Bots used in the paper: ChatMatch: Evaluating Chatbots by Autonomous Chat Tournaments
https://aclanthology.org/2022.acl-long.522.pdf

1. **PLATO-2**: trained on The Pushshift Reddit Dataset containing 5,601,331,385 comments.

2. **Blender Bot**: trained on the Blended skill talk dataset. Contains 7k English conversations. Link: https://huggingface.co/datasets/blended_skill_talk

3. **CS** (Seq2Seq model): trained on the Persona-Chat dataset, which contains roughly 11k dialogues in English with over 165k utterances. This dataset contains different personas. Paper about the dataset: https://aclanthology.org/P18-1205.pdf

4. **UG** (large pretrained Seq2Seq model): trained on the the same data as PLATO-2.

5. **DialoGPT medium:** trained on Reddit data from the period 2005 to 2017, total of 147million conversations. Same kind as PLATO-2 and UG.

6. **Image Seq2Seq model: t**rained on a mixture of 12 datasets, some good picks: ConvAI2, Pushshift Reddit Dataset (same as PLATO-2), Wizard of wikipedia. The latter contains 22k+ dialogues, with a total of 202k turns.

Additional Datasets and Information:

1. Multi-domain Wizard of Oz dataset. Contains 40.2k English dialogues for multi-domain. Link: https://huggingface.co/datasets/meta_woz

List of used datasets in the paper: *The Dialogue Dodecathlon:* Open-Domain Knowledge and Image Grounded Conversational Agents. Link: https://aclanthology.org/2020.acl-main.222.pdf

Name	Ask Questions	Answer Questions	Respond to Statements	Persona Grounding	Knowledge Grounding	Situation Grounding	Image Grounding	Train	Valid	Test	# Turns	Resp. Length
ConvAI2	✓	✓	✓	✓				131,438	7,801	6,634	14.8	11.9
DailyDialog	✓	✓	✓					87,170	8,069	7,740	7.9	14.6
Wiz. of Wikipedia	✓	✓	✓		✓			74,092	3,939	3,865	9.0	21.6
Empathetic Dialog	✓	✓	✓			✓		40,252	5,736	5,257	4.3	15.2
Cornell Movie	✓	✓	✓					309,987	38,974	38,636	4.0	15.0
LIGHT	✓	✓	✓	✓		✓		110,877	6,623	13,272	13.0	18.3
ELI5		✓			✓			231,410	9,828	24,560	2.0	130.6
Ubuntu	✓	✓	✓					1,000,000	19,560	18,920	2.0	18.9
Twitter	✓	✓	✓					2,580,428	10,405	10,405	2.0	15.7
pushshift.io Reddit	✓	✓	✓					∼ 2200 M	10,000	10,000	2.0	35.0
Image Chat	✓	✓	✓	✓			✓	355,862	15,000	29,991	3.0	11.4
IGC	✓	✓					✓	4,353	486	7,773	3.0	8.6

Table 1: The 12 *dodeca*Dialogue subtasks, their sizes (number of train, valid, test utterances), and average number of turns and response length (words).