

元智大學資管系

學術類畢業專題頂石課程(二)

期末報告

針對住宿設施的日文評論進行基於面向的情感分析
(ABSA)

高田朋希、甲斐陸都、溝上大翔、

佐佐木華、小笠原伶晏

指導教授：禹良治 老師

中華民國 114 年 11 月

Nov, 2025

Abstract

本報告書總結了針對日文飯店評論的維度面向情感分析 (Dimensional Aspect-Based Sentiment Analysis, DimABSA) 的研究成果。本研究旨在超越傳統的定性情感分類，建立一套運用 Valence (V: 愉悅-不快) 和 Arousal (A: 覺醒-鎮靜) 二維連續值 (VA Score) 的定量分析方法。本研究最大的貢獻在於奠定了兩大基石：

首先，透過嚴格的手動標註作業，我們在日文評論中完成了「面向詞」、「評價詞」的結構化提取，並賦予 VA 連續值，成功建立了涵蓋 DimABSA 全部三個子任務所需的高品質數據集。

其次，我們針對複雜的結構化提取任務（三元組及四元組提取）引入了 LLM (Qwen 8B) 的指令微調方法，並成功地在 Subtask 2 實現了 cF1 0.4284，在 Subtask 3 實現了 cF1 0.3479，有力地證明了 LLM 在處理日文多維情感結構提取任務中的實用性。

儘管回歸任務 (Subtask 1) 中的最佳模型仍面臨 VA 預測趨勢相關性極低的挑戰，本研究成功將日文評論分析的軸心轉移至情感強度的精確定量把握，為未來 DimABSA 領域的應用與模型重構建立了實證基礎。

Contents

Abstract	i
Contents	ii
Chapter 1. 緒論	1
1.1 研究背景	1
1.2 問題意識	1
1.3 研究目的	1
1.4 研究架構	1
Chapter 2. 相關技術與研究	2
2.1 情感分析與 ABSA 的基礎	2
2.1.1 情感分析的概要	2
2.1.2 ABSA (基於面向的情感分析)	2
2.2 ABSA 的課題與展望	2
Chapter 3. 研究方法	3
3.1 研究設定	3
3.2 資料前處理	3
3.3 標註作業	5
3.4 全體標註流程與一致率提升策略	7
3.4.1 第一次全體標註與配對組形成	7

3.4.2 暑期會議與第二次標註	7
3.4.3 工具改良與規則追加	8
3.4.4 進階標註與 Final 準備	9
3.4.5 Final 作業與修正	9
3.4.6 Value Annotation (VA) 作業	9
3.5 DimABSA 競賽提交與評估方法	10
Chapter 4. 實驗結果或系統展示	12
4.1 面向詞抽取標註的實施方法	12
4.2 評價詞 (Opinion) 標註規則之制定	12
4.3 抽取結果之範例與觀察	12
4.4 標註作業之課題、反思與解決策略	14
4.4.1 克服多人標註的一致性問題	14
4.4.2 導入情感價 Valence/Arousal (VA) 以解決多層次評價問題	14
4.4.3 技術挑戰與流程改善	15
4.5 最終成果：高品質日文 ABSA + VA 數據集之完成	15
4.6 最終結果 (Final Results)	15
4.6.1 評估概述與分數	15
4.6.2 子任務分析	16
4.7 小結：系統優勢與限制	18
Chapter 5. 結論	19

5.1 目標達成與系統驗證.....	19
5.2 核心成果與實證.....	19
5.3 主要貢獻	19
5.4 未來展望	19
Reference	21
附錄 A. 專題工作內容	22
附錄 B. 專題心得與建議	24

Chapter 1. 緒論

1.1 研究背景

近年來，網路上每天都有大量關於商品與服務的感想與意見被發布，這些評論與口碑對消費者而言是寶貴的參考資訊。同時，對企業來說，分析顧客的聲音也有助於改善服務與制定新策略。因此，能夠自動從評論中抽取與分類評價與情感的「情感分析（Sentiment Analysis）」其重要性日益提升。

在情感分析之中，能夠針對文章中所包含的各種觀點（面向）進行細緻情感判斷的「基於面向的情感分析（Aspect-Based Sentiment Analysis, ABSA）」受到矚目，因其有助於更具體地掌握顧客需求與評估服務品質。例如，在飯店評論中若寫道「地點很方便，但房間太小」，ABSA 便能分別判斷為「地點：正面」「房間：負面」，相比於單純的整體評價，能提供更精細的分析。

1.2 問題意識

目前，ABSA（基於面向的情感分析）主要在英文領域已有相當多的研究成果，並且針對餐廳、筆記型電腦、電影評論等主題也已建立起大規模的資料集。然而，相較之下，在日文領域中，雖然一般性的情感分析資料集已經存在，專門用於 ABSA 的公開資料卻極為有限。這成為日本語 ABSA 研究與技術應用難以推展的重要原因之一。

此外，日文語言本身具有省略主詞與受詞的傾向，常出現模糊或委婉的表達方式，這使得在評論中進行面向判斷與情感分類比英語更加困難。例如，在實際的評論中常見「如果房間能再大一點就好了」這類迴避直接否定的表述。這些日語特有的語言特性，也成為建立高精準度 ABSA 模型所面臨的重要挑戰。

1.3 研究目的

本專題旨在針對日文評論文本，建立一套能夠分類文中所包含之面向（Aspect）及其對應情感的模型，並驗證其有效性。透過活用有限的日文資料與已習得的自然語言處理知識，實作 ABSA 模型並進行分析，以探討日文情感分析的潛力與面臨的課題。

1.4 研究架構

第二章將概括本研究的基礎，包括 ABSA 的基本概念、技術背景及現有研究的趨勢。接著，在第三章中，將介紹以該數據集為基礎所開發的能同時抽取 Aspect、Opinion 及 VA 的模型的構築方法與具體研究方法。第四章將展示所構建模型的實驗結果的分析。尤其，透過對實驗結果的分析，我們將闡述最終課題的發現以及達成目標前的解決方法。最後，在結論章中，我們將對本研究在提升日文評論分析精度與實用性方面所做的貢獻進行總結。

Chapter 2. 相關技術與研究

在本章中，將介紹本專題所探討之基於面向的情感分析（Aspect-Based Sentiment Analysis, ABSA）相關的基本技術，以及近年情感分析領域中的主要方法。特別聚焦於自然語言處理領域中廣泛應用的深度學習模型與前處理方法等，說明實作過程中所需的關鍵要素。

