A Commonsense-Enhanced Document-Grounded Conversational Agent: A Case Study on Task-Based Dialogue

Carl Strathearn and Dimitra Gkatzia

Abstract This paper argues that future dialogue systems must retrieve relevant information from multiple structured and unstructured data sources in order to generate natural and informative responses as well as exhibit commonsense capabilities and flexibility in dialogue management. To this end, we firstly review recent methods in document-grounded dialogue systems (DGDS) and commonsense-enhanced dialogue systems and then demonstrate how these techniques can be combined in a unified, commonsense-enhanced document-grounded dialogue system (CDGDS). As a case study, we use the `Task2Dial` dataset,[1] a newly collected dataset which contains instructional conversations between an information giver (IG) and information follower (IF) in the cooking domain. We then propose a novel architecture for commonsense-enhanced document-grounded conversational agents, demonstrating how to incorporate various sources to synergistically achieve new capabilities in dialogue systems. Finally, we discuss the implications of our work for future research in this area.

1 Introduction

Much of the work in dialogue systems has focused on developing task- and goal-oriented conversational agents that are capable of completing tasks, such as making restaurant reservations, ordering transport services and booking travel [1]. Traditionally, dialogue systems utilise domain-specific database schemas [2] and focus on slot-filling response generation. However, encoding all available information can be prohibitive in most domains, as the majority of domain knowledge exists in

[1] https://huggingface.co/datasets/cstrathe435/Task2Dial/tree/main.

C. Strathearn (✉) · D. Gkatzia
Edinburgh Napier University, Edinburgh, UK
e-mail: c.strathearn@napier.ac.uk; d.gkatzia@napier.ac.uk

M. Abbas (ed.), *Analysis and Application of Natural Language and Speech Processing*, Signals and Communication Technology,
https://doi.org/10.1007/978-3-031-11035-1_6

some unstructured format, such as documents [3]. DGDS can provide opportunities for dialogue systems that were not possible before, such as answering questions based on the information provided in documents and imitating the human capacity to possess background knowledge. Recent work on DGDS has focused on question-answering (Q&A) and machine reading comprehension. For instance, CoQA [4], a Q&A task between two interlocutors who have access to the same passage, requires the receiver to comprehend the passage in order to ask questions. Other tasks have focus on commonsense reasoning. For instance, QuAC [5] follows a similar setting as CoQA; however, only the receiver has access to the passage, and the questioner asks questions based on the title of the passage alone.

Here, we focus on `Task2Dial` [6], a new task for CDGDS, which aims at generating instructions grounded in a document so that the receiver of the instructions can complete a task. Task2Dial is similar to QuAC in that the information giver (IG) has access to the underlying document. However, Task2Dial differs from QuAC, because the information follower (IF) can ask questions for answers which are not grounded to a specific document, requiring commonsense capabilities by both IG and IF. Task2Dial requires following steps in a pre-specified order, invoking every day communication characteristics, such as asking for clarification, questions or advice, which may require the use of commonsense knowledge to answer. The proposed task differs from existing document-grounded tasks, as answers may require commonsense knowledge generated from the underlying information that may not be present in the document. Inspired by previous work on document-grounded dialogue [3, 7, 8], commonsense-enhanced natural language generation (NLG) [9, 10] and Q&A [4, 5], neural referring expression generation [11], concept acquisition [12], and task-based/instructional dialogue [13], we aim to capture two different types of knowledge: (1) document-level procedural context, i.e. what is the next step, and (2) commonsense, i.e. answering questions that are not available in the document, as demonstrated in Fig. 1. Our task is designed as an instruction-following scenario with an information giver (IG) and an information follower (IF), inspired partly by the GIVE challenge [13]. The IG has access to the recipe and gives instructions to the IF. The IG might choose to omit irrelevant information, simplify the content or provide it as is. The IF will either 'follow' the task by confirming that they have understood an instruction or ask for further information. For this, the IG may have to rely on information outside the given document; in other words, the IG will rely on their commonsense to enhance understanding and success of the task. To explore this, we propose a novel conversational agent, `ChefBot`, to structure and control the flow and type of information provided to the user from the documents. This requires a cumulative approach to formatting additional data from documents, i.e. structuring and extracting metadata to create additional knowledge databases that contain information such as the utility of objects and alternative ingredients, while retaining the underlying sequential structure of the instructional document [14].

The `Task2Dial` dataset introduces new challenges for dialogue systems: (1) generating instructions for task completion requires a flexible dialogue manager, as following specific steps in the form of a checklist might invoke discourse

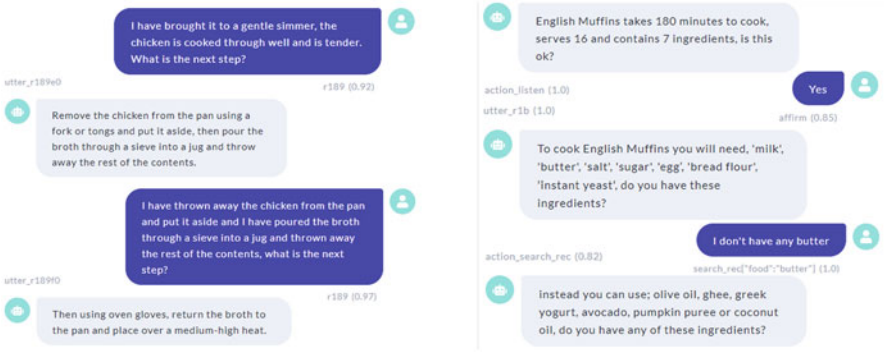

Fig. 1 Excerpt from dialogue showing the commonsense handling of hot objects using the Task2Dial dataset and ChefBot. Left: commonsense handling of objects. Right: swapping ingredients for appropriate alternatives using custom actions

phenomena not present in other dialogue styles, such as paraphrasing, as instructional responses may have been modified from the underlying document, and interlocutors may ask for clarification or alternative steps; (2) hence, this task requires commonsense knowledge, since questions may not necessarily be grounded in the document; (3) generating requires planning based on context, as task steps need to be provided in order; and finally (4) Task2Dial's human reference texts show more lexical richness and variation than other document-grounded dialogue datasets. The Task2Dial dataset contains dialogues with an average 18.15 number of turns and 19.79 tokens per turn, as compared to 12.94 and 12, respectively, in existing datasets. Therefore, developing a conversational agent based on this new task requires flexible dialogue management with global and domain-specific intents to enhance natural communication, custom actions to swap ingredients and explain unknown objects and rule-based state tracking for sequential and non-sequential information giving. For instance, it is not enough for the agent to just 'read' the next recipe instruction—the conversation might briefly diverge from the current plan to provide information about an ingredient replacement, and then it will have to correctly resume the previous conversation.

To this end, our contributions to the field can be summarised as follows:

- We propose a new task, Task2Dial, for commonsense-enhanced document-grounded dialogue.
- We present a novel dataset for commonsense-enhanced document-grounded dialogue.
- We propose a novel conversational agent architecture which considers how elements of the documents are represented within the dialogue manager, i.e. intents, utterances, entities and actions, and how the data is labelled to enable the system to follow the sequential logic of a given recipe while remaining flexible in terms of topic switch.

