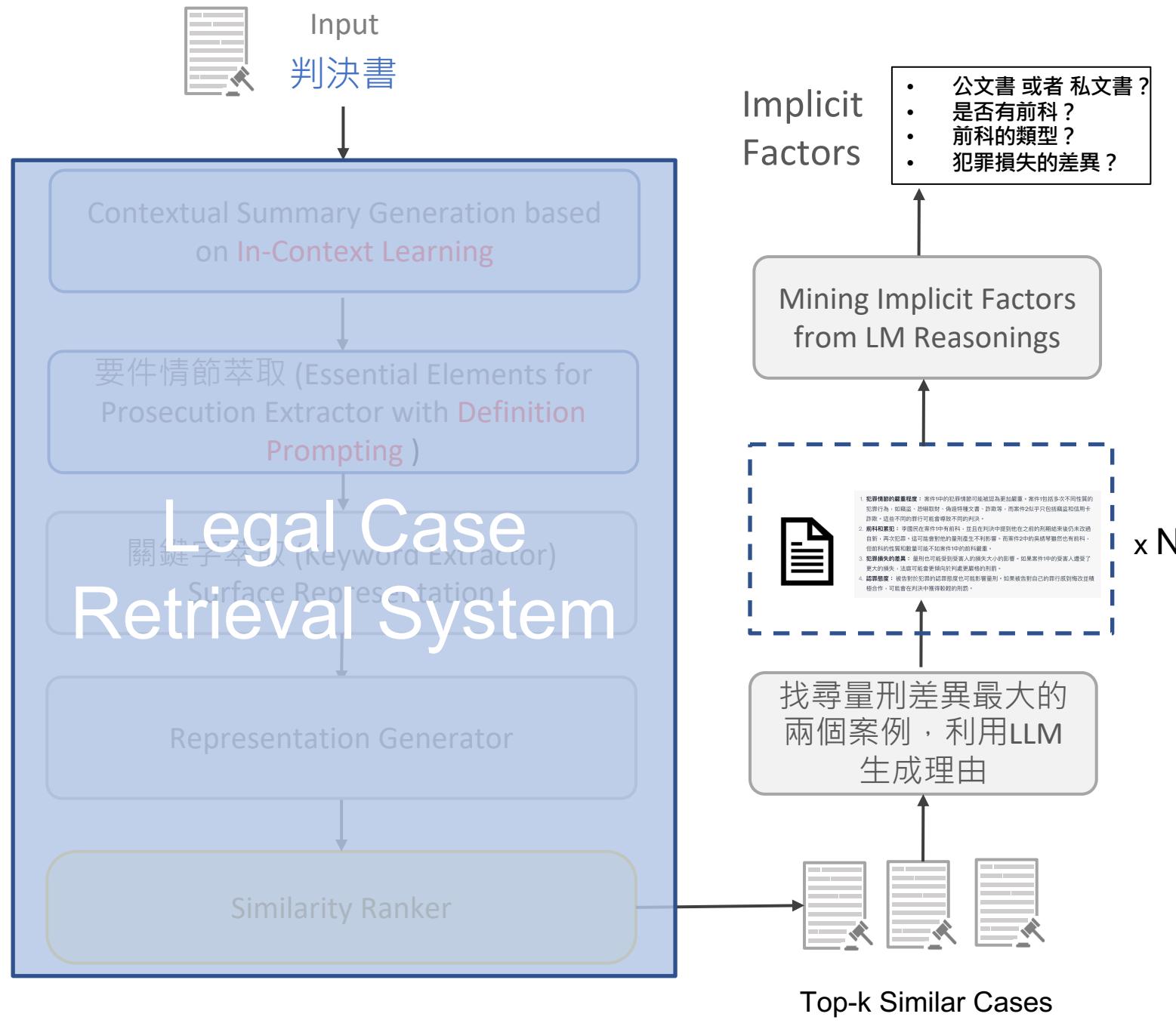
How to Discover Implicit Factors ?

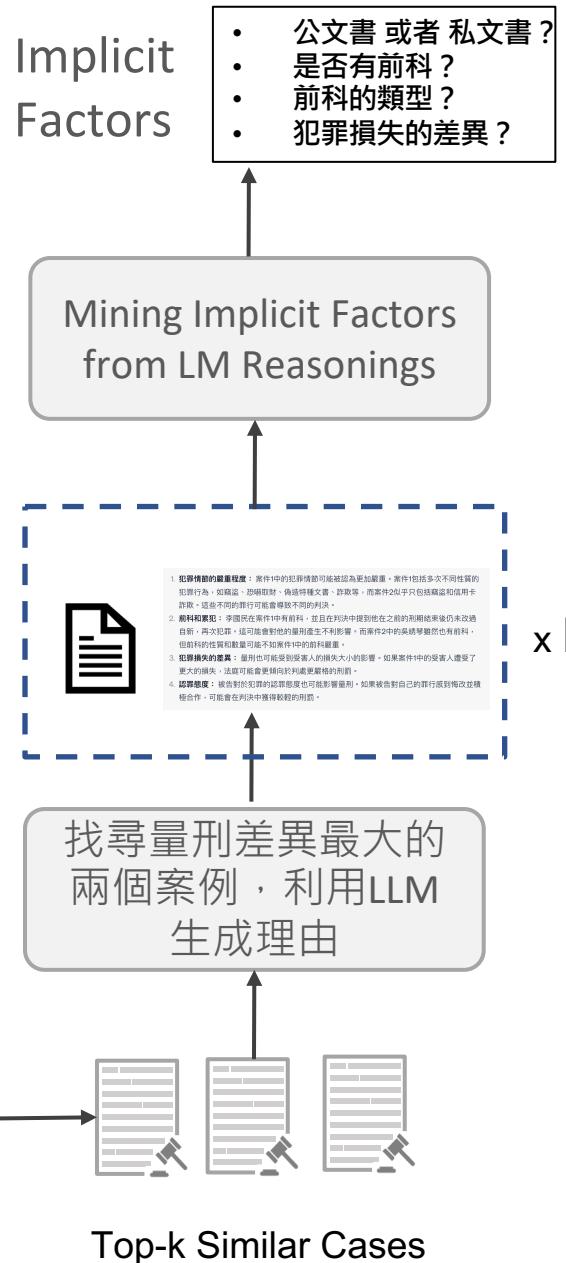
- Idea: Implicit Factor Discovering
 - Using LLM to generate reasons
 - Hand two cases to LLM and prompt the LLM to explain why different outcome



1. **犯罪情節的嚴重程度**: 案件1中的犯罪情節可能被認為更加嚴重。案件1包括多次不同性質的犯罪行為，如竊盜、恐嚇取財、偽造特種文書、詐欺等，而案件2似乎只包括竊盜和信用卡詐欺。這些不同的罪行可能會導致不同的判決。
2. **前科和累犯**: 李國民在案件1中有前科，並且在判決中提到他在之前的刑期結束後仍未改過自新，再次犯罪。這可能會對他的量刑產生不利影響。而案件2中的吳綉琴雖然也有前科，但前科的性質和數量可能不如案件1中的前科嚴重。
3. **犯罪損失的差異**: 量刑也可能受到受害人的損失大小的影響。如果案件1中的受害人遭受了更大的損失，法庭可能會更傾向於判處更嚴格的刑罰。
4. **認罪態度**: 被告對於犯罪的認罪態度也可能影響量刑。如果被告對自己的罪行感到悔改並積極合作，可能會在判決中獲得較輕的刑罰。



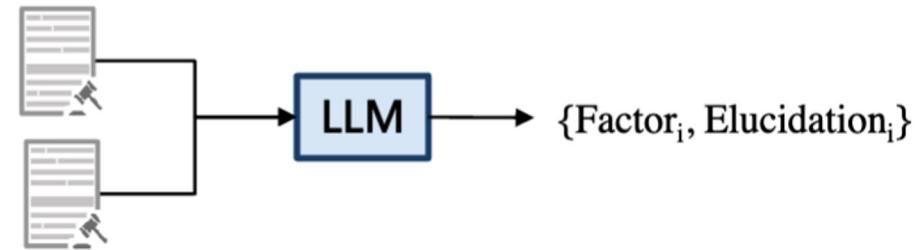




Baseline Prompting

Direct Prompt:

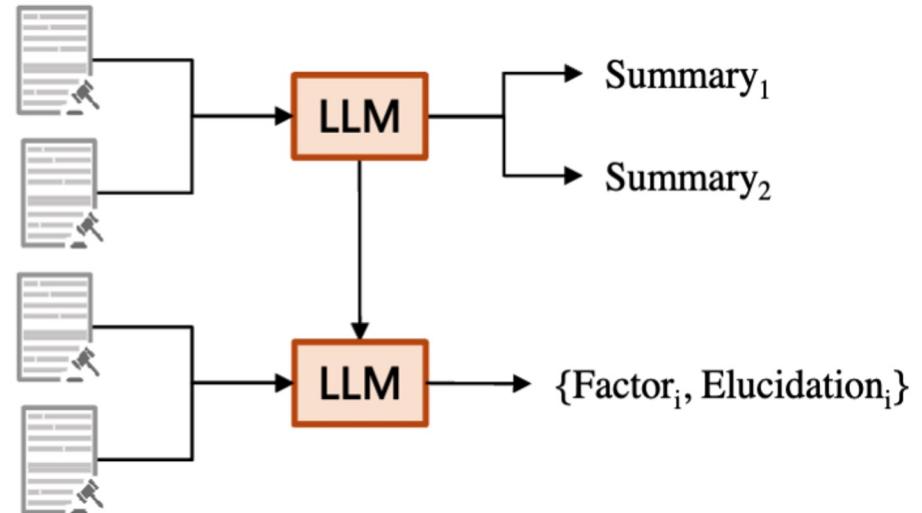
Please explain why the two legal cases have resulted in different judicial outcomes.



SSR Prompting

Step 1:
Ask LLM to read the legal documents and generate summary.

Step 2:
Further prompt LLM to generate implicit factors.



An overview of discovered implicit factors

Baseline Prompting

No	Factor (Translated)
1	損害程度 (Degree of Damage)
2	犯罪手法不同 (Different Methods of Crime)
3	目的和動機 (Purpose and Motive)
4	犯罪行爲的性質 (Nature of Criminal Behavior)
5	參與人員不同 (Different Participants)

Summary Steered Prompting

No	Factor (Translated)
1	造文書的方式和數量 (Methods and Quantity of Forged Documents)
2	涉及毒品的案件類型 (Types of Cases Involving Drugs)
3	偽造私文書的時間間隔 (Time Intervals for Forging Private Documents)
4	犯罪者的前科紀錄不同 (Differences in the Criminal Record of Offenders)
5	涉案的人數和角色不同 (Different Number and Roles of Involved Individuals)

EAR Prompting

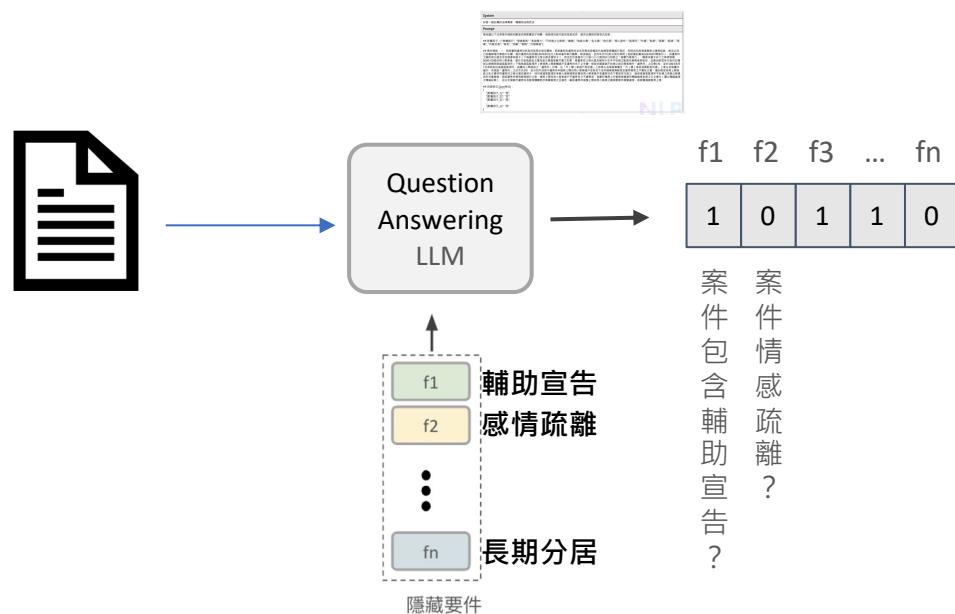
No	Factor (Translated)
1	犯罪手段和程度-對多個金融機構的損害 (Means and Degree of Crime - Damage to Multiple Financial Institutions)
2	詐騙方式不同-冒充公職人員 (Different Fraud Methods - Impersonating Public Officials)
3	損害程度-涉及的受害者數量 (Degree of Damage - Number of Victims Involved)
4	前科累積程度-犯罪次數 (Accumulation of Criminal History - Number of Criminal Offenses)
5	偽造文書情節的嚴重程度-職官章盜刻 (Severity of the Plot of Forging Documents - Counterfeiting Official Seals)



How to validate the utility of the discovered implicit factors ?