2.1 情感分析與 ABSA 的基礎

2.1.1 情感分析的概要

情感分析是一種從文字中讀取意見或感想，並判斷其屬於「正面（Positive）」、「負面（Negative）」或「中立（Neutral）」的技術。分析對象為評論網站內容。

2.1.2 ABSA（基於面向的情感分析）

ABSA 是情感分析中一種更為細緻的分析方法，其目的是針對文本中所包含的各個面向（評價對象），判斷其分別所對應的情感傾向。例如，在一則飯店評論中提到：「地點很方便，但早餐有些令人失望」，則可分別分析為「地點＝正面」、「早餐＝負面」，如此便能從多個觀點進行情感評估。

本專題聚焦於此類句子中，自動抽取如「地點」「早餐」等面向詞（Aspect Term）的部分。

2.2 ABSA 的課題與展望

ABSA 在技術上仍面臨數項挑戰。例如，有些評論文中並未明確指出面向詞（Aspect），必須依靠上下文進行推測。此外，日語中常見的模糊表達、委婉的否定或間接的建議語句等，也會使模型的判斷變得更加困難。為了解決這些問題，不僅需要提升模型的準確度，還必須在標註方法上加以創新，並強化模型的語境理解能力。

本專題即是以提升面向詞抽取的準確性為目標，同時兼顧日語中獨特的語言表達與句法特性。

Chapter 3. 研究方法

3.1 研究設定

本研究在使用飯店評論資料以訓練情感分析模型的過程中，依據指導教授的建議，採用 ABSA（基於面向的情感分析）方法。透過 ABSA，模型得以自評論文本中抽取面向詞（Aspect Term），並判斷其所對應的情感極性。

原先本研究向樂天提出資料集使用申請，希望能取得官方飯店評論資料。然而，由於未獲得正式回覆，故改以自製網路爬蟲蒐集樂天旅遊網站上所公開的評論內容。最終共取得約 4,500 則飯店評論。

3.2 資料前處理

在資料蒐集後，本研究進行面向（Aspect）、評價詞（Opinion）、情感極性（Sentiment）等標註之資料前處理作業。此外，為進行句子層級分析，我們使用 Python 進行評論句子切分及屬性標註。

例：

● 原始資料

```
1 {
2   "hotel_id": "127",
3   "prefecture": "hokkaido",
4   "stay_date": "2024年09月",
5   "review_text": "札幌にも法華クラブを見つけて即決で予約しました。朝食の石狩鍋やザンギなどご当地おかず、
6   福岡で食べておいしかったナンコツの醤油煮もあり、よくばって食べてしまいました。お部屋が古いためか少し匂い
   が気になりましたが、清潔にしてありゆっくり過ごす事ができました。"
```

● 前處理後的資料

```
1 {
2   "aspect": "部屋",
3   "opinion": "清潔",
4   "sentiment": "Positive",
5   "entity": "Room",
6   "sentence": "お部屋が古いためか少し匂いが気になりましたが、清潔にしてありゆっくり過ごす事ができました。",
7   "hotel_id": "127"
8 },
9
10 {
11   "aspect": "部屋",
12   "opinion": "古い",
13   "sentiment": "Negative",
14   "entity": "Room",
15   "sentence": "お部屋が古いためか少し匂いが気になりましたが、清潔にしてありゆっくり過ごす事ができました。",
16   "hotel_id": "127"
17 }
18
```

透過前處理，我們成功為每句評論新增 aspect（面向）、opinion（評價詞）、

sentiment（情感極性）、entity（實體）、sentence（句子）、hotel_id（飯店編號）等資訊。然而，指導教授指出資料仍存在下列問題：

（1）Aspect（面向）與 Opinion（評價詞）的對應不穩定

例如：

● 原始資料

```
1 {
2   "hotel_id": "109107",
3   "prefecture": "hokkaido",
4   "stay_date": "2024年09月",
5   "review_text": "部屋は広くて良かったですけどベッドの頭のほうの壁？にガリガリやられたような傷がたくさんあったりお風呂のカーテンの裏にながーい髪の毛が張り付いてたりちょっと嫌でした...でも朝食も美味しいし、コンビニもあって駅チカ、すすきのから歩いても意外と近くて良かったです。"
6 }
7
8
```

● 處理後的資料

```
1 {
2   "aspect": "朝食",
3   "opinion": "美味しい",
4   "sentiment": "Positive",
5   "entity": "Breakfast",
6   "sentence": "でも朝食も美味しいし、コンビニもあって駅チカ、すすきのから歩いても意外と近くて良かったです。",
7   "hotel_id": "109107"
8 }
9
10 {
11   "aspect": "駅",
12   "opinion": "美味しい",
13   "sentiment": "Positive",
14   "entity": "Location",
15   "sentence": "でも朝食も美味しいし、コンビニもあって駅チカ、すすきのから歩いても意外と近くて良かったです。",
16   "hotel_id": "109107"
17 }
18
```