In the next section (Sect. 2), we refer to the related work. The proceeding sections cover the task formulation and data curation methodology (Sect. 6) and present an analysis of the Task2Dial dataset and a comparison to related datasets (Sect. 4). Finally, Sect. 5 proposes a novel conversational agent architecture for addressing the task of CDGDS, and finally, in Sect. 6, we discuss the implications and challenges for the development of instruction-giving dialogue systems for real-world tasks.

2 Related Work

The work presented in this paper focuses on the development of a CDGDS conversational agent for instruction-giving task-based dialogue, which is relevant to several areas of research on task- and goal-oriented dialogue, state tracking, document-grounded dialogue, commonsense reasoning and dialogue management. Next, we review each of these areas.

2.1 Task- and Goal-Oriented Dialogue

In dialogue management, task-oriented approaches focus on the successful completion of the individual stages of a task, towards achieving an end goal [15]. Comparatively, goal-oriented approaches focus on comparing the outcome or overall performance against a gold standard [16]. Task- and goal-oriented dialogue systems are common in domains such as booking and reservation systems for businesses [17]. However, business models are typically goal-oriented as the instructions are minimal and the focus is on the outcome [18]. Instead, the Task2Dial task is formulated as a task-oriented dialogue paradigm to imitate real-world practical scenarios that can vary in complexity and require adaptability, additional information, clarification and natural conversation in order to enhance understanding and success.

2.2 Dialogue State Tracking and Planning

Task-based dialogue systems require the user and artificial agent to work synergistically by following and reciting instructions to achieve a goal. Human-bot conversational models are defined as follows [19]:

- **Single intent and single turn policy:** relies solely on question and answer pairs assuming that the user provides all slot values in a single utterance. This type of task does not require dialogue state tracking.

- **Single intent and multi-turn policy:** extends the previous conversational model; however, this model can include multiple turns, to fill in missing information. Historic information is then extracted from all turns and used to structure data.
- **Multi-intent and multi-turn policy:** the intents can change depending on the context.

Instruction-giving scenarios follow the *multi-intent multi-turn* conversational framework, since they must accommodate knowledge and variability outside of a linear deterministic model as practical tasks can vary in complexity and the conversation can vary based on the interlocutors' prior knowledge and experience. In addition, there is no restriction on the amount of variability introduced into a task, such as introducing alternate methods, commonsense knowledge and concepts that change the structure and information within the dialogue. Variability is often reduced in human-machine scenarios as systems are limited in knowledge and their ability to respond to questions not seen in training [20], which can result in shortened responses and fewer questions asked on aspects of the task [21]. This reduces the system's ability to ensure that the IF has understood the IG's directions, which may produce irregular outcomes or result in an incomplete task. Therefore, capturing and emulating natural variability within the dialogue is crucial for creating robust and reliable conversational systems for instruction-giving scenarios.

Existing datasets such as the Multi-Domain Wizard-of-Oz (MultiWOZ) [22], Taskmaster-1 [21], Doc2dial [3] and Action-Based Conversations Dataset (ABCD) [23] strictly follow the sequential logic of an instructional document. However, in addition to grounded information in documents, Task2Dial aims to accommodate questions and clarification on different aspects of a task that might not be grounded in the document. In previous work, the user is limited to the path of the subroutine; however, in Task2Dial, the IF can ask the IG questions at any stage of the task, regardless of the position within a given sequence, and then return to that position after the question is fulfilled. For example, in a cooking scenario, the IF may ask the IG how to use a certain kitchen utensil. The IG would need to answer this question and then return to the correct stage in the recipe in order to continue the sequence. This introduces additional challenges for state tracking. The conversational agent must not only generate instructions sequentially, based on the schema of a document, but also request confirmation to ensure that the user has understood the task and answer questions outside its predefined script. Using document-grounded subroutines to capture intents that change the direction of a task broadens the interaction between the IG and IF [23] and introduces new challenges for dialogue state tracking.

2.3 Document-Grounded Dialogue

DGDS classify unstructured, semi-structured and structured information in documents to aid in understanding human knowledge and interactions, creating greater

naturalistic human-computer interactions (HCI) [24]. The aim of DGDS is to formulate a mode of conversation from the information (utterances, turns, context, clarification) provided in a document(s) [25]. DGDS are particularly useful in task-oriented and goal-oriented scenarios as they emulate the natural dialogue flow between the IG and IF. A recent example of DGDS and closest to our work is Doc2Dial, a multi-domain DGDS dataset for goal-oriented dialogue modelled on hypothetical dialogue scenes (dialogue act, a role such as user or agent and a piece of grounding content from a document) and dialogue flows (a sequence of dialogue scenes) to simulate realistic interactions between a user and machine agent in information seeking settings [3]. DoQA [26] contains domain-specific Q&A dialogues in three domains including cooking, where users can ask for recommendations/instructions regarding a specific task, although the task does not involve providing steps for completing it. Other document-grounded tasks have been proposed such as MultiWOZ [22], Taskmaster-1 [21] and ABCD [23] which demonstrate how DGDS can be configured in end-to-end pipelines for task-driven dialogue in virtual applications such as online booking systems. Here, we follow a similar setup as Doc2Dial; however, in our proposed task, we allow users to ask clarification questions, the answers to which are not necessarily grounded in the document. This consideration is vital in the development of instruction-giving conversational agents as it has implications for the dialogue pipeline.

2.4 Commonsense-Enhanced Dialogue

Commonsense reasoning is a general understanding of our surroundings, situations and objects, which is essential for many AI applications [27]. Simulating these perceptual processes in task- and goal-oriented DGDS generates greater context and grounding for more human-like comprehension. An example of commonsense dialogue in a practical task-based scenario is understanding the common storage locations of objects or the safe handling and use of objects from their common attributes, i.e. a handle, knob or grip. Commonsense dialogue is highly contextual: in Question Answering in Context (QuAC) [5], dialogues are constructed from Wikipedia articles interpreted by a teacher. A student is given the title of the article and asks the teacher questions on the subject from prior knowledge, and the teacher responds to the students' questions using the information in the document. This mode of question answering (Q&A) development is more naturalistic and grounded than previous methods as the challenges of understanding the information are ingrained in the dialogue from the underlying context. Similarly, the Conversational Question Answering Challenge (CoQA) dataset [4] is formulated on a rationale, scenario and conversation topic, and the Q&A pairs are extracted from this data. This methodology is used in the Task2Dial dataset as it provides greater co-reference and pragmatic reasoning within the dialogue for enhanced comprehension as shown in Fig. 1.

In human-human IG/IF tasks, the IG may have prior knowledge of appropriate alternative methods, components and tools that can be used in a task that are not mentioned in the instructions. This information is vital if the IF has missing components or requires clarification on aspects of the task that are not clearly represented in the document. Variability is problematic to capture in DGDS alone as hypothetical scenarios in documents cannot account for all the potential issues in practice [28]. Thus, the ability to ask questions that are not available in the document is crucial when conducting real-world tasks due to the changeable conditions, complexity of the task and availability of components. This is particularly important in cooking tasks (as well as other instruction-giving tasks) as the user may not have all the ingredients stated in a recipe but may have access to alternative items that can be used instead. This approach can also be used in other domains such as maintenance or construction tasks if the user does not have a specific tool but has access to a suitable alternative tool without knowing it. This inevitably introduces new challenges for dialogue systems as commonsense-related intents and actions need to be introduced in the dialogue system. Task2Dial moves away from the closed knowledge base(s) in DGDS into incorporating multiple sources of information to broaden the adaptability and application of DGDS. This is achieved by developing additional resources that list alternative ingredients to those mentioned in the metadata from the original recipes, as well as instructions on how to use cookery tools. Appropriate alternative ingredients were collected and verified using certified online cooking resources that provide food alternatives.

