Academic Paper Writing in Artificial Intelligence

CGU-ILALab at FoodBench-QA 2026: Comparing Traditional and LLM-based Approaches for Recipe Nutrient Estimation

Wei-Chun Chen[a], Yu-Xuan Chen[b], I-Fang Chung[a], Ying-Jia Lin[c,d]
[a]Institute of Biomedical Informatics, National Yang Ming Chiao Tung University, Taipei, Taiwan
[b]Institute of Health Data Science, Chang Gung University, Taoyuan, Taiwan
[c]Department of Artificial Intelligence, Chang Gung University, Taoyuan, Taiwan
[d]Artificial Intelligence Research Center, Chang Gung University, Taoyuan, Taiwan
{andybox11111.md14, ifchung}@nycu.edu.tw, {m1462004, yjlin}@cgu.edu.tw

Abstract

Accurate nutrient estimation from unstructured recipe text is an important yet challenging problem in dietary monitoring, due to ambiguous ingredient terminology and highly variable quantity expressions. We systematically evaluate models spanning a wide range of representational capacity, from lexical matching methods (TF-IDF with Ridge Regression), to deep semantic encoders (DeBERTa-v3), to generative reasoning with large language models (LLMs). Under the strict tolerance criteria defined by EU Regulation 1169/2011, our empirical results reveal a clear trade-off between predictive accuracy and computational efficiency. The TF-IDF baseline achieves moderate nutrient estimation performance with near-instantaneous inference, whereas the DeBERTa-v3 encoder performs poorly under task-specific data scarcity. In contrast, few-shot LLM inference (e.g., Gemini 2.5 Flash) and a hybrid LLM refinement pipeline (TF-IDF combined with Gemini 2.5 Flash) deliver the highest validation accuracy across all nutrient categories. These improvements likely arise from the ability of LLMs to leverage pre-trained world knowledge to resolve ambiguous terminology and normalize non-standard units, which remain difficult for purely lexical approaches. However, these gains come at the cost of substantially higher inference latency, highlighting a practical deployment trade-off between real-time efficiency and nutritional precision in dietary monitoring systems.

Keywords: Recipe Nutrient Estimation, Large Language Models

1. Introduction

Accurate nutrient estimation from recipe text is a practically relevant yet technically challenging task in the food and nutrition domain. While large language models (LLMs) have demonstrated remarkable capabilities in general-purpose reasoning and knowledge retrieval, it remains unclear how well they handle the structured, quantitative reasoning required for food applications. Recipe texts present unique difficulties: ambiguous ingredient terminology, non-standard measurement units, and the need to map diverse culinary expressions to precise nutritional values. These challenges make recipe nutrient estimation a compelling testbed for evaluating whether different modeling paradigms — from traditional lexical methods to LLM-based inference — can produce reliable nutritional predictions.

The FoodBench-QA 2026 shared task[1] asks systems to answer food and nutrition questions based on unstructured recipe text. The complexity of the food domain stems from its inherent heterogeneity. Unlike standardized datasets in other fields, culinary data is characterized by ambiguous ingredient terminology, non-standard and highly variable measurement units (e.g., "a pinch," "a medium bunch"), and the challenge of mapping these varied expressions to structured food composition databases. A model must not only perform accurate entity disambiguation to distinguish between semantically similar but nutritionally distinct ingredients (e.g., "coconut milk" vs. "coconut water") but also execute complex reasoning to normalize quantities and aggregate nutritional values from disparate sources.

Case 1 Prompt: Assess the nutritional profile of 180 ml of pineapple juice, canned or bottled, unsweetened, without added ascorbic acid.
Case 1 Answer: The nutrient values demonstrated here are: energy - 99.79, fat - 0.23, protein - 0.68, salt - 3.78, saturates - 0.02, sugars - 18.86.

Case 2 Prompt: Check the nutritional values per 100 g in a recipe that comprises these ingredients: 2 teaspoons corn, sweet, white, raw, 2 teaspoons corn, sweet, white, raw, 2 teaspoons peppers, sweet, green, raw, 1 teaspoon spices, coriander seed, 1/2 cup butter, without salt.
Case 2 Answer: Nutrient details in 100 g: energy - 713.49, fat - 80.53, protein - 0.86, salt - 0.04, saturates - 50.10, sugars - 0.09.

Table 1: Examples in the FoodBench-QA 2026 shared task. Prompt: nutritional queries; Answer: response that should be provided by a system.

In response to these challenges, the FoodBench-

[1] https://www.codabench.org/competitions/12112/

Proceedings of the Third Workshop on Patient-Oriented Language Processing (CL4Health), pages 362–368
12 May 2026. ©ELRA Language Resources Association (ELRA), 2026

Basic Skills in Academic Writing / Course Introduction

Instructor: Ying-Jia Lin

September 10th, 2026

About Me

CGU
Bachelor in Biomedical Sciences, 2017

NYMU (NYCU)
Master in Biomedical Informatics, 2019

NCKU
Ph.D. in Computer Science and Information Engineering, 2024

Main research direction: Natural Language Processing (NLP)

About Me

2016-FALL	LS1101	21633	Department of Biomedical Sciences	3	Scientific Writing	劉世東	2

CGU
Bachelor in Biomedical Sciences, 2017

NYMU (NYCU)
Master in Biomedical Informatics, 2019

NCKU
Ph.D. in Computer Science and Information Engineering, 2024

Main research direction: Natural Language Processing (NLP)

About Me

2016-FALL	LS1101	21633	Department of Biomedical Sciences	3	Scientific Writing	劉世東	2

CGU
Bachelor in Biomedical Sciences, 2017

NYMU (NYCU)
Master in Biomedical Informatics, 2019

NCKU
Ph.D. in Computer Science and Information Engineering, 2024

Main research direction: Natural Language Processing (NLP)

Published papers: Journals: 5; Conferences: 18

Before Enrolling in This Course

- Please get your advisor's approval (upload to E-Learning by Sep. 17th)

Consent Form for the Use of Research Data↵

Academic Paper Writing in Artificial Intelligence↵

Instructor: Ying-Jia Lin, Assistant Professor (CGUAI)↵

Semester: First Semester, Academic Year 115 (2026 Fall)↵

---↵

This course trains graduate students to write academic papers in artificial intelligence. Each week, students write a specific section of their own paper during class. They will therefore use their ongoing research in your laboratory — dataset descriptions, methods, experimental results, and figure drafts — as the material they write about. We would be grateful for your consent.↵

Why do we still need this course?

Why do we still need this course?

1. AI generates mysterious things.

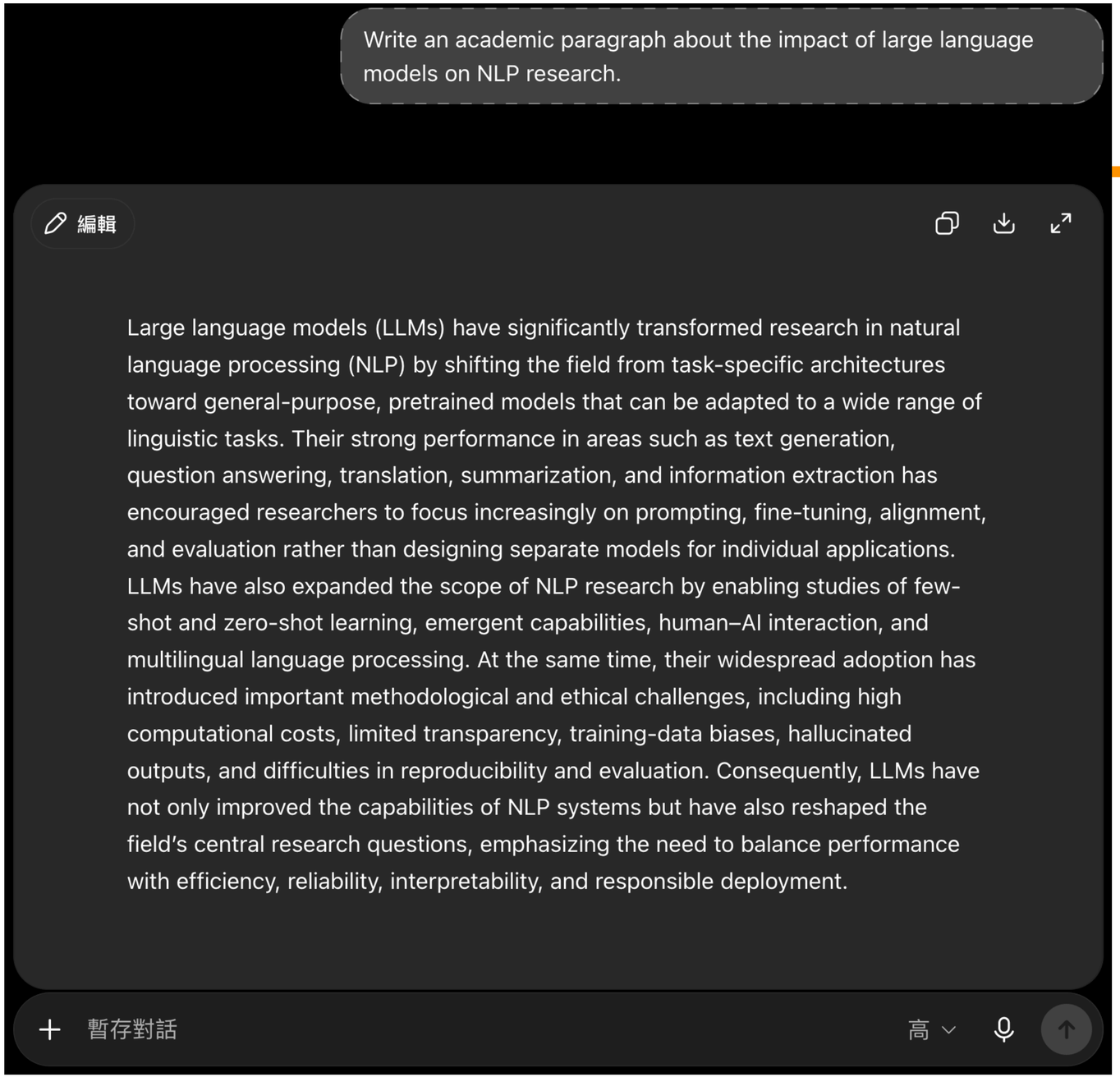
Write an academic paragraph about the impact of large language models on NLP research.