如例所示，「車站（location）」的面向在標註後被錯誤歸類為「美味」這類評價詞，顯示僅依賴詞語共現容易造成錯誤分類。

（2）Entity 與 Attribute 未被定義

本次初步標註並未明確界定評論中的 Entity（評價對象）與 Attribute（評價角度）。教授因此要求：

- 重新定義并統一 Entity/Attribute
- 情感極性（Sentiment）須人工標註
- 根據教授提供之先行研究 Python 工具重新調整程式

研究方針之調整

- Entity、Attribute 依教授指定之論文分類方式

Appendix C. Hotel Aspect Categories

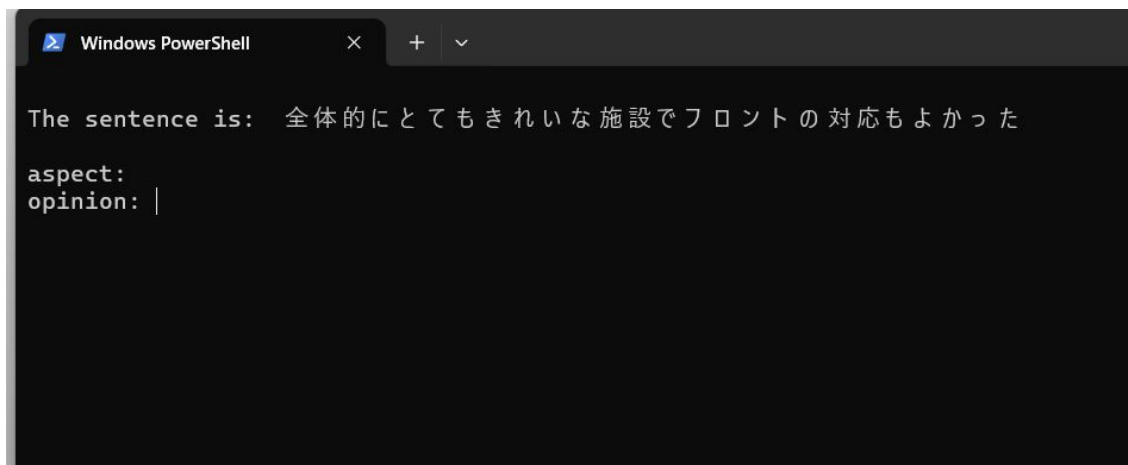
Entity Labels	Attribute Labels
1. HOTEL	A. GENERAL
2. ROOMS	B. PRICE
3. FACILITIES	C. COMFORT
4. ROOM AMENITIES	D. CLEANLINESS
5. SERVICE	E. QUALITY
6. LOCATION	F. DESIGN & FEATURES
7. FOOD & DRINKS	G. STYLE & OPTIONS
	H. MISCELLANEOUS

- 調整教授提供之程式，作為手動 ABSA 標註工具
- 兩人一組，個別標註 50 筆，並相互比對一致性
- 一致率（相符率）目標至少為 70%

3.3 標註作業

使用改良後的工具，我們對評論資料逐句進行人工標註並比較兩名成員之結果。

- 工具輸入畫面



例：

● 標註結果 A

```
{
  "ID": "IGEMFCR5sNrM3ZhBml3RRQ",
  "Sentence": "露天風呂が最高ですねーサウナがあれば100点満点",
  "Aspect": [
    "露天風呂",
    "サウナ"
  ],
  "Category": [
    "Facilities#Quality",
    "Facilities#General"
  ],
  "Opinion": [
    "最高ですねー",
    "100点満点"
  ]
}
```

● 標註結果 B

```
{
  "ID": "IGEMFCR5sNrM3ZhBml3RRQ",
  "Sentence": "露天風呂が最高ですねーサウナがあれば100点満点",
  "Aspect": [
    "露天風呂",
    "サウナ"
  ],
  "Category": [
    "Facilities#Comfort",
    "Facilities#Style & Options"
  ],
  "Opinion": [
    "最高",
    "あれば100点満点"
  ]
}
```

比較後發現下列不一致情形：

標註不一致的主要原因

1. 助動詞、接尾詞、敬語等是否應納入標註缺乏統一標準
2. 日文常見省略句型導致 Aspect 判斷不一致
3. Aspect/Opinion 是否應以單字或片語標註存在歧見
4. 「Quality」「General」等分類過於寬泛，易產生不同解讀

新增標註規則

- 統一單詞擷取形式
- 若無適當的 Aspect，統一標註為 NULL

第二次標註後一致率變化

- Aspect : 52% → 59%
- Opinion : 20% → 50%
- Category : 20% → 39%

此結果顯示，標註規範的統一能顯著提升一致性。

3.4 全體標註流程與一致率提升策略

3.4.1 第一次全體標註與配對組形成

在教授指示下，全員共同標註相同的 100 句評論，並依一致率組成 7 組搭配組合。同時，高田撰寫工具整合 7 組 JSON 結果轉為 CSV，據此進行第一次一致率分析。

```
=== Krippendorff's alpha (nominal) ===
Aspect          : 0.820
Opinion         : 0.771
Category - Entity : 0.881
Category - Attribute: 0.557
```

其中，小笠原與甲斐在 Category 標註上需改善，兩人重新進行標註。

```
=== Krippendorff's alpha (nominal) ===
Aspect          : 0.805
Opinion         : 0.710
Category - Entity : 0.831
Category - Attribute : 0.430
Category (Entity#Attribute): 0.479
```

本階段發現之問題

- 資料 ID 重複或遺漏
- Aspect 欄位空白（未填 NULL）造成欄位位移
- 手動輸入造成表達不一致與錯誤
- 應拆分為多個 Aspect 卻被誤合併

→ 全員再度確認標註規則並修訂作業流程。

3.4.2 暑期會議與第二次標註

暑假期間，每週皆召開會議以統一如下規範：

- 不使用「Miscellaneous」分類
- 明確定義「Comfort」

- 區分「Quality」與「Style & Options」
- 釐清「Design & Features」與「Comfort」差異

再標註後，「Category」一致率改善有限。教授指出計算方式可能導致重複計算，須重新修正演算法。

第二次結果：

```

=== Krippendorff's alpha (nominal) ===
Aspect                : 0.691
Opinion               : 0.668
Category - Entity     : 0.744
Category - Attribute  : 0.493
Category (Entity#Attribute): 0.489
-----
Overall (weighted mean) : 0.615

```

3.4.3 工具改良與規則追加

- 排除如「ありがとう」「グレードアップ」等非 Opinion 評價詞
- 高田開發 JSON→Excel、JSON→JSONL 等工具
- 溝上新增自動輸出一致率 CSV 功能

教授新指示：

- 每組標註 300 句
- 先進行 50 句試標註
 - 一致率 <70%：討論後再標註
 - ≥70%：每次續標註 100 句

之後：

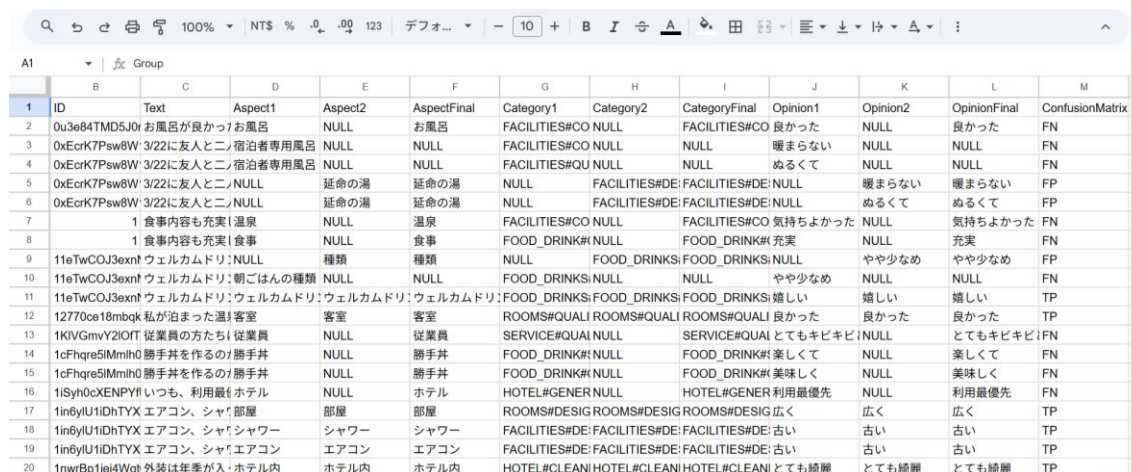
- 「Quad」改為「Quadruplet」
- Category 名稱統一修正
- 一致率較低之組別須召開追加會議

