

PersonaChat	<ul style="list-style-type: none">Dataset contains multi-turn dialogs with each speaker adopting a persona8,939 training dialogs, 1000 for validation, 968 for testing955 possible personas in training set, 100 in validation, 100 in testingConversations cover a broad range of topics such as work, family, personal interestDiscussion topics change frequently during the course of the conversationsThese properties make the conversations particularly suitable for learning smooth, natural transitions <i>at each turn</i>																																													
CONVAI2	<ul style="list-style-type: none">Similar to Persona ChatTest set extended by 100 new personas and over 1,015 new dialogs																																													
Intent Datasets	<ul style="list-style-type: none">1 conversational datasets, 6 intent datasets, 1 short text clustering dataset are used in the paperThe paper uses its technique to discover and classify intents in all the datasetsBoth the raw and intent labelled versions are availableAirlines Twitter Conversation is a conversational dataset and Stackoverflow Dataset is a short text clustering dataset. The remaining are intent datasets <table><tr><th>Dataset</th><th>Utterances</th><th>Intents</th><th>Max Intent count</th><th>Min Intent count</th></tr><tr><td>Airlines Twitter Conversation</td><td>491</td><td>14</td><td>107</td><td>10</td></tr><tr><td>FinanceData</td><td>10003</td><td>77</td><td>187</td><td>35</td></tr><tr><td>AskUbuntuCorpus</td><td>162</td><td>5</td><td>57</td><td>8</td></tr><tr><td>ChatbotCorpus</td><td>206</td><td>2</td><td>128</td><td>78</td></tr><tr><td>WebApplicationCorpus</td><td>88</td><td>7</td><td>23</td><td>5</td></tr><tr><td>ATIS</td><td>4972</td><td>17</td><td>3666</td><td>6</td></tr><tr><td>Personal Assistant</td><td>25312</td><td>64</td><td>1218</td><td>171</td></tr><tr><td>Stackoverflow Dataset</td><td>20000</td><td>20</td><td>1000</td><td>1000</td></tr></table>	Dataset	Utterances	Intents	Max Intent count	Min Intent count	Airlines Twitter Conversation	491	14	107	10	FinanceData	10003	77	187	35	AskUbuntuCorpus	162	5	57	8	ChatbotCorpus	206	2	128	78	WebApplicationCorpus	88	7	23	5	ATIS	4972	17	3666	6	Personal Assistant	25312	64	1218	171	Stackoverflow Dataset	20000	20	1000	1000
Dataset	Utterances	Intents	Max Intent count	Min Intent count																																										
Airlines Twitter Conversation	491	14	107	10																																										
FinanceData	10003	77	187	35																																										
AskUbuntuCorpus	162	5	57	8																																										
ChatbotCorpus	206	2	128	78																																										
WebApplicationCorpus	88	7	23	5																																										
ATIS	4972	17	3666	6																																										
Personal Assistant	25312	64	1218	171																																										
Stackoverflow Dataset	20000	20	1000	1000																																										
Topical Chat	<ul style="list-style-type: none">Open domain, knowledge grounded, human to human chat datasetDataset consists of conversations on particular topics in which neither participant is given a roleDataset covers 8 topics. Fashion, politics, books, sports, general entertainment, music, science and technology, movies <table><tr><th></th><th>Train</th><th>Valid Freq.</th><th>Valid Rare</th><th>Test Freq.</th><th>Test Rare</th><th>All</th></tr><tr><td># conversations</td><td>8628</td><td>539</td><td>539</td><td>539</td><td>539</td><td>10784</td></tr><tr><td># utterances</td><td>188378</td><td>11681</td><td>11692</td><td>11760</td><td>11770</td><td>235434</td></tr><tr><td>average # turns per conversation</td><td>21.8</td><td>21.6</td><td>21.7</td><td>21.8</td><td>21.8</td><td>21.8</td></tr><tr><td>average length of utterance</td><td>19.5</td><td>19.8</td><td>19.8</td><td>19.5</td><td>19.5</td><td>19.6</td></tr></table>		Train	Valid Freq.	Valid Rare	Test Freq.	Test Rare	All	# conversations	8628	539	539	539	539	10784	# utterances	188378	11681	11692	11760	11770	235434	average # turns per conversation	21.8	21.6	21.7	21.8	21.8	21.8	average length of utterance	19.5	19.8	19.8	19.5	19.5	19.6										
	Train	Valid Freq.	Valid Rare	Test Freq.	Test Rare	All																																								
# conversations	8628	539	539	539	539	10784																																								
# utterances	188378	11681	11692	11760	11770	235434																																								
average # turns per conversation	21.8	21.6	21.7	21.8	21.8	21.8																																								
average length of utterance	19.5	19.8	19.8	19.5	19.5	19.6																																								
Wizard of Wikipedia	<ul style="list-style-type: none">Open domain, knowledge grounded, human to human chat datasetDiffers from Topical Chat in that the speakers are given the role of a knowledgeable expert and a curious learner																																													