編輯

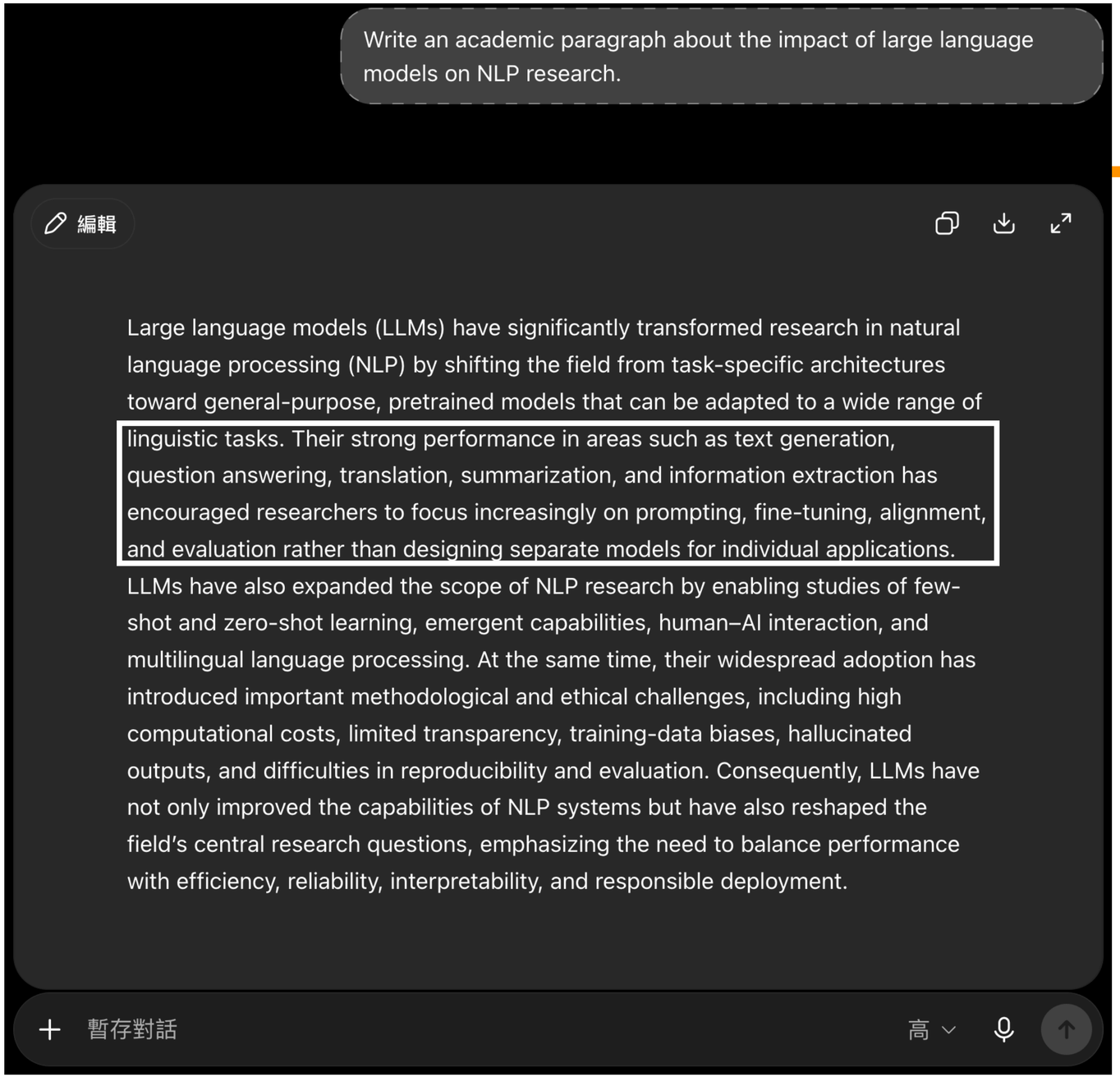
Large language models (LLMs) have significantly transformed research in natural language processing (NLP) by shifting the field from task-specific architectures toward general-purpose, pretrained models that can be adapted to a wide range of linguistic tasks. Their strong performance in areas such as text generation, question answering, translation, summarization, and information extraction has encouraged researchers to focus increasingly on prompting, fine-tuning, alignment, and evaluation rather than designing separate models for individual applications. LLMs have also expanded the scope of NLP research by enabling studies of few-shot and zero-shot learning, emergent capabilities, human–AI interaction, and multilingual language processing. At the same time, their widespread adoption has introduced important methodological and ethical challenges, including high computational costs, limited transparency, training-data biases, hallucinated outputs, and difficulties in reproducibility and evaluation. Consequently, LLMs have not only improved the capabilities of NLP systems but have also reshaped the field's central research questions, emphasizing the need to balance performance with efficiency, reliability, interpretability, and responsible deployment.