2.5 Dialogue Management

Dialogue managers are used to structure data and control the flow of a conversation and the way in which information is delivered to the user [29]. There are numerous DM tools for DGDS; however, it is important to consider the structure of the dataset and the complexity of the task [30]. Due to the complexity of our cooking scenario, the DM must be able to read multiple documents, intents, state tracking, paths, entities, rules and actions to generate responses logically and coherently [14]. The ability to deploy a DM on different platforms, channels and servers is also an important consideration for accessibility, usability, data protection and security [31]. Open-source DM tools such as RASA X[2] are particularly useful for task-based dialogue as the natural language understanding and core dialogue manager libraries are highly configurable for different tasks [32]. This is an important consideration for handling structured and unstructured data; flexibility in dialogue management, i.e. customisation of features; configuring classifiers; interpreter pipelines for training; conversation history; and managing interaction. This cannot be achieved

[2] rasa.com/docs/rasa-x/.

with DM tools such as Amazon Lex[3] and Google Dialogflow[4] due to system limitations and restricted user access [33, 34].

3 Task2Dial

The proposed task considers the recipe-following scenario with an information giver (IG) and an information follower (IF), where the IG has access to the recipe and gives instructions to the IF. The IG might choose to omit irrelevant information, simplify the content of a recipe or provide it as is. The IF will either follow the task or ask for further information. The IG might have to rely on information outside the given document (i.e. commonsense) to enhance understanding and success of the task. In addition, the IG decides on how to present the recipe steps, i.e. split them into sub-steps or merge them together, often diverting from the original number of recipe steps. The task is regarded successful when the IG has successfully followed/understood the recipe. Hence, other dialogue-focused metrics, such as the number of turns, are not appropriate here. Formally, *Task2Dial* can be defined as follows: given a recipe R_i from $R = R_1, R_2, R_3, \ldots, R_n$, an ontology or ontologies $O_i = O_1 1, O_2, \ldots, O_n$ of cooking-related concepts, a history of the conversation h, predict the response r of the IG.

The `Task2Dial` dataset includes (1) a set of recipe documents and (2) conversations between an IG and an IF, which are grounded in the associated recipe documents. Figure 2 presents sample utterances from a dialogue along with the associated recipe. It demonstrates some important features of the dataset, such as mentioning entities not present in the recipe document, re-composition of the

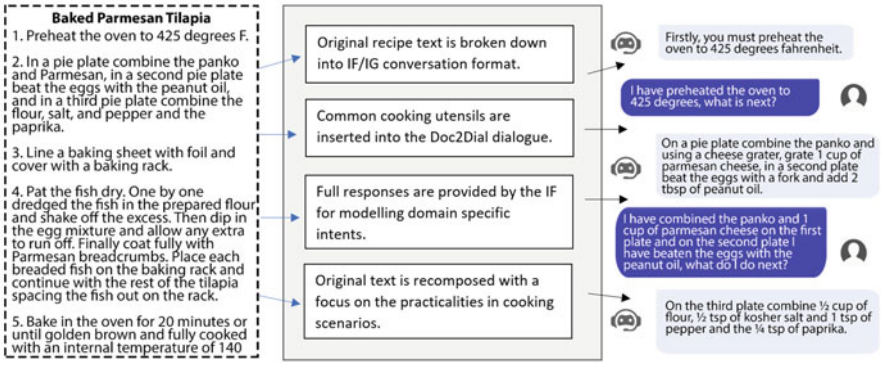

Fig. 2 Original recipe text converted to Task2Dial dialogue

[3] https://aws.amazon.com/lex/.

[4] https://cloud.google.com/dialogflow.

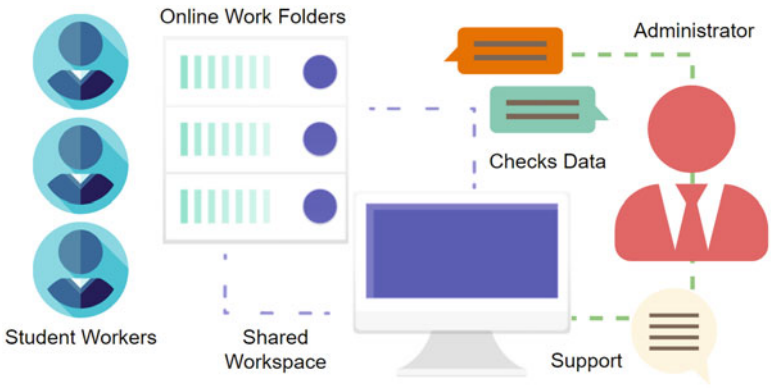

Fig. 3 Overview of the Task2Dial dataset collection

original text to focus on the important steps and the breakdown of the recipe into manageable and appropriate steps. Following recent efforts in the field to standardise NLG research [35], we have made the dataset freely available.[5]

3.1 Data Collection Methodology

The overall data collection methodology is shown in Fig. 3 and is described in detail below.

Pilot Data Collection Prior to data collection, we performed three pilot studies. In the first, two participants assumed the roles of IG and IF, respectively, where the IG had access to a recipe and provided recipe instructions to the IF (who did not have access to the recipe) over the phone, recording the session and then transcribing it. Next, we repeated the process with text-based dialogue through an online platform following a similar setup; however, the interaction was solely chat-based. The final study used *self-dialogue* [21], with one member of the team who wrote the entire dialogues assuming both the IF and IG roles. We found that self-dialogue results were proximal to the results of two-person studies. However, time and cost were higher for producing two-person dialogues, with additional time needed for transcribing and correction; thus, we opted to use self-dialogue.

Creation of a Recipe Dataset Three open-source and creative commons licensed cookery websites[6] were identified for data extraction, which permit any use or non-commercial use of data for research purposes [36, 37]. As content submission to the

[5] www.huggingface.co/datasets/cstrathe435/Task2Dial.

[6] (a) www.makebetterfood.com, (b) www.cookeatshare.com and (c) www.bbcgoodfood.com.

cooking websites was unrestricted, data appropriateness was ratified by the ratings and reviews given to each recipe by the public, and highly rated recipes with positive feedback were given preference over recipes with low scores and poor reviews [38]. From this, a list of 353 recipes was compiled and divided amongst the annotators for the data collection. As mentioned earlier, annotators were asked to take on the roles of both IF and IG, rather than a multi-turn WoZ approach, to allow flexibility in the utterances. This approach allowed the annotators additional time to formulate detailed and concise responses.

Participants Research assistants (RAs) from the School of Computing were employed on temporary contracts to construct and format the dataset. After an initial meeting to discuss the job role and determine suitability, the RAs were asked to complete a paid trial, and this was evaluated, and further advice was given on how to write dialogues and format the data to ensure high quality. After the successful completion of the trial, the RAs were permitted to continue with the remainder of the data collection. To ensure high quality of the dataset, samples of the dialogues were often reviewed, and further feedback was provided.