3.4.4 進階標註與 Final 準備

教授提出新要求：

- 計算刪除後剩餘句數
- 計算全組一致率 (F1 分數)
- 7 組加權平均

高田修正 JSONL 結構，溝上配布 8 月資料集並整理 7 月成果。教授於例會說明 Final (以兩人標註結果產生正解資料) 之作業流程並分配任務。



	B	C	D	E	F	G	H	I	J	K	L	M
1	ID	Text	Aspect1	Aspect2	AspectFinal	Category1	Category2	CategoryFinal	Opinion1	Opinion2	OpinionFinal	ConfusionMatrix
2	0u3e84TMD5J0r	お風呂が良かった	お風呂	NULL	お風呂	FACILITIES#CO	NULL	FACILITIES#CO	良かった	NULL	良かった	FN
3	0xEcrK7Psw8W	3/22に友人と二人宿泊者専用風呂	NULL	NULL	NULL	FACILITIES#CO	NULL	NULL	暖まらない	NULL	NULL	FN
4	0xEcrK7Psw8W	3/22に友人と二人宿泊者専用風呂	NULL	NULL	NULL	FACILITIES#QU	NULL	NULL	ぬるくて	NULL	ぬるくて	FN
5	0xEcrK7Psw8W	3/22に友人と二人	NULL	延命の湯	延命の湯	NULL	FACILITIES#DE	FACILITIES#DE	NULL	暖まらない	暖まらない	FP
6	0xEcrK7Psw8W	3/22に友人と二人	NULL	延命の湯	延命の湯	NULL	FACILITIES#DE	FACILITIES#DE	NULL	ぬるくて	ぬるくて	FP
7	1	食事内容も充実	温泉	NULL	温泉	FACILITIES#CO	NULL	FACILITIES#CO	気持ちよかった	NULL	気持ちよかった	FN
8	1	食事内容も充実	食事	NULL	食事	FOOD_DRINK#	NULL	FOOD_DRINK#	充実	NULL	充実	FN
9	11eTwCOJ3exn	ウェルカムドリ	種類	NULL	種類	NULL	FOOD_DRINKS	FOOD_DRINKS	NULL	やや少なめ	やや少なめ	FP
10	11eTwCOJ3exn	ウェルカムドリ	朝ごはんの種類	NULL	NULL	FOOD_DRINKS	NULL	NULL	やや少なめ	NULL	NULL	FN
11	11eTwCOJ3exn	ウェルカムドリ	ウェルカムドリ	ウェルカムドリ	ウェルカムドリ	FOOD_DRINKS	FOOD_DRINKS	FOOD_DRINKS	嬉しい	嬉しい	嬉しい	TP
12	12770ce18mbqk	私が泊まった温	客室	客室	客室	ROOMS#QUALI	ROOMS#QUALI	ROOMS#QUALI	良かった	良かった	良かった	TP
13	1KIVGmvY2iOrT	従業員の方たち	従業員	NULL	従業員	SERVICE#QUAI	NULL	SERVICE#QUAI	とてもキビキビ	NULL	とてもキビキビ	FN
14	1cFhqr5IMmh0	勝手井を作るの	勝手井	NULL	勝手井	FOOD_DRINK#	NULL	FOOD_DRINK#	楽しくて	NULL	楽しくて	FN
15	1cFhqr5IMmh0	勝手井を作るの	勝手井	NULL	勝手井	FOOD_DRINK#	NULL	FOOD_DRINK#	美味しく	NULL	美味しく	FN
16	1Syh0cXENPYf	いつも、利用最	ホテル	NULL	ホテル	HOTEL#GENER	NULL	HOTEL#GENER	利用最優先	NULL	利用最優先	FN
17	1in6yIU1IDhTYX	エアコン、シャ	部屋	部屋	部屋	ROOMS#DESIG	ROOMS#DESIG	ROOMS#DESIG	広く	広く	広く	TP
18	1in6yIU1IDhTYX	エアコン、シャ	シャワー	シャワー	シャワー	FACILITIES#DE	FACILITIES#DE	FACILITIES#DE	古い	古い	古い	TP
19	1in6yIU1IDhTYX	エアコン、シャ	エアコン	エアコン	エアコン	FACILITIES#DE	FACILITIES#DE	FACILITIES#DE	古い	古い	古い	TP
20	1nwrBp1iei4Wat	外装は年季が入	ホテル内	ホテル内	ホテル内	HOTEL#CLEAN	HOTEL#CLEAN	HOTEL#CLEAN	とても綺麗	とても綺麗	とても綺麗	TP

3.4.5 Final 作業與修正

Final 進行期間多次發現資料錯誤，經反覆修正後最終完成全部 7 名成員之 Final 標註。

3.4.6 Value Annotation (VA) 作業

教授講解 VA (Value Annotation) 並啟用專屬網站，全員註冊後先進行 35 句試標註。後續依指示完成：

- 142 句 + 額外 15 句
- 修正後每日 115 句 × 9 日
- 追加 50 句
- 每日 100 句 × 7 日

最終每人完成 550 句 VA 標註。

標註錯誤紀錄由高田依教授指示紀錄於 Notion。

專屬網站：

Annotation System 1111661

已標記數量: 186

Sentence: 窓からの景色は楽しめませんが、**広くて清潔な部屋**、**ゆったり**と足を伸ばせる**お風呂**、**快適な空間**でくつろげて**良かった**です