How to validate the utility of the discovered implicit factors ?

- Learn an Outcome Prediction Model as follows



System

你是一個台灣的法律專家，精通刑法與民法

Prompt

請依據以下法律案件情節判斷是否與影響因子相關，每個項目都只能回答是或否，請完全遵照回答格式回答：

影響因子：["車輛過戶", "酒後駕車", "家庭暴力", "不知情之公務員", "離婚", "偽造文書", "私文書", "身分證", "個人證件", "監理所", "外遇", "勒索", "吸毒", "超速", "酒駕", "內線交易", "背信", "恐嚇", "簽賭", "代辦業者"]

案件情節：一、郭家豪與盧秀玲前為同居男女朋友關係。郭家豪明知盧秀玲並未同意或授權其代為辦理車輛過戶登記，然因自知有酒後駕車之違規紀錄，無法以自己名義辦理汽車過戶手續，竟於盧秀玲自民國106年8月1日入勒戒處所執行觀察、勒戒後起，至同年月7日前之某日某時（起訴書記載為106年8月間某日），持盧秀玲之國民身分證及全民健康保險卡（下稱盧秀玲之身分證及健保卡），前往位於高雄市○○區○○○路0000○00號之「威霸汽車商行」，購買洪韻文名下之車牌號碼0000-00號自用小客車後，基於行使偽造私文書及使公務員登載不實之犯意，將盧秀玲之身分證及健保卡交予不知情之監理代辦業者郭宜珍，並委由郭宜珍代為向交通部公路總局高雄區監理所（下稱高雄區監理所）辦理將上開車輛過戶至盧秀玲名下之手續。郭宜珍遂委請不知情之刻印業者製作「盧秀玲」之印章1枚，並於106年8月7日某時前往高雄區監理所，接續持上開偽刻之「盧秀玲」印章，在「汽（機）車過戶登記書」上新車主名稱簽章欄及「汽（機）車各項異動登記書」上車主名稱欄內蓋印，而偽造「盧秀玲」之印文共2枚，並分別作成表示盧秀玲申請將上開自用小客車過戶至其名下及申請補發異動登記書等意思之不實私文書。復由郭宜珍將上開偽造之私文書併同盧秀玲之身分證及健保卡，持向高雄區監理所承辦人員辦理將前揭自用小客車過戶至盧秀玲名下事宜而行使之，致使高雄區監理所不知情之承辦公務員為形式審查後，誤認盧秀玲確有辦理過戶之意，遂將上開自用小客車過戶予盧秀玲之不實事項，登載於職務上所掌管車籍資料電腦檔案系統之公文書中（屬以電腦處理之電磁紀錄），足以生損害於盧秀玲及監理機關對於車籍管理之正確性。嗣因盧秀玲接獲上開自用小客車之違規罰單而報警處理，始經警循線查悉上情

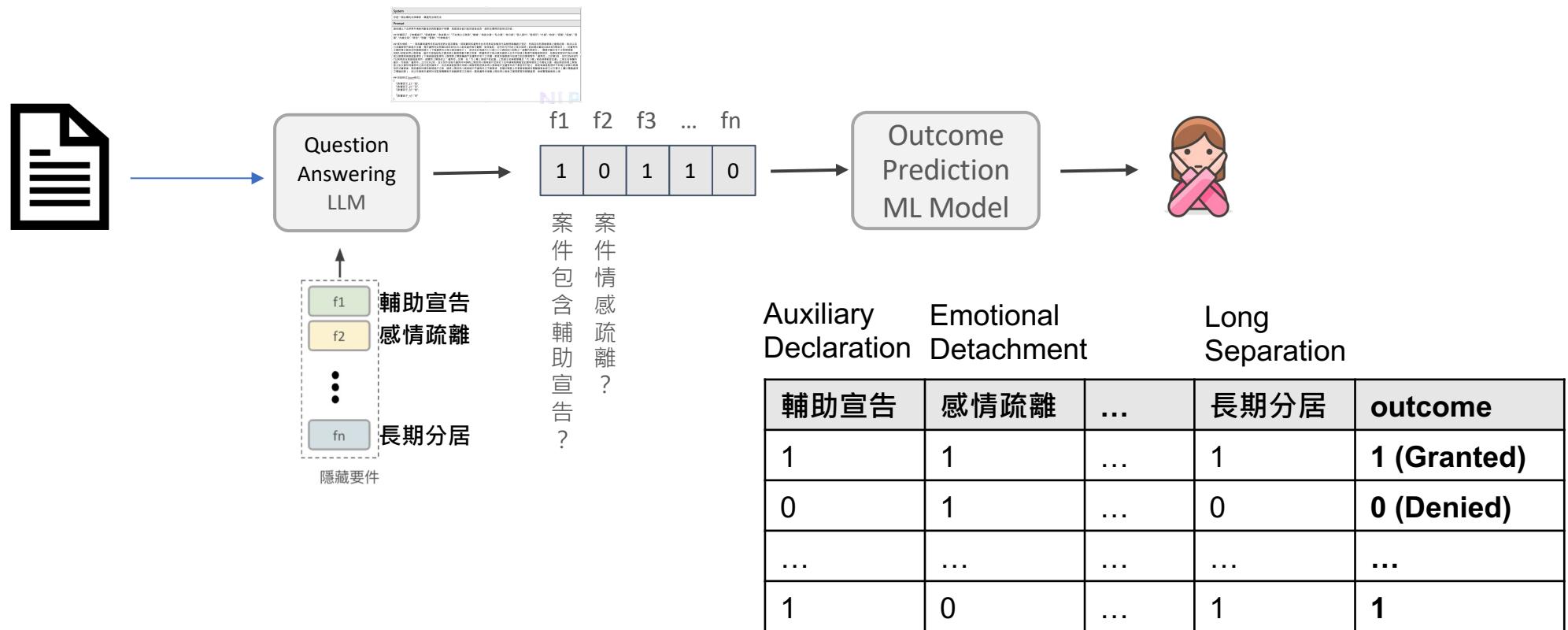
回答格式 (json格式)：

```
{
    "[影響因子_1)": "是",
    "[影響因子_2)": "否",
    "[影響因子_3)": "是",
    ...
    "[影響因子_n)": "否"
}
```



How to validate the utility of the discovered implicit factors ?

- Learn an Outcome Prediction Model as follows



Experiment Result on Divorce

Method	Model	Accuracy	Precision	Recall	F1	Specificity	NPV	F1 (Neg)	TP	TN	FP	FN
Baseline	LogisticRegression	0.73	0.70	0.73	0.69	0.25	0.54	0.34	1097	114	349	99
	SVC	0.73	0.71	0.73	0.67	0.13	0.61	0.22	1157	62	401	39
	GradientBoosting	0.74	0.72	0.74	0.71	0.26	0.60	0.37	1113	122	341	83
	MLP	0.63	0.62	0.63	0.63	0.30	0.33	0.31	912	139	324	284
	RandomForest	0.66	0.63	0.66	0.64	0.27	0.35	0.31	967	126	337	229
	GaussianNB	0.36	0.72	0.36	0.29	0.96	0.30	0.45	150	444	19	1046
	DecisionTree	0.57	0.61	0.57	0.58	0.39	0.29	0.33	764	179	284	432
	KNeighbors	0.69	0.66	0.69	0.67	0.29	0.41	0.34	1006	133	330	190
SSR	LogisticRegression	0.90	0.90	0.90	0.90	0.69	0.71	0.70	3672	538	245	221
	SVC	0.90	0.90	0.90	0.90	0.74	0.69	0.72	3638	578	205	255
	GradientBoosting	0.90	0.90	0.90	0.90	0.71	0.70	0.71	3660	556	227	233
	MLP	0.87	0.87	0.87	0.87	0.59	0.62	0.60	3611	461	322	282
	RandomForest	0.89	0.89	0.89	0.89	0.68	0.69	0.68	3655	530	253	238
	GaussianNB	0.78	0.87	0.78	0.80	0.85	0.42	0.56	2957	667	116	936
	DecisionTree	0.83	0.84	0.83	0.83	0.53	0.49	0.51	3465	415	368	428
	KNeighbors	0.89	0.89	0.89	0.89	0.60	0.70	0.65	3693	473	310	200
EAR	LogisticRegression	0.89	0.88	0.89	0.88	0.50	0.70	0.59	3942	396	391	170
	SVC	0.89	0.88	0.89	0.88	0.50	0.71	0.59	3947	396	391	165
	GradientBoosting	0.89	0.88	0.89	0.88	0.54	0.70	0.61	3927	428	359	185
	MLP	0.85	0.84	0.85	0.84	0.49	0.53	0.51	3762	387	400	350
	RandomForest	0.88	0.87	0.88	0.88	0.50	0.69	0.58	3937	396	391	175
	GaussianNB	0.78	0.84	0.78	0.80	0.68	0.39	0.50	3297	532	255	815
	DecisionTree	0.83	0.83	0.83	0.83	0.47	0.46	0.47	3669	373	414	443
	KNeighbors	0.87	0.86	0.87	0.86	0.42	0.66	0.51	3939	332	455	173

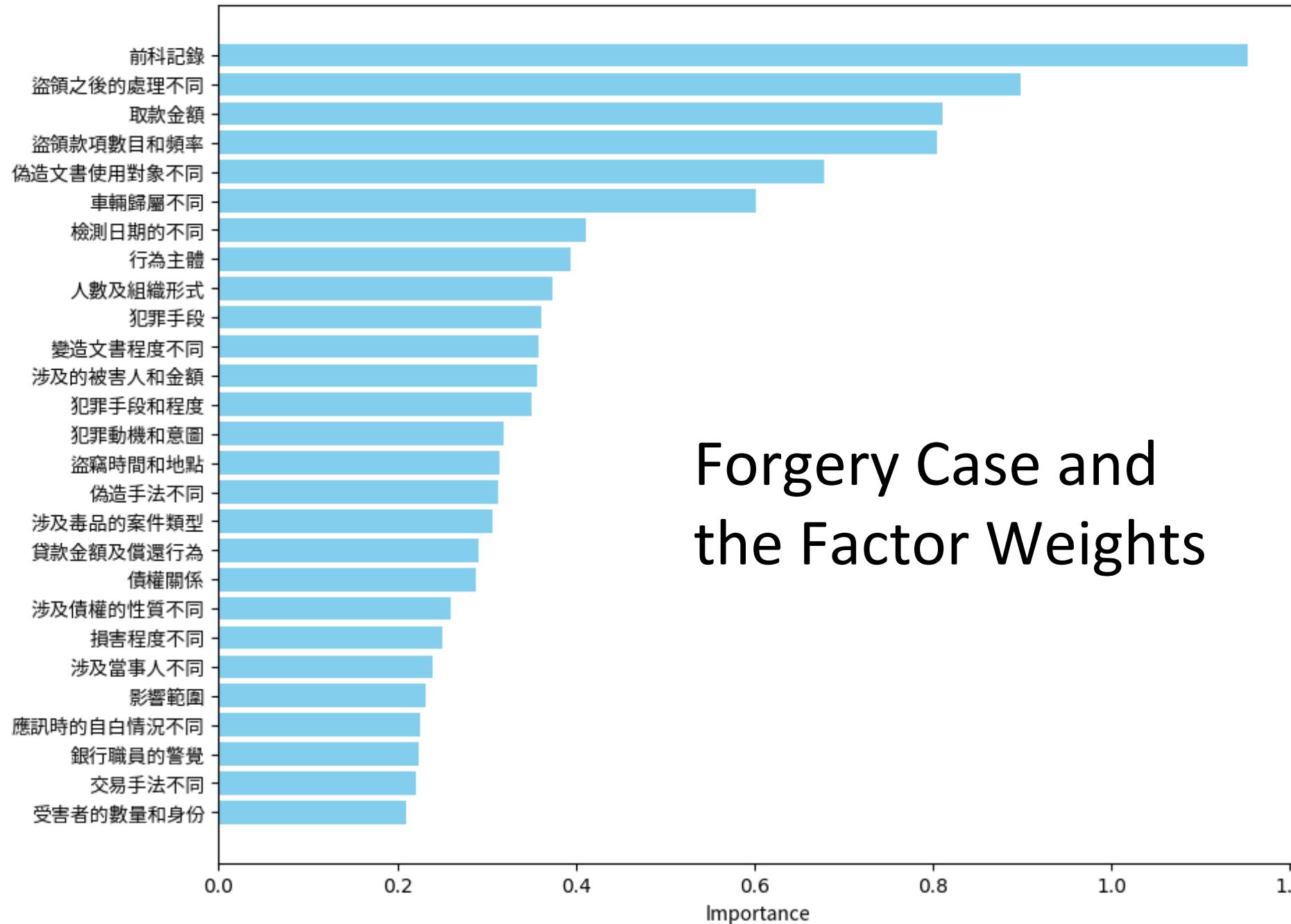
Table 14: Model Evaluation Metrics (Weighted Average) for Three Implicit Factor Discovery Methods based on 5,000 Divorce Judgments, with Factors clustered via Agglomerative Clustering. Specificity denotes recall for class 0, while NPV denotes precision for class 0

