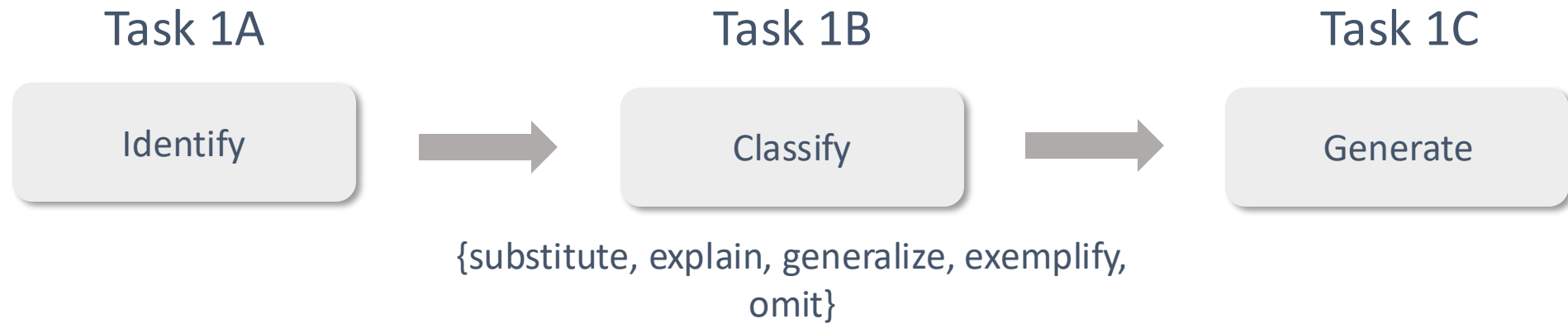# PLABA Task 1 Overview – TREC 2024

Organizers: Brian Ondov, Bill Xia, Ishita Unde,
Hoa Dang, Dina Demner-Fushman

# The Task

# Term Replacement Subtasks

Task 1A

Identify

Task 1B

Classify

{substitute, explain, generalize, exemplify, omit}

Task 1C

Generate

# Replacement Types

1.  **Substitution:** The term is jargon for a common alternative (e.g., "myocardial infarction" → "heart attack").

2.  **Explained:** The term has no alternative or is important to the topic. For example:
    "This study looked at treatments for <u>sleep apnea</u> (*when you stop breathing while sleeping*)."

3.  **Generalized:** The term can be replaced with a more general category without losing its significance. For example:
    "Clearing of the infection is confirmed with a ~~Nucleic Acid Amplification Test (NAAT)~~ common lab test."

4.  **Exemplified:** The term is a general category that can be better represented by a well-known example. For example:
    "Depression is common in people with <u>neurodegenerative diseases</u> (*like Parkinson's*)."

5.  **Omitted:** The term is not relevant to understanding the sentence, or too technical to explain and does not need to appear in a consumer version.
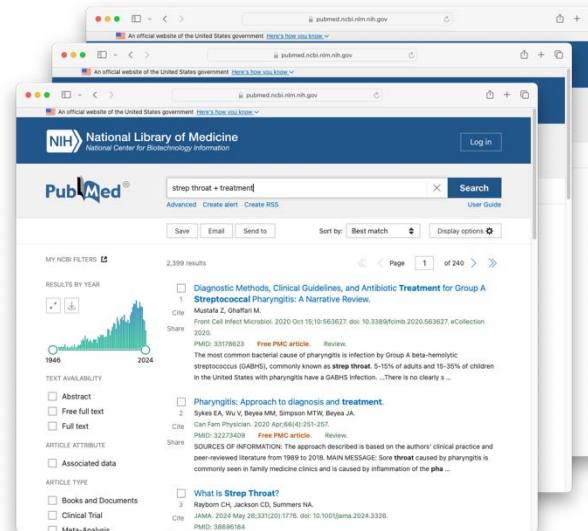
# The Data

# Source Data

40
MedlinePlus Questions



400
Abstracts

10 train
30 test

X

10
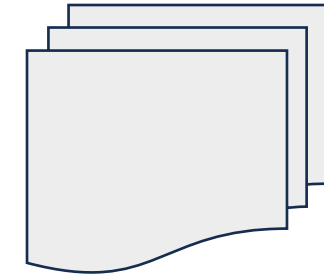PubMed
Results

=

100 train
300 test

# Translation to Plain Language

Median change from baseline was evaluated for procoagulant, anticoagulant, and fibrinolytic parameters, and for sex hormone-binding globulin (SHBG).