暫存對話

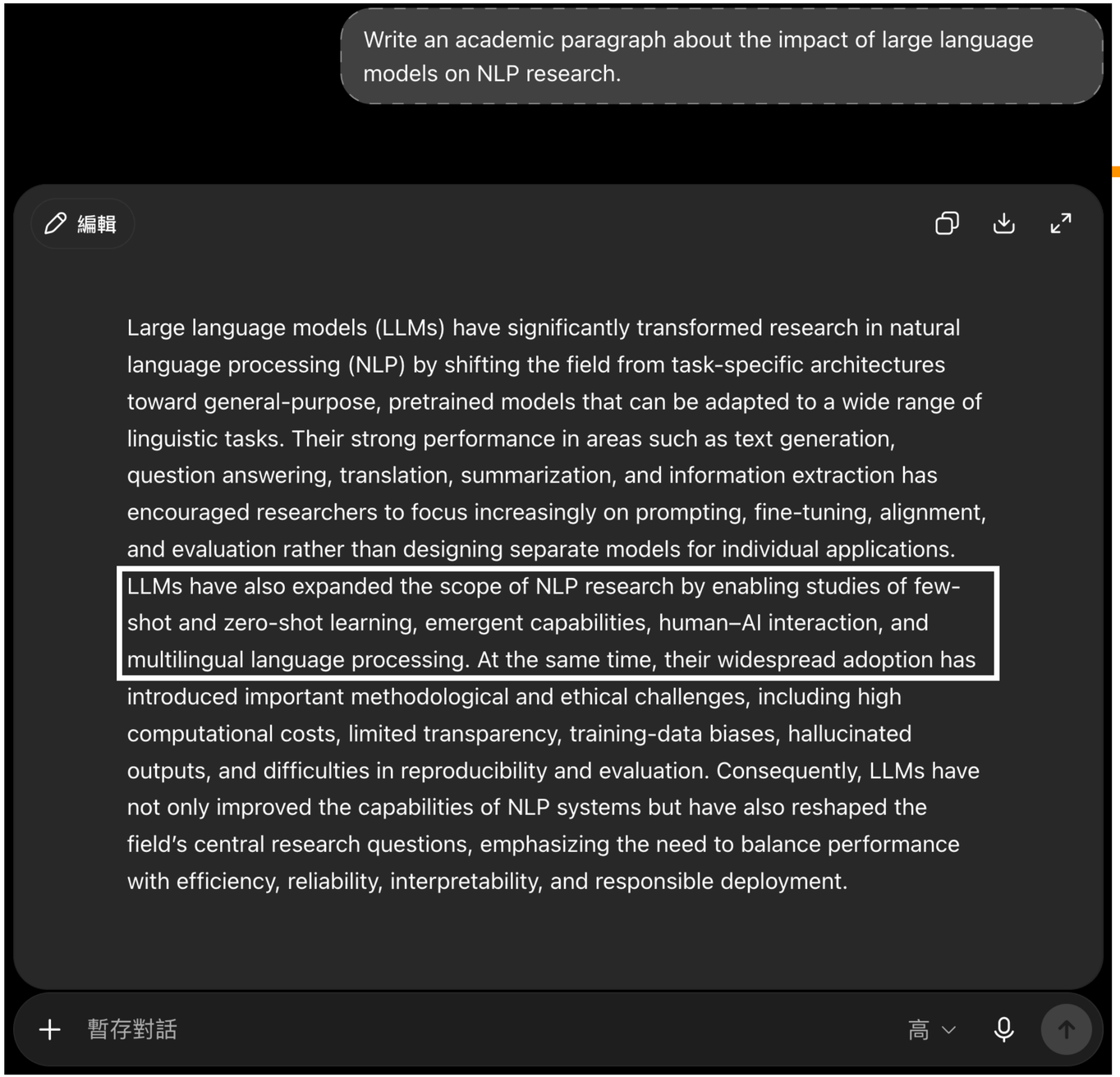
高



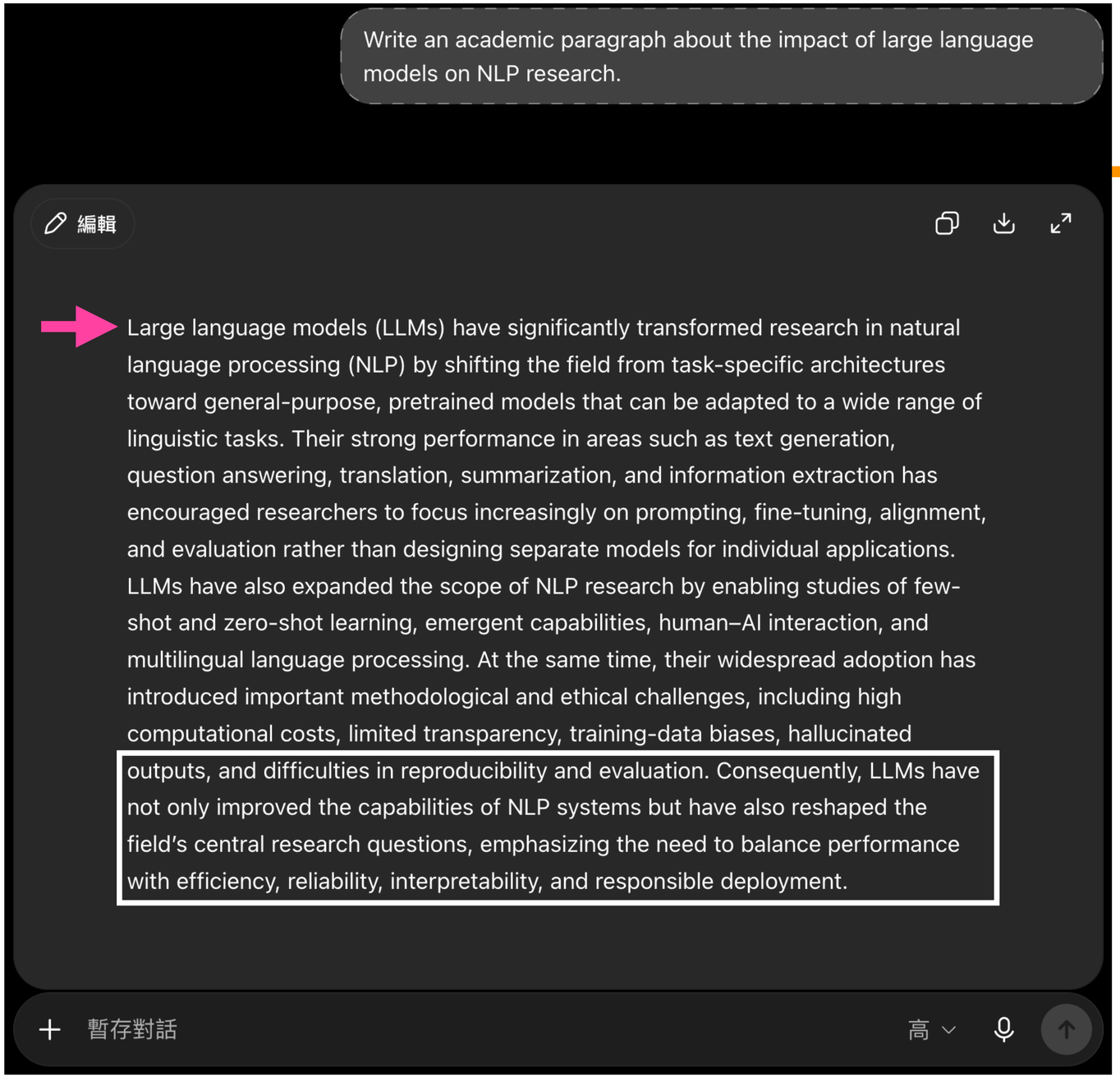
- At least three problems:



- At least three problems:
  1. No evidence (references?).



- At least three problems:
  1. No evidence (references?).
  2. No details. Just more things that LLMs can do.



- At least three problems:
  1. No evidence (references?).
  2. No details. Just more things that LLMs can do.
  3. Not supporting the claim. Universally fit.

Why do we still need this course?

1. AI generates mysterious things.

Why do we still need this course?

1. AI generates mysterious things.
2. You don’t know how to write. Just let AI do.

Write paper, or click accepts from Claude?

```
● Update(NEXT_STEPS.md)
  └ Added 4 lines, removed 3 lines
      116     (abstract, keywords, Discussion skeleton) replaced with "object detection"; the best
      117     model is now RetinaNet (95.6% test mAP) and YOLOv3 is out of the paper.
      118
      119 -5. **Cut to 8 display items.** Currently 7 tables + 2 figures = 9. Send the IoU sweep,
      120 -   the semi-supervised pseudo-label result, the model-agreement matrix and the FFT
      121 -   supplement to Supplementary.
      119 +5. **[DONE 2026-08-17] Cut to 8 display items.** Achieved by deleting the ensemble
      120 +   table: now 6 tables + 2 figures = 8, exactly at the MP limit. Note this leaves no
      121 +   headroom — any new display item (e.g., supplementary material promoted back, or a
      122 +   YOLOv11 table) requires demoting something else.
      123
      124  6. **[RESOLVED 2026-08-17] RT-DETR's fate.** Kept as the third candidate detector in
      125     Methods (with rationale), reported in Tables 2–3, excluded from the operating-point

● 改完了,編譯乾淨(26 頁、無 undefined reference),未 commit。彙總:

  已執行

[0] 0:[tmux]*                                                  "✳ 檢查專案分支狀態" 19:26 17-Aug-26
```

Central Guideline of Academic Writing

- Good academic writing = Visible clear **thinking**
- A good paper does three things:
    1. Makes a clear claim
    2. Provides convincing evidence
    3. Guides the reader through the reasoning

What is a claim?

claim

KK[/klem/]

v. 聲稱；主張；要求（權利、財產等）

n. 聲明；主張；要求（權利、財產等）

What is a claim?

claim

KK[/klem/]

v. 聲稱；主張；要求（權利、財產等）

n. 聲明；主張；要求（權利、財產等）

- A claim is a sentence that someone could disagree with.

What is a claim?

claim

KK[/klem/]

v. 聲稱；主張；要求（權利、財產等）

n. 聲明；主張；要求（權利、財產等）

- A claim is a sentence that someone could disagree with.
- ❌ Not a claim:

What is a claim?

claim

KK[/klem/]

v. 聲稱；主張；要求（權利、財產等）

n. 聲明；主張；要求（權利、財產等）

- A claim is a sentence that someone could disagree with.
- ❌ Not a claim:
  - We use GPT-5 to perform medical report generation.

What is a claim?

claim

KK[/klem/]

v. 聲稱；主張；要求（權利、財產等）

n. 聲明；主張；要求（權利、財產等）

- A claim is a sentence that someone could disagree with.
- ❌ Not a claim:
  - We use GPT-5 to perform medical report generation.
    - Nobody can disagree. This is your method.

What is a claim?

claim

KK[/klem/]

v. 聲稱；主張；要求（權利、財產等）

n. 聲明；主張；要求（權利、財產等）

- A claim is a sentence that someone could disagree with.
- ❌ Not a claim:
  - We use GPT-5 to perform medical report generation.
    - Nobody can disagree. This is your method.
- ✅ A claim:

What is a claim?

claim

KK[/klem/]

v. 聲稱；主張；要求（權利、財產等）

n. 聲明；主張；要求（權利、財產等）

- A claim is a sentence that someone could disagree with.
- ❌ Not a claim:
  - We use GPT-5 to perform medical report generation.
    - Nobody can disagree. This is your method.
- ✅ A claim:
  - Our method only needs 50 training examples, while others require 50,000.

What is a claim?

claim

KK[/klem/]

v. 聲稱；主張；要求（權利、財產等）

n. 聲明；主張；要求（權利、財產等）

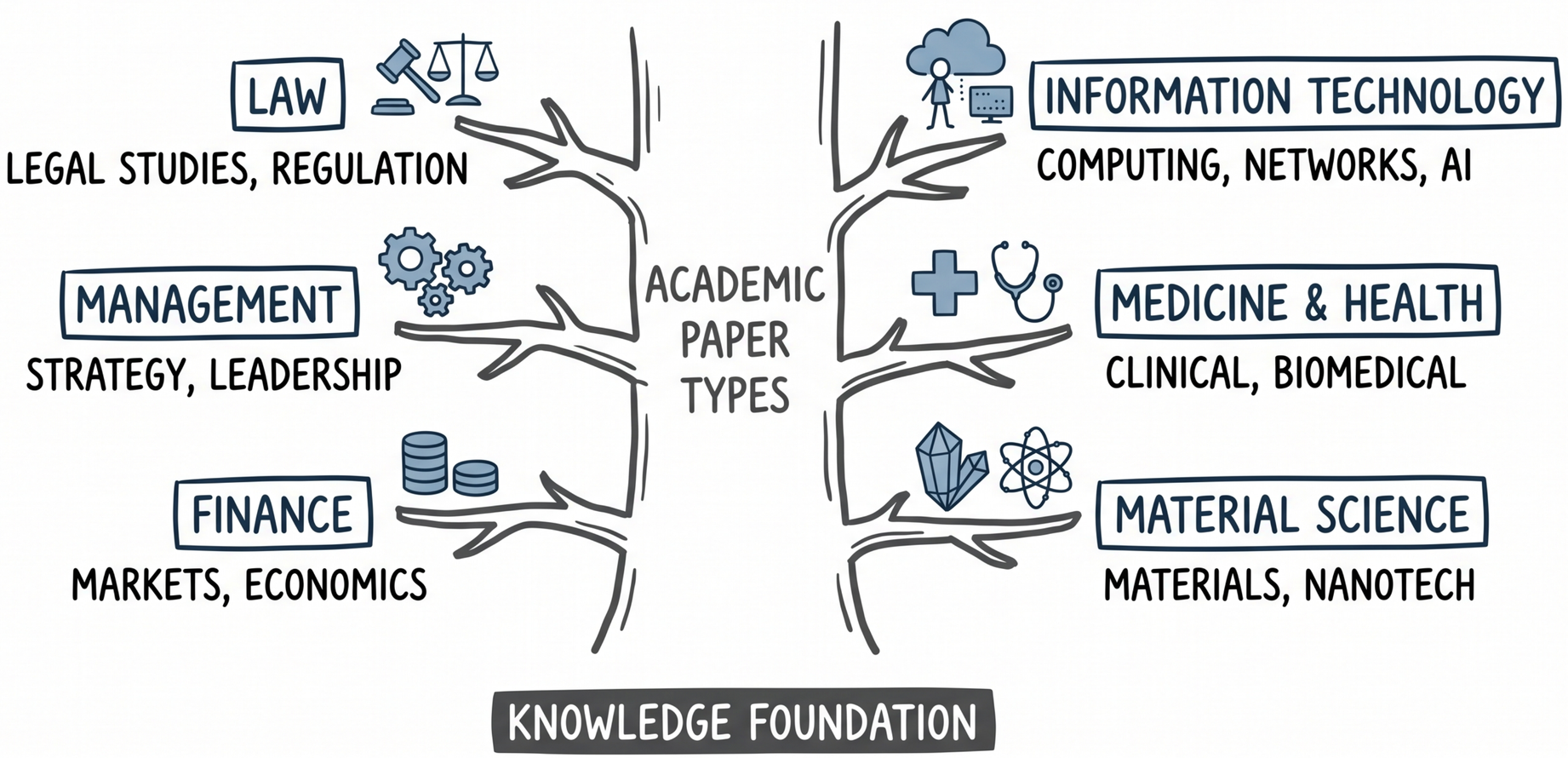
- A claim is a sentence that someone could disagree with.
- ❌ Not a claim:
  - We use GPT-5 to perform medical report generation.
    - Nobody can disagree. This is your method.
- ✅ A claim:
  - Our method only needs 50 training examples, while others require 50,000.
  - Radiology text report alone is enough to label liver tumor findings.