Outcome	# Samples
Divorce Granted	4214
Divorce Denied	786



Table 12: Distribution of Sentencing Outcomes for 5,000 Divorce Dataset

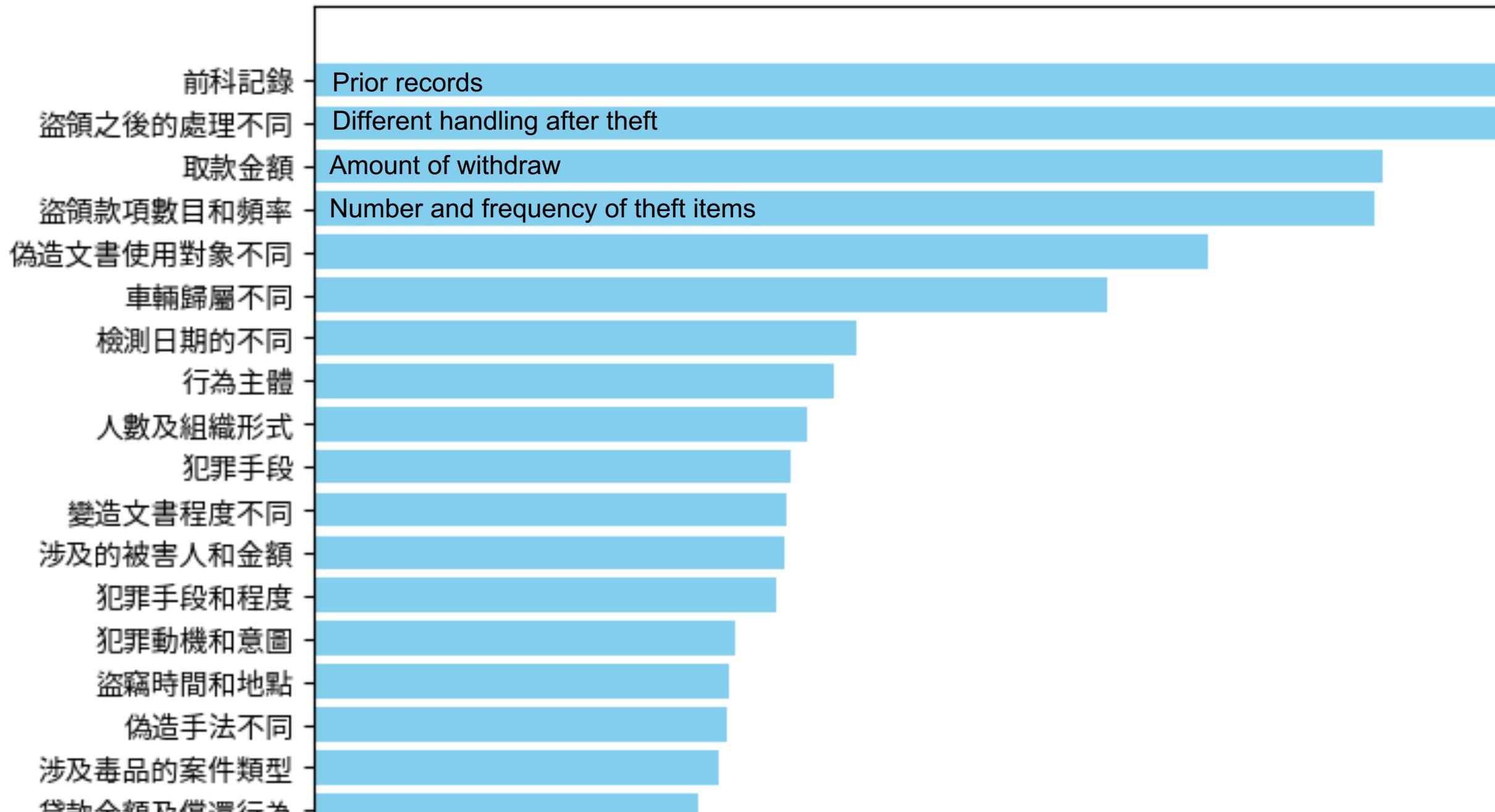
Keyword Importance Visualization



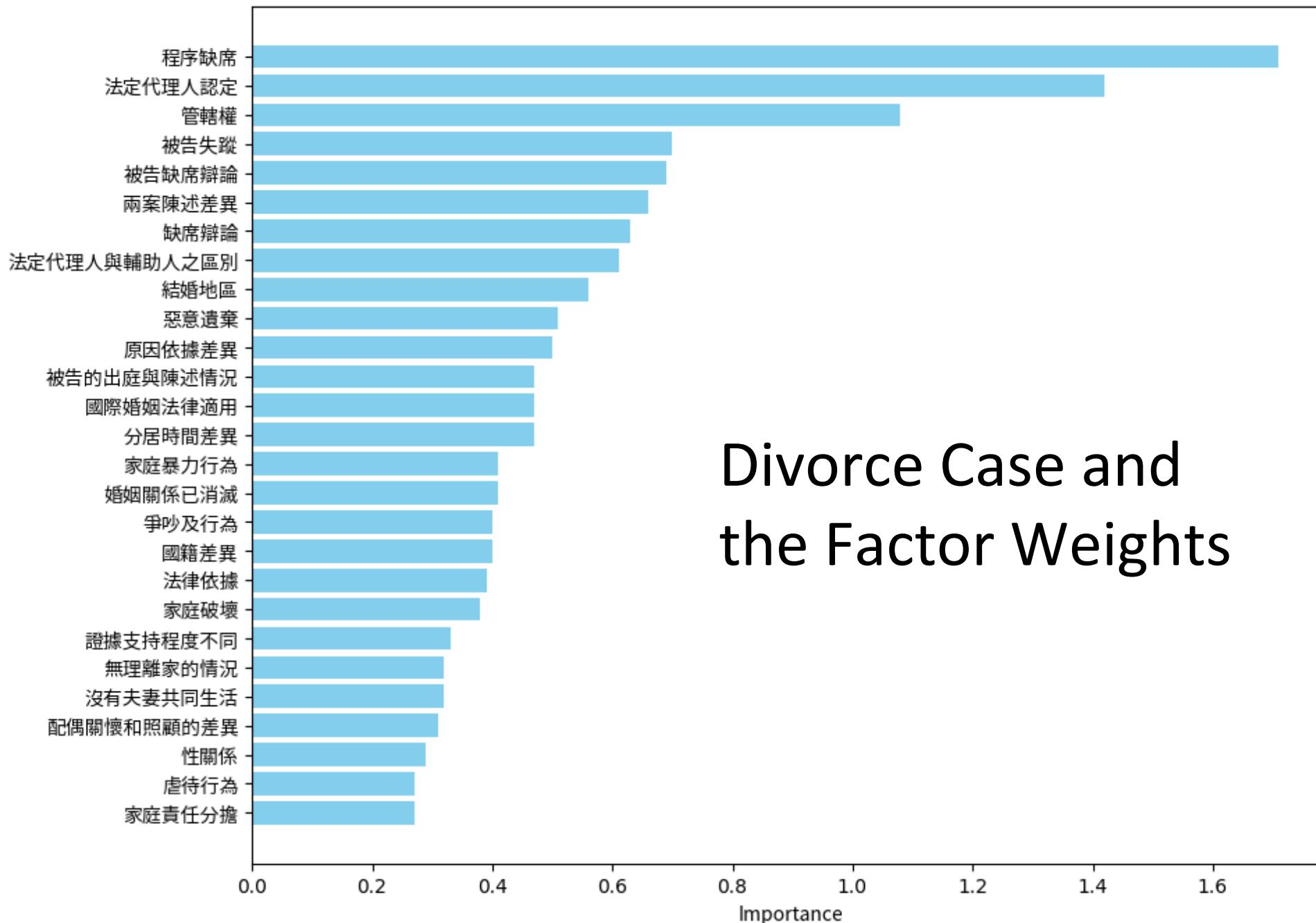
Forgery Case and the Factor Weights



Keyword Importance Visualization



Keyword Importance Visualization



Divorce Case and the Factor Weights



Keyword Importance Visualization



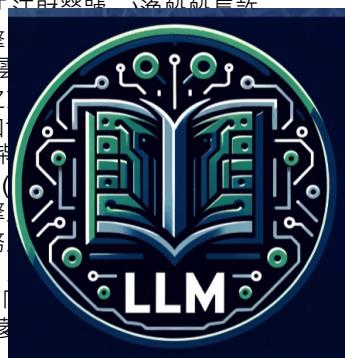
Short Conclusion

- This study underscores the potential of LLMs in legal analysis !
- Many interesting applications ahead !



Legal Document Structuring System Design

吳雲雀，負責南臺灣公司之業務，以從事漁船引擎暨相關零件販賣並開立所販售之船用機器證明書為業，其明知船用機器證明書係表彰確實交易引擎之文書，且將做為船長(主)向漁業署申請變更引擎登記之文件，應依據交易事實正確填載後始得交付他人，竟為下列犯行：(一)於民國93年11月20日，明知其販售與「興財發2號」(編號:CT4-1505號)漁船船長陳政利引擎主機為750匹馬力，而非「三菱牌、型號S6R2-MPTK2、1000匹馬力」之引擎，竟仍基於行使業務上登載不實文書之概括犯意，開立「三菱牌、型號S6R2-MPTK2、1000匹馬力」之「船用機器證明書」之業務上不實之文書，交與不知情之陳政利，使其得憑此不實之船用機械證明，於93年12月17日向漁業署提出「興財發2號」之「動力漁船改、增裝主、副機申請書」，將上揭主機換裝成三菱牌6缸、1000匹馬力柴油機，並經漁業屬於94年1月20日核准，足生損害於漁政與港務機關對漁船管理之正確性。(二)於94年4月10日，明知其販售與「勝鴻昌號」(編號：CT2-6124號)漁船船長呂強進之引擎1部係非省能源型野馬牌6缸190匹馬力(型號：6CHK-DT)，仍承前之行使業務上登載不實文書之概括犯意，在業務上掌管之船用機器證明書將呂強進購買之引擎型號登載為6CH-DTE(省能源型引擎)，並持以行使交付不知情之呂強進。呂強進取得該證明書後持以交由不知情之船務代辦商陳瑞全於94年5月3日向澎湖縣政府提出新建漁船「勝鴻昌」及漁業執照申請，並於94年5月4日經核准，足生損害於漁政與港務機關對漁船管理之正確性。(三)於94年6月15日間，明知其販售與「新吉豐6號」(編號：CT4-1180號)漁船船長兼船長郭寬令引擎主機非「三菱牌、型號S6R2-MPTK2、1000匹馬力」之引擎，竟承前之行使業務上登載不實文書之概括犯意，並與郭寬令共同基於犯意之聯絡，推由吳雲雀開立「三菱牌、型號S6R2-MPTK2、1000匹馬力」之「船用機器證明書」之業務上不實文書交予知情之郭寬令，郭寬令再持此不實之船用機械證明，於94年7月10日向漁業署提出「新吉豐6號」之「動力漁船改、增裝主、副機申請書」，欲將上揭主機換裝成三菱牌6缸、1000匹馬力柴油機，並經漁業署於94年7月11日核准，足生損害於漁政與港務機關對漁船管理之正確性。(四)於94年10月1日間，明知其販售與「江和發號」(編號:CT4-1500號，現改名為「江財發號」)漁船船長許詠賢引擎主機非「三菱牌、型號S6R2-MPTK2、1000匹馬力」之引擎，登載不實文書之概括犯意，並與許詠賢共同基於犯意之聯絡，推由吳雲雀開立「三菱牌、型號S6R2-MPTK2、1000匹馬力」之「船用機器證明書」之業務上不實之文書，詠賢持此不實之船用機械證明，於94年10月25日向漁業署提出「江和發號」之「動力漁船改、增裝主、副機申請書」，將上揭主機換裝成三菱牌6缸、1000匹馬力柴油機，並經漁業署於94年10月26日核准，足生損害於漁政與港務機關對漁船管理之正確性。(五)於94年10月9日間，明知其販售與「進金和發號」(編號:CT4-2207號)漁船船長呂光平引擎主機非「三菱牌、型號S6R2-MPTK2、1000匹馬力」之引擎，竟與呂光平共同基於行使業務上不實文書，呂光平再持此不實之船用機械證明，於97年1月9日向漁業署提出「進金和發號」之「動力漁船改、增裝主、副機申請書」，將上揭主機換裝成三菱牌6缸、1000匹馬力柴油機。



