

Transitioning from Open-Domain to Task-Related Topics

- ❑ Task: Topic Transition Through Dialogue
- ❑ Data: Generate from BlenderBot by yourselves
- ❑ Evaluation: Keyword hit rate & Quality of dialogue
- ❑ Submission:
 - ❑ Model weight + inference script
 - ❑ Test set output
 - ❑ (later) Human evaluation & report

- **simulator.py**
 - Can be used to interact with the simulator
 - Usage: `python3 simulator.py`
 - `model_name_or_path`: model to chat with simulator
 - `output`: file to save the dialogs
 - for more detail, you can see the help message.
- **hit.py**
 - Calculate the hit rate from your output dialogues and keywords
 - Usage: `python3 hit.py`
 - `prediction`: file that save the prediction dialogs

Important Note of the Released Scripts

1. **DON'T use the test set to train/tune your model**
2. **DON'T modify the seed when generating test set output**
3. **DO modify the script to inference your model**
4. **DON'T modify keywords.json or hit.py**

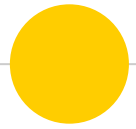
- ❑ Keyword hit rate (on test set)
 - ❑ Automatic, generate your output on test split and run `hit.py` to get the score, write in your report
- ❑ Quality of dialogue
 - ❑ Human evaluation (you'll need to help us for this!)
 - ❑ Naturalness, Relevance, Aggressiveness, Overall

Human Evaluation

- ❑ Each *student* would receive ~30 pairs of dialogue,
 - ❑ Each pair shares the same starting sentence from bot
 - ❑ You need to judge which one is better in the four directions (in the previous page)
- ❑ We will release sample output file for your judgement
 - ❑ Ex. (a,b,a,a) for the 4 directions for dialogue pair (a, b)

7 Quality of your Assessment

- ❑ Please give consistent judgements
 - ❑ Others' scores depend on you!
- ❑ We'll use pairwise win rate as the final ranking (1 ranking for each directions)
- ❑ If you correctly select the winner (agree with the final ranking), you'll get some bonus score!



Grading

Grading

- Report (35%)
- Performance (25%)
 - Keyword Hit Rate (10%) ← also based on ranking
 - Rankings (15%)
- Annotation (15% + bonus 5%) ← by student, not by team
- Oral Presentation (25%)
- Participation (bonus 5%)

Report Format

- Abstract
- Introduction
- Approach
- Experiment / Evaluation
- Conclusion
- Work Distribution

Oral Presentation

- Video Recording (~8 min)
 - NOTE: introduce yourselves in the beginning
 - Presenting by all members is recommended
- Submission: submit Youtube link to COOL
- Grading will be based on
 - Novelty, Clarity, and Result Analysis

Important Dates

- 6/12 model & output due
- 6/17 annotation due
- 6/19 report & oral video due
- 6/24 participation due

Mon	Tue	Wed	Thu	Fri	Sat	Sun
5/30	5/31	1	2	3	4	5
6	7	8	9	10	11	12 Model & Output
13	14	15	16	17 Annotation	18	19 Report & Oral
20	21	22	23	24 Participation	25	26