# The Role of Personal Perspectives in Open-Domain Dialogue

Selene Báez Santamaría*[0000-0003-3808-2614], Lea Krause*[0000-0001-7187-5224], Lucia Donatelli[0000-0002-5974-7454], and Piek Vossen[0000-0002-6238-5941]

Vrije Universiteit Amsterdam
{firstname_initial.lastname}@vu.nl

**Abstract.** As Conversational AI becomes integral to various digitally enhanced services, there is a growing need for fluent and personalized dialogue agents. Yet, recent state-of-the-art neural dialogue systems face challenges in transparency and memory precision, often leading to incoherent and factually incorrect interactions. In this position paper, we argue that open-domain dialogue plays a crucial role in building effective dialogue agents by fostering shared memories and capturing personal perspectives and social information. Specifically, we see the need for (i) datasets that capture long-term interaction; and (ii) a representation scheme that enables tracking of factoids and perspectives over time. We present an analysis of several existing open-domain dialogue datasets and highlight the richness and complexity of conversational data. We further present a data modeling scheme that represents and tracks factoids and perspectives over time and makes use of shared, long-term memories crucial for facilitating fluid and meaningful interaction. Our scheme is designed to improve long-term memory, increase transparency, and can serve as inspiration for designing neuro-symbolic dialogue agents capable of long-term social interaction.

**Keywords:** Open-domain Dialogue · Dialogue Systems · Dialogue Representations · Knowledge Graph

## 1 Introduction

As Conversational AI becomes a generic interface for various digitally enhanced services, reliability and accuracy are crucial for assisting users effectively. Task-based dialogue systems, such as Q&A, recommender systems, and e-commerce assistants, have a clear and measurable objective. However, the goal of open-domain dialogue systems is less defined.

In contrast to [46]'s suggestion of using open-domain dialogue as a supplement to task-oriented conversations, we argue that open-domain dialogue has an independent role in conveying personal perspectives and social information, fostering the development of shared memories that are crucial for nurturing social connections [10, 19]. As such, open-domain conversational agents have the

* Equal contribution

potential to exploit non-factual information, to enhance the social aspect of their interactions.

Conversations involve a nuanced blend of *factoids*, *non-factoids*, and *perspectives*. A *factoid* typically represents a concise, factual piece of information that can be straightforwardly verified [44]. On the other hand, *non-factoids* encompass broader and often subjective areas of dialogue, demanding deeper understanding, reasoning, or interpretation. They are not always easily verifiable, and may include explanations, reasoned arguments, or interpretative answers. Interwoven with these are *perspectives*, which embody an individual's beliefs, sentiments, and personal viewpoints on a given topic. These perspectives are pivotal, adding a subjective richness to conversations, highlighting the personal and emotive dimensions that set them apart from the more objective nature of factoids. Table 1 presents excerpts from five dialogue datasets, highlighting instances of personal perspective expressions.

| | |
|---|---|
| HP | "Butterbeer!" said Harry, without thinking.<br>"Yeah, <u>I like that stuff!</u>" |
| MZ | There are 5 restaurants in the area. <u>I like Frankie and Bennys</u> in the south. Can I make you a reservation? |
| OD | Yes <u>i like him too</u> did you know he also was in Zodiac a crime fiction film. |
| DRE | "Speaker 1: I never did anything with Adrienne Turner."<br>"Speaker 2: Oh please, and you knew how much <u>I liked her</u>."<br>"Speaker 1: I don't know what... you're talking about." |
| PEDC | Right. I get really self-conscious about wearing the same tie like more than once in a couple weeks.<br><u>I also just really like ties</u>. |

**Table 1.** Examples of utterances in which personal perspectives are expressed (<u>underlined</u>), MZ = MultiWOZ, OD = OpenDialog, HP = Harry Potter, DRE = DialogRE, PEDC = Personal Events in Dialogue

. Extracting both factoids and perspectives, and tracking them over time, is essential for constructing explicit memories from conversations that also reflect personal experiences and viewpoints which can be referenced [24]. Such personal memories can be leveraged to improve communication in terms of understanding and generation [23], to decide on relevance and preferences, and to be more sensitive to (changing) personal values instead of relying on generalized biases. Memories go beyond user profiling as they combine explicit factoids with personal values and are grounded in time [29]. Furthermore, the ability to reference shared memories can bolster social ties and foster trust [10].