	<ul style="list-style-type: none">1365 open domain dialog topics are covered <table><tr><th>Wizard of Wikipedia Task</th><th>Train</th><th>Valid</th><th>Test Seen</th><th>Test Unseen</th></tr><tr><td>Number of Utterances</td><td>166,787</td><td>17,715</td><td>8,715</td><td>8,782</td></tr><tr><td>Number of Dialogues</td><td>18,430</td><td>1,948</td><td>965</td><td>968</td></tr><tr><td>Number of Topics</td><td>1,247</td><td>599</td><td>533</td><td>58</td></tr><tr><td>Average Turns per Dialogue</td><td>9.0</td><td>9.1</td><td>9.0</td><td>9.1</td></tr><tr><td>Knowledge Database</td><td colspan="2">5.4M articles</td><td colspan="2">93M sentences</td></tr></table>	Wizard of Wikipedia Task	Train	Valid	Test Seen	Test Unseen	Number of Utterances	166,787	17,715	8,715	8,782	Number of Dialogues	18,430	1,948	965	968	Number of Topics	1,247	599	533	58	Average Turns per Dialogue	9.0	9.1	9.0	9.1	Knowledge Database	5.4M articles		93M sentences	
Wizard of Wikipedia Task	Train	Valid	Test Seen	Test Unseen																											
Number of Utterances	166,787	17,715	8,715	8,782																											
Number of Dialogues	18,430	1,948	965	968																											
Number of Topics	1,247	599	533	58																											
Average Turns per Dialogue	9.0	9.1	9.0	9.1																											
Knowledge Database	5.4M articles		93M sentences																												
Facebook Task Oriented Parsing (TOP)	<ul style="list-style-type: none">Dataset with complex hierarchical and nested queries45K annotations with 25 intents and 36 slots randomly split into 31K training, 5K validation and 9K test utterancesDataset mainly consists of user queries about navigation and various public events																														
MultiWOZ 2.1	<ul style="list-style-type: none">It is a multi-domain human to human conversation corpusIt contains labelled, goal oriented conversationsIt has 10k dialogs each with an average of 13.67 turnsThere are 7 domains and 37 slots																														
Amazon Commonsense Dialogs Dataset Related Blog	<ul style="list-style-type: none">Multi-turn, open domain dialog dataset focused on commonsense knowledge11k dialogs created by humans upon being shown a promptApproximately 6 turns per dialog																														
DailyDialog	<ul style="list-style-type: none">Dataset of multi-turn, chitchat dialogs with strong annotations for topic, emotion and utterance actDialogs reflect daily communication and cover topics about daily lifeContains 13,118 multi-turn dialogs each with about 8 speaker turns																														
Switchboard Dialog Act Corpus	<ul style="list-style-type: none">Collection of 1155 five-minute telephone conversations between 2 participantsIn the conversations, callers question receivers on provided topics such as child care, recyclingIt has 221,616 utterances (or 122,646 utterances when consecutive utterances from the same speaker are combined)Speech act tags are annotated																														
Poly AI conversational datasets Related Paper	<ul style="list-style-type: none">3 very large conversational datasets providedReddit Dataset: 3.7 billion comments structured as threaded conversations. 727 million context-response pairs are extractedOpenSubtitles: Dataset uses movie and television subtitles in 62 languages. Consecutive lines of subtitles are used to create conversation data. No guarantee that consecutive lines correspond to different speakers or that consecutive lines belong to the same scene or even the same show/movie. 316 million valid context-response pairsAmazon QA: Question Answer pairs from the ecommerce domain (collected from Amazon product QA). 3.6 million pairs are available																														