1. Aspect: お風呂, Opinion: ゆったり, Category: FACILITIES#COMFORT

Valence	☐ 1	☐ 1.5	☐ 2	☐ 2.5	☐ 3	☐ 3.5	☐ 4	☐ 4.5	☐ 5	☐ 5.5	☐ 6	☐ 6.5	☒ 7	☐ 7.5	☐ 8	☐ 8.5	☐ 9
Arousal	☐ 1	☐ 1.5	☐ 2	☐ 2.5	☐ 3	☐ 3.5	☐ 4	☐ 4.5	☐ 5	☐ 5.5	☒ 6	☐ 6.5	☐ 7	☐ 7.5	☐ 8	☐ 8.5	☐ 9

2. Aspect: 空間, Opinion: くつろげて, Category: HOTEL#COMFORT

Valence	☐ 1	☐ 1.5	☐ 2	☐ 2.5	☐ 3	☐ 3.5	☐ 4	☐ 4.5	☐ 5	☐ 5.5	☐ 6	☐ 6.5	☒ 7	☐ 7.5	☐ 8	☐ 8.5	☐ 9
Arousal	☐ 1	☐ 1.5	☐ 2	☐ 2.5	☐ 3	☐ 3.5	☐ 4	☐ 4.5	☐ 5	☐ 5.5	☐ 6	☒ 6.5	☐ 7	☐ 7.5	☐ 8	☐ 8.5	☐ 9

3. Aspect: 部屋, Opinion: 広くて, Category: ROOMS#COMFORT

Valence	☐ 1	☐ 1.5	☐ 2	☐ 2.5	☐ 3	☐ 3.5	☐ 4	☐ 4.5	☐ 5	☐ 5.5	☐ 6	☐ 6.5	☒ 7	☐ 7.5	☐ 8	☐ 8.5	☐ 9
Arousal	☐ 1	☐ 1.5	☐ 2	☐ 2.5	☐ 3	☐ 3.5	☐ 4	☐ 4.5	☐ 5	☐ 5.5	☒ 6	☐ 6.5	☐ 7	☐ 7.5	☐ 8	☐ 8.5	☐ 9

3.5 DimABSA 競賽提交與評估方法

本研究最終將全部標註資料整理為 JSONL 檔，提交至 SemEval-2026 主辦競賽平台 Codabench，參加 Task 3: Dimensional Aspect-Based Sentiment Analysis (DimABSA) Track A。

DimABSA 特點

- 不使用「正面／負面」等類別標籤
- 採用 Valence（愉悅—不快）、Arousal（覺醒—鎮靜）之連續值進行情感量化

本研究處理的三個子任務

1. Subtask 1 (DimASR)：情感回歸（預測 VA 分數）
2. Subtask 2 (DimASTE)：三元組抽取 (Aspect, Opinion, VA)
3. Subtask 3 (DimASQP)：四元組抽取 (Aspect, Category, Opinion, VA)

利用模型, 設計

- Subtask 1：以 Transformer（如 BERT-base）為基礎，添加迴歸層進行微調
 - cl-tohoku/bert-base-japanese-whole-word-masking
 - Bert-base-multilingual-cased
 - cl-tohoku/bert-base-japanese-char-v3
- Subtask 2/3：採用指令微調 (Instruction Tuning) 之大型語言模型 (LLM)

- unsloth/Qwen3-4B-Instruct-2507-bnb-4bit
- unsloth/Qwen3-8B-unsloth-bnb-4bit
- unsloth/gemma-2-9b-it-bnb-4bit

評估方式

- Subtask 1 : RMSE_VA , PCC V / PCC A
- Subtask 2/3 : cF1, cPrecision , cRecall

Chapter 4. 實驗結果或系統展示

4.1 面向詞抽取標註的實施方法

本研究以刊載於樂天旅遊網站之日文飯店評論作為資料來源，進行 Aspect-Based Sentiment Analysis (ABSA) 的基礎工作，並將重點放在 面向詞 (Aspect Term) 之人工標註作業。評論內容由實際住宿的顧客撰寫，語言自然、表達方式多元，具有高度真實性。本次標註作業由兩名成員共同執行，針對 100 筆評論文本進行人工標註。

標註作業主要聚焦於以下四項元素：

1. 面向詞 (Aspect Term)
作為評價對象之名詞或名詞片語，如：「房間」、「溫泉」、「早餐」等。
2. 評價詞 (Opinion)
表達情感傾向之詞語，包括形容詞、副詞、形容動詞等，如：「美味」、「乾淨」、「遺憾」等。
3. 面向分類 (Aspect Category)
依照屬性將面向詞劃分為七類（如：HOTEL、ROOMS、FACILITIES 等）。
4. 評價詞分類 (Opinion Category)
依屬性將評價詞劃分為八類（如：GENERAL、PRICE、COMFORT 等）。

4.2 評價詞 (Opinion) 標註規則之制定

為提升標註一致性，本研究制定以下統一規範：

1. 連體形保持原樣
例：「楽しくて」保留為「楽しくて」，維持語境自然結構。
2. 敬語去除，採用基本形紀錄
例：「楽しかったです」→「楽しかった」，以抽象化形式提升客觀性。
3. 副詞修飾一併保留
例：「とても楽しい」→「とても楽しい」，保留強調語氣之分析價值。

藉由此規範，可避免人為表述差異，提升後續自動處理與模型訓練之資料品質。

4.3 抽取結果之範例與觀察

以下列出實際進行標註的評論句，以及從中所抽取出的面向詞 (Aspect) 與評價詞 (Opinion) 的範例：

```

1  {
2
3      "ID": "uXXlQeqCeXkhdAdaKG1LXg",
4      "Sentence": "アメニティが揃っていて良かった。",
5      "Aspect": [
6          "アメニティ"
7      ],
8      "Category": [
9          "Room Amenities#Style & Options"
10     ],
11     "Opinion": [
12         "良かった"
13     ]
14 },
15 {
16     "ID": "3f4ciChAa1m053eXKE0-XQ",
17     "Sentence": "周辺にはコンビニもあり便利でした。",
18     "Aspect": [
19         "コンビニ"
20     ],
21     "Category": [
22         "Location#General"
23     ],
24     "Opinion": [
25         "便利"
26     ]
27 },
28 {
29     "ID": "qYXUlwT6Zyfnz0J5-PdTNA",
30     "Sentence": "大浴場、露天風呂は最高でした！",
31     "Aspect": [
32         "大浴場",
33         "露天風呂"
34     ],
35     "Category": [
36         "Facilities#Quality",
37         "Facilities#Quality"
38     ],
39     "Opinion": [
40         "最高",
41         "最高"
42     ]
43 },
44
45
46 {
47     "ID": "kZpWab_YRRedUESiPaXcBw",
48     "Sentence": "温泉も良く、食事も良く、快適にリラックスして過ごすことができました。",
49     "Aspect": [
50         "温泉",
51         "食事",
52         "NULL",
53         "NULL"
54     ],
55     "Category": [
56         "Facilities#Quality",
57         "Food & Drinks#Quality",
58         "Hotel#Comfort",
59         "Hotel#Comfort"
60     ],
61     "Opinion": [
62         "良く",
63         "良く",
64         "快適",
65         "リラックス"
66     ]
67 },
68
69 },
70