A claim and its evidence

- Experimental results demonstrate that SCURank outperforms other methods.

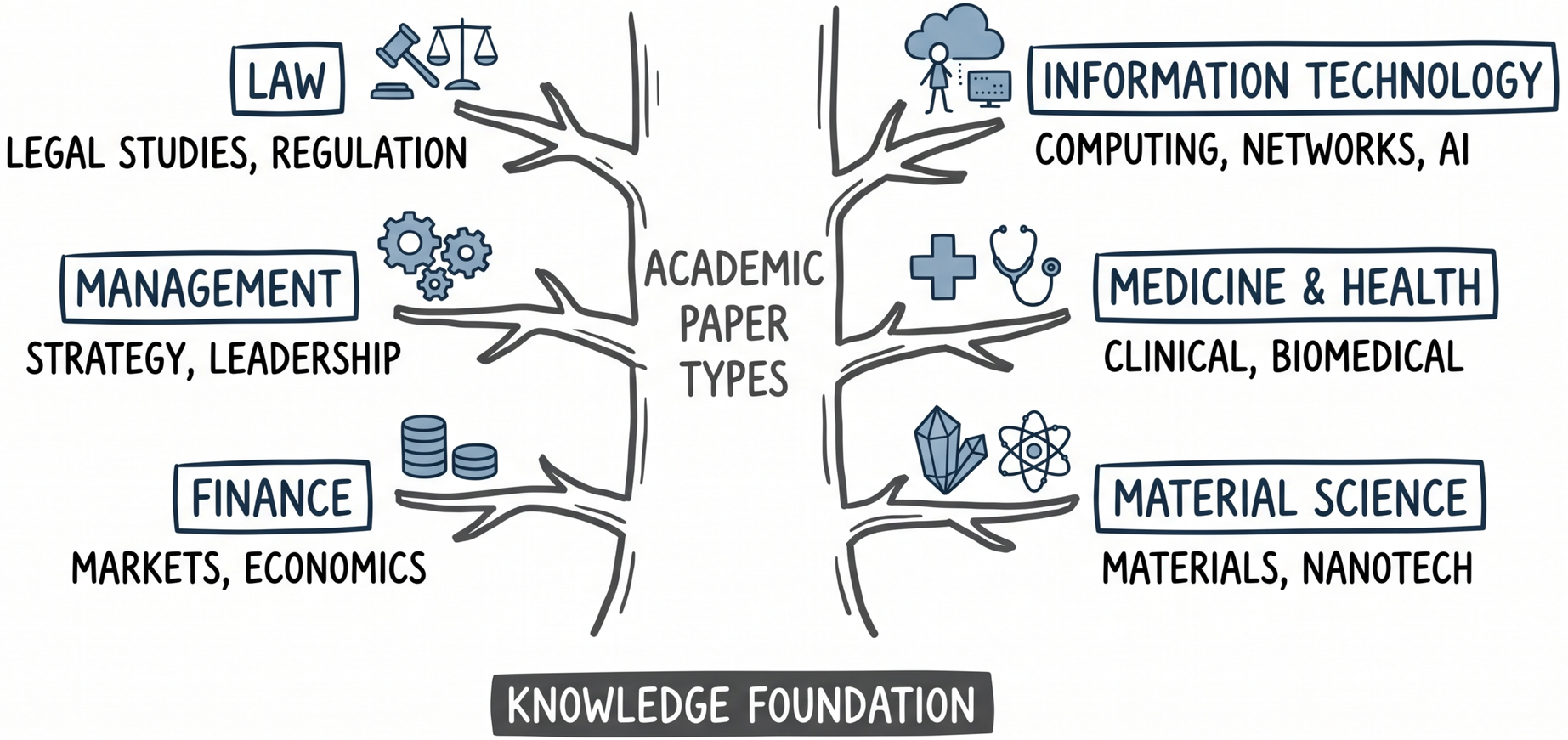
	Dataset: BASE					
UnRank	$42.3_{\pm 0.11}$	$19.4_{\pm 0.09}$	$29.6_{\pm 0.15}$	$49.8_{\pm 0.36}$	$68.6_{\pm 0.08}$	$-2.73_{\pm 0.85}$
ROUGE	$43.3_{\pm 0.56}$	$20.1_{\pm 0.42}$	$30.3_{\pm 0.42}$	$50.9_{\pm 0.37}$	$69.0_{\pm 0.30}$	$-2.62_{\pm 0.02}$
BERTScore	$43.4_{\pm 0.55}$	$20.3_{\pm 0.48}$	$30.1_{\pm 0.76}$	$\underline{51.4}_{\pm 0.30}$	$69.0_{\pm 0.37}$	$-2.59_{\pm 0.02}$
BLANC	$\underline{43.9}_{\pm 0.68}$	$\underline{20.6}_{\pm 0.67}$	$\underline{30.6}_{\pm 0.62}$	$\underline{51.4}_{\pm 0.93}$	$\underline{69.5}_{\pm 0.29}$	$\mathbf{-2.45}_{\pm 0.09}$
GPTRank	$43.0_{\pm 0.35}$	$19.8_{\pm 0.75}$	$29.6_{\pm 0.62}$	$51.2_{\pm 0.81}$	$69.1_{\pm 0.38}$	$\underline{-2.52}_{\pm 0.08}$
SCURank	$\mathbf{44.3}_{\pm 0.87}$	$\mathbf{20.8}_{\pm 0.96}$	$\mathbf{30.9}_{\pm 1.02}$	$\mathbf{51.7}_{\pm 0.67}$	$\mathbf{69.9}_{\pm 0.50}$	$\mathbf{-2.45}_{\pm 0.10}$

Academic Paper Fields



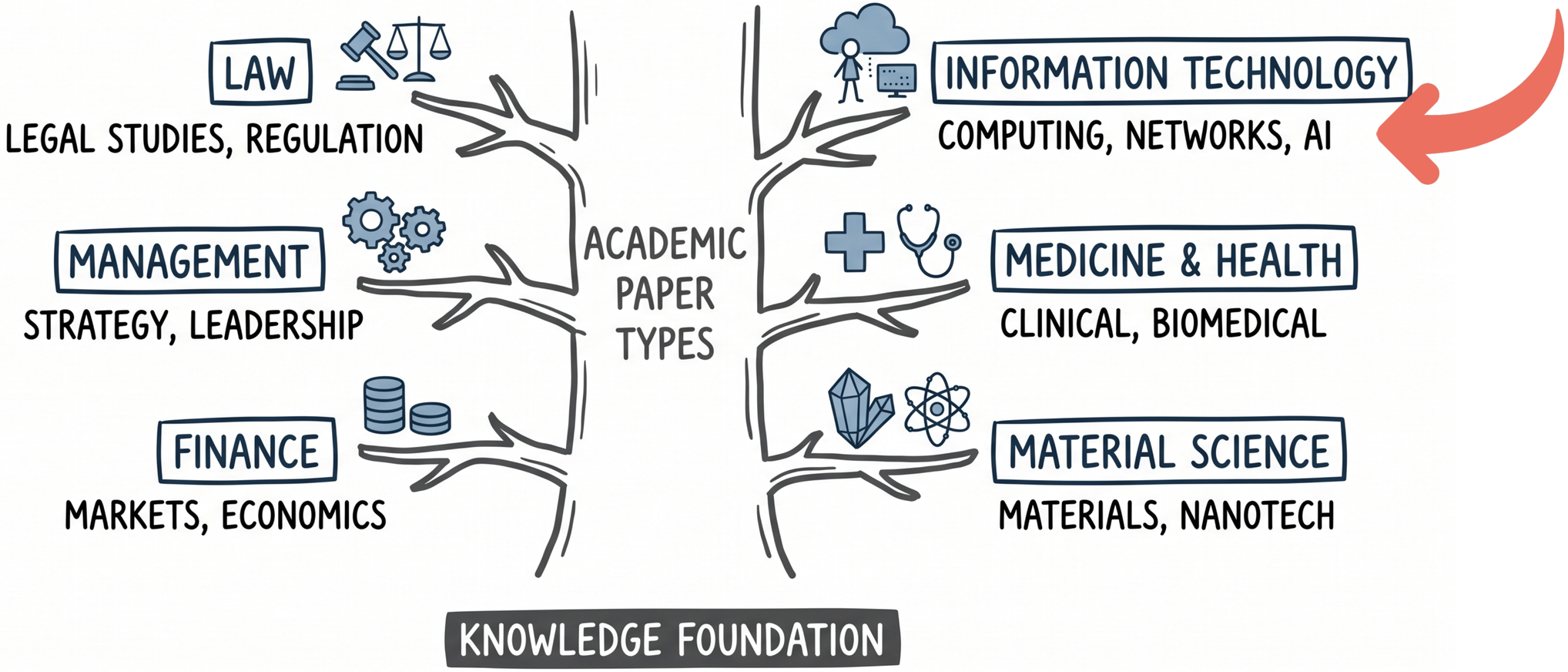
Academic Paper Fields

This course mainly focuses on papers in: (1) AI and (2) AI + Medicine.