Forgery Implicit Factors

Forgery Implicit Factors

隱藏要件	對應情節
犯罪手段	
損害程度	
影響範圍	
法律適用	
涉及債權的性質不同	
應訊時的自白情況不同	
偽造文件的種類和使用方式	
累犯	

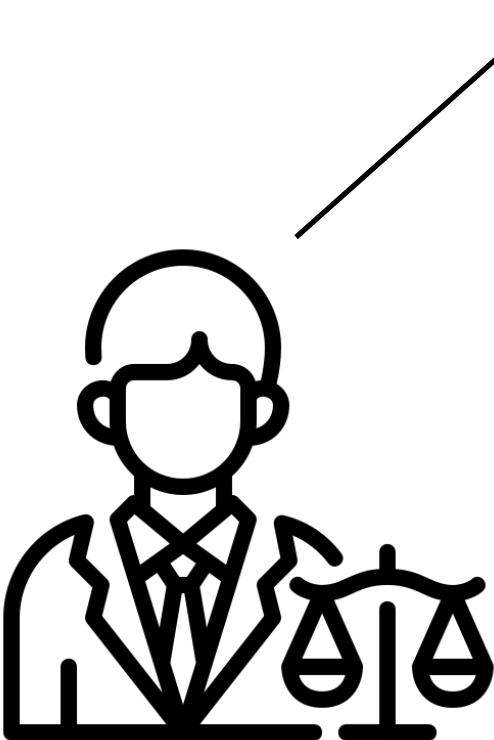
Q: 上述案件的犯罪手段？

Q: 上述案件的損害程度？

:

Q: 上述案件是否為累犯？

System Usage Scenario



Q : 已故個人的遺產引發了爭議。配偶與子女是合法繼承人，但配偶卻企圖透過詐欺手段來侵占共同財產。她冒用已故人的身份，偽造文件並盜用印章，讓金融機構誤以為是合法的交易。最終，金融機構執行了這些虛偽指示，導致繼承人的權益受損。

Delta-NCHU Legal Retrieval System

Top-K Similar Law Cases

第1篇 第2篇 第3篇 第4篇 第5篇

一、劉士毅於任職於薌園生技股份有限公司（下稱薌園公司），其在該公司民國98年度之薪資總額為新臺幣（下同）23萬7,360元，且其於99年9月間之財力、資力狀況均不佳，經洽詢銀行相關人員而得知以其上開經濟條件將無法通過銀行核准貸款之標準，其可預見若將其98年度各類所得扣繳暨免扣繳憑單（下稱扣繳憑單）交予代辦業者辦理貸款，該代辦業者可能變造其扣繳憑單上之薪資總額，再持變造之扣繳憑單向銀行申請貸款，銀行恐因而誤信其有較高之收入而具償債能力即同意貸款，猶意圖為自己不法之所有，與身分不詳之成年代辦業者，基於行使變造私文書及詐欺取財不確定故意之犯意聯絡，提供其98年度扣繳憑單予某身分不詳之成年代辦業者，由該代辦業者在不詳時、地，以不詳方式變造劉士毅所提供之98年度扣繳憑單上薪資給付總額欄內之金額為43萬7,360元，復於99年9月3日前某日，在板橋火車站，由劉士毅在個人信用貸款申請書之申請人欄簽名，並提供其身分證件影本予該代辦業者，嗣由代辦業者在該申請書上填載劉士毅在薌園公司月收入3萬6,000元、年收入43萬7,000元等不實內容，復由該代辦業者以前揭變造之98年度扣繳憑單、劉士毅之個人信用貸款申請書及身分證件影本等文件，持向大眾商業銀行股份有限公司（下稱大眾銀行）申請36萬元之信用貸款而行使之，致大眾銀行承辦本件貸款之桃竹區個金無擔業務中心（址設桃園市○○區○○路○段○○號○○樓）放款、授信部門員工均陷於錯誤，誤認劉士毅有年收入43萬餘元之穩定收入，具備償債能力，而核准貸款18萬元予劉士毅，並於99年9月15日核撥18萬元至劉士毅指定之玉山商業銀行股份有限公司朴子分行000000000000號帳戶，致生損害於薌園公司對於員工薪資扣繳憑單核發之正確性、大眾銀行對於核撥貸款審查之正確性。

What is the difference between w/o LLM

- 以犯罪情境進行檢索



犯罪情境範例

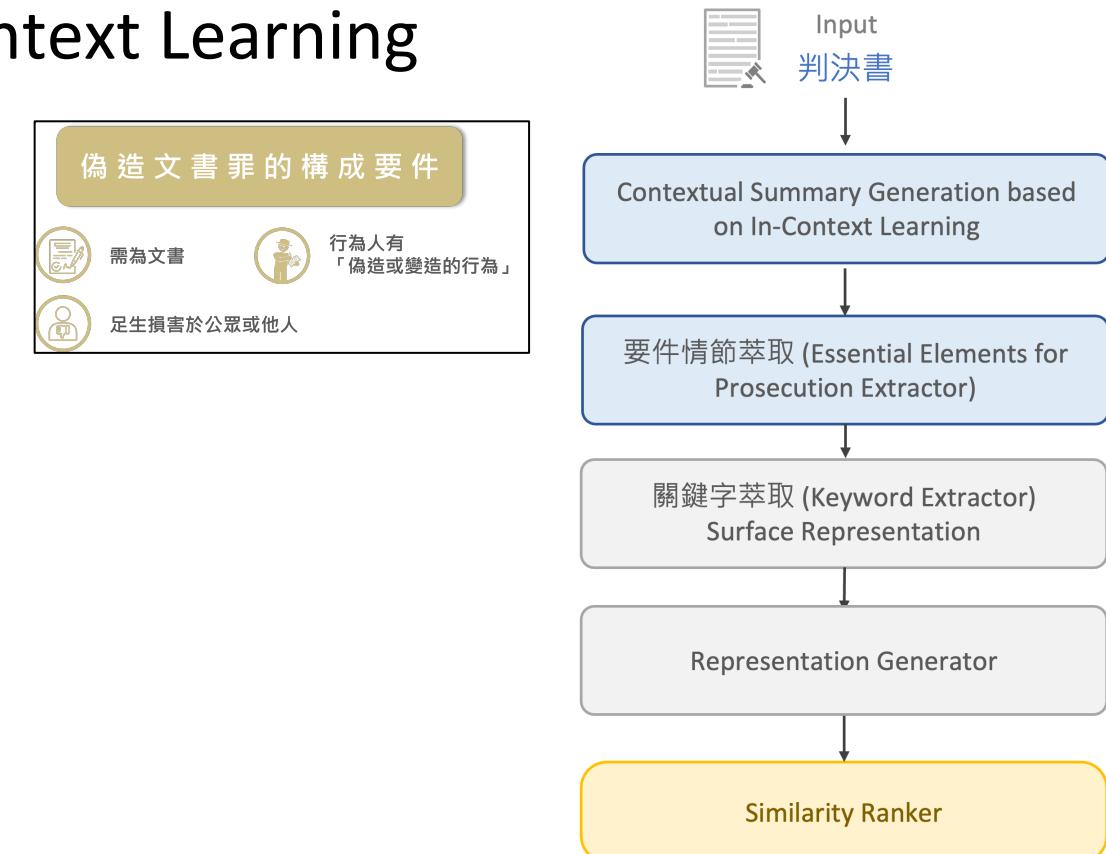
孫桂蘭與孟奇於民國83年7月10日結婚，孫桂蘭係孟奇之子女即孟慶明、孟慶虹、孟慶經、孟慶緯、孟蓉之繼母。孟奇於96年1月22日死亡，孫桂蘭明知孟奇死亡後，其名下財產均為遺產，於遺產分割前屬孫桂蘭、孟慶明、孟慶虹、孟慶經、孟慶緯、孟蓉全體繼承人公同共有，未經全體繼承人同意，任一繼承人不得擅自處分，竟意圖為自己不法所有，基於詐欺取財及行使偽造私文書之犯意，於附表各編號所示之時間，至附表各編號所示金融機構，冒用孟奇之名義，在空白取款憑證填寫如附表各編號所示之日期、帳號及金額，並於「取款印鑑」欄，使用孟奇之印章盜蓋印文1枚，表示係孟奇欲自附表各編號所示金融機構帳戶內提領附表各編號所示提款金額，而偽造孟奇名義之取款憑證私文書後，再連同上開帳戶存摺持之向不知情之附表各編號所示金融機構承辦人員辦理取款而行使之，致該金融機構承辦人員陷於錯誤，誤認孫桂蘭係經孟奇本人授權而為上開手續，如數交付如附表各編號所示金額，均足以生損害於孟慶明、孟慶虹、孟慶經、孟慶緯（已歿）、孟蓉之權益及日盛國際商業銀行股份有限公司《下稱日盛商業銀行》、合作金庫商業銀行股份有限公司竹塹分行《下稱合作金庫商業銀行竹塹分行》、中華郵政股份有限公司清華大學郵局《下稱清華大學郵局》、永豐商業銀行股份有限公司《下稱永豐商業銀行》等金融機構對於孟奇帳戶資料、存款管理之正確性。

• Design Consideration

- 人名、日期、地址、公司名等等雜訊將影響相似度判斷。
- Semantic Matching
 - 印鑑 v.s. 印章 v.s. 圖章
- 判決書有對應之起訴要件，細部區分要件情節來進行相似度比對，也應該納入考量。

We propose: A Semantic Retrieval Pipeline for Legal Documents

- Context Summarization based on In-Context Learning
 - 情境摘要
- Essential Element for Prosecution
 - 要件情節萃取
- Surface Representation Generation
 - 取得關鍵字
- Representation Generation
 - 取得 Sentence Vector
- Similarity Ranker
 - 判決書相似度計算



Contextual Summary v.s. Plot Summary

犯罪情境全文

孫桂蘭與孟奇於民國83年7月10日結婚，孫桂蘭係孟奇之子女即孟慶明、孟慶虹、孟慶經、孟慶緯、孟蓉之繼母。孟奇於96年1月22日死亡，孫桂蘭明知孟奇死亡後，其名下財產均為遺產，於遺產分割前屬孫桂蘭、孟慶明、孟慶虹、孟慶經、孟慶緯、孟蓉全體繼承人公同共有，未經全體繼承人同意，任一繼承人不得擅自處分，竟意圖為自己不法所有，基於詐欺取財及行使偽造私文書之犯意，於附表各編號所示之時間，至附表各編號所示金融機構，冒用孟奇之名義，在空白取款憑證填寫如附表各編號所示之日期、帳號及金額，並於「取款印鑑」欄，使用孟奇之印章盜蓋印文1枚，表示係孟奇欲自附表各編號所示金融機構帳戶內提領附表各編號所示提款金額，而偽造孟奇名義之取款憑證私文書後，再連同上開帳戶存摺持之向不知情之附表各編號所示金融機構承辦人員辦理取款而行使之，致該金融機構承辦人員陷於錯誤，誤認孫桂蘭係經孟奇本人授權而為上開手續，如數交付如附表各編號所示金額，均足以生損害於孟慶明、孟慶虹、孟慶經、孟慶緯（已歿）、孟蓉之權利及日盛國際商業銀行股份有限公司《下稱日盛商業銀行》、合作金庫商業銀行股份有限公司竹塹分行《下稱合作金庫商業銀行竹塹分行》、中華郵政股份有限公司清華大學郵局《下稱清華大學郵局》、永豐商業銀行股份有限公司《下稱永豐商業銀行》等金融機構對於孟奇帳戶資料、存款管理之正確性。

But How?