Instructions to Annotators Each annotator was provided with a detailed list of instructions, an example dialogue and an IF/IG template (see Appendix A). The annotators were asked to read both the example dialogue and the original recipe to understand the text, context, composition, translation and annotation. The instructions included information handling and storage of data, text formatting, metadata and examples of high-quality and poor dialogues. An administrator was on hand throughout the data collection to support and guide the annotators. This approach reduced the amount of low-quality dialogues associated with large crowdsourcing platforms that are often discarded post evaluation, as demonstrated in the data collection of the Doc2Dial dataset [3].

Time Scale The data collection was scheduled over 4 weeks. This was to permit additional time for the annotators to conduct work and study outside of the project. Unlike crowdsourcing methods, the annotators were given the option to work on the project flexibly in their spare time and not commit to a specific work pattern or time schedule.

Ethics An ethics request was submitted for review by the board of ethics at our university. No personal or other data that may be used to identify an individual was collected in this study.

Task2Dial Long-Form Description Unlike previous task- and goal-oriented DGDS, the Task2Dial corpus is unique as it is configured for practical IF/IG scenarios as demonstrated in Fig. 2. Following [39], we provide a long-form description of the Task2Dial cooking dataset here.

Curation Rationale Text selection was dependent on the quality of information provided in the existing recipes. Too little information and the transcription and interpretation of the text became diffused with missing or incorrect knowledge. Conversely, providing too much information in the text resulted in a lack of

creativity and commonsense reasoning by the data curators. Thus, the goal of the curation was to identify text that contained all the relevant information to complete the cooking task (tools, ingredients, weights, timings, servings) but not in such detail that it subtracted from the creativity, commonsense and imagination of the annotators.

Language Variety The recipes selected for this dataset were either written in English or translated into English prior to data collection for ease of the annotators, language understanding and future training for language models. This made the dataset accessible to all contributors involved in the curation, support and administration framework.

Speaker Demographics The recipes are composed by people of different race/ethnicity, nationalities, socioeconomic status, abilities, age, gender and language with significant variation in pronunciations, structure, language and grammar. This provided the annotators with unique linguistic content for each recipe to interpret the data and configure the text into an IF/IG format. To help preserve sociolinguistic patterns in speech, the data curators retained the underlying language when paraphrasing, to intercede social and regional dialects with their own interpretation of the data to enhance lexical richness [40].

Annotator(s) Demographics Undergraduate RAs were recruited through email. The participants were paid an hourly rate based on a university pay scale which is above the living wage and corresponds to the real living wage, following ethical guidelines for responsible innovation [41]. The annotation team was composed of two male and one female data curators, under the age of 25 years of mixed ethnicity with experience in AI and computing. This minimised the gender bias that is frequently observed in crowdsourcing platforms [42].

Speech Situation The annotators were given equal workloads, although workloads were adjusted accordingly over time per annotator availability to maximise data collection. The linguistic modality of the dialogue is semi-structured, synchronous interactions as existing recipes were used to paraphrase the instructions for the IG. Following this, the IF responses were created spontaneously following the logical path of the recipe in the context of the task. The intended audience for the Task2Dial dataset is broad, catering for people of different ages and abilities. Thus, the dataset is written in plain English with no jargon or unnecessary commentary to maximise accessibility.

Text Characteristics The structural characteristics of the Task2Dial dataset are influenced by real-world cooking scenarios that provide genre, texture and structure to the dialogues. This provides two important classifications, utterances and intents, that are universal for all task-based datasets and domain-specific text that is only relevant for certain tasks. This data is used when training language models as non-domain-specific sample utterances such as 'I have completed this step' can be used to speed up the development of future task-based DGDS.

Recording Quality As mentioned previously, the dialogues in Task2Dial are text-based.

4 Dataset Analysis

This section presents an overall statistics of the Task2Dial dataset. We compare our dataset to the Doc2Dial dataset, although the latter focuses on a different domain. Employing research assistants to collect and annotate data rather than using crowdsourcing platforms meant that no dialogues were discounted from the dataset. However, a pre-evaluation check was performed on the dataset before statistical analysis to reduce spelling and grammatical errors that may affect the results of the lexical analysis.

Size Table 1 summarises the main descriptive statistics of Task2Dial and Doc2Dial. The dialogues in Task2Dial contain a significantly higher number of turns than Doc2Dial dialogues (18.15 as opposed to 12.94). In addition, Task2Dial utterances are significantly longer than in Doc2Dial, containing on average more than seven tokens.

Lexical Richness and Variation We further report on the lexical richness and variation [43], following [44] and [45]. We compute both type-token ratio (TTR), i.e. the ratio of the number of word types to the number of words in a text, and the mean segmental TTR (MSTTR), which is computed by dividing the corpus into successive segments of a given length and then calculating the average TTR of all segments to account for the fact the compared datasets are not of equal size.[7] All results are shown in Table 1. We further investigate the distribution of the top 25 most frequent bigrams and trigrams in our dataset as seen in Fig. 4. The majority of both trigrams (75%) and bigrams (59%) is only used once in the dataset, which creates a challenge to efficiently train on this data. For comparison, in Doc2Dial, 54% of bigrams and 70% of trigrams are used only once. Infrequent words and phrases pose a challenge for the development of data-driven dialogue systems as handling out-of-vocabulary words is a bottleneck.

Table 1 Size and lexical richness of the dataset

Dataset	#docs	#Turns	#Tkns/turn	TTR	MSTTR
TASK2DIAL	353	**18.15**	**19.79**	**0.025**	0.84
DOC2DIAL	487	12.94	12	0.011	0.86

[7] TTR and MSTTR have been computed using https://github.com/LSYS/LexicalRichness.

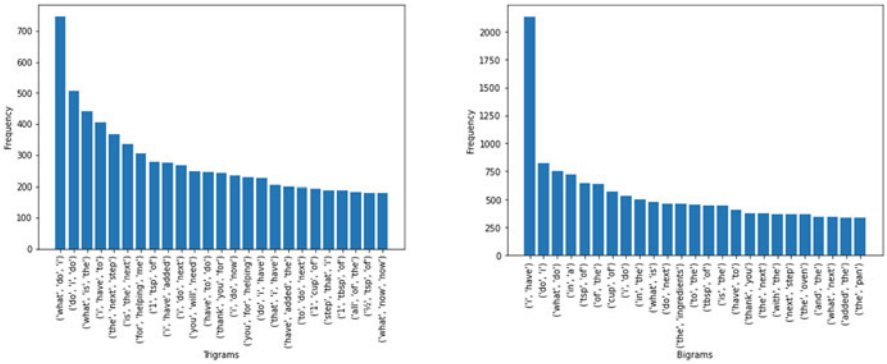

Fig. 4 Frequencies of trigrams and bigrams in the Task2Dial dataset

5 The ChefBot Conversational Agent

ChefBot was created using the RASA X dialogue manager[8] to control the dialogue flow and access external databases for swapping ingredients, object explanations, intents, utterances and entities modelled from the dialogues in the Task2Dial dataset, as shown in Fig. 5.