Neural dialogue systems are proficient in matching user needs with appropriate responses [33, 55]. However, they often lack transparency and struggle with memory precision, which can lead to erroneous responses, known as hallucinations [25]. This challenge is exacerbated when it comes to accurately modeling and referencing complex long-term memories. Implicit modeling tends to falter in intricate contexts [28], and limited computational research exists on long-term interactions. These limitations make it challenging for dialogue agents to establish precise long-term memory for social exchanges. Graph representations of dialogue as memories may help as their structures can effectively represent the volume, diversity, and provenance of information over time [1, 54] while providing connectivity links across these layers and high interpretability [36, 45, 47].

In this paper, we argue that open-domain dialogue plays a critical role in fostering shared memories and conveying social information in addition to factoids. We analyze current open-domain datasets for conversational traits, showcasing their diversity, complexity, and limitations in capturing the intricate social aspects of open conversations. We highlight three key aspects essential for Open-Domain Conversational AI datasets: 1) coverage of long-term interactions, arguably needing explicit memory representations, 2) inclusion of factoid and non-factoid information, and 3) accurately reflective of human-machine interactions. We discuss how graph representations can be leveraged during communication to improve long-term memory and prevent unwanted hallucinations and biases from conversational systems. We recognize the need for a new dataset that captures the three components above and propose a representation scheme that enables the recovery of shared memory through underlying explicit representation.

## 2 Related work

*Open-domain dialogue.* Open-domain dialogue in NLP focuses on developing personable conversational agents capable of discussing various topics, in a wide range of contexts [41]. While existing methods look into augmenting memory to enhance the persona or knowledge of conversational agents, these primarily neural methods suffer from various errors that impact the ease of communication [27]. Generated responses can be non-sensical or ungrammatical [40]; answers can be inconsistent or not reflect full dialogue history [57]; it is difficult to incorporate context and world knowledge [48]; and agents can reflect biases or use offensive language [12].

*Text-based dialogue datasets.* A number of dialogue datasets has been published as research resources, in the form of textual transcripts of dyadic dialogues. For example, PersonaChat [56] (162k utterances, 1155 personas) is a crowd-sourced dataset focusing on personal topics and speaker profiles. ConvAI2 [13] augments PersonaChat with additional personas, and reworded utterances (194k utterances, 1250 topics). DailyDialog [31] simulates everyday conversations with human-written, noise-minimized dialogues. Encompassing various life topics, it also provides communication intention and emotion labels. EmpatheticDialogues [37] (50k utterances) introduces a benchmark for assessing the capacity to discuss emotional situations and respond empathically. BlendedSkillTalk [42] (5k conversations) combines persona, knowledge, and empathy by using different samples of the previous datasets for training and testing.

*Structured representations of dialogue.* To compensate for errors that may not be explainable in neural systems, several datasets provide structured data in addition to text. Such structure often serves to model the global context of the dialogue [13, 20, 50, 51]. Harry Potter Dialogue [8] provides background information about scenes, relations, and attributes to better model character and story development. OpenDialKG [34] pairs each system dialogue turn with paths through a knowledge graph to support an open-domain response. However, the graph path does not necessarily express all the information of the response.

Structured dialogue may also focus on the linguistic signal itself. The Personal Events in Dialogue Corpus [16] presents podcast transcripts in which personal events are extracted from surface-level tokens. DialogueRE [52] and PELD [49], based on the TV series “Friends,” represent entities and their relations (e.g. *person:siblings*) through argument pairs and lexical triggers. Finally, Dialogue AMR (DialAMR) [3] extends graph-based meaning representations for semantic parsing to the dialogue level by incorporating speech acts for task-based human-robot interaction.

*Intent annotation in dialogue.* Linguistic research has addressed the goal of intent in dialogue through dialogue act and speech act annotation [4–7, 39]. These annotations serve as units of a conversation transmitting specific communicative functions; yet they fall short when discovering long-term phenomenons across utterances, dialogues, and encounters over time. These phenomenons commonly relate to *perspectives* or attitudes, representing a cognition (‘I know X is good’), affective reactions (e.g., of pleasure), and/or behavioral tendencies (e.g., to approach or avoid) [2], which are common to change over time. Annotations of intents should thus ideally consider longer conversational contexts but also long-term memory and relational contexts.