Contextual Summary

在這個案件中，涉及到一段婚姻關係，並有繼承人之間的財產分割。一名已故的個人的遺產成為了糾紛的焦點。在其去世後，其配偶及子女成為了繼承人。然而，配偶明知其去世，卻企圖以詐欺和偽造私文書的方式來獲取他們的共同財產。她冒用已故人的名義，填寫了取款憑證並偽造了他的印章，使金融機構誤以為是合法要求。最終，金融機構執行了這些虛偽的指示，繼承人之間的權益受損，也對涉及的金融機構造成這些行為涉及偽造私文書、詐欺取財，並對金融機構產生了不良影響。



Plot Summary

孫桂蘭與孟奇於民國83年結婚，並成為孟奇的後母。孟奇於96年1月去世，留下孫桂蘭以及孟奇的子女繼承他的財產。孟奇去世後，其財產為遺產，需等到遺產分割後才能確定各繼承人的份額。然而，孫桂蘭明知孟奇已故，卻在未經其他繼承人同意的情況下，企圖以不法手段獲得財產。

In-Context Learning for Contextual Summary Generation

- Guide ChatGPT to generate Contextual Summary via In-context Learning

Prompt : I will show you several examples. Please follow the examples to generate contextual summary for a new input

Example Text 1 :

孫桂蘭與孟奇於民國83年7月10日結婚。孫桂蘭係孟奇之子女即孟慶明、孟慶虹、孟慶鑑、孟蓉之繼母。孟奇於96年1月22日死亡，孫桂蘭明知孟奇死亡後，其名下財產均為遺產，於遺產分割前屬孫桂蘭、孟慶明、孟慶虹、孟慶鑑、孟蓉全體繼承人共同共有，未經全體繼承人同意，任一繼承人不得擅自分割，竟意圖為自己不法所有，基於詐欺取財及行使偽造私文書之犯意，於附表各編號所示之時間，至附表各編號所示金融機構，冒用孟奇之名義，在空白取款憑證填寫如附表各編號所示之日期、帳號及印鑑，並於「取款印鑑」欄，使用孟奇之印章蓋印文一枚，表示係孟奇欲自附表各編號所示金融機構帳戶內提款附表各編號所示提款金額，而偽造孟奇名義之取款憑證私文書後，再連同上開帳戶存摺得之不知情之附表各編號所示金融機構承辦人員陷於錯誤，誤認該孫桂蘭係經孟奇本人授權而為上開手續，如數交付附表各編號所示提款金額，均足以生損害於孟慶明、孟慶虹、孟慶鑑（已歿）、孟蓉之權益及日盛國際商業銀行股份有限公司（下稱日盛商業銀行）等金融機構對於孟奇帳戶資料、存款管理之正確性。**Example 1** 輸出：在這個案件中，涉及到一段婚姻關係，並有繼承人之間的財產分割。一名已故的個人的遺產成為了糾紛的焦點。在其去世後，其配偶及子女成為了繼承人。然而，配偶明知其去世，卻企圖以詐欺和偽造私文書的方式來獲取他們的共同財產。她冒用已故人的名義，填寫了取款憑證並偽造了他的印章，使金融機構誤以為是合法的取款要求。最終，金融機構執行了這些虛偽的指示，導致繼承人之間的權益受損，也對涉及的金融機構造成了損害。這些行為涉及偽造私文書、詐欺取財，並對繼承人和金融機構產生了不良影響。

Example Summary 1 :

在這個案件中，涉及到一段婚姻關係，並有繼承人之間的財產分割。一名已故的個人的遺產成為了糾紛的焦點。在其去世後，其配偶及子女成為了繼承人。然而，配偶明知其去世，卻企圖以詐欺和偽造私文書的方式來獲取他們的共同財產。她冒用已故人的名義，填寫了取款憑證並偽造了他的印章，使金融機構誤以為是合法的取款要求。最終，金融機構執行了這些虛偽的指示，導致繼承人之間的權益受損，也對涉及的金融機構造成了損害。這些行為涉及偽造私文書、詐欺取財，並對繼承人和金融機構產生了不良影響。

Example Text 2:

吳文錦係吳清海（民國108年2月2日死亡）之長子，吳芸安則係吳清海次子吳文標（於103年3月22日死亡）之女，為吳文標之代位繼承人，吳淑呢、吳金桂、吳淑惠、吳月嬌亦為吳清海之子女，惟吳文錦於吳清海過世後，為辦理吳清海之後事相關事宜時，明知不得未經全體繼承人同意或授權，須經全體繼承人同意或授權，始得提款存款，竟未經全體繼承人同意或授權，為行使偽造私文書之犯意，接續於108年2月11日、108年2月20日，持其所保管吳清海之雲林縣○○○鄉○○○號：0000000000000000號，下稱四湖鄉農會帳戶之存摺、印鑑，前往雲林縣○○○村○○○路○○○號之雲林縣四湖鄉農會，冒用吳清海之名義，在該農會之提款憑條簽章欄上，填寫領取新臺幣（下同）337,600元、7,290元，並偽簽「吳清海」之署押1枚，盜蓋「吳清海」印文2枚（詳如附表所示），而偽造吳清海同意提款337,600元、7,290元之私文書，並持該偽造之提款憑條交予上開農會之不知情承辦人員而行使之，使吳文錦自前開帳戶提共344,890元，足生損害於吳清海全體繼承人及四湖鄉農會對客戶存款管理之正確性。

Example Summary 2:

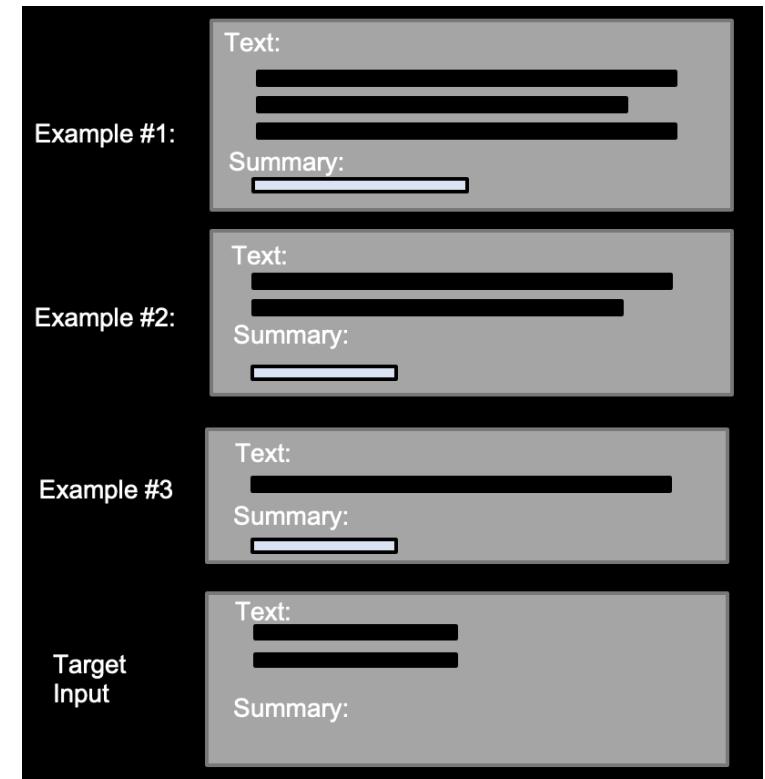
這宗案件涉及家族關係和遺產問題。已故個人的子女是合法繼承人，但其中一人明知不得未經全體繼承人同意或授權就提款存款。他使用偽造文件，包括存摺和印鑑，前往當地農會充當已故父親的身份，提領款項，造成損害。這涉及偽造文件和可能的詐騙行為，對繼承人和農會的存款管理產生不良影響。

Text :

廖志遠與廖娟娟為廖梓彬之子女，邱美蘭為2人之母。廖梓彬於民國106年9月7日因車禍過世，遺有包括坐落新竹市○○段00號地號土地及其上門牌號碼為新竹市○○路000號房屋（下稱本案房地）之遺產，因廖梓彬生前未立遺囑，依法本應由其子女即廖志遠、廖娟娟及配偶邱美蘭共同繼承，應繼分各1/3。記廖志遠竟基於行使偽造私文書及使公務人員登載不實之犯意，先於106年11月2日，在新竹市○○路000號，向廖娟娟佯稱要辦理相關產權繼承登記等事宜所屬，取得廖娟娟交付之印章、印鑑證明等資料後，未經廖娟娟之同意或授權，旋於106年11月2日至該月10日前不時詢問，前往台北市○○區○○路000號之6，委託不知情之邱美蘭製作產權分割協議書，並由不知情之件書將廖娟娟之印章，接續於該產權分割協議書蓋印文共計4枚，用以表示廖娟娟同意放棄繼承廖梓彬之遺產，並同意包括本案房地及其他動產之遺產均由邱美蘭單獨繼承之意，由該代書以郵寄方式將土地登記申請書、上開遺產分割協議書、繼承系統表、廖志遠、廖娟娟、邱美蘭三人之戶籍謄本、印鑑證明及跨縣市代收地政案件申請單，郵寄至新竹市地政事務所，以辦理登記事宜，將本案房地登記於邱美蘭名下。該事務所公務人員，於同年10月10日收到包括上開申請文件為形式審查後，於106年11月16日將此不實資料登載於其職務上所掌之土地登記謄本及相關電腦資料等公文書上而辦理登記完畢，足生損害於廖娟娟及新竹市地政事務所對不動產申辦繼承及分割繼承登記核管理之正確性。廖志遠於本案房地登記在邱美蘭名下後，再於107年1月19日委由劉富明所屬地政事務所，將本案房地應有部分各1/15，共計應有部分4/15，以贈與為原因，向新竹市地政事務所辦理移轉登記予廖志遠及廖志遠之子廖偉丞、廖士堯名下，於107年1月24日辦理登記完畢；嗣於107年5月21日，將邱美蘭就本案房地剩餘之應有部分各4/15，共計應有部分4/15，以贈與為原因，移轉登記至廖志遠、廖偉丞、廖士堯名下，於107年5月25日辦理登記完畢。

Summary :

在這個案件中涉及了家庭關係和遺產爭議。已故個人的子女和配偶成為了合法的繼承人，應該按照法定份額分割遺產。然而，其中一個繼承人採取了欺詐手段，包括偽造文件和冒用他人印章，使其他繼承人放棄他們的繼承份額。他還通過偽造文件的方式將不動產登記在他母親名下，然後進行了多次贈與交易。這些行為對其他繼承人和地政事務所的不動產管理造成了損害。



in-context learning



Input
判決書

Contextual Summary Generation based
on In-Context Learning

要件情節萃取 (Essential Elements for
Prosecution Extractor with Definition
Prompting)