ChefBot System Architecture The system architecture for ChefBot is depicted in Fig. 6, and the technical details of the system are described in this section. The data folder contains the files that the ChefBot is trained on, and these include the IF dialogues and recipe sequences from the Task2Dial dataset. The rules file contains the directives for intents, paths and state tracking. The actions folder holds the entities files which are the external datasets and rules for alternative ingredients and object explanations. When a model is trained, it is stored in the models folder. Similarly, if a path is changed or corrected during a session, i.e. using RASA interactive, it is stored in the test folder. The domain file contains the IG dialogues from the Task2Dial dataset configured into utterances. This file also contains the classifications for the intents, entities and actions. The credential's file contains the parameters for deploying the system on channels and servers. Similarly, the endpoints file is the data for the custom actions server for entity extraction. The config file is the interpreter pipeline for the NLU model that includes the classifiers and policies for training the ChefBot. When a trained model is loaded into a terminal (such as Anaconda[9] or similar), it can be deployed using the RASA shell or RASA X commands to load the RASA user interface (UI) on a channel or server, allowing the user to interact with ChefBot.

[8] rasa.com/blog/dialogue-policies-rasa-2/.

[9] www.anaconda.com/.

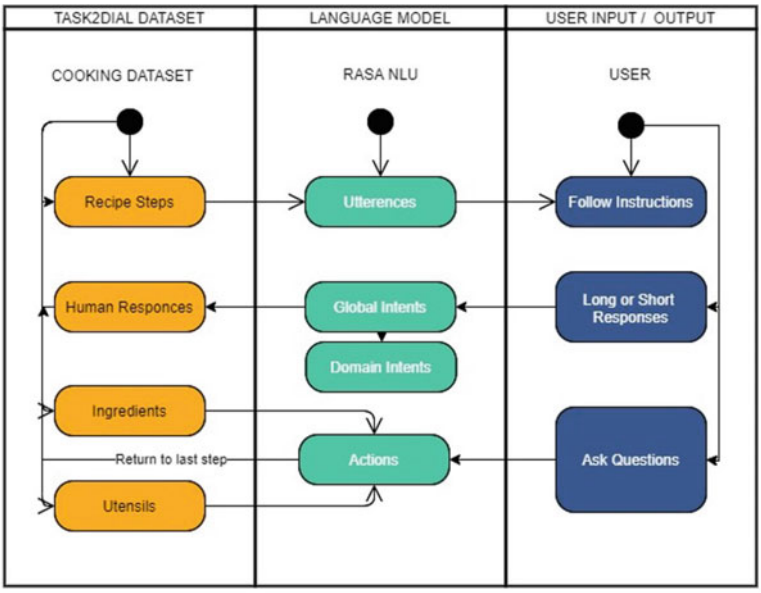

Fig. 5 The pipeline from Task2Dial to ChefBot and the user

Data Entry and Formatting Data entry was conducted over a 6-week period by project members. Due to the restructuring of data, manual entry was the most effective way to ensure data from the Task2Dial dataset was formatted and entered correctly in ChefBot. All 353 recipe documents, alternative food and object databases from the Task2Dial dataset were successfully uploaded into ChefBot within the designated 6-week period.

Modelling Intents and Utterances ChefBot uses non-domain-specific 'user' responses from the Task2Dial dataset to model *global intents* in the dialogue manager, such as 'I have done this', 'OK what's next' and 'What is the next step'. These global intents can be used in other task-based dialogue scenarios, such as cleaning and maintenance tasks. *Domain-specific intents* are modelled from the user responses which contain information that is only relevant to the cooking domain. For example, 'I have put the cake in the oven' or 'I have mixed the ingredients in a bowl'. This approach is important for enhancing natural communication between the IG and IF as it allows the IF to give both short and full responses to the IG, proximal to a genuine human conversation. Within the domain file, the instructions from the IG were turned into utterances and numerically labelled depending on the position of the instruction within the sequence of a recipe, i.e. r1a, r1b, r1c, etc., as shown in the example below. This approach creates a sequential order for each recipe which can be tracked in the DM. This data is used both for state tracking and creating a dialogue pathway for each recipe.

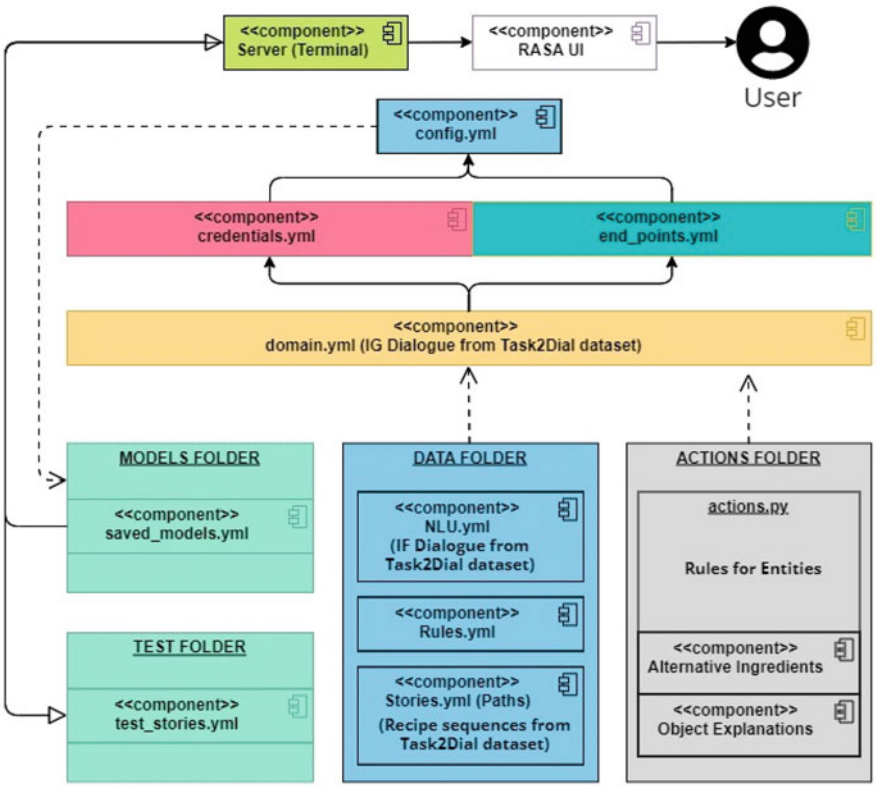

Fig. 6 ChefBot system architecture

Example of Modelling User Utterances in ChefBot
utter_r1a:—text: English Muffins takes 180 minutes to cook, serves 16 and contains 7 ingredients, is this ok?
utter_r1b:—text: To cook English Muffins you will need, 'milk', 'butter', 'salt', 'sugar', 'egg', 'bread flour', 'instant yeast', do you have these ingredients?
utter_r1c:—text: To start with combine 1 and three quarter cups of lukewarm milk, 3 tablespoons of soft butter, one and a half teaspoons of salt, 2 tablespoons of sugar, one egg, 5 cups of bread flour, and 2 teaspoons of yeast in a large mixing bowl of an electric stand mixer.