In the last sixth cycle the blood clotting chemicals changed significantly with the largest change in the EE/DRSP group, closely followed by the EE/LNG group.

Results: Median change of endogenous thrombin potential (ETP) based activated protein C sensitivity resistance (APCr) at cycle 6 was +30% for E4/DRSP, +165% for EE/LNG (p-value <0.05 vs E4/DRSP), and +219% for EE/DRSP (p-value <0.05 vs E4/DRSP).

For the study's results, the average change for a coagulation or bleeding-prevention test at cycle 6 was +30% for E4/DRSP, +165% for EE/LNG, which was much higher than E4/DRSP, and +219% for EE/DRSP, which was much higher than vs E4/DRSP.

Changes to prothrombin fragment 1 + 2 and SHBG for E4/DRSP, EE/LNG, and EE/DRSP were +23%, +71%, and +64% (p-value <0.05 vs E4/DRSP); and +55%, +74% and +251% (p-value <0.05 vs E4/DRSP), respectively.

Change of internal thrombin potential (ETP) based activated protein C sensitivity resistance (APCr) at cycle 6 was increased for all groups.  APCr is an increased tendency of the blood to clot.
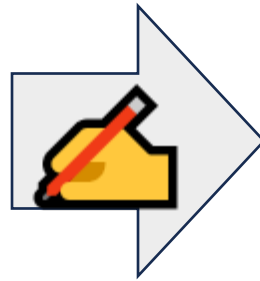
# Plain Language Data

400
Abstracts

~1,200
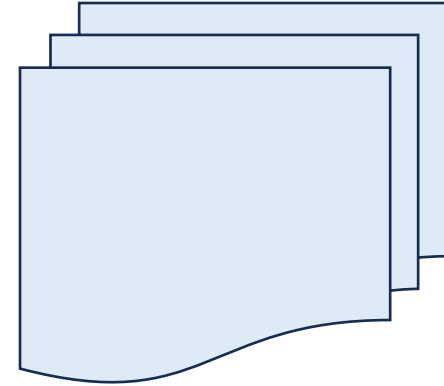Sentence Pairs
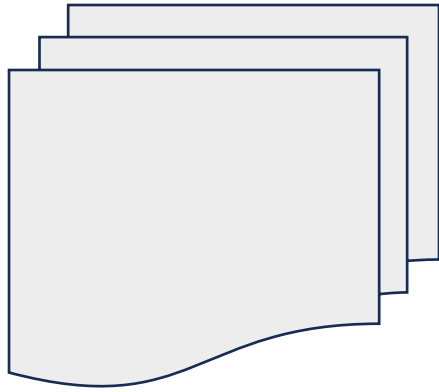
Academic Language

3
Domain Experts

Plain Language

# Annotation Types

**Expert Term**

**Synonym**

**Substitution**

**Explanation**

## Source (PMID 35511781)

| |
|---|
| Purpose: To find predictive markers for the visual potential in optical coherence tomography (OCT) one month after surgical repair of macula-involving rhegmatogenous retinal detachment (miRD) with and without internal limiting membrane (ILM) peeling. |
| Methods: This retrospective single-center, single-surgeon cohort study included 74 patients who underwent pars plana vitrectomy (PPV) for primary miRD between January 2013 and August 2020 with follow-up examinations for at least 6 months. |
| Patients developing recurrent detachments, media opacities, or with an axial length over 27 mm were excluded from the analysis. |

**Omission**

## Adaptation by Domain Expert

| |
|---|
| This study aimed to find out if vision can be predicted by diagnostic imaging one month after surgery to repair common retinal detachment (the damage to light-sensitive layer at the back of the eye (retina) when it is pulled away from its normal position). In some surgeries, the tissues next to retina were pulled to repair the detachment. |
| This study looked at past cases at a single health care center, and that had surgery done by a single surgeon. |
| Patients who had other retina detachments in the past or media opacities (such as cataracts) were not included in the study. |

**Generalization**

**Exemplification**

# Annotation Example 1



Original
term
term
[expert_term]
However, Arruga's ring suture method induced cataract more frequently than did other surgical procedures.