```

從標註結果中，我們觀察到：

- 面向詞多集中於「早餐」、「房間」、「地點」、「服務」等評價重點。
- 單一句子同時包含多個面向詞之情況相當普遍，如：「房間很乾淨，溫泉也很不錯」。
- 評價詞的表現形式多樣，包括形容詞（「方便」、「乾淨」、「滿意」、「狹窄」、「失望」）及副詞結構（「非常」、「有點」、「超級」）。

此類語言特徵可作為後續建立 情感強度 (Valence/Arousal) 分析 的重要基礎資訊。

4.4 標註作業之課題、反思與解決策略

4.4.1 克服多人標註的一致性問題

在期中階段，本研究面臨下列主要挑戰：

- 面向詞・評價詞邊界模糊
- 抽象面向詞之分類標準不明確
- 類別依賴語境而變動
- 難以處理中立、間接或委婉的評價表達

為解決此問題，我們採行以下措施：

1. 持續調整與統一規則

於暑期每週召開會議，釐清以下分類之定義與運用方式：

- Miscellaneous
- Comfort
- Quality
- Style & Options
- Design & Features

以上規則反覆更新，使標註者逐步形成共同理解。

2. 結構化討論與品質管理 (QC)

- 逐批測量一致率並進行原因分析
- 引入客觀指標 (F1 Score、加權平均)
- 對一致率較低之組別進行強制討論與共識形成

最終，標註一致性明顯提升，並收斂至無需討論即可提交之標準。

4.4.2 導入情感價 Valence/Arousal (VA) 以解決多層次評價問題

針對「難以處理中立、間接、諷刺或多層次情感」之困境，本研究導入：

Valence (V：愉悅—不快) Arousal (A：覺醒—鎮靜)

藉此補充情感強度資訊，使原先僅有正負向之 ABSA 架構得以升級為更深層的 量化情感分析。

VA 標註過程也同步修正了多處 Aspect 錯誤，確保資料品質。

4.4.3 技術挑戰與流程改善

1. 資料處理流程自動化

高田開發：

- JSON → Excel/CSV 轉換工具
- JSON → JSONL 結構轉換工具

使大量資料的處理流程更順暢。

2. 標註程式之改善

溝上負責：

- 修正標註程式之空值 (NULL)
- 錯誤規格更新 (如 Quad → Quadruplet)
- Category 名稱統一調整

3. 標註環境之改善

教授提供 VA 專用網站，使大量標註作業得以高效率進行。

4.5 最終成果：高品質日文 ABSA + VA 數據集之完成

經過系統性標註流程與 QA 管控，本研究最終成功完成：

- 大規模日文 ABSA 資料集
- 含 VA (Valence/Arousal) 情感強度之完整標註
- 可支援 Triplet / Quadruplet 結構化任務之資料格式

此成果為後續模型訓練奠定高度實用的基礎。

4.6 最終結果 (Final Results)

4.6.1 評估概述與分數

本系統提交至 Codabench (SemEval-2026 Task 3 Track A)，於開發集 (Dev Set) 中取得以下分數：

子任務	課題名稱	排名指標
Subtask1(DimASR)	情感回歸(VA 預測)	Normalized RMSE

Subtask2(DimASTE)	三元組提取	Continuous F1 (cF1)
Subtask3(DimASQP)	四元組提取	Continuous F1 (cF1)

4.6.2 子任務分析

(1) Subtask 1：情感回歸

模型	使用的模型名稱 (ID)	RMSE VA	PCC V (↑)	PCC A (↑)
日文 (單詞級別)	cl-tohoku/bert-base-japanese-whole-word-masking	1.9908	0.1026	0.0761
多語言	bert-base-multilingual-cased	2.0561	0.0198	0.0850
日文 (字符級別)	cl-tohoku/bert-base-japanese-char-v3	2.0866	0.0610	0.0219

考察 (子任務 1)

- 為什麼「日文 (單詞級別)」(1.9908) 優於「多語言」(2.0561)?
 - 「bert-base-multilingual-cased」學習了 104 種語言，模型的知識相對分散。
 - 相比之下，「cl-tohoku/bert-base-japanese-whole-word-masking」專注於日語，能更深入地理解日語細微的語義和上下文。
 - 結果是，日語情感分數 (VA) 的預測誤差 (RMSE) 降低，準確度得到提升。
- 為什麼「日文 (單詞級別)」(1.9908) 優於「日文 (字符級別)」(2.0866)?
 - 這歸因於「任務的特性」。子任務 1 要求模型專注於預測情感。
 - 單詞級別模型**：Mecab 事先將文章分割成「單詞」（例如：「ホテル」/「飯店」），模型可以直接識別「飯店」這個概念，並將資源集中於情感預測。
 - 字符級別模型**：必須先從「ホ」、「テ」、「ル」等字符序列中重建「飯店」這個概念，

然後才能進行情感預測，這是一個困難的兩階段處理。

(2) Subtask 2：三元組抽取

模型	使用的模型名稱 (ID)	cF1 (↑)	cPrecision (↑)	cRecall (↑)
Qwen (8B)	unsloth/Qwen3-8B-unsloth-bnb-4bit	0.4359	0.4273	0.4449
Qwen (4B)	unsloth/Qwen3-4B-Instruct-2507-bnb-4bit	0.4027	0.3679	0.4447
Gemma (9B)	unsloth/gemma-2-9b-it-bnb-4bit	0.1078	0.0613	0.4485

考察（子任務 2）

- **規模擴展成功**：通過將模型從默認的 Qwen-4B 擴展到 Qwen-8B，cF1 分數從 0.4027 提升至 0.4359。
- **Gemma 失敗（重要）**：Gemma-9B 未能學習任務格式。儘管其找回能力（cRecall 0.4485）較高，但 FP（錯誤預測數）暴增至 2492 筆，導致 cPrecision（精確率）降至 0.0613（即 94% 的預測是錯誤的），使得 cF1 分數遭受毀滅性打擊。

1.