While structured representations prove 1) useful for representing information implicit in the linguistic signal, and, 2) relevant to interpreting the context of the dialogue, these benefits are rarely leveraged together. Additionally, structured representations have yet to focus on more dynamic aspects of conversation such as persona, memory, and knowledge state in a way that neural methods have. In what follows, we look at properties of existing open-domain datasets and outline a path for developing a holistic structured representation that captures all of the above characteristics.

## 3 Comparative analysis of dialogue datasets

To review and assess the extend of factoid, non-factoid, and perspective information, we investigate five dialogue datasets: four open-domain dialogue datasets, and one commonly used task-based dataset: The Personal Events in Dialogue Corpus (PEDC) has long personal conversations, DialogRE (DRE) and Harry Potter (HP) contain fictional dialogues, OpenDialKG (OD) has instructed Q&A between crowd workers, and the task-based MultiWOZ2.2 (MZ) with targeted services in various e-commerce domains. We expect MZ to differ significantly from the open conversation datasets, with OD being the most similar to it and PEDC the least.

*Volume of data* MZ and OD are the largest datasets in the number of turns and conversations, while PEDC is significantly smaller (Table 2). Instead, PEDC has more extended conversations, almost six times longer, hinting that daily life conversations are expected to be more complex.

| | PEDC | DRE | HP | OD | MZ | Avg. |
|---|---|---|---|---|---|---|
| Conversations | 14 | 1788 | 1262 | **13802** | 10437 | 5461 |
| Turns | 1035 | 23129 | 16891 | 96942 | **143040** | 56207 |
| Turns/Conversation | **73.93** | 12.94 | 13.38 | 7.02 | 13.71 | 24.20 |
| Tokens/Turn | 10.64 | 8.21 | 11.41 | **14.00** | 9.32 | 10.72 |

**Table 2.** Volume statistics for conversational datasets.

*Two-token utterances* Most datasets present a high number of two-token sentences, which are commonly one-word utterances followed by punctuation (Table 3). While none explicitly convey factual properties, they offer perspectives (*What?*), confirmations (*Yes.*) and negations (*No.*) about the knowledge communicated before. Remarkably in PEDC, hardly any of these utterances occur.

| | DRE | HP | OD | MZ | Avg. |
|---|---|---|---|---|---|
| | n=7587 | n=3458 | n=8562 | n=11450 | |
| Yes. | 0.79% | 0.52% | **7.94%** | 7.23% | 4.12% |
| Goodbye. | 0.04% | 0.00% | 0.11% | **14.89%** | 3.76% |
| What? | 5.79% | **8.07%** | 0.00% | 0.02% | 3.47% |
| Thanks! | 0.34% | 0.09% | **6.81%** | 5.46% | 3.17% |
| Sure! | 0.41% | 0.00% | **7.60%** | 3.47% | 2.87% |

**Table 3.** Most frequent two-token turns across four datasets sorted on the average proportional frequency.

*Dialogue acts* Utilizing MIDAS [53], we explored the distribution of dialogue acts as illustrated in Table 4. While *Statement* is the prevalent act across datasets, *opinions* stand out as the second most frequent, signifying the presence of personal perspectives. However, their occurrence is markedly less frequent in the MZ dataset. In contrast, MZ displays a pronounced inclination towards *commands*, as expected for task-based dialogue, where users typically guide and direct the system.

| Dialog act | PEDC | DRE | HP | OD | MZ | Avg. |
|---|---|---|---|---|---|---|
| statement | 39.08% | 34.87% | **44.80%** | <u>24.77%</u> | 37.36% | 36.18% |
| opinion | 24.66% | 26.19% | 22.89% | **27.77%** | <u>14.77%</u> | 23.25% |
| command | 7.52% | 8.73% | <u>5.72%</u> | 14.01% | **21.13%** | 11.42% |

**Table 4.** Most frequent dialogue acts across datasets sorted on the average proportional frequency. **Highest** and <u>lowest</u> values per metric are marked.