Rewrite 1
term
term
[EXPLAINS]
However, a particular procedure called "Arruga's ring suture method" caused cataract more often than other surgical procedures.

Rewrite 2
term
[GENERALIZES]
However, a specific suture method caused cataract more frequently than did other surgeries.

Rewrite 3
However, Arruga's ring suture method caused more cataracts than other surgeries

Inter-annotator Agreement: **52.03%**

# Annotation Example 2



**Original**

Preoperative SFFH was measured on two occasions an hour apart using optical coherence tomography (OCT) by two independent observers.

**Rewrite 1**

We measured the height of the fluid under the fovea before surgery two times an hour apart.

**Rewrite 2**

We made these measurements with cross-sectional imaging techniques and made sure two different people took them.

**Rewrite 3**

The height of the liquid behind the retina was measure twice an hour apart before surgery.

# Evaluations

# Subtask Evaluation

- ## Task 1A – Identification

  - **F1 against union of annotations (≥75% overlap to match)**

- ## Task 1B – Classification

  - **F1 against 3-fold annotator labels**

- ## Task 1C – Generation

  - **Manual (3-point Likert scales, mapped to 0-1)**

    - **Accuracy:** Outputs should contain the accurate information.

    - **Completeness:** Outputs should seek to minimize information lost from the original text.

    - **Simplicity:** Outputs should be easy to understand.

    - **Brevity:** Outputs should be concise.

```
"ictal EEGs": [
    [
        "GENERALIZE",
        "Measurements"
    ],
    [
        "SUBSTITUTE",
        "The electrical brain rhythm"
    ],
    [
        "OMIT",
        ""
    ]
],
"generalized bilaterally synchronous": [
    [
        "SUBSTITUTE",
        "across both sides of the brain"
    ],
    [
        "OMIT",
        ""
    ]
],
```
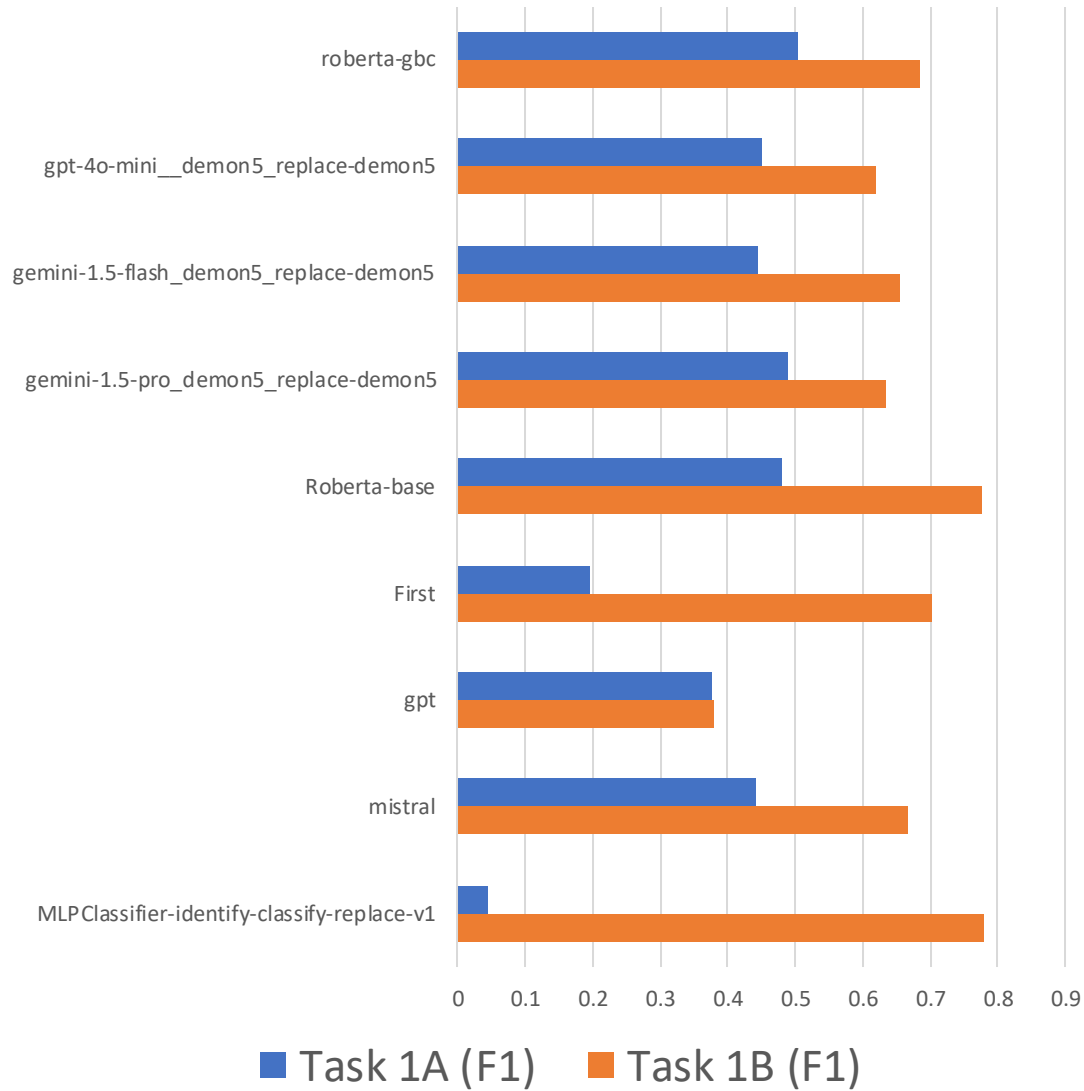
# Participation