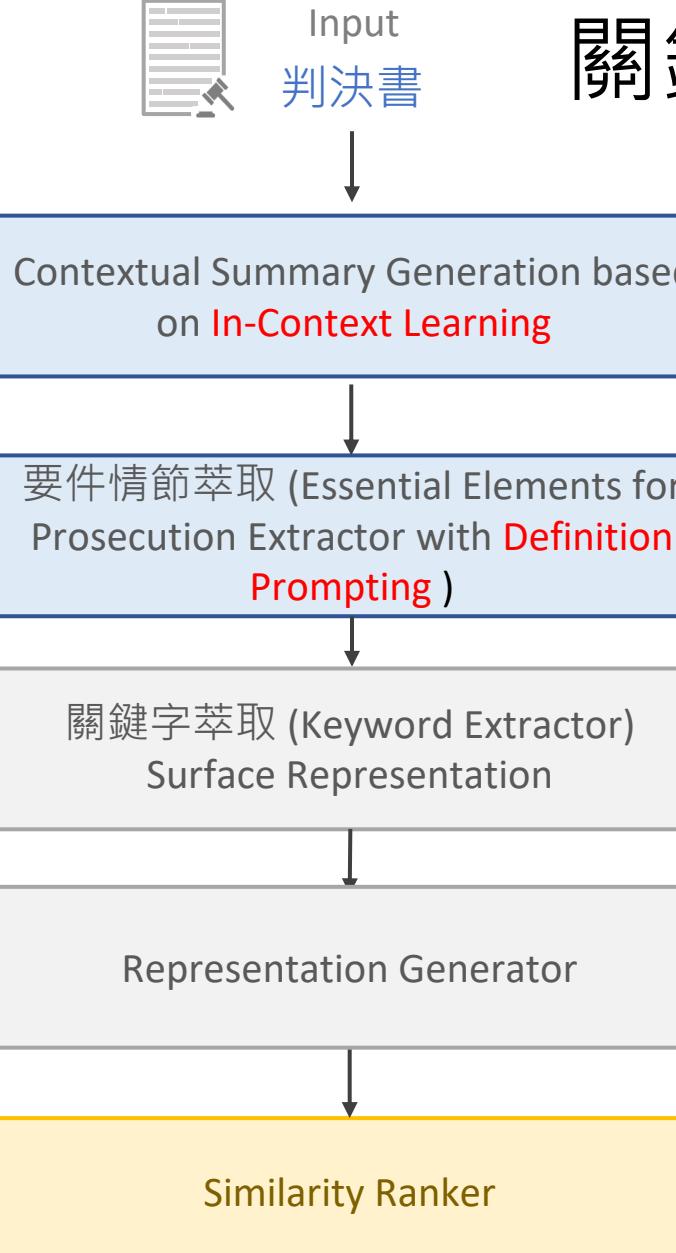
Definition Prompting for 要件萃取

Input : Contextual Summary

案件涉及家族關係與遺產問題。已故個人的遺產引發了爭議。他的配偶與子女是合法繼承人，但配偶卻企圖透過詐欺手段來侵占共同財產。她冒用已故人的身份，偽造文件並盜用印章，讓金融機構誤以為是合法的交易。最終，金融機構執行了這些虛偽指示，導致繼承人的權益受損，也給金融機構造成損失。這些行為牽涉到偽造文件和詐騙，對繼承人和金融機構都帶來不良影響。

Output:

- **需為文書**：案件中涉及的是偽造的取款憑條、同意提款的私文書，這些都是具有書面形式的文件，符合「文書」的要件。
- **行為人有偽造或變造的行為**：配偶在這個案件中偽造了已故配偶的名義，填寫了偽造的取款憑條，並且冒用已故人的印章，這些行為都屬於偽造或變造的行為。
- **足生損害於公眾或他人**：配偶的偽造行為導致金融機構誤以為是合法的交易，最終金融機構執行了這些虛偽指示。這使得合法繼承人的權益受損，也給金融機構造成了損失。因此，這些行為足以對繼承人和金融機構造成損害。



關鍵字萃取 (Surface Representation)

Definition Prompt : 偽造文書犯罪的構成要件: 一、需為文書，二、行為人有偽造或變造的行為，三、足生損害於公眾或他人。

Input : Contextual Summary from Previous Stage

案件涉及家族關係與遺產問題。已故個人的遺產引發了爭議。他的配偶與子女是合法繼承人，但配偶卻企圖透過詐欺手段來侵占共同財產。她冒用已故人的身份，偽造文件並盜用印章，讓金融機構誤以為是合法的交易。最終，金融機構執行了這些虛偽指示，導致繼承人的權益受損，也給金融機構造成損失。這些行為牽涉到偽造文件和詐騙，對繼承人和金融機構都帶來不良影響。

Output:

- **需為文書**：案件中涉及的是偽造的取款憑條、同意提款的私文書，這些都是具有書面形式的文件，符合「文書」的要件。
- **行為人有偽造或變造的行為**：配偶在這個案件中偽造了已故配偶的名義，填寫了偽造的取款憑條，並且冒用已故人的印章，這些行為都屬於偽造或變造的行為。
- **足生損害於公眾或他人**：配偶的偽造行為導致金融機構誤以為是合法的交易，最終金融機構執行了這些虛偽指示。這使得合法繼承人的權益受損，也給金融機構造成了損失。因此，這些行為足以對繼承人和金融機構造成損害。

- **文書**：偽造取款憑條、同意提款的私文書
- **行為**：配偶偽造已故配偶名義，填寫偽造取款憑條，冒用已故人印章
- **損害**：金融機構執行虛偽指示，合法繼承人權益受損，金融機構損失

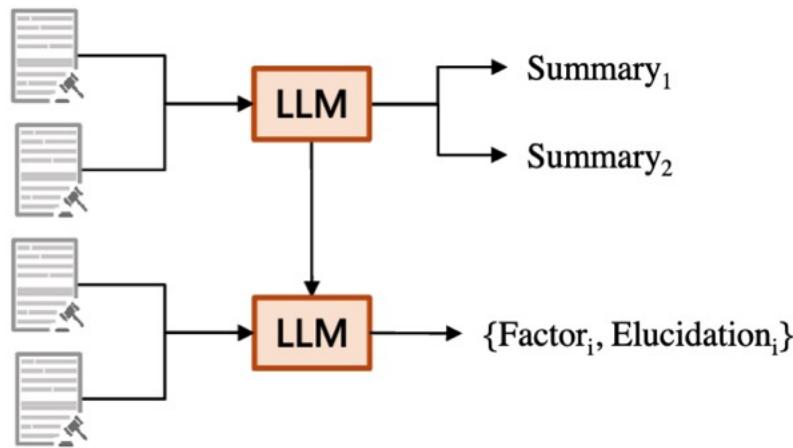
機會在哪裡？

- 法律判決書分析
 - Prompting
 - 懂得跟AI做溝通

SSR Prompting

Step 1:
Ask LLM to
read the legal documents
and generate summary.

Step 2:
Further prompt LLM
to generate implicit factors.



Indigenous Languages Understanding using LLM

Fine-Tuning LLM for Extremely Low Resource Language Translation?

我也很好，謝謝 →



?



南勢阿美語

→ Kapah:tu kaku, aray.

dataset
Zhiwei Tong
sequence
pre-train
BERT
general
train
learn
fine-tune
prompt
chain
prompt chain
self-supervised learning
text
layer
model
language model
tokens
BART

dataset
Zhiwei Tong
sequence
pre-train
BERT
general
train
learn
fine-tune
prompt
chain
prompt chain
self-supervised learning
text
layer
model
language model
tokens
BART

dataset
Zhiwei Tong
sequence
pre-train
BERT
general
train
learn
fine-tune
prompt
chain
prompt chain
self-supervised learning
text
layer
model
language model
tokens
BART

Let the LLM learn to use tools and examples for translation?

範例語句 (Parallel Examples) →
字典 (Word Dictionary) →
我也很好，謝謝 →



→ Kapah:tu kaku, aray.



範例語句

Fetches contextually similar examples

[zh]: 是的，我讀過，這本書很好看。
(En: Yes, I've read it, and the book is very interesting.)
[amis]: Hay , nam'i asiptu kaku , kapah kina cudad .

[zh]: 我也很好，謝謝。
(En: I'm doing well too, thank you.)
[amis]: Kapah:tu kaku , aray .

[zh]: 郵差也感到喜悅與滿足。
(En: The postman also feels joy and satisfaction.)
[amis]: U yu-cay satu , mikihiatiya a lipahak , a mi'edem tu ulah nu valucu' .

.....