Modelling Dialogue Paths and Conversation History Within each pathway are the global and domain-specific intents for each recipe that are activated using the 'or' variable in multi-intent, multi-turn policy, as outlined in the literature review. This information is important for training the DM to determine the next logical step in a sequence from the history of the conversation and the path. Custom actions for alternative ingredients activate if the user answers 'no' to 'do you have all the ingredients?'. This initiates the search_rec function and lists the alternative

ingredients for each recipe. The paths are modelled using the IG and IF sequences from the Task2Dial dataset, as demonstrated below.

Example of Recipe Path in ChefBot

- story:
  strawberrypienopath //Name of path
  steps:
- intent: strawberrypie //Name of recipe from Task2Dial dataset
- action: utter_r2a //First line of the recipe sequence
- or:
- intent: globint //Global intents
- intent: r2 //Domain/recipe specific intents
- action: utter_r2b //Second line of the recipe sequence
- intent: nomode
- action: utter_ingredients_strawberrypie //Identify missing ingredient's
- intent: search_rec_r2a // Perform a search for alternative ingredients

Rule-Based Tracking and Entity Extraction The frequently asked questions (FAQ) rule in RASA allows the user to ask questions that may not be represented within a given form or path. The DM will then answer the question using specific 'FAQ' labelled intents and then return to the next or previous step in a sequence, as shown below. The FAQ labelling facility can also be used to create a list of intents for context-aware entity extraction, i.e. 'how do I use a' with entities [cheese knife] (utensil) within a given FAQ function. This method is less formulaic than using RASA forms, which requires specified slots to be filled at each stage of a sequence or sub-sequence, which is important in ChefBot as we aim to capture the natural flow of conversation between the IG and IF from the Task2Dial dataset, to enhance user understanding and accessibility.

Rule-Based Tracking Example

- rule: respond to IF questions
  steps:
- intent: utter_faq_questions
- action: search_utensils

External Databases for Alternative Ingredients and Object Descriptions In ChefBot, additional commonsense knowledge is modelled in two external databases. The first is the ability to swap ingredients for appropriate alternatives. It is important that the alternative ingredients do not alter the procedural context of the recipe. For example, swapping olive oil for sunflower oil will not change how a recipe is prepared or cooked. Conversely, changing chicken breast for beef fillet would require a significant change in the recipe instructions. This would have an impact on the cooking situation, including times, food preparation, servings, steps and utensils, and may require additional ingredients for cooking or preparation. Therefore, to avoid unnecessary complications, all alternative ingredients must not significantly affect the sequence and instructions within a given recipe.

Utensil Explanations

```
def name(self) -> Text:
        return "action_utensilexplan"
            if e['entity'] =='utensils':
if name == "bowl":
                message = "A cooking
container that is usually larger than a cup
and kept in a kitchen cupboard, it is
typically made from glass, ceramic,
plastic."
if name == "spoon":
                message = "an eating or
cooking utensil consisting of a small
shallow bowl with a relatively long handle
made from metal, plastic, wood it is usually
kept in a kitchen drawer."

if name == "sieve":
                message = "A metal or plastic
device with a wire mesh or perforations
through which finer particles of a mixture
are sifted, it is usually kept in a kitchen
and stored in a cupboard."
```

Alternative Ingredients

```
 r1: Better English Muffins ['milk', 'butter',
'salt', 'sugar', 'egg', 'bread flour',
'instant yeast']


 if name == "plain flour":

            message ="instead you can use; oat
flour, bread flour, cake flour or coconut
flour"
        if name == "olive oil":
            message ="instead you can use;
peanut oil, walnut oil, sunflower oil, canola
oil or vegetable oil"

        if name == "salt":
            message ="instead you can use;
mint, rosemary, nutmeg, basil, cardamon,
chili, cinnamon or chives"
```

Fig. 7 Examples of how the additional datasets were handled as custom actions in ChefBot. Left: utensil explanations. Right: alternative ingredients

Metadata containing information on the ingredients and utensils used in each recipe from the Task2Dial dataset was extracted. The first dataset was created using the list of ingredients from each recipe. A Google search using cooking and food health websites was performed to find appropriate alternatives for each ingredient. Similarly, a list of cooking utensils and kitchen devices was constructed using the same approach. However, the second dataset also contains object descriptions, object comparisons, alternative names for objects, appropriate handling methods and common storage locations. This data is important as it may not be grounded in the original documents, but vital for enhancing user understanding. This approach allows the IG to simplify the content or provide additional information depending on the needs of the IF. The two datasets are transformed into custom actions in the dialogue manager as shown in Fig. 7.

Using these databases as custom actions allows the user to trigger an action at any stage of the task from keyword recognition. For instance, in Fig. 6, the keyword or entity extraction is the names of the ingredients and objects. In the intent list, these entities are given context, for example, 'how do I use a (fish slice) [object-name]' or 'what does a (lemon zester) [object-name] look like'. This is important as the user's response may consist of more than one named entity. For instance, 'I do not know where my (fish slice) is kept or what a (lemon zester) looks like'. Here context awareness is important for relaying information back to the user in a meaningful way. This was achieved by using the multi-intent function in rasa to handle more than one intent per turn.

ChefBot Demo and Repository Training ChefBot takes approximately 2–3 hours, so a trained model is supplied in a GitHub repository within the system files for

ease of demonstration.[10] A description of the libraries and system requirements needed to run ChefBot are located in the 'requirements.txt' file. The provided video demonstrates how the ChefBot generates dialogue, swaps ingredients, uses global and domain-specific intents and explains the utility of objects and state tracking, using a random recipe selected from the Task2Dial dataset.[11]

6 Conclusions and Future Work

This paper demonstrates how commonsense-enhanced document-grounded dialogue can be modelled for task-based dialogue. As a case study, we used the Task2Dial, a task-based document-grounded conversation dataset, modelled as an interaction between an IG and an IF during a cooking task. In this domain, commonsense is the ability to provide alternative ingredients and provide recommendations on object utility, both of which are not present in the cooking instruction dialogues and require additional knowledge in the form of a database or domain ontology. We then presented a novel conversational agent architecture, ChefBot, which is able to flexibly adapt to the changes in dialogue flow. With this research, we extend previous work in DGDS in order to emulate the unpredictability of human-human conversations in instruction-giving tasks that do not necessarily follow a tight schema as the sequential structure of instructional documents. Instead, other discourse and dialogue phenomena might take place such as clarification questions and explanations. We further considered the aforementioned challenges of modelling dialogue for instruction-giving tasks with a focus on state tracking, task planning and commonsense reasoning and proposed a new task, model and associated dataset. With this, we demonstrate a more robust approach for DGDS called CDGDS to more effectively handle real-world task-based scenarios and open the door to tasks outside the cooking domain, such as general maintenance and furniture assembly.

6.1 Future Work and Open Questions

Our proposed task aims to motivate research for modern dialogue systems that address the following challenges. Firstly, modern dialogue systems should be flexible and allow for 'off-script' scenarios in order to emulate real-world phenomena, such as the ones present in human-human communication. This will require new ways of encoding user intents and new approaches to dialogue management in general. Secondly, as dialogue systems find different domain applications, the

[10] github.com/carlstrath/ChefBot.