*Linguistic variation* We extract root predicates, subjects, and objects, and report the Type/Token and Turn/Type ratios as proxies for linguistic variation and word repetition (Table 5). MZ shows low lexical variation, while PEDC a large variation, supporting the notion that personal conversations are complex and varied, while task-based communication is more systematic. In line with this finding, words hardly re-occur in PEDC, whereas they often re-occur in MZ.

| | | PEDC | DRE | HP | OD | MZ | Avg. |
|---|---|---|---|---|---|---|---|
| PRE | types | 362 | 2598 | 2698 | 2587 | 1814 | 7131 |
| | tokens | 1914 | 40605 | 32207 | 96938 | 237046 | 81742 |
| | type/token | **0.19** | 0.06 | 0.08 | 0.03 | <u>0.01</u> | 0.07 |
| | turn/type | <u>2.86</u> | 8.90 | 6.26 | 37.47 | **78.85** | 26.87 |
| SUB | types | 159 | 1075 | 1309 | 5643 | 2788 | 2195 |
| | tokens | 1913 | 40567 | 31959 | 96924 | 237043 | 81681 |
| | type/token | **0.08** | 0.03 | 0.04 | 0.06 | <u>0.01</u> | 0.04 |
| | turn/type | <u>6.51</u> | 21.52 | 12.90 | 17.18 | **51.31** | 21.88 |
| OBJ | types | 336 | 1897 | 2020 | 7374 | 3228 | 2971 |
| | tokens | 1912 | 40495 | 31784 | 96881 | 237038 | 81622 |
| | type/token | **0.18** | 0.05 | 0.06 | 0.08 | <u>0.01</u> | 0.08 |
| | turn/type | <u>3.08</u> | 12.19 | 8.36 | 13.15 | **44.31** | 16.22 |

**Table 5.** Type/Token and Turn/Type ratios for the use of predicates (PRE), subjects (SUB) and objects (OBJ). Type measures the unique instances of root predicates, subjects, and objects. **Highest** and <u>lowest</u> values per metric are marked.

*Discourse content* First and second person pronouns are dominant subjects and objects (Table 6), indicating strong interpersonal communication [17]. We identify predicates expressing cognitive and perceptual perspectives [38, 43], such as *think*, *like*, *know*, and *love* among the most frequent to occur across the datasets.

To summarize, MZ and PEDC represent two extremes on a spectrum. On one side, MZ is characterized by short and systematic Q&A, task-centric dialogue acts, plenty of two-token utterances for confirmations and closings, and the utilization of perspective predicates associated with users' needs. On the other extreme, PEDC compiles lengthy, diverse conversations that lack a specific task or

| | | HP | MZ | OD | DRE | PEDC | Avg. |
|---|---|---|---|---|---|---|---|
| | | n=31959 | n=237043 | n=96924 | n=40567 | n=1913 | |
| SUB | I | 16.99% | 29.40% | 30.30% | 22.09% | 21.54% | 24.06% |
| | you | 13.02% | 18.28% | 21.36% | 13.20% | 13.07% | 15.79% |
| | it | 5.67% | 3.37% | 6.09% | 4.85% | 9.46% | 5.89% |
| | | n=31784 | n=237038 | n=96881 | n=40495 | n=1912 | |
| OBJ | you | 2.85% | 8.59% | 2.46% | 2.88% | 1.26% | 3.61% |
| | it | 3.68% | 1.78% | 4.33% | 2.47% | 3.61% | 3.17% |
| | what | 2.49% | 0.34% | 2.01% | 2.10% | 2.67% | 1.92% |

**Table 6.** Most frequent subjects and objects in terms of averaged proportion. Punctuation excluded.

service focus, and contain perspective predicates associated with the speaker's open views. Despite the differences, all dialogues exchange information in the form of factoids ***and*** perspectives, albeit to varying degrees. Conversational systems sensitive to user needs and aiming to sustain long-term relationships should be equipped to model this multifaceted information and retain a memory that can be utilized in future interactions.

## 4 Dataset requirements

The proposed outlook on open-domain dialogue addresses the exchange of factoids and social information through prolonged and repeated interactions. Under such conditions, establishing common ground and personalization become critical for leveraging built-up knowledge. Therefore, we propose to generate a dialogue dataset pertaining to communication that is complex in its content, form, and social context. This dataset must include aligned explicit representations, addressing the problems explored in Section 1. In this section, we describe different data requirements, while Appendix A shows further specifics on the annotation scheme and a detailed example of how such a dataset would look.