字典

Retrieves the individual translations of each word

[zh]: 很好學 (En: Easy to learn.)
[amis]: kapah

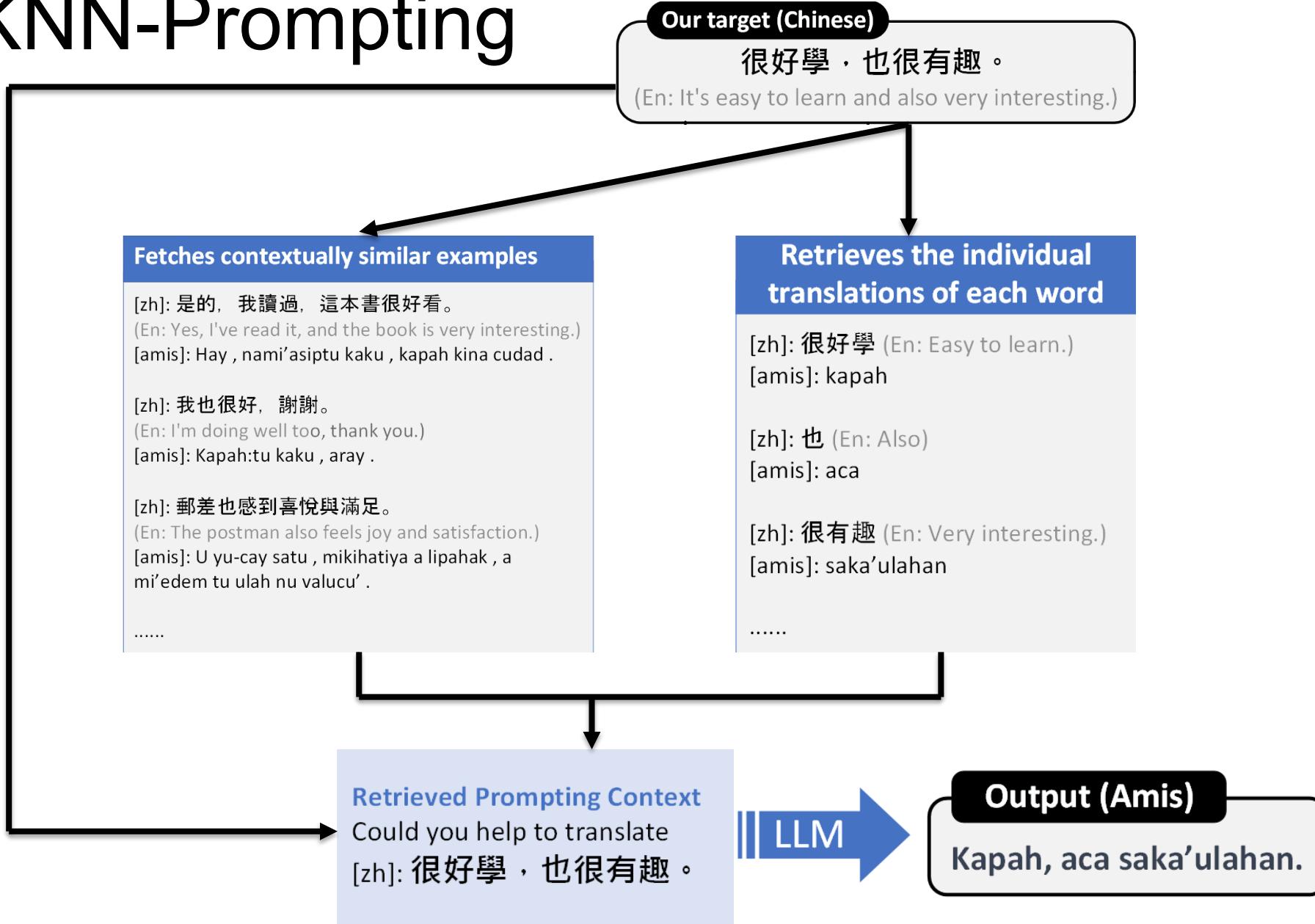
[zh]: 也 (En: Also)
[amis]: aca

[zh]: 很有趣 (En: Very interesting.)
[amis]: saka'ulahan

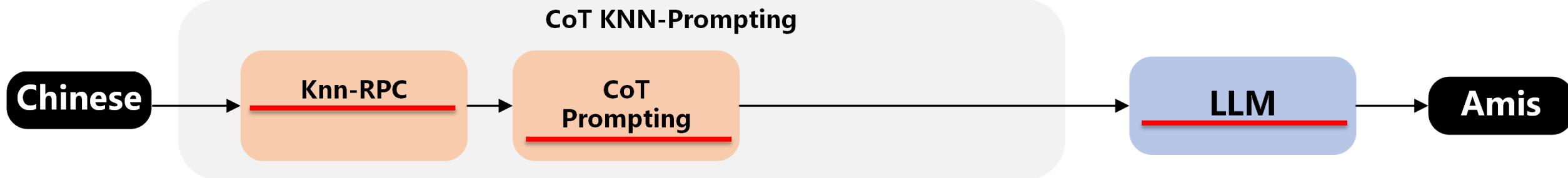
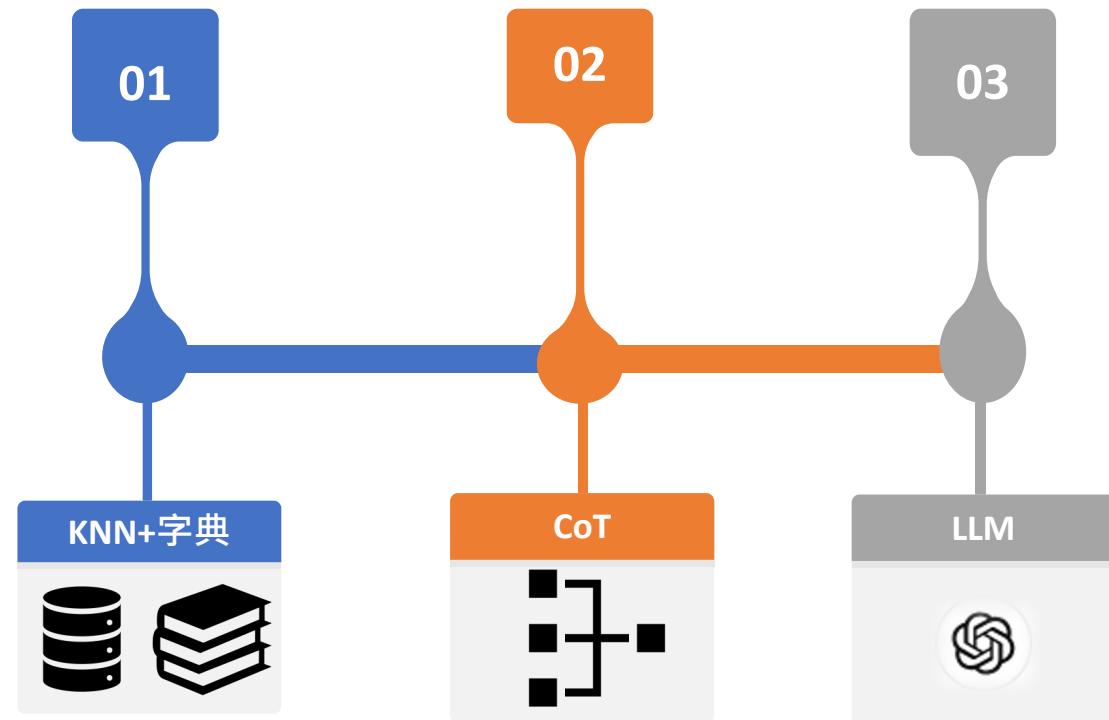
.....