[11] https://youtu.be/XoTXraGs5rA.

complexity of the dialogues might increase as well as the reliance of domain knowledge that can be encoded in structured or unstructured ways, such as documents, databases, etc. Many applications might require access to different domain knowledge sources in a course of a dialogue, and as such, context selection might prove beneficial in choosing 'what to say' [46]. Finally, as we design more complex dialogue systems, commonsense will play an essential part, with models required to perform reasoning with background commonsense knowledge, and generalise to tackle unseen concepts, similar to [9]. In the future, we aim to benchmark and evaluate a dialogue system based on the Task2Dial dataset and the ChefBot [14] and extend this approach to a human-robot interaction (HRI) scenario. Other interesting directions can include the exploration of pre-trained models as part of a conversational agent architecture to eliminate the need to encode knowledge or design domain ontologies [47].

Acknowledgments The research is supported under the EPSRC projects CiViL (EP/T014598/1) and NLG for low-resource domains (EP/T024917/1).

References

1. Chen, H., Liu, X., Yin, D., Tang, J.: A survey on dialogue systems: Recent advances and new frontiers. SIGKDD Explor. Newsl. **19**(2), 25–35 (2017). https://doi.org/10.1145/3166054.3166058
2. Shah, P., Hakkani-Tür, D., Liu, B., Tür, G.: Bootstrapping a neural conversational agent with dialogue self-play, crowdsourcing and on-line reinforcement learning. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 3 (Industry Papers), pp. 41–51. Association for Computational Linguistics, New Orleans - Louisiana (2018). https://doi.org/10.18653/v1/N18-3006. https://www.aclweb.org/anthology/N18-3006
3. Feng, S., Wan, H., Gunasekara, C., Patel, S., Joshi, S., Lastras, L.: doc2dial: A goal-oriented document-grounded dialogue dataset. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 8118–8128. Association for Computational Linguistics, Online (2020). https://doi.org/10.18653/v1/2020.emnlp-main.652. https://www.aclweb.org/anthology/2020.emnlp-main.652
4. Reddy, S., Chen, D., Manning, C.D.: CoQA: A conversational question answering challenge. Trans. Assoc. Comput. Linguist. **7**, 249–266 (2019). https://doi.org/10.1162/tacl_a_00266. https://aclanthology.org/Q19-1016
5. Choi, E., He, H., Iyyer, M., Yatskar, M., tau Yih, W., Choi, Y., Liang, P., Zettlemoyer, L.: Quac: Question answering in context (2018)
6. Strathearn, C., Gkatzia, D.: The Task2Dial dataset: A novel dataset for commonsense-enhanced task-based dialogue grounded in documents. In: Proceedings of The Fourth International Conference on Natural Language and Speech Processing (ICNLSP 2021), pp. 242–251. Association for Computational Linguistics, Trento, Italy (2021). https://aclanthology.org/2021.icnlsp-1.28
7. Hu, Z., Dick, M., Chang, C.N., Bowden, K., Neff, M., Fox Tree, J., Walker, M.: A corpus of gesture-annotated dialogues for monologue-to-dialogue generation from personal narratives. In: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), pp. 3447–3454. European Language Resources Association (ELRA), Portorož, Slovenia (2016). https://aclanthology.org/L16-1550

8. Stoyanchev, S., Piwek, P.: Constructing the CODA corpus: A parallel corpus of monologues and expository dialogues. In: Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10). European Language Resources Association (ELRA), Valletta, Malta (2010). http://www.lrec-conf.org/proceedings/lrec2010/pdf/127_Paper.pdf
9. Lin, B.Y., Zhou, W., Shen, M., Zhou, P., Bhagavatula, C., Choi, Y., Ren, X.: CommonGen: A constrained text generation challenge for generative commonsense reasoning. In: Findings of the Association for Computational Linguistics: EMNLP 2020, pp. 1823–1840. Association for Computational Linguistics, Online (2020). https://doi.org/10.18653/v1/2020.findings-emnlp.165. https://aclanthology.org/2020.findings-emnlp.165
10. Clinciu, M.A., Gkatzia, D., Mahamood, S.: It's commonsense, isn't it? demystifying human evaluations in commonsense-enhanced NLG systems. In: Proceedings of the Workshop on Human Evaluation of NLP Systems (HumEval), pp. 1–12. Association for Computational Linguistics, Online (2021). https://aclanthology.org/2021.humeval-1.1
11. Panagiaris, N., Hart, E., Gkatzia, D.: Generating unambiguous and diverse referring expressions. Comput. Speech Lang. **68**, 101184 (2021). https://doi.org/10.1016/j.csl.2020.101184. https://www.sciencedirect.com/science/article/pii/S0885230820301170
12. Gkatzia, D., Belvedere, F.: "what's this?" comparing active learning strategies for concept acquisition in hri. In: Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction, HRI '21 Companion, p. 205–209. Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3434074.3447160
13. Gargett, A., Garoufi, K., Koller, A., Striegnitz, K.: The GIVE-2 corpus of giving instructions in virtual environments. In: Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10). European Language Resources Association (ELRA), Valletta, Malta (2010). http://www.lrec-conf.org/proceedings/lrec2010/pdf/532_Paper.pdf
14. Strathearn, C., Gkatzia, D.: Chefbot: A novel framework for the generation of commonsense-enhanced responses for task-based dialogue systems. In: Proceedings of the 14th International Conference on Natural Language Generation, pp. 46–47. Association for Computational Linguistics, Aberdeen, Scotland, UK (2021). https://aclanthology.org/2021.inlg-1.5
15. Hosseini-Asl, E., McCann, B., Wu, C.S., Yavuz, S., Socher, R.: A simple language model for task-oriented dialogue. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M.F., Lin, H. (eds.) Advances in Neural Information Processing Systems, vol. 33, pp. 20179–20191. Curran Associates, Inc. (2020). https://proceedings.neurips.cc/paper/2020/file/e946209592563be0f01c844ab2170f0c-Paper.pdf
16. Ham, D., Lee, J.G., Jang, Y., Kim, K.E.: End-to-end neural pipeline for goal-oriented dialogue systems using GPT-2. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 583–592. Association for Computational Linguistics, Online (2020). https://doi.org/10.18653/v1/2020.acl-main.54. https://aclanthology.org/2020.acl-main.54
17. Zhang, Z., Takanobu, R., Huang, M., Zhu, X.: Recent advances and challenges in task-oriented dialog system. CoRR **abs/2003.07490** (2020). https://arxiv.org/abs/2003.07490
18. Ilievski, V., Musat, C., Hossmann, A., Baeriswyl, M.: Goal-oriented chatbot dialog management bootstrapping with transfer learning. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence, IJCAI'18, p. 4115–4121. AAAI Press (2018)
19. Zamanirad, S., Benatallah, B., Rodriguez, C., Yaghoubzadehfard, M., Bouguelia, S., Brabra, H.: State machine based human-bot conversation model and services. In: Dustdar, S., Yu, E., Salinesi, C., Rieu, D., Pant, V. (eds.) Advanced Information Systems Engineering, pp. 199–214. Springer International Publishing, Cham (2020)
20. Shum, H.Y., He, X., Li, D.: From eliza to xiaoice: Challenges and opportunities with social chatbots (2018)
21. Byrne, B., Krishnamoorthi, K., Sankar, C., Neelakantan, A., Duckworth, D., Yavuz, S., Goodrich, B., Dubey, A., Cedilnik, A., Kim, K.: Taskmaster-1: Toward a realistic and diverse dialog dataset. CoRR **abs/1909.05358** (2019). http://arxiv.org/abs/1909.05358