### 4.1 Conversational requirements

The dataset will focus on *open-domain* dialogues since task-oriented dialogue has extensively explored the use of structured representations to boost performance [32]. The data should include conversations spread over time that display long-term social relationships and highlight the need for long-term memory. Preferably, the dialogues should be human-machine; however, currently, there is no system for complex long-term social conversation, hence the need to fall back on human-human data. Yet, this data might not be fully representative of the social relations that would arise between humans and machines [18].

The dialogues' average turn length would range between 5 and 50. While shorter lengths might resemble standard conversational Q&A tasks, longer sequences risk complicating data collection and annotation without significantly enhancing the representation of recurrent interactions. Interlocutors should express perspectives freely so that language variation is reflected and sufficient complexity is represented, for example, *beliefs*, *sentiments*, and *certainty*.

### 4.2 Explicit representations requirements

Graphs have become a popular representation choice for structured semantic content. At the sentence level, labeled, directed graphs have become a focus of semantic parsing tasks: general graphs are semantically more expressive than surface syntax, and their discrete and hierarchical nature allows for explainable interpretation of meaning in natural language [30, 35]. Yet, the current structured analysis of conversations where each utterance or adjacency pair is evaluated as an independent graph (particularly tree structures) misses the opportunity to connect and accumulate signals over larger conversational units.

Knowledge graphs are structured ontologies representing relations between entities in subject-predicate-object triples. Going beyond the sentence level, these types of graphs can serve as contextual grounding for interpreting content in interaction [15]. We propose a knowledge graph structure for open-domain dialogue that integrates sentence-level semantics with contextual information. The graph should represent the information exchange occurring in the dialogue instead of enhancing or supporting background information. A suitable representation model should be able to store information related to a dialogue's content, form, and metadata. This structure has four key features:

*1. Enhanced representation of memory.* The structure of graphs not only allows for long-term tracking of discourse entities and simple representation of coreference via reentrancy but also facilitates the accumulation of knowledge over time, capturing its current state as well as its history. See Appendix A, where the entity *RobertDowneyJr* is referred to throughout the dialogue.

*2. Dual representation of knowledge and speaker persona.* The graph should illustrate both the linguistic form of information communicated and the perspectives of the interlocutors [26]. Hence, the graph can operate as a Theory of Mind model, facilitating reasoning over different world views. See Appendix A, where the triple *assistant_affirm_AssistantLikeIronMan* expresses a perspective on the triple *AssistantLikeIronMan*.

*3. Generalizability.* Structured semantic representations have been shown to aid tasks related to compositional generalization, the ability of a model to predict the meaning of unseen sentences by recombining training instances in novel ways, similar to how humans acquire and use language [14]. Such generalization is essential in open-domain dialogue, where new topics, linguistic structures, and personas are constantly introduced. See Appendix A, where the predicate *like* operates on objects of type *person*.

*4. Transparency.* Transparent intermediate representations offer key advantages. They lower error and hallucination risks, simplify their detection, and aid in recognizing and mitigating possible biases in the model or data [25]. Transparent explicit memories of the conversation allow to communicate about the communication itself, so that alignments between interlocutors can be restored when needed. See Appendix A, where the utterance "I heard that" is represented as an explicit reference to the previous triple *AntmanNotHasActorRobertDowneyJr*.

## 5 Challenges for dataset creation

### 5.1 Data collection

The first choice to face is whether to enrich an existing dataset or to collect a new dataset altogether. From the analysis in Section 3, we note that current datasets 1) do exhibit factoid and social information, but 2) may lack long-term relationships or 3) do not accurately exhibit natural and interactive [24] open-domain human-machine dialogue. As such, the first step in building this dataset would be to alter existing dialogue datasets to fulfill 2). With regards to 3), we recognize that open-domain dialogue datasets largely comprise human-human interactions due to the current lack of systems for long-term social conversations between humans and machines. Overall, research indicates that humans tend to interact with social robots in similar ways as they interact with humans, with anthropomorphic features facilitating this effect [22]. However, we should be cautious in our interpretations, as the Unique Agent Hypothesis [11] suggests that using only human-human dialogues may not capture the full spectrum of biases and expectations, such as the prevailing belief in automation's flawlessness, which influences human-machine dynamics.