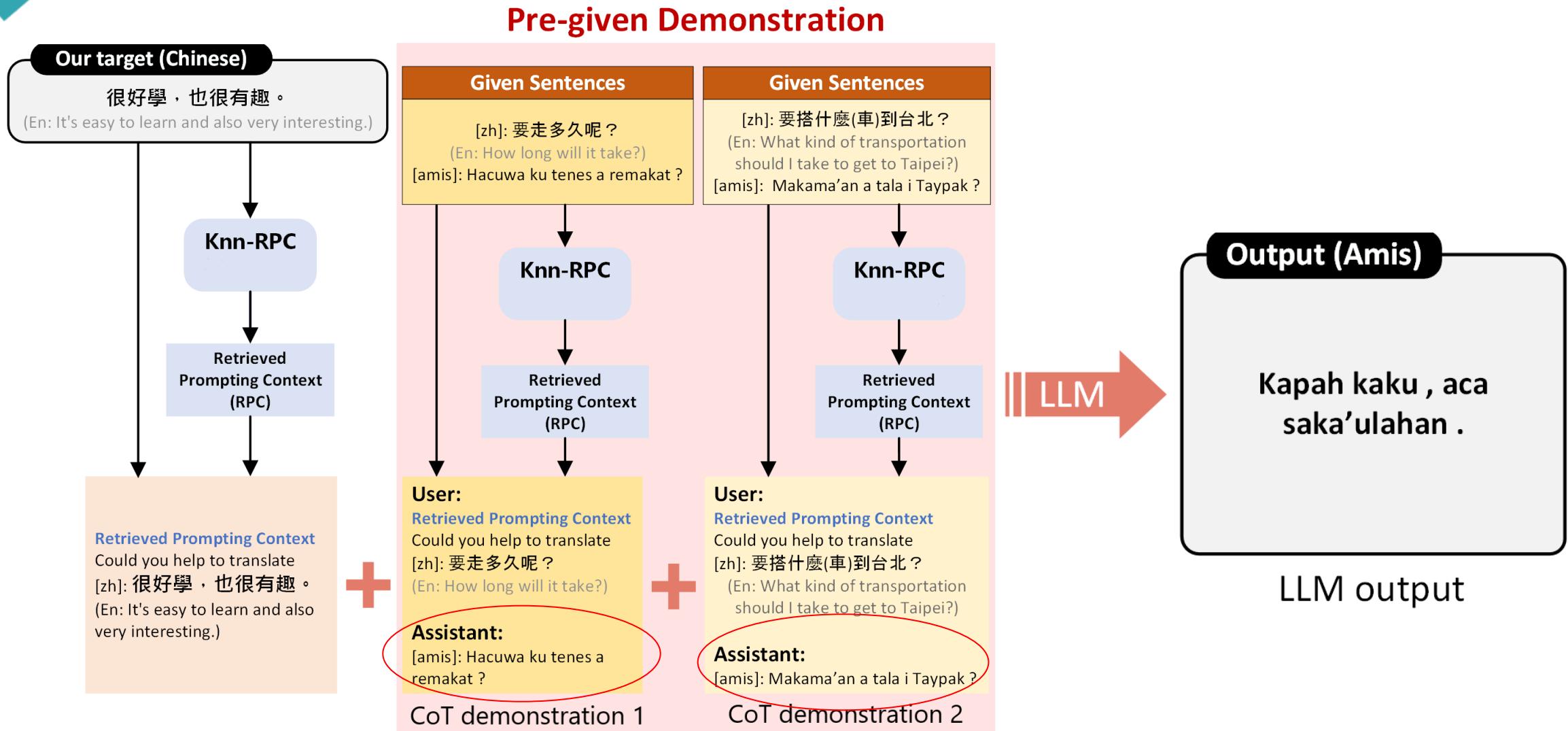
KNN-Prompting



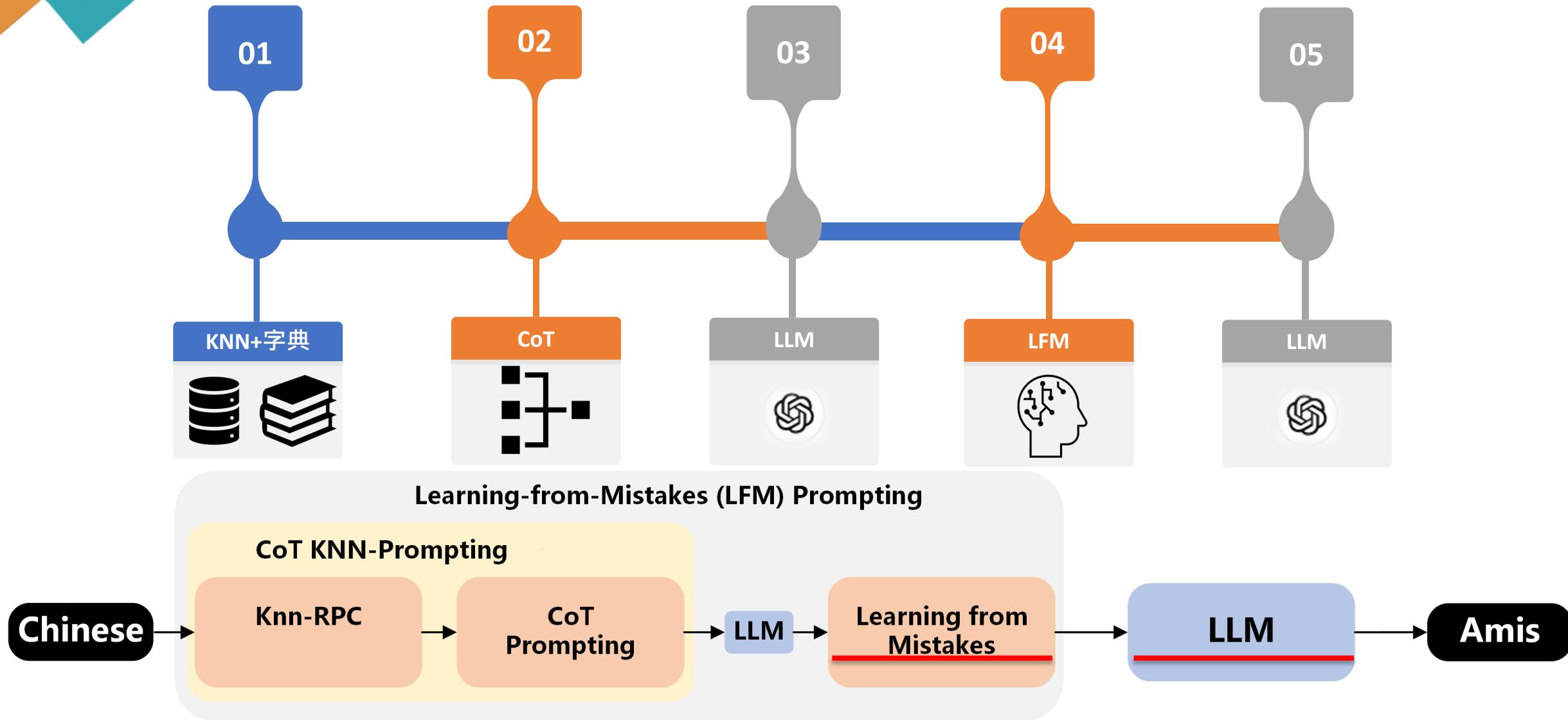
CoT-Prompting



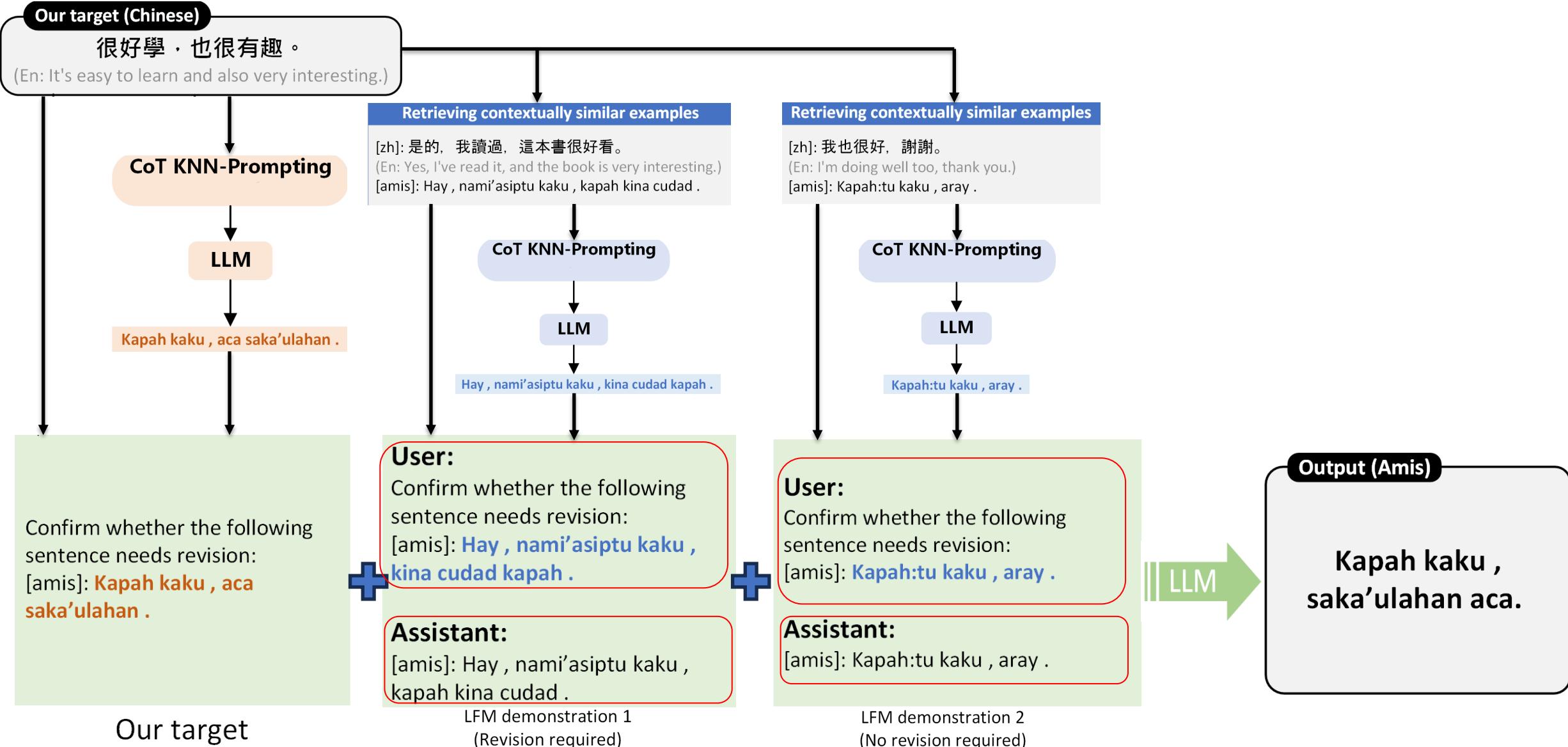
CoT-Prompting



LFM (Learning From Mistake) -Prompting



LFM-Prompting



Performance Overview

Coastal Amis				
Methods	BLEU1_{STD}	BLEU2_{STD}	BLEU3_{STD}	chrF++_{STD}
Knn-Prompting w. RPC (k=5)	42.9 _{1.8}	11.8 _{0.9}	4.7 _{1.1}	45.4 _{0.8}
Knn-Prompting w. RPC (k=10)	43.3 _{1.2}	13.4 _{0.6}	5.8 _{0.8}	44.8 _{1.1}
CoT Prompting	44.5 _{2.8}	11.9 _{3.0}	4.7 _{2.3}	45.7 _{1.6}
LFM Prompting	44.1 _{2.0}	12.6 _{2.9}	5.7 _{2.5}	46.1 _{1.8}

Wanda Tayal				
Methods	BLEU1_{STD}	BLEU2_{STD}	BLEU3_{STD}	chrF++_{STD}
Knn-Prompting w. RPC (k=5)	41.5 _{2.5}	13.0 _{2.1}	4.8 _{1.6}	42.5 _{2.4}
Knn-Prompting w. RPC (k=10)	42.1 _{2.2}	13.6 _{2.7}	5.7 _{1.4}	42.8 _{2.6}
CoT Prompting	46.3 _{1.8}	14.4 _{2.6}	5.8 _{1.2}	44.7 _{2.1}
LFM Prompting	45.2 _{2.0}	14.0 _{1.7}	5.8 _{0.9}	43.9 _{2.0}

Siji Tayal				
Methods	BLEU1_{STD}	BLEU2_{STD}	BLEU3_{STD}	chrF++_{STD}
Knn-Prompting w. RPC (k=5)	44.3 _{3.2}	14.6 _{2.1}	4.9 _{1.7}	39.3 _{2.0}
Knn-Prompting w. RPC (k=10)	44.4 _{3.0}	14.5 _{2.0}	5.4 _{2.3}	40.9 _{1.8}
CoT Prompting	47.5 _{2.7}	16.0 _{1.2}	5.9 _{1.4}	41.2 _{1.0}
LFM Prompting	50.0 _{1.2}	20.0 _{1.4}	9.3 _{2.0}	43.4 _{2.0}

Duda Seediq				
Methods	BLEU1_{STD}	BLEU2_{STD}	BLEU3_{STD}	chrF++_{STD}
Knn-Prompting w. RPC (k=5)	45.0 _{1.2}	16.2 _{1.5}	5.4 _{0.8}	38.2 _{0.8}
Knn-Prompting w. RPC (k=10)	45.7 _{1.2}	17.1 _{1.4}	6.7 _{1.5}	39.3 _{1.6}
CoT Prompting	46.1 _{1.6}	17.5 _{1.4}	6.9 _{1.0}	38.9 _{1.1}
LFM Prompting	46.3 _{1.5}	17.3 _{2.1}	6.9 _{1.4}	39.3 _{1.2}

Coastal Amis				
Models	BLEU1_{STD}	BLEU2_{STD}	BLEU3_{STD}	BLEU4_{STD}
Breeze-7b	45.6 _{1.2}	17.1 _{0.2}	9.0 _{0.7}	4.4 _{0.6}
GPT-3.5	47.1 _{1.6}	16.5 _{1.3}	7.5 _{2.1}	2.8 _{1.1}
GPT-4	50.9 _{1.3}	21.9 _{2.0}	11.6 _{2.5}	5.7 _{1.7}

Wanda Tayal				
Models	BLEU1_{STD}	BLEU2_{STD}	BLEU3_{STD}	BLEU4_{STD}
Breeze-7b	44.5 _{1.6}	18.6 _{2.0}	8.5 _{2.3}	3.7 _{1.8}
GPT-3.5	47.6 _{1.6}	18.0 _{1.2}	8.7 _{1.5}	4.6 _{1.0}
GPT-4	49.8 _{0.7}	20.6 _{1.8}	10.0 _{2.0}	4.5 _{1.1}

Siji Tayal				
Models	BLEU1_{STD}	BLEU2_{STD}	BLEU3_{STD}	BLEU4_{STD}
Breeze-7b	47.9 _{2.7}	21.5 _{3.0}	10.3 _{2.8}	5.7 _{2.6}
GPT-3.5	51.6 _{1.2}	22.1 _{1.2}	9.6 _{1.3}	5.0 _{1.5}
GPT-4	54.4 _{1.8}	26.1 _{2.8}	13.2 _{2.1}	7.4 _{1.7}

Southern Amis				
Models	BLEU1_{STD}	BLEU2_{STD}	BLEU3_{STD}	BLEU4_{STD}
Breeze-7b	44.3 _{1.1}	18.9 _{0.9}	8.5 _{1.5}	4.2 _{1.2}
GPT-3.5	48.2 _{1.5}	20.4 _{2.0}	9.6 _{1.3}	4.8 _{1.4}
GPT-4	51.5 _{2.2}	28.1 _{3.8}	17.1 _{3.0}	10.1 _{2.2}



Research Goal

Can LLM understand unknown languages with few-shot examples?

This study investigates LLMs' ability to learn from

1. Parallel language pairs and 
2. Individual word mappings

To construct coherent sentences in target languages.

The focus is on enabling these models to handle extremely low-resourced languages, including indigenous and ancient dialects.

Research Goal

Can LLM understand unknown languages with few-shot examples?

This study investigates LLMs' ability to learn from

1. Parallel language pairs and
2. Individual word mappings 

To construct coherent sentences in target languages.

The focus is on enabling these models to handle extremely low-resourced languages, including indigenous and ancient dialects.

Using Context-Free Grammar to Create an Unseen Language

```

 $S \rightarrow VP\ NP$ 
 $VP \rightarrow V\ NP \mid V\ NP\ PP \mid V$ 
 $NP \rightarrow Det\ N \mid Det\ N\ PP \mid Pronoun$ 
 $PP \rightarrow Prep\ NP$ 

```

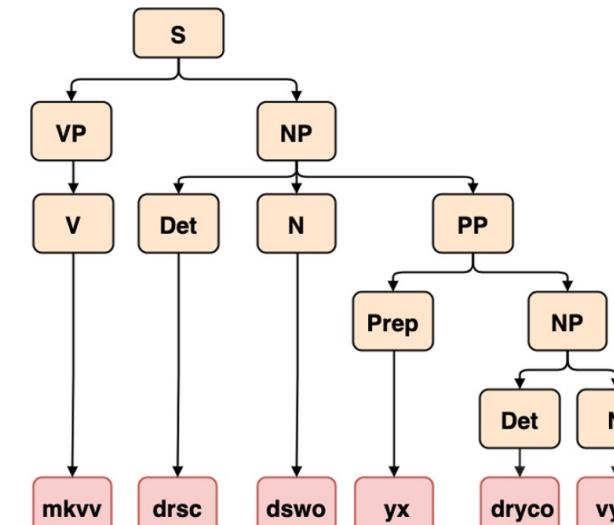


Figure 1: Parse Tree of a Sentence of Source Language

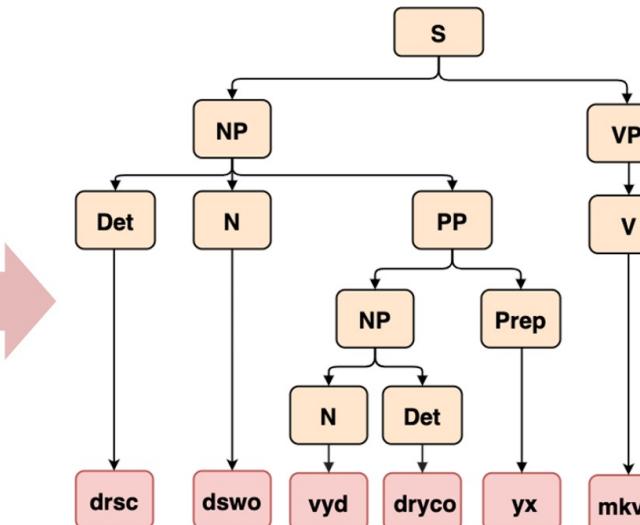


Figure 2: Parse Tree of Target Language with respect to Source Sentence in Fig 1



Performance Evaluation

100 Sentences				
	BLEU1	BLEU2	BLEU3	BLEU4
GPT3.5	0.9805	0.5610	0.2877	0.0104
GPT4	0.9912	0.6210	0.2877	0.4408

1000 Sentences				
	BLEU1	BLEU2	BLEU3	BLEU4
GPT3.5	0.9918	0.6679	0.3782	0.1692
GPT4	0.9941	0.7817	0.5682	0.3367

10000 Sentences				
	BLEU1	BLEU2	BLEU3	BLEU4
GPT3.5	0.9900	0.6498	0.3686	0.1733
GPT4	0.9946	0.8023	0.6404	0.4308

Table 1: In-context Learning Performance over Various Size of Sentence Set

Short Conclusion

- LLM as an Intelligence to understand the language structure through examples



Thanks and Questions ?

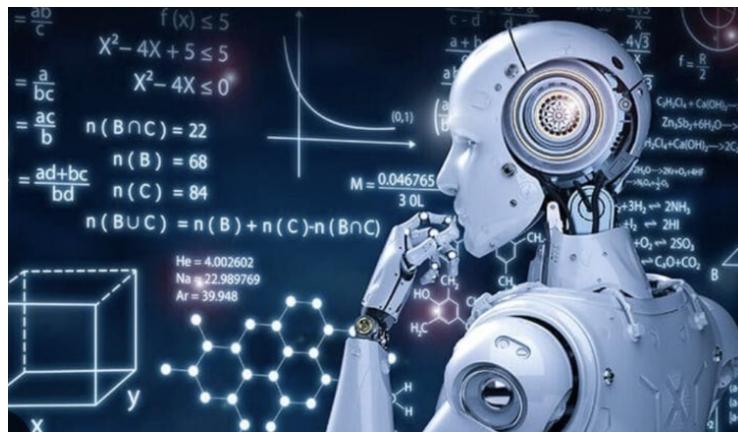


Mr Hendricks. Are you sure
you are in the right room?
This is MachineLearning 101.



Conclusion

- Be Aware of Singularity Now



Accelerating Growth in Technology

(condensed)