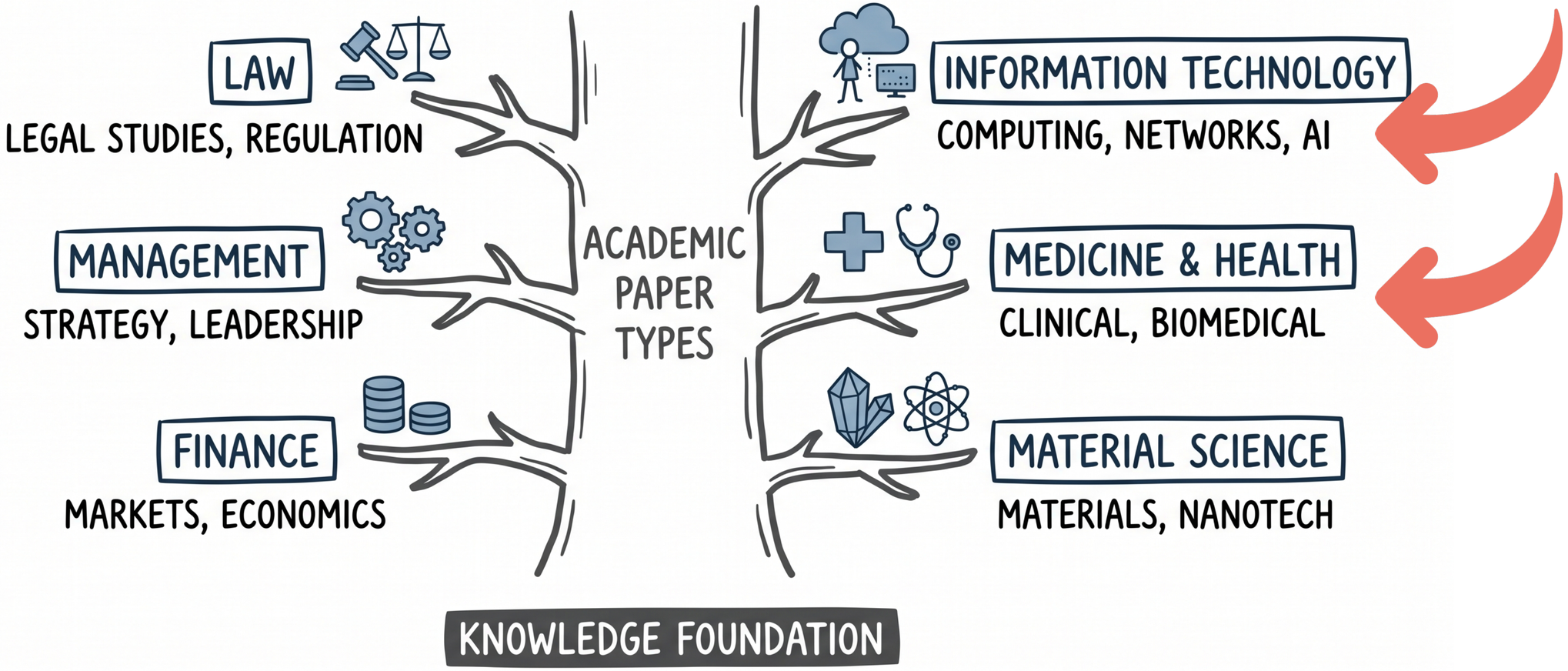
Academic Paper Fields

This course mainly focuses on papers in: (1) AI and (2) AI + Medicine.



Academic Paper Fields

This course mainly focuses on papers in: (1) AI and (2) AI + Medicine.



Journal Papers vs. Conference Papers

- Before you start to write a paper, you need to decide the place to submit.

	AI Conference	**AI / CS Journal**	**Medical Informatics Journal**
Length	shorter (usually up to 9 pages)	longer (10+ pages)	longer (10+ pages)
Review cycle	deadline-driven	at least one rounds for revision	at least one rounds for revision
Related Work	Independent	Independent	Usually combined in Introduction
Discussion	Usually integrated in Results	Usually integrated in Results	Explicit

Do not memorize this table. Open three recent papers from the venue you are targeting and look at their section headings.

Overall Structure

	Main Body (Your Main Paper)
	Abstract
	Introduction
Depends on where you submit	Related Work (Literature Review)
	Methods
	Results
Depends on where you submit	Discussion
	Conclusion

Overall Structure

Main Body (Your Main Paper)
Abstract
Introduction
Related Work (Literature Review)
Methods
Results
Discussion
Conclusion

Depends on where you submit 

Depends on where you submit 

Abstract

- Abstract is usually placed at one of the most identifiable places in a paper.

CGU-ILALab at FoodBench-QA 2026: Comparing Traditional and LLM-based Approaches for Recipe Nutrient Estimation

Wei-Chun Chen[a], Yu-Xuan Chen[b], I-Fang Chung[c], Ying-Jia Lin[c,d]
[a]Institute of Biomedical Informatics, National Yang Ming Chiao Tung University, Taipei, Taiwan
[b]Institute of Health Data Science, Chang Gung University, Taoyuan, Taiwan
[c]Department of Artificial Intelligence, Chang Gung University, Taoyuan, Taiwan
[d]Artificial Intelligence Research Center, Chang Gung University, Taoyuan, Taiwan
{andybox11111.md14, lfchung}@nycu.edu.tw, {m1462004, yjlin}@cgu.edu.tw

Abstract

Accurate nutrient estimation from unstructured recipe text is an important yet challenging problem in dietary monitoring, due to ambiguous ingredient terminology and highly variable quantity expressions. We systematically evaluate models spanning a wide range of representational capacity, from lexical matching methods (TF-IDF with Ridge Regression), to deep semantic encoders (DeBERTa-v3), to generative reasoning with large language models (LLMs). Under the strict tolerance criteria defined by EU Regulation 1169/2011, our empirical results reveal a clear trade-off between predictive accuracy and computational efficiency. The TF-IDF baseline achieves moderate nutrient estimation performance with near-instantaneous inference, whereas the DeBERTa-v3 encoder performs poorly under task-specific data scarcity. In contrast, few-shot LLM inference (e.g., Gemini 2.5 Flash) and a hybrid LLM refinement pipeline (TF-IDF combined with Gemini 2.5 Flash) deliver the highest validation accuracy across all nutrient categories. These improvements likely arise from the ability of LLMs to leverage pre-trained world knowledge to resolve ambiguous terminology and normalize non-standard units, which remain difficult for purely lexical approaches. However, these gains come at the cost of substantially higher inference latency, highlighting a practical deployment trade-off between real-time efficiency and nutritional precision in dietary monitoring systems.

Keywords: Recipe Nutrient Estimation, Large Language Models

1. Introduction

Accurate nutrient estimation from recipe text is a practically relevant yet technically challenging task in the food and nutrition domain. While large language models (LLMs) have demonstrated remarkable capabilities in general-purpose reasoning and knowledge retrieval, it remains unclear how well they handle the structured, quantitative reasoning required for food applications. Recipe texts present unique difficulties: ambiguous ingredient terminology, non-standard measurement units, and the need to map diverse culinary expressions to precise nutritional values. These challenges make recipe nutrient estimation a compelling testbed for evaluating whether different modeling paradigms — from traditional lexical methods to LLM-based inference — can produce reliable nutritional predictions.

The FoodBench-QA 2026 shared task[1] asks systems to answer food and nutrition questions based on unstructured recipe text. The complexity of the food domain stems from its inherent heterogeneity. Unlike standardized datasets in other fields, culinary data is characterized by ambiguous ingredient terminology, non-standard and highly variable measurement units (e.g., "a pinch," "a medium bunch"), and the challenge of mapping these varied expressions to structured food composition databases. A model must not only perform accurate entity disambiguation to distinguish between semantically similar but nutritionally distinct ingredients (e.g., "coconut milk" vs. "coconut water") but also execute complex reasoning to normalize quantities and aggregate nutritional values from disparate sources.

Case 1 Prompt: Assess the nutritional profile of 180 ml of pineapple juice, canned or bottled, unsweetened, without added ascorbic acid.
Case 1 Answer: The nutrient values demonstrated here are: energy - 99.79, fat - 0.23, protein - 0.68, salt - 3.78, saturates - 0.02, sugars - 18.86.

Case 2 Prompt: Check the nutritional values per 100 g in a recipe that comprises these ingredients: 2 teaspoons corn, sweet, white, raw, 2 teaspoons corn, sweet, white, raw, 2 teaspoons peppers, sweet, green, raw, 1 teaspoon spices, coriander seed, 1/2 cup butter, without salt.
Case 2 Answer: Nutrient details in 100 g: energy - 713.49, fat - 80.53, protein - 0.86, salt - 0.04, saturates - 50.10, sugars - 0.09.

Table 1: Examples in the FoodBench-QA 2026 shared task. Prompt: nutritional queries; Answer: response that should be provided by a system.

In response to these challenges, the FoodBench-

[1]https://www.codabench.org/competitions/12112/

Proceedings of the Third Workshop on Patient-Oriented Language Processing (CL4Health), pages 362–368
12 May 2026. ©ELRA Language Resources Association (ELRA), 2026

SCURank: Ranking Multiple Candidate Summaries with Summary Content Units for Enhanced Summarization

Bo-Jyun Wang[1,2], Ying-Jia Lin[2,3], Hung-Yu Kao[4]
[1]Department of Computer Science and Information Engineering, National Cheng Kung University,
[2]Artificial Intelligence Research Center, Chang Gung University,
[3]Department of Artificial Intelligence, Chang Gung University,
[4]Department of Computer Science, National Tsing Hua University
bojyun.wang@cgu.edu.tw, yjlin@cgu.edu.tw, hykao@cs.nthu.edu.tw

Abstract

Small language models (SLMs), such as BART, can achieve summarization performance comparable to large language models (LLMs) via distillation. However, existing LLM-based ranking strategies for summary candidates suffer from instability, while classical metrics (e.g., ROUGE) are insufficient to rank high-quality summaries. To address these issues, we introduce **SCURank**, a framework that enhances summarization by leveraging **Summary Content Units (SCUs)**. Instead of relying on unstable comparisons or surface-level overlap, SCURank evaluates summaries based on the richness and semantic importance of information content. We investigate the effectiveness of SCURank in distilling summaries from multiple diverse LLMs. Experimental results demonstrate that SCURank outperforms traditional metrics and LLM-based ranking methods across evaluation measures and datasets. Furthermore, our findings show that incorporating diverse LLM summaries enhances model abstractiveness and overall distilled model performance, validating the benefits of information-centric ranking in multi-LLM distillation. The code for SCURank is available at https://github.com/IKMLab/SCURank.