SNIPS	<ul style="list-style-type: none"> • Used for training and testing semantic parsing models • Has utterances from 7 intents, SearchCreativeWork, GetWeather, BookRestaurant, PlayMusic, AddToPlaylist, RateBook, and SearchScreeningEvent • Each intent has around 2000 training examples and 100 for testing • Dataset has only simple queries with simple intents and flat slots
TimeDial	<ul style="list-style-type: none"> • Crowdsourced challenge set for temporal commonsense reasoning • Dataset formulated as a multiple choice cloze task (task of predicting masked words) • Around 1.5k dialogs derived from DialyDialog
Simulated Dialog Dataset	<ul style="list-style-type: none"> • 3000 dialogs for movie ticket and restaurant reservation • Restaurant <ul style="list-style-type: none"> ◦ Slots are price_range, location, restaurant_name, category, num_people, date, time ◦ Train: 1116, Validation: 349, Test: 775 • Movie <ul style="list-style-type: none"> ◦ Slots are theatre_name, movie, date, time, num_people ◦ Train: 384, Validation: 120, Test: 264
ATIS in tree format and Simulated Ordering Dataset	<ul style="list-style-type: none"> • ATIS dataset contains audio recordings of people requesting flight reservations, with 21 intent types and 120 slot labels • 4478 utterances in the training set, 893 in the test set and 500 utterances in the validation set • ATIS dataset in bracketed tree format is also given • Simulated dataset contains 2 intents and 7 slot labels, 4767 training examples, 1362 test examples and 681 validation examples • Simulated dataset has complex hierarchical queries
Disfl-QA	<ul style="list-style-type: none"> • Dataset for contextual disfluencies during question answering • Disfluencies are interruption in the regular flow of the conversation • Disfl-QA extends SQuAD by adding disfluencies to every question in the validation set • There are around 12k disfluent question-answer pairs • Over 90% of disfluencies are corrections or restarts
	Blog post for TimeDial and Disfl-QA
ASNQ	<ul style="list-style-type: none"> • Dataset for answer sentence selection • It has 57,242 questions in training set and 2672 in validation set • For each question there is a paragraph with sentences labelled as positive if they make up the answer and negative if the sentence is not part of the answer to the question
WikiQA	<ul style="list-style-type: none"> • The WikiQA corpus is a set of question and sentence pairs, collected and annotated for research on open-domain question answering • Similar to ASNQ it contains a paragraph for each question in which

	<p>some sentences constitute the answer</p> <ul style="list-style-type: none"> There are 3047 questions and 29,258 sentences in the dataset, where 1473 sentences were labelled as answer sentences to their corresponding questions
TREC-QA	<ul style="list-style-type: none"> TREC-QA is also a question-answer dataset where for each question there are positive(correct) answers and negative(wrong) answers
SQuAD	<ul style="list-style-type: none"> Reading comprehension dataset consisting of questions posed by crowdworkers on Wikipedia articles Answer for a question is either a span of text in the article or the question is unanswerable with the given article SQuAD has 100,000 answerable questions and 50,000 unanswerable questions (which are framed to be similar to answerable questions)

Links To Datasets

1. PersonaChat : <https://github.com/facebookresearch/ParlAI/tree/main/projects/personachat>
2. Simulated Dialog Dataset : <https://github.com/google-research-datasets/simulated-dialogue>
3. Intent Datasets: <https://github.com/ajaychatterjee/IntentMining>
4. Topical Chat: <https://github.com/alexa/Topical-Chat>
5. Wizard of Wikipedia: https://parl.ai/projects/wizard_of_wikipedia/
6. ATIS in tree format and simulated ordering dataset: <https://aclanthology.org/2020.acl-main.186/>
7. Facebook Task Oriented Parsing (TOP): <http://fb.me/semanticparsingdialog>
8. SNIPS: <https://github.com/sonos/nlu-benchmark>
9. ASNQ: https://github.com/alexa/wqa_tanda
10. WikiQA: <https://www.microsoft.com/en-us/download/details.aspx?id=52419>
11. TREC-QA: <https://github.com/mcrisc/lexdecomp/tree/master/trec-qa>
12. MultiWOZ 2.1: <https://github.com/budzianowski/multiwoz/tree/master/data>
13. CONVAI2: <https://github.com/facebookresearch/ParlAI/tree/main/parlai/tasks/convai2>
14. DailyDialog: <http://yanran.li/dailydialog>
15. Switchboard Dialog Act Corpus: <https://catalog.ldc.upenn.edu/LDC97S62>
16. TimeDial: <https://github.com/google-research-datasets/timedial>
17. SQuAD: <https://rajpurkar.github.io/SQuAD-explorer/>
18. Disfl-QA: <https://github.com/google-research-datasets/disfl-qa>
19. Poly AI conversational datasets: <https://github.com/PolyAI-LDN/conversational-datasets> (Paper: <https://aclanthology.org/W19-4101.pdf>)
20. Amazon Commonsense Dialogs Dataset: <https://github.com/alexa/commonsense-dialogues> (Blog: <https://www.amazon.science/blog/amazon-releases-new-dataset-for-commonsense-dialogue>)