22. Budzianowski, P., Wen, T.H., Tseng, B.H., Casanueva, I., Ultes, S., Ramadan, O., Gašić, M.: MultiWOZ - a large-scale multi-domain Wizard-of-Oz dataset for task-oriented dialogue modelling. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 5016–5026. Association for Computational Linguistics, Brussels, Belgium (2018). https://doi.org/10.18653/v1/D18-1547. https://www.aclweb.org/anthology/D18-1547
23. Chen, D., Chen, H., Yang, Y., Lin, A., Yu, Z.: Action-based conversations dataset: A corpus for building more in-depth task-oriented dialogue systems. In: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 3002–3017. Association for Computational Linguistics, Online (2021). https://www.aclweb.org/anthology/2021.naacl-main.239
24. Zhou, K., Prabhumoye, S., Black, A.W.: A dataset for document grounded conversations. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 708–713. Association for Computational Linguistics, Brussels, Belgium (2018). https://doi.org/10.18653/v1/D18-1076. https://aclanthology.org/D18-1076
25. Ma, L., Zhang, W., Li, M., Liu, T.: A survey of document grounded dialogue systems (DGDS). CoRR **abs/2004.13818** (2020). https://arxiv.org/abs/2004.13818
26. Campos, J.A., Otegi, A., Soroa, A., Deriu, J., Cieliebak, M., Agirre, E.: Doqa—accessing domain-specific faqs via conversational qa (2020)
27. Ilievski, F., Oltramari, A., Ma, K., Zhang, B., McGuinness, D.L., Szekely, P.: Dimensions of commonsense knowledge (2021)
28. Li, Z., Niu, C., Meng, F., Feng, Y., Li, Q., Zhou, J.: Incremental transformer with deliberation decoder for document grounded conversations. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 12–21. Association for Computational Linguistics, Florence, Italy (2019). https://doi.org/10.18653/v1/P19-1002. https://aclanthology.org/P19-1002
29. Galitsky, B., Ilvovsky, D.: Chatbot with a discourse structure-driven dialogue management. In: Proceedings of the Software Demonstrations of the 15th Conference of the European Chapter of the Association for Computational Linguistics, pp. 87–90 (2017)
30. Ma, L., Zhang, W.N., Li, M., Liu, T.: A survey of document grounded dialogue systems (dgds) (2020)
31. Hasal, M., Nowaková, J., Ahmed Saghair, K., Abdulla, H., Snášel, V., Ogiela, L.: Chatbots: Security, privacy, data protection, and social aspects. Concurr. Comput. Pract. Exp. **33**(19), e6426 (2021). https://doi.org/10.1002/cpe.6426. https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.6426
32. Bocklisch, T., Faulkner, J., Pawlowski, N., Nichol, A.: Rasa: Open source language understanding and dialogue management (2017)
33. Williams, S.: Hands-On Chatbot Development with Alexa Skills and Amazon Lex: Create Custom Conversational and Voice Interfaces for Your Amazon Echo Devices and Web Platforms. Packt Publishing Ltd. (2018)
34. Sabharwal, N., Agrawal, A.: Cognitive Virtual Assistants Using Google Dialogflow: Develop Complex Cognitive Bots Using the Google Dialogflow Platform. Apress (2020)
35. Gehrmann, S., Adewumi, T.P., Aggarwal, K., Ammanamanchi, P.S., Anuoluwapo, A., Bosselut, A., Chandu, K.R., Clinciu, M., Das, D., Dhole, K.D., Du, W., Durmus, E., Dusek, O., Emezue, C., Gangal, V., Garbacea, C., Hashimoto, T., McMillan-Major, A., Mille, S., van Miltenburg, E., Nadeem, M., Narayan, S., Nikolaev, V., Niyongabo, R.A.: The GEM benchmark: Natural language generation, its evaluation and metrics. CoRR **abs/2102.01672** (2021). https://arxiv.org/abs/2102.01672
36. Bień, M., Gilski, M., Maciejewska, M., Taisner, W., Wisniewski, D., Lawrynowicz, A.: RecipeNLG: A cooking recipes dataset for semi-structured text generation. In: Proceedings of the 13th International Conference on Natural Language Generation, pp. 22–28. Association for Computational Linguistics, Dublin, Ireland (2020). https://www.aclweb.org/anthology/2020.inlg-1.4

37. Marin, J., Biswas, A., Ofli, F., Hynes, N., Salvador, A., Aytar, Y., Weber, I., Torralba, A.: Recipe1m+: A dataset for learning cross-modal embeddings for cooking recipes and food images. IEEE Trans. Pattern Anal. Mach. Intell. (2019). arXiv:1810.06553
38. Wang, Y., Kim, J.: Interconnectedness between online review valence, brand, and restaurant performance. J. Hosp. Tour. Manag. **48**, 138–145 (2021). https://doi.org/10.1016/j.jhtm.2021.05.016. https://www.sciencedirect.com/science/article/pii/S1447677021000851
39. Bender, E.M., Friedman, B.: Data statements for natural language processing: Toward mitigating system bias and enabling better science. Trans. Assoc. Comput. Linguist. **6**, 587–604 (2018). https://doi.org/10.1162/tacl_a_00041
40. Zampieri, M., Nakov, P., Scherrer, Y.: Natural language processing for similar languages, varieties, and dialects: A survey. Nat. Lang. Eng. **26**(6), 595–612 (2020). https://doi.org/10.1017/S1351324920000492
41. Silberman, M.S., Tomlinson, B., LaPlante, R., Ross, J., Irani, L., Zaldivar, A.: Responsible research with crowds: Pay crowdworkers at least minimum wage. Commun. ACM **61**(3), 39–41 (2018). https://doi.org/10.1145/3180492
42. Goodman, J.K., Cryder, C., Cheema, A.: Data collection in a flat world: Strengths and weaknesses of mechanical turk samples. J. Behav. Decis. Making (2012, Forthcoming)
43. Van Gijsel, S., Speelman, D., Geeraerts, D.: A variationist, corpus linguistic analysis of lexical richness, pp. 1–16 (2005)
44. Novikova, J., Dušek, O., Rieser, V.: The E2E dataset: New challenges for end-to-end generation. In: Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue, pp. 201–206. Association for Computational Linguistics, Saarbrücken, Germany (2017). https://doi.org/10.18653/v1/W17-5525. https://aclanthology.org/W17-5525
45. Perez-Beltrachini, L., Gardent, C.: Analysing data-to-text generation benchmarks. In: Proceedings of the 10th International Conference on Natural Language Generation, pp. 238–242. Association for Computational Linguistics, Santiago de Compostela, Spain (2017). https://doi.org/10.18653/v1/W17-3537. https://aclanthology.org/W17-3537
46. Gkatzia, D.: Content selection in data-to-text systems: A survey. CoRR **abs/1610.08375** (2016). http://arxiv.org/abs/1610.08375
47. Petroni, F., Rocktäschel, T., Riedel, S., Lewis, P., Bakhtin, A., Wu, Y., Miller, A.: Language models as knowledge bases? In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp. 2463–2473. Association for Computational Linguistics, Hong Kong, China (2019). https://doi.org/10.18653/v1/D19-1250. https://aclanthology.org/D19-1250