### 5.2 Data annotation

Converting conversations, whether in speech or text form, into structured data presents substantial challenges. Firstly, the representation of communication content is commonly addressed via closed/open information extraction (IE). Yet, these methods do not provide explicit semantics support to inference, and fall short when supporting question representation [9]. Moreover, these methods primarily focus on factoids, thus overlooking interaction-level information, including the viewpoints of the conversation participants, such as their attitude, sentiment, or level of certainty. Incorporating these elements can deepen our comprehension of the utterance, enabling the discernment of more nuanced meanings.

We propose manual annotation as automatic extraction methods cannot generate the required graph. We suggest an open IE approach, where the knowledge base schema emerges from, as opposed to being imposed on, the data. When creating a new triple-predicate, linguistic guidance should be given to ensure consistency and semantics. Having detailed, specific guidelines can significantly enhance the accuracy and consistency of the annotation process, ultimately improving the data quality. Furthermore, annotators should discuss and revise annotations, address disagreements and ambiguity, and document the definition and scope of existing triple-predicates and types [21].

## 6 Impact on conversational NLP tasks

A dual-aligned representation of dialogue, incorporating both text and graph modalities, offers enriched context and content understanding, enhancing various conversational NLP tasks. This approach is beneficial in *response generation* by considering perspectives alongside factual elements, facilitates more context-aware translations in *machine translation* systems by capturing interpersonal nuances, and enables *summarization* systems to produce comprehensive summaries that represent both factual content and evolving interpersonal dynamics.

## 7 Conclusion

Conversations of a casual nature are often dismissed as inconsequential, yet they can transmit critical information about social contexts, personal beliefs, and individual perspectives, even when the topics appear casual or routine. It is crucial to acknowledge and research these nuances of dialogue as they hold significance in especially long-term interaction. Our proposed annotation scheme is a step in that direction and aims to capture the complexities of these dialogues in the combination of both factoids and perspectives. Building on our vision for open-domain dialogue, our annotation scheme was designed to enhance the representation of extended interactions that blend both factoids and social nuances. While many datasets focus on the mere exchange of information, our scheme underscores the importance of capturing the layers of evolving perspectives and identities. Another distinction is the emphasis on data accumulation over time. While several annotation methodologies advocate for graph-like designs, they often miss out on the continuous evolution of dialogues — a facet our scheme underscores. Given the challenge of prioritizing certain data elements as memories in neural representations, our model posits that explicit representations can provide clear referencing and interpretability. The inclusion of nested spans and cross-references in our approach further refines the dialogue dataset, ensuring it will remain attuned to complex content, form, and social contexts, as required in long-form open dialogue.

Our future work centers on the application of our annotation scheme to an established dialogue dataset. This will be facilitated by the creation of exhaustive annotation guidelines and the development of a custom annotation tool. Upon completion, we intend to release the annotated data under an open-source license for public access.

## Limitations

Our analysis, while comprehensive, has several limitations. Primarily, we only examined datasets that are in English, thereby potentially limiting the broader applicability and universality of our conclusions.

Ideally, our data would encompass conversations that span across time, showcasing the development of long-term social relationships and preferably, these dialogues should be between humans and machines. However, given the current state of technology, no system available is robust enough to handle complex, long-term social conversations to function as a data source. This limitation necessitates our reliance on human-human conversation data. It is worth noting that while human-human data provides valuable insights, it might not fully encapsulate the unique dynamics that might arise in human-machine interactions.

The interpretation of predicates can be a challenging task, especially when the context available to annotators is inadequate, which may lead to misinterpretation. Similarly, the annotation of speaker intention can also be complex, with difficulties in accurately discerning the speaker's intent, such as differentiating between a statement and an indirect request, which can change based on the cultural background of the speaker. Moreover, the comparison between written and spoken dialogues is an important consideration. While spoken dialogue is inherently more challenging due to the presence of noise, it also carries additional cues such as pauses and ellipsis, which can provide valuable insight into the conversation dynamics.