| Team | Run | Priority | Task 1A | Task 1B | Task 1C |
|------|-----|----------|---------|---------|---------|
| **BU** | MLPClassifier-identify-classify-replace-v1 | 1 | X | X | X |
| **CLAC** | mistral | 1 | X | X | X |
| **CLAC** | gpt | 2 | X | X | |
| **IITH** | First | 1 | X | X | |
| **UM** | Roberta-base | 1 | X | X | |
| **ntu_nlp** | gemini-1.5-pro_demon5_replace-demon5 | 1 | X | X | X |
| **ntu_nlp** | gemini-1.5-flash_demon5_replace-demon5 | 2 | X | X | |
| **ntu_nlp** | gpt-4o-mini__demon5_replace-demon5 | 3 | X | X | |
| **Yseop** | roberta-gbc | 1 | X | X | |

# Evaluation Results Preview (not final)

| Team | Run | Priority | Task 1A (F1) | Task 1B (F1) | Task 1C (manual judgments, 0-1) | | | | |
|------|-----|----------|--------------|--------------|------|------|------|------|------|
| | | | | | Accy | Comp | Simp | Brev | Avg. |
| BU | MLPClassifier-identify-classify-replace-v1 | 1 | 0.0459 | **0.7788** | 0.7500 | 0.7258 | 0.9274 | **0.9435** | 0.8367 |
| CLAC | mistral | 1 | 0.441 | 0.6661 | **0.9505** | **0.9483** | **0.9738** | 0.6702 | **0.8857** |
| CLAC | gpt | 2 | 0.3767 | 0.3795 | | | | | |
| IITH | First | 1 | 0.1956 | 0.7014 | | | | | |
| UM | Roberta-base | 1 | 0.4787 | 0.7765 | | | | | |
| ntu_nlp | gemini-1.5-pro_demon5_replace-demon5 | 1 | 0.4885 | 0.6335 | 0.9117 | 0.9010 | 0.9511 | 0.5871 | 0.8377 |
| ntu_nlp | gemini-1.5-flash_demon5_replace-demon5 | 2 | 0.4431 | 0.6544 | | | | | |
| ntu_nlp | gpt-4o-mini__demon5_replace-demon5 | 3 | 0.4518 | 0.6197 | | | | | |
| Yseop | roberta-gbc | 1 | **0.5036** | 0.6838 | | | | | |

# Task 1A/B

- roberta-gbc
- gpt-4o-mini__demon5_replace-demon5
- gemini-1.5-flash_demon5_replace-demon5
- gemini-1.5-pro_demon5_replace-demon5
- Roberta-base
- First
- gpt
- mistral
- MLPClassifier-identify-classify-replace-v1

Axis: 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9

■ Task 1A (F1)   ■ Task 1B (F1)

# Task 1C

— MLPClassifier-identify-classify-replace-v1
— mistral
— gemini-1.5-pro_demon5_replace-demon5

Completeness, Simplicity, Brevity, Accuracy