1 Introduction

Following the advent of ChatGPT (Ouyang et al., 2022), a paradigm shift has occurred in the domain of Natural Language Processing (NLP), marked by substantial advancements, including the field of summarization. In the wake of this development, GPT-4 (OpenAI, 2024a) and numerous Large Language Models (LLMs) have emerged (Google, 2024; Anthropic, 2024; MistralAI, 2024), demonstrating superior performance. These models are easily accessible via APIs, allowing users to obtain responses with minimal effort. However, while these models are powerful, their resource demands, such as the cost of local deployments, are significant (Hsieh et al., 2023) due to their incredibly large model size. This has led to a growing trend of distilling these models into smaller ones, optimized for specific tasks such as summarization (Hinton et al., 2015; Jiang et al., 2024). The distilled models exhibit great resource efficiency without significant performance degradation, and in some cases, they demonstrate a capability to outperform LLMs in the summarization task (Liu et al., 2024).

Previous work by Liu et al. (2024) leverages BRIO (Liu et al., 2022) as a contrastive learning framework for model distillation. In BRIO, positive and negative samples are constructed by ranking candidate summaries. Consequently, the quality of the ranking function is critical for effective contrastive learning. To address this challenge, they introduced GPTRank (Liu et al., 2024), a novel method that ranks high-quality summaries generated by LLMs, ensuring more reliable supervision in BRIO training. However, there are two significant challenges to this approach. First, studies by Shen et al. (2023); Wang et al. (2024) indicate that LLMs remain unreliable and inconsistent in text comparison and candidate ranking. Second, relying on summaries from a single LLM introduces the risk of model-specific bias (e.g. content selection) and limits the diversity of generation patterns. Thus, we explore the use of multiple LLMs to generate candidate summaries for distillation, and investigate effective ranking methods for these summaries.

To bypass the instability of direct LLM ranking, we propose shifting the evaluation focus back to the core goal of summarization: information retention. To capture the information from summaries, we draw upon the concept of Summary Content Units (SCUs) (Nenkova and Passonneau, 2004). Every SCU presents simple, standalone, and unique information in the summary (Shapira et al., 2019). Generally, annotating SCUs requires manual efforts, thus expensive and not reproducible (Zhang and Bansal, 2021). Nawrath et al. (2024)

Findings of the Association for Computational Linguistics: ACL 2026, pages 38980–38997
July 2-7, 2026 ©2026 Association for Computational Linguistics

International Journal of Medical Informatics 111 (2018) 159–164
Contents lists available at ScienceDirect
International Journal of Medical Informatics
journal homepage: www.elsevier.com/locate/ijmedinf

Predicting post-stroke activities of daily living through a machine learning-based approach on initiating rehabilitation

Wan-Yin Lin[a,1], Chun-Hsien Chen[b,c,1], Yi-Ju Tseng[b,d,1], Yu-Ting Tsai[e], Ching-Yu Chang[e], Hsin-Yao Wang[f,*], Chih-Kuang Chen[e,f,**]
[a] Department of Physical Medicine & Rehabilitation, Chang Gung Memorial Hospital at Linkou, Taoyuan City, Taiwan
[b] Department of Information Management, Chang Gung University, Taoyuan City, Taiwan
[c] Department of Neurology, Chang Gung Memorial Hospital at Taoyuan, Taoyuan City, Taiwan
[d] Department of Laboratory Medicine, Chang Gung Memorial Hospital at Linkou, Taoyuan City, Taiwan
[e] School of Medicine, Chang Gung University, Taoyuan City, Taiwan
[f] Department of Physical Medicine & Rehabilitation, Chang Gung Memorial Hospital at Taoyuan, Taoyuan City, Taiwan

ARTICLE INFO
Keywords:
Stroke
Activities of daily living
Rehabilitation
Machine learning
Computer-assisted decision making

ABSTRACT
Objectives: Prediction of activities of daily living (ADL) is crucial for optimized care of post-stroke patients. However, no suitably-validated and practical models are currently available in clinical practice.
Methods: Participants of a Post-acute Care-Cerebrovascular Diseases (PAC-CVD) program from a reference hospital in Taiwan between 2014 and 2016 were enrolled in this study. Based on 15 rehabilitation assessments, machine learning (ML) methods, namely logistic regression (LR), support vector machine (SVM), and random forest (RF), were used to predict the Barthel index (BI) status at discharge. Furthermore, SVM and linear regression were used to predict the actual BI scores at discharge.
Results: A total of 313 individuals (men: 208, women: 105) were enrolled in the study. All the classification models outperformed single assessments in predicting the BI statuses of the patients at discharge. The performance of the LR and RF algorithms was higher (area under ROC curve (AUC): 0.79) than that of SVM algorithm (AUC: 0.77). In addition, the mean absolute errors of both SVM and linear regression models in predicting the actual BI score at discharge were 9.86 and 9.95, respectively.
Conclusions: The proposed ML-based method provides a promising and practical computer-assisted decision making tool for predicting ADL in clinical practice.

1. Introduction

In 2016, a new onset of stroke events occurred every 40 s in the United States, which caused a heavy disease burden on the health care system [1]. In Taiwan, in 2012, approximately 230 patients were hospitalized per day because of acute stroke [2]. Long-term disabilities in patients after stroke may create an enormous physical, mental, and financial burden for the patients, their families, and society. Early rehabilitation of patients improves recovery and reduces disabilities [3]. The ultimate goal of stroke rehabilitation is to ensure that the patients achieve a high quality of life by training them to perform activities of daily living (ADL) as independently as possible. In the initial stage of rehabilitation, accurate prediction of ADL for the subsequent few months is essential. Accurate prediction of ADL is crucial for optimizing stroke management during the first months following a stroke attack. Moreover, it guides realistic goal-setting, facilitates the creation of an early discharge plan, and provides accurate information to patients and their caregivers [3]. Based on accurate prediction of ADL, adequate socioeconomic support and health care resources can be effectively provided by either the government or the patient's family.

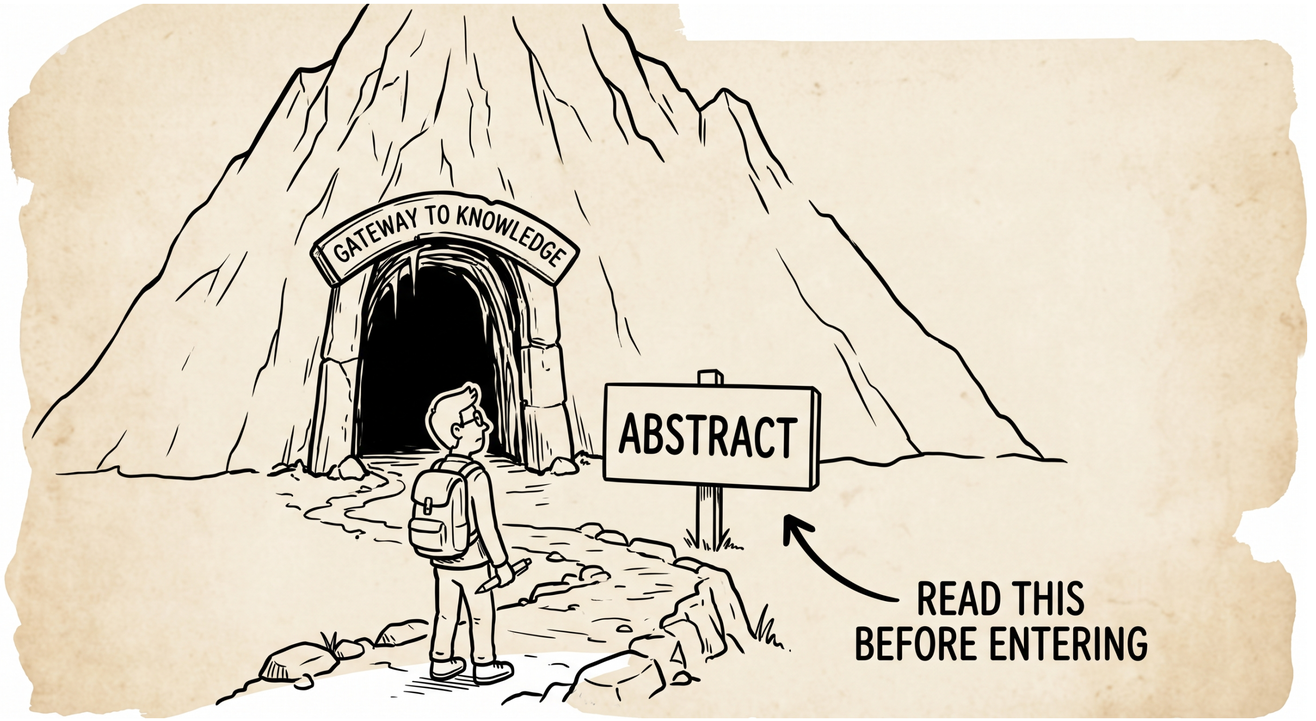
In Taiwan, a Post-acute Care-Cerebrovascular Diseases (PAC-CVD) program was implemented since 2014 to improve the quality of care by providing condensed rehabilitation programs to the patients with stroke [4,5]. Various assessments should be evaluated for patients of PAC-CAD program to provide more detailed information from various aspects [6]. The comprehensive assessments include modified Rankin scale (MRS),

* Corresponding author at: Department of Laboratory Medicine, Chang Gung Memorial Hospital at Linkou, No.5, Fuxing St., Guishan Dist., Taoyuan City 333, Taiwan.
** Corresponding author at: Department of Physical Medicine & Rehabilitation, Chang Gung Memorial Hospital at Taoyuan, School of Medicine, Chang Gung University, No.123, Dinghu Rd., Guishan Dist. Taoyuan City, Taiwan.
E-mail addresses: hsinyaowang@cgmh.org.tw (H.-Y. Wang), leonard@cgmh.org.tw (C.-K. Chen).
[1] These authors contributed equally to this manuscript.

https://doi.org/10.1016/j.ijmedinf.2018.01.002
Received 19 August 2017; Received in revised form 22 November 2017; Accepted 2 January 2018
1386-5056/

The Content of Abstract

- Abstract provides **essential** information for readers to decide if they should read the full paper.
  - Background & Goal
  - Methods
  - Results
  - Conclusion



Introduction

- There are three main functions of Introduction:
  - Tell readers more comprehensive backgrounds (than Abstract)
  - Carefully point out the problem to be fixed (may be absent in Abstract)
  - Tell readers the objective in a more complete way (than Abstract)

The Content of Introduction

Background
Problem
Objective
Methodology
Results
Key Findings / Conclusion

Related Work (Literature Review)

- Tell readers about previous works
- You usually find many cited papers in this section.
- This section also tells readers why our paper differs.

