Timeline for shared tasks:

1. Choose Shared Task from the following 4 tasks, Form your teams, Do necessary registration (1-2 per team)
   a. For shared task 2, please do remember to register for this task when you are ready.
2. Write a 1 page proposal about your initial understanding of the task. **(Due on April 3)**
   a. Team Name, and Team Members
   b. Briefly describe Your Chosen Task
      i. For all the tasks, you could specifically have your own research interests on any tasks.
      ii. For the SharedTask3, you could at least choose 2 over the 5 tracks to participate.
      iii. If you choose SharedTask 4, list the paper you hope to reimplement or improve, describe the tasks in your own words.
   c. Proposed Models
      i. Baseline Methods
      ii. At least 3 methods you hope to try
         1. have learned or will learn from this class.
         2. Or other methods you know in the literature.
   d. Your timeline for the project, different roles for the team members.
   e. I will give you more feedback by **April 4 Office hours or by Appointment.**
3. Final Report **(Due on May 8)**
   a. 5 pages (at most)
   b. Including a standard research paper sections:
      i. Abstract, Introduction, Related Work, Methods, Results, Discussion, Conclusion
   c. For SharedTask 2, you are encouraged to submit your result for evaluations **by April 23, you may be selected by the BEA workshop to publish a paper.**

Shared Task 1: Speaker-aware Dialogue Act Classification

- Baseline Paper: Speaker Turn Modeling for Dialogue Act Classification, https://aclanthology.org/2021.findings-emnlp.185/
- Baseline Code: https://github.com/zihaohe123/speak-turn-emb-dialog-act-clf
- Datasets
  - 3 datasets already Summerized here.https://github.com/zihaohe123/speak-turn-emb-dialog-act-clf/blob/master/data.zip

- ○ |C| are the number classes to predict for each utterance.
- ○ |P| are the number of speakers(parties) in the dialogue

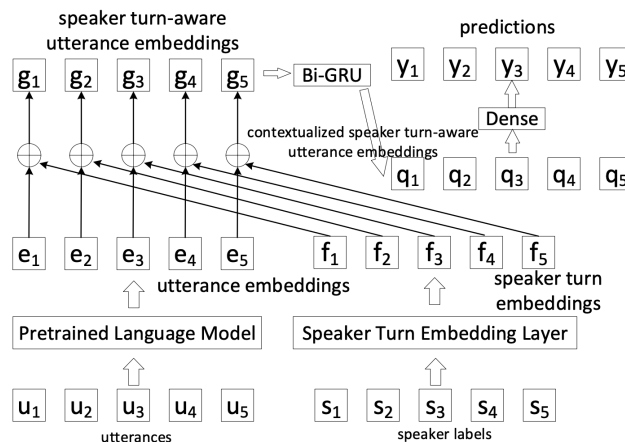| Dataset	$|C|$	$|P|$	Train	Val	Test
SwDA	43	2	1003/193K	112/20K	19/4.5K
MRDA	5	multiple	51/75k	11/15.3K	11/15K
DyDA	5	2	11K/87.1K	1K/8K	1K/7.7K

Figure 1: The overall framework of our proposed method. In this toy example, the conversation consists of five utterances.

Please read the paper and run their code, openly investigate your own research questions(not limited the following ones):

1. How to improve the speaker encodings and fuse them with utterance encoding?
2. Any other public online dataset could help on the speaker-aware dialogue?
3. Can HMM or CRF help on this sequence labeling task of dialogue act prediction?
4. How well will the LLM perform on this?

SharedTask 2: Math Tutor Response Evaluation

https://sig-edu.org/sharedtask/2025

Task Goals & Description (See more details in the above task page, choose at least 2 over 5 tracks)

You need to build multiple classification models to detect whether the tutor's response(already generated form LLMs, will provided to you) are "Yes","To some extent", "No" to student's utterance accordin to the following metrics:

- Mistake_Identification marking whether the tutor recognized a mistake in a student's response
- Mistake_Location marking whether the tutor's response accurately points to a genuine mistake and its location
- Providing_Guidance marking whether the tutor offers correct and relevant guidance, such as an explanation, elaboration, hint, examples, and so on
- Actionability marking if it is clear from the tutor's feedback what the student should do next

Important Timeline if you plan to participate in this competition.
- March 12, 2025: Development data release
- April 9, 2025: Test data release
- April 23, 2025: System submissions from teams due
- April 30, 2025: Evaluation of the results by the organizers
- May 21, 2025: System papers due
- May 28, 2025: Paper reviews returned
- June 9, 2025: Final camera-ready submissions
- July 31 and August 1, 2025: BEA 2025 workshop at ACL

SharedTask 3: BRAINTEASER: A Novel Task Defying Common Sense

More details on the task page here: https://brainteasersem.github.io/#data
You need to work on both 2 subtasks.

SharedTask 4: Open Research Task

You could choose any research topic, but the research must already have **the existing datasets and evaluation methods available** for you to play. Survey-only based tasks are not allowed.

You are encouraged to reimplement/improve a published work **at NLP Conferences from the last 5 years (since 2019)**:

ACL: https://aclanthology.org/venues/acl/
NAACL:https://aclanthology.org/venues/naacl/
EMNLP: https://aclanthology.org/venues/emnlp/