(3)Subtask 3：四元組抽取

模型	使用的模型名稱 (ID)	cF1 (↑)	cPrecision (↑)	cRecall (↑)
Qwen (4B)	unsloth/Qwen3-4B-Instruct-2507-bnb-4bit	0.3342	0.3360	0.3323
Qwen (8B)	unsloth/Qwen3-8B-unsloth-bnb-	0.2942	0.3155	0.2756

	4bit			
Gemma (9B)	unsloth/gemma-2-9b-it-bnb-4bit	0.1144	0.0978	0.1378

考察（子任務 3）

- **規模擴展失敗（重要）**：結果與子任務 2 截然相反。將模型從 **Qwen-4B** (cF1: 0.3342) 擴展到 **Qwen-8B** (cF1: 0.2942) 後，分數惡化。
- **失敗原因分析**：**Qwen-8B** 相較於 **Qwen-4B**，cRecall 從 0.3323 大幅下降至 0.2756，FN（漏檢數）從 235 增加至 258。這表明在提取 4 個元素這個更複雜的任務中，模型未能有效學習，或者遺漏的案例有所增加。
- **Gemma 失敗（再次）**：**Gemma-9B** 與子任務 2 類似，FP（錯誤預測數）高達 460 筆，不適用於該任務。

4.7 小結：系統優勢與限制

- 子任務 1 中，模型 **cl-tohoku/bert-base-japanese-whole-word-masking** 在總體誤差（RMSE）方面表現最佳。
- 子任務 2 中，模型 **unsloth/Qwen3-8B-unsloth-bnb-4bit** 表現最佳。
- 子任務 3 中，模型 **unsloth/Qwen3-4B-Instruct-2507-bnb-4bit** 表現最佳。
- 情感回歸（Subtask 1）缺乏趨勢相關性，需改善模型結構
- LLM 能處理日文三元組結構（Subtask 2）
- 四元組（Subtask 3）仍受 Category 影響而困難

Chapter 5. 結論

5.1 目標達成與系統驗證

本研究以日文飯店評論為對象，成功建立 ABSA (Aspect-Based Sentiment Analysis) 之基礎資料集，涵蓋 Aspect 與 Opinion，並進一步導入 Valence (V) 與 Arousal (A) 的連續值情感標註，形成完整之 DimABSA 基礎資料集。

透過跨學期協作，本研究完成：

- 高品質 ABSA+VA 標註資料
 - 混合式系統 (PLM 回歸模型 + LLM 指令微調)
 - 三大子任務 (DimASR、DimASTE、DimASQP) 之完整解決方案
- 並於 Codabench 平台取得定量驗證分數。

5.2 核心成果與實證

本研究的主要成果包括：

1. 構建日本語飯店評論之完整 ABSA+VA 數據集
2. 成功完成三項任務之模型設計與提交
3. 於國際評估平台取得穩定之 cF1 與 RMSE 分數
4. 將 ABSA 從定性提升至定量分析的重大進步

本系統不僅能回答：

- 「針對什麼進行評價？」
- 「評價是正面還是負面？」

更能量化回答：

- 「評價強度有多高？」

這為旅宿產業之評論分析提供更強的決策基礎。

5.3 主要貢獻

1. 完整且標準化之 ABSA+VA 標註流程
2. 高品質日文情感分析資料集
3. 結構化情感抽取 (Triplet/Quadruplet) 之技術驗證
4. 具實用價值之旅宿評論分析模型

5.4 未來展望

為提升整體性能，未來將專注於：

- 提升 Category 辨識能力，以改善 Subtask 3 之 cF1
- 重構情感回歸模型，提高 V/A 與真實趨勢的相關性

- 改良指令提示 (Prompting) 以增強 LLM 的抽取穩定性
- 擴增資料規模，使模型具更佳泛化能力

Reference

1. Nakayama, Y., Murakami, K., et al. “A Large-Scale Japanese Dataset for Aspect-Based Sentiment Analysis.” LREC 2022. Pontiki, M., Galanis, D., et al. “SemEval-2015 Task 12: Aspect Based Sentiment Analysis.” SemEval 2015.
2. <https://www.codabench.org/competitions/10918/#/pages-tab>
3. <http://nlp.innobic.yzu.edu.tw/resources/ChineseEmoBank.html>
4. travel.rakuten.co.jp

附錄 A. 專題工作內容

● 溝上大翔

- 負責主導研究與管理各項任務。
- 對樂天飯店評論資料進行網路爬蟲 (Scraping)。
- 執行標註作業。
- 建立資料集。
- 團隊管理與任務分配
- 為制定標註規則，建立測試用資料集。
- 使用 Python 構建對話式標註工具，整備標註作業的操作環境。
- 比較標註結果並分析一致率，進一步整理出需改善之處。
- 制定標註作業的詳細規則。
- 資料品質控管與數據統整。
- 用於報告書的簡報製作。
- 執行模型比較實驗。

● 高田朋希

- 負責整體掌握、任務管理及工作分配。
- 調整經由網路爬蟲取得的資料。
- 建立並調整資料集。
- 執行標註作業。
- 撰寫報告書。
- 制定標註作業的詳細規則

● 甲斐陸都

- 率先進行標註作業，並撰寫已完成標註的程式碼。
- 提出並確定用於資料的標籤 (Label)，確立目前的標註方針。
- 執行實際的標註作業。
- 撰寫報告書。
- 用於報告書的簡報製作。

● 佐佐木華

- 調整經由網路爬蟲取得的資料。

- 執行標註作業。
- 制定標註作業的詳細規則。
- 撰寫報告書。
- 用於報告書的海報製作。
- 用於報告書的簡報製作。
- 制定標註作業的詳細規則。

● 小笠原伶晏

- 用於報告書的網頁製作
- 調整經由網路爬蟲取得的資料。
- 執行標註作業。
- 撰寫報告書。

附錄 B. 專題心得與建議

● 溝上大翔

在這次的畢業專題中，我擔任團隊協調者與專題推動者，從研究啟動、資料集建構到模型實驗，深度參與了所有階段。我對自然語言處理領域一直抱有濃厚興趣，而禹良治教授從研究方法到相關論文，都給予我細心而精準的指導，在此由衷表示感謝。

在團隊合作過程中，難免遇到意料之外的困難或意見分歧。透過這個專題，我培養了與成員協作、靈活應對並解決問題的能力。

這次研究讓我在技術和團隊協作能力上都獲得了顯著成長。這不僅是大學四年學習的總結，更為我未來的研究道路奠定了堅實基礎。我將以此為基礎，在研究所階段繼續投入評論相關的自然語言處理研究。