Methods

- Tell readers what exactly did you do.
- Write the details as more as possible.
  - Because we have to ensure reproducibility.
  - This section is like a **recipe** of your paper.

Results

- Results from experiments are described in this section.
- The most important part of your paper.

Task (metric)	System	Score (%)
Doc retrieval (Recall)	Our baseline	87.65
	BEVERS[a]	92.60
Sent retrieval (Recall)	Our baseline	76.65
	BEVERS[a]	86.60
RTE (Accuracy)	Our baseline	61.17
	BEVERS[a]	69.73
	GPT-3.5 (*zeroshot*)	43.17
	GPT-3.5 (*3-shot*)	44.20
	GPT-4 (*zeroshot*)	47.23
	GPT-4 (*3-shot*)	48.40
Full pipeline (FEVER Score)	Our baseline	52.47
	BEVERS[a]	64.80

Table 4: Results on the different stages. *a*: DeHaven and Scott (2023).

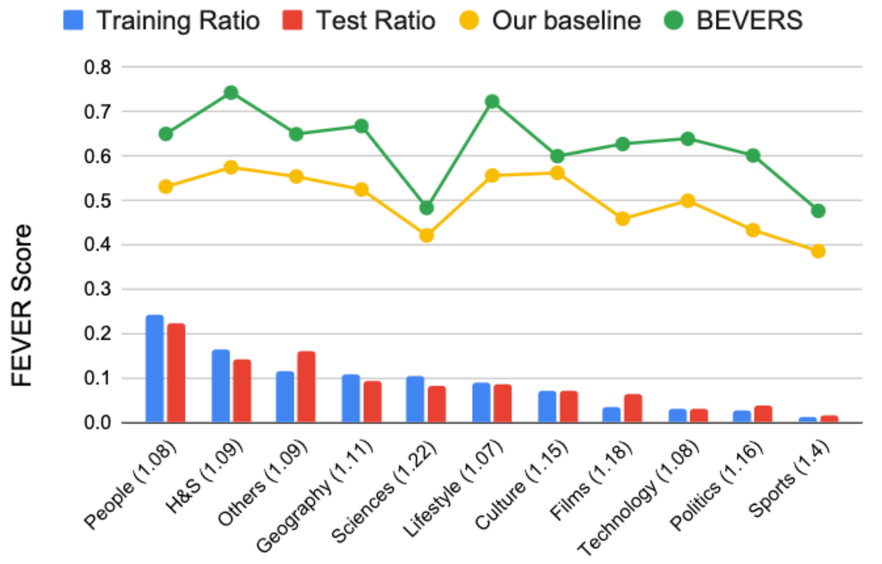


Figure 2: Performance comparisons for the claims of different domains in the full-pipeline setting. H&S refers to the domain of Humanities & Social Sciences. Values in the parentheses are the average number of evidence pages.

Writing Results is like taking IELTS Task 1

Academic Writing Sample Task – 1B

WRITING TASK 1

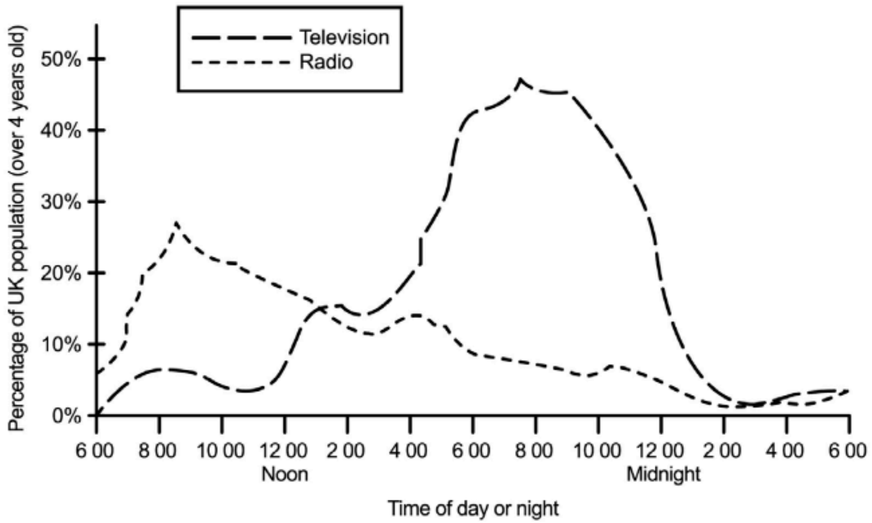
You should spend about 20 minutes on this task.

The graph below shows radio and television audiences throughout the day in 1992.

Summarise the information by selecting and reporting the main features, and make comparisons where relevant.

Write at least 150 words.

Radio and television audiences in UK, October – December 1992



https://ielts.org/cdn/Sample-tests/ielts-academic-writing-sample-tasks-2023.pdf

Discussion

- Write your findings based on the results.
- If no Discussion section, write your findings based in the **Results** section.

Conclusion

- Key findings of your paper.

Syllabus

Week	Topic	LaTeX	Note
1	Course Introduction, How to write your paper **title**?	LaTeX environment (env)	
2	**Datasets**	BibTeX, table env	
3	**Evaluation Metrics**, **Baselines**	\citet & \citep	
4	Paper Figures: Method Overview (including captions)	figure env & caption	Diagramming
5	**Method** Section (I): Problem Formulation and Model Architecture	Math symbols	
6	**Method** Section (II): Algorithms, Loss Functions, and Training Details	Equations	
7	How to Write Related Work (Literature Review)	BibTeX (clean)	
8	**Results** Section (I): Tables (including captions)	table env, \ref	
9	**Results** Section (II): Figures (including captions)	figure/subfigure env, \ref	
10	How to Write **Discussion**		
11	**Abstract** and **Title** (more formal than the first week)		
12	**Introduction** (I): Motivation and Problem Statement		
13	**Introduction** (II): Contributions and Paper Organization	\begin{enumerate}, \ref	
14	Paper Optimization		
15	Academic Writing with AI		
16	Final Paper Presentation and Discussions		Presentation

Why is the course structured this way?

Week	Topic
1	Course Introduction, How to write your paper **title**?
2	**Datasets**
3	**Evaluation Metrics**, **Baselines**
4	Paper Figures: Method Overview (including captions)
5	**Method** Section (I): Problem Formulation and Model Architecture
6	**Method** Section (II): Algorithms, Loss Functions, and Training Details
7	How to Write Related Work (Literature Review)
8	**Results** Section (I): Tables (including captions)
9	**Results** Section (II): Figures (including captions)
10	How to Write **Discussion**
11	**Abstract** and **Title** (more formal than the first week)
12	**Introduction** (I): Motivation and Problem Statement
13	**Introduction** (II): Contributions and Paper Organization
14	Paper Optimization
15	Academic Writing with AI 🤖
16	Final Paper Presentation and Discussions

W1-W14: We do not use AI.

Once we know how to write, we can steer AI to help us write.

Course Mode

- Each week:
    1. Short lecture
    2. In-class writing: this week's new section, and revise last week's draft according to the teacher's feedback (submit in class or by 23:59 on the same day)
    3. Revise last week's draft
    4. Individual guidance
    5. Feedback from the instructor after class

Evaluation

1. In-class writing activities and participation: 15% (1% for each week)
    - You can get the scores if you come and write.
2. Weekly section drafts and revisions: 60% (4% for each week)
3. Final Paper Presentation and Discussions: 25%

No homework. All the writing happens in class.

- If you cannot finish in class, submit by 23:59 on the same day (no more delay). You still get the full score.

No exam in this course.

- No quizzes or exams.
- We just need to practice writing.

Weekly section drafts and revisions

Score	3%	2%	1%	0%
This week	Evaluated by the teacher based on the draft quality.			Not received.
Next week	-	-	Correctly revised.	Not received.

This week you write in class. Next week you revise it in class, on top of that week's new writing.

Final Paper Presentation and Discussions (W16)

Items	Oral Presentation (60 min)	Peer-review (50 min)
Details	Present the biggest changes in your Paper Writing during this semester.	Read your classmates' papers in class and **check** whether the paper contains essential elements you learn in this course.
Score	15%	10%

The teacher will summarize the review comments after oral presentations and peer-review (30 min).
Last week revision: 10 min.

Textbook

試閱

理科英文論文寫作

作者： R. Lewis、N. Whitby、 E. Whitby 追蹤作者

譯者： 劉華珍、李珮華

出版社：眾文 訂閱出版社新書快訊

出版日期：2009/03/10

語言：繁體中文

定價：~~350~~元

優惠價： **95** 折 **333** 元

本商品單次購買10本9折 315 元

優惠折扣 8/17-8/19樂購日-鑽石會員結帳滿1200再88折,部分除外

優惠折扣 8/17-8/19樂購日-白金/黃金/一般會員結帳滿1200再9折,部分除外

點數回饋 8/17-8/19滿$5000享2%點數回饋,上限200點

點數回饋 8/17-8/19加碼｜買就送最高3%OPENPOINT(部份除外)

運送方式： 臺灣與離島 海外

可配送點： 台灣、蘭嶼、綠島、澎湖、金門、馬祖

可配送點：台灣、蘭嶼、綠島、澎湖、金門、馬祖

https://www.books.com.tw/products/0010429760

Any questions?