● 高田朋希

這次的畢業研究，是一項既充滿樂趣，同時也極具挑戰性的研究。

確定本研究的專題主題時，我們團隊成員首先探討了「我們最感興趣的領域」，最終，基於成員的興趣，我們定下了「透過語言分析來建構 AI 輔助用的數據集，並驗證其有效性」這一目標。

隨著研究的深入，我們體認到主題的奧妙之處，同時也因為是多人的團隊，我們面臨了極其困難的挑戰，例如整體任務的管理、以及發生延遲時如何擬定挽回計劃等。尤其在考量所有成員的個別狀況的同時，要維持研究進度非常困難，因此，我們頻繁地向指導教授請教，不斷地摸索解決方案。

然而，多虧了大家的堅持不懈與努力，雖然研究成果的評估上仍有許多改善空間，但我們最終還是取得了巨大的成果——實際完成了數據集的建立，並成功進行了 AI 模型的實作與運行。

透過這次的經驗，我的專業知識獲得了增長，這是理所當然的；除此之外，我更培養了在多數人團隊中的協調性，以及當問題發生時，能邏輯性地找出解決方案的實踐能力。我強烈地感覺到，這次畢業研究所獲得的寶貴經驗與成長，必將在未來的活動中發揮作用。

● 甲斐陸都

這次的研究過程，是一段既艱辛又充滿成就感的旅程。尤其是在標註作業中，最讓我感到挑戰的，是「一致率始終難以提升」這件事。明明大家都依照規則進行標註，卻因為日文表達的模糊性、語感上的差異、長句或委婉語的判斷不同，常常導致兩個人標注的結果差異很大。

於是我們不斷回到原點、檢討標準——哪些詞應該排除、哪些形式應該統一，甚

至為了小小的語尾辭修正規則。每次以為「應該搞定了」，卻又在檢查中發現新的不一致，只能再重來。有時候是一整份檔案重新命名、格式調整，有時候是一百筆資料從頭標一次。深感這份作業真的不是只靠耐心就能解決，而是需要每一位成員共同磨合、持續改善。

但是正因為有這樣不斷重做、反覆修正的過程，我們團隊在七個人一起工作、互相討論、互相支持的過程中逐漸形成默契。大家一起分享心得、解決困難，努力讓這個專案不僅完成，更可以說是「共同堆疊出來的成果」。

回顧整個過程，再怎麼辛苦，都換來一份無可取代的成就感。從一開始什麼都沒有，到現在能確實呈現出一份針對日文情感分析的高品質資料集，我真心覺得這是一段非常值得珍惜的經驗。

● 佐佐木華

這次的畢業研究，從期中發表到最終發表，推動多項作業的過程，說實話，真的很辛苦。然而，多虧了指導教授們和團隊成員的協助，以及為了解決作業困難所做的各種努力，我們團隊才能全員一起，順利完成這項研究。

在本研究中，能夠以「在台灣的大學就讀的日本學生」這個獨特的視角，去發現並著手解決「日本數據集不足」這個挑戰，我感到非常驕傲。在製作最終報告書、海報、簡報（投影片）和網頁等資料時，我們四年級的全體成員也都是團結一心，共同努力到最後才得以完成。

這次的畢業研究不僅是專業知識的學習，在「如何作為團隊的一員參與其中，並進行順暢的溝通」這種更具實踐性的層面上，我也認為自己獲益良多，成長許多。

● 小笠原伶晏

本研究歷時兩個學期，由 5 名四年級與 2 名三年級學生組成的 7 人團隊共同推動。我本人是在暑假開始前的 6 月左右加入團隊，因此在初期階段，因為年級差異與經驗落差，面臨了「知識量不同」、「作業節奏不一致」、「對標註規則的理解不一致」等協作層面的挑戰。特別是在 ABSA 的人工標註過程中，Aspect（面向詞）、Opinion（評價詞）、Category（分類）的判斷標準與邊界，由於每位成員的解讀不同，若缺乏共通認知便容易造成結果的嚴重分歧。因此，在教授的指導下，我們多次展開討論，並於暑假期間每週固定召開會議，以統一標準並建立一致的作業體系。最終，全體成員在保有自身作業風格的同時，也能彼此調整步伐，順利協力完成大規模的標註任務，這讓我深刻體會到大型專案中「協作」與「溝通」的關鍵性。

在技術面方面，透過這次專題，我深切感受到大量語料的人工標註是一項耗時且極度需要專注與細緻判斷的工作。尤其是日文評論常具有主語省略、委婉語氣、多

層次情感、強烈的語境依賴等特性，使得情感判斷比英文更加困難。例如，「もう少し広いとよかった（如果能再大一點就好了）」表面上語氣溫和、甚至接近中立，但實際上隱含負向情緒；又如「美味しくいただきました（非常美味地享用了）」這種帶有敬語成分的表述是否應視為 Opinion，也需制定一致的標準。面對這些語言特性，我們透過反覆的標註比較與錯誤分析，逐步掌握日文評論中情感的微妙波動，並重新認識到「情緒並非正負二分，而是具有連續強度」這一重要概念。這樣的理解在後續進行 VA (Valence/Arousal) 標註時發揮了極大作用，使我更能精準地判斷語氣與情感強度。

展望未來，若要進行類似的大型標註專案，應在初期階段投入更多時間於「規則制定」與「標註者訓練」，以便更早形成共同理解。此外，若能在部分情境下導入半自動或輔助式標註工具，將有助於減輕人工作業負擔，同時維持資料品質。我期望本次專題累積的經驗能成為後續研究與實務工作的參考，並促進未來日本語 ABSA 以及情感分析領域的發展。

透過長時間參與本專題的標註作業，我深刻體會到自然語言處理 (NLP) 領域中資料品質的重要性。然而，同時我也強烈感受到自己希望能有更多機會直接參與資料整備之外的技術性工作，例如系統開發、模型建構與調整等。特別是模型設計策略、特徵處理、提示詞 (prompt) 最佳化、錯誤分析等實作層面的工夫，往往直接影響模型性能，這些都是我未來希望能更深入參與的領域。接下來，我期望能挑戰更多技術範疇，例如分析流程 (pipeline) 構築、模型訓練與推論、提示詞設計、結果可視化等，並以同時理解「資料」與「系統」兩個面向的開發者為目標持續成長。