LaTeX Environment

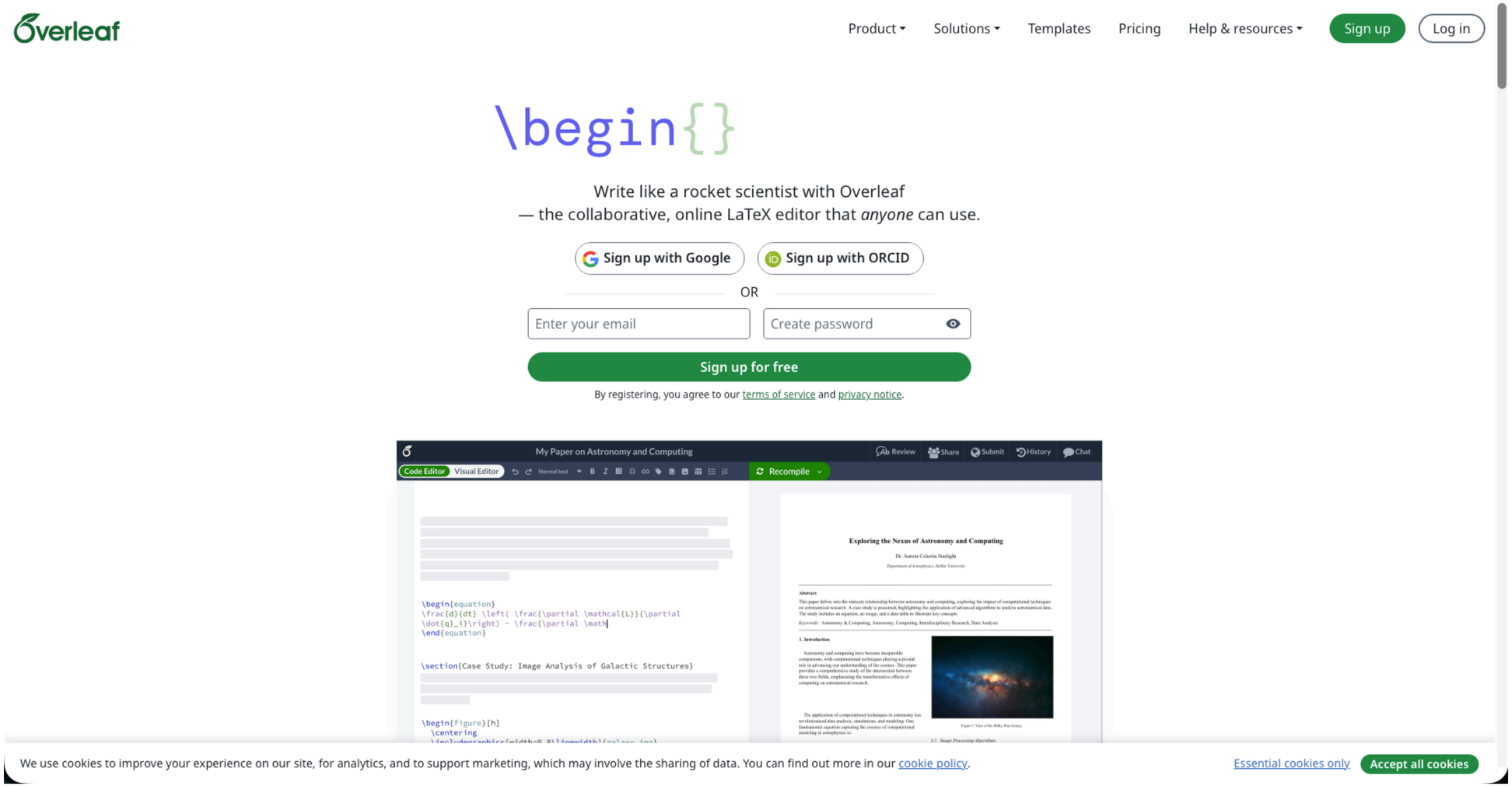
Overleaf

https://www.overleaf.com/

- Online LaTeX editor.
- Free usage for most of the time.
- Can be synced with GitHub

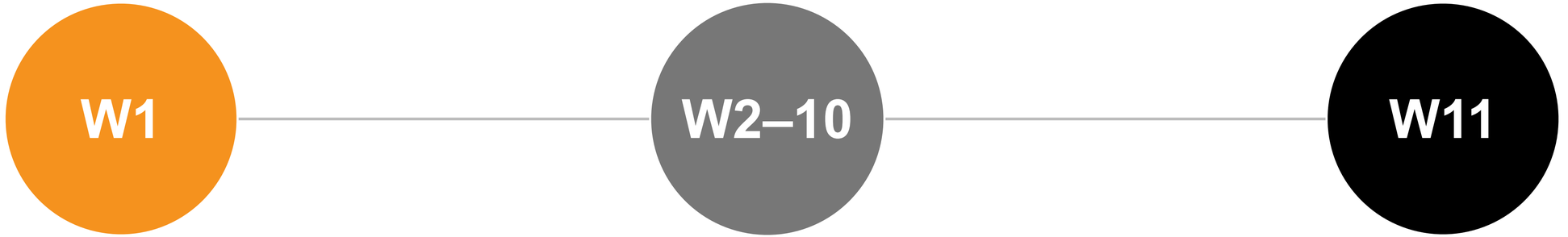
Overleaf

https://www.overleaf.com/



Your First Writing Task: Title

Today you write a working title



Write a working title
one title, today

Check it every week
does what I wrote still match the title?

Write the real title
naming, wording, searchability

No results yet?

Write the title you hope to earn.
Every week you can check whether you are earning it.

Why write a title before you have results?

- A title is the shortest possible statement of what your work claims.
  - One line only. Re-check every week.
  - When your work drifts, the title stops matching first.
  - Forces you to name your contribution (make your contribution visible).

The Function of “Title”

- A title is the first part that readers read from a paper.
- A title needs to be **attractive**.

英文	Resident Evil	Backrooms
臺灣翻譯	2002惡靈古堡	後室
中國翻譯	生化危機	後室
香港翻譯	生化危機之變種生還者	嚇房
新加坡翻譯	妖獸工廠	-

We usually find the translated movie title distort the beauty and attractiveness in its source language.

How to write a title?

- (1) Length: Concise (short). Aim for 8-12 words.
- (2) Title Structure: A phrase, not a sentence.
- (3) Abbreviation: No abbreviations your readers do not already know.

How to write a title? (1) Length

- Concise (short). Aim for 8-12 words.

❌ A Novel Deep Learning Based Approach for the Automatic Classification of Liver Lesions in Computed Tomography Images

How to write a title? (1) Length

- Concise (short). Aim for 8-12 words.

Every paper people write should be novel. We don’t need to mention “novel” again.

❌ A Novel Deep Learning Based Approach for the Automatic Classification of Liver Lesions in Computed Tomography Images

How to write a title? (1) Length

- Concise (short). Aim for 8-12 words.

Every paper people write should be novel. We don't need to mention "novel" again.

❌ A Novel Deep Learning Based Approach for the Automatic Classification of Liver Lesions in Computed Tomography Images

In 2026, deep learning is assumed in default.

How to write a title? (1) Length

- Concise (short). Aim for 8-12 words.

Every paper people write should be novel. We don't need to mention "novel" again.

❌ A Novel Deep Learning Based Approach for the Automatic Classification of Liver Lesions in Computed Tomography Images

In 2026, deep learning is assumed in default.

Your readers definitely know it.

How to write a title? (1) Length

- Concise (short). Aim for 8-12 words.

Every paper people write should be novel. We don't need to mention "novel" again.

❌ A Novel Deep Learning Based Approach for the Automatic Classification of Liver Lesions in Computed Tomography Images

In 2026, deep learning is assumed in default.

Your readers definitely know it.

✅ Automatic Classification of Liver Lesions in CT Images

How to write a title? (2) Title Structure

- [Traditional] A phrase, not a sentence.

✅	Deep Residual Learning for Image Recognition
❌	We Propose a Deep Residual Learning Method

How to write a title? (2) Title Structure

- You can use a sentence if your title is the finding itself.

✅	Attention Is All You Need.
✅	Language Models are Few-shot Learners

How to write a title? (3) Abbreviation

- No abbreviations the readers do not already know.

❌	An **MLM**-Based **PTM** for **FSL** on Radiology Reports
✅	**Few-shot** Radiology Report Labeling with **Masked Language Modeling**

*PTM (pre-trained Transformer models) is unnecessary for this title.

How to write a good title

- I did some pre-training using Transformers.

Pre-training of Transformers

🤔 Which kind of transformers? What's your novelty?

Pre-training of Deep Bidirectional Transformers

🤔 What's your final goal/task?

Pre-training of Deep Bidirectional Transformers for Language Understanding

Your Turn Now

How to write a title?

- (1) Length: Concise (short). Aim for 8-12 words.
- (2) Title Structure: A phrase, not a sentence.
- (3) Abbreviation: No abbreviations your readers do not already know.

1. Write three (~5) candidate titles for your own upcoming paper.
2. Check each one against the three rules.
3. Pick the best one.

NEXT WEEK ATTENTION!!

- We are going to write the **Datasets** section.
- Please prepare your materials, including:
  - (Name: the dataset name you use) <- optional
  - Source: where does the dataset you use come from?
  - Statistics: number of instances, class distribution, average length (NLP), image sizes (CV), and so on.
  - Preprocessing & Filtering: Did you remove some weird data or perform normalization? Tell readers all.
  - Splits: how to split your dataset into training, validation and test sets?
  - Ethical Considerations:

Thank you for your participation 🙂