## Ethics Statement

With regard to ethics, the sensitive nature of the information that can potentially be shared during open dialogue poses a considerable challenge. The potential for privacy issues during data collection is high, especially when dealing with personal or intimate conversations. Therefore, it's critical that any data collection or annotation methods adhere to strict privacy standards to ensure that the personal information of individuals involved is adequately protected. This research is designed to facilitate and promote the study of long-term personal relationships, an area that necessitates particular ethical sensitivity due to its personal nature. To ensure ethical conduct, special emphasis will be put on developing comprehensive annotation guidelines and providing extensive training for annotators. This will help to capture a wide array of perspectives, and to minimize conscious or unconscious bias, thereby ensuring that the research is conducted in a manner that respects the rights and dignity of all participants.

## Acknowledgements

This research was funded by the Vrije Universiteit Amsterdam and the Netherlands Organisation for Scientific Research (NWO) via the *Spinoza* grant (SPI 63-260) awarded to Piek Vossen, and the *Hybrid Intelligence Centre* via the Zwaartekracht grant (024.004.022).

## References


1. Adolphs, L., Shuster, K., Urbanek, J., Szlam, A., Weston, J.: Reason first, then respond: Modular generation for knowledge-infused dialogue. In: Findings of the Association for Computational Linguistics: EMNLP 2022. pp. 7112–7132. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Dec 2022), https://aclanthology.org/2022.findings-emnlp.527
2. Amodio, D.M.: Social cognition 2.0: An interactive memory systems account. Trends in Cognitive Sciences **23**(1), 21–33 (2019)
3. Bonial, C., Donatelli, L., Abrams, M., Lukin, S.M., Tratz, S., Marge, M., Artstein, R., Traum, D., Voss, C.: Dialogue-AMR: Abstract Meaning Representation for dialogue. In: Proceedings of the Twelfth Language Resources and Evaluation Conference. pp. 684–695. European Language Resources Association, Marseille, France (May 2020), https://aclanthology.org/2020.lrec-1.86
4. Bruner, J.S.: The ontogenesis of speech acts. Journal of child language **2**(1), 1–19 (1975)
5. Bunt, H.: Towards an analysis of dialogue organization principles. IPO Annual Progress Report **12**, 105–114 (1977), https://api.semanticscholar.org/CorpusID:151039658
6. Bunt, H., Alexandersson, J., Choe, J.W., Fang, A.C., Hasida, K., Petukhova, V., Popescu-Belis, A., Traum, D.R.: Iso 24617-2: A semantically-based standard for dialogue annotation. In: International Conference on Language Resources and Evaluation (2012), https://api.semanticscholar.org/CorpusID:6242409
7. Bunt, H., Petukhova, V., Gilmartin, E., Pelachaud, C., Fang, A., Keizer, S., Prévot, L.: The ISO standard for dialogue act annotation, second edition. In: Proceedings of the Twelfth Language Resources and Evaluation Conference. pp. 549–558. European Language Resources Association, Marseille, France (May 2020), https://aclanthology.org/2020.lrec-1.69
8. Chen, N., Wang, Y., Jiang, H., Cai, D., Chen, Z., Wang, L., Li, J.: What would harry say? building dialogue agents for characters in a story (2022). https://doi.org/10.48550/ARXIV.2211.06869, https://arxiv.org/abs/2211.06869
9. Chiang, T.R., Ye, H.T., Chen, Y.N.: An empirical study of content understanding in conversational question answering. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 7578–7585 (2020)
10. Clark, L., Pantidi, N., Cooney, O., Doyle, P., Garaialde, D., Edwards, J., Spillane, B., Gilmartin, E., Murad, C., Munteanu, C., Wade, V., Cowan, B.R.: What makes a good conversation? challenges in designing truly conversational agents. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. p. 1–12. CHI '19, Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3290605.3300705, https://doi.org/10.1145/3290605.3300705
11. de Visser, E.J., Monfort, S.S., McKendrick, R., Smith, M.A.B., McKnight, P.E., Krueger, F., Parasuraman, R.: Almost human: Anthropomorphism increases trust resilience in cognitive agents. Journal of Experimental Psychology: Applied **22**(3), 331–349 (2016). https://doi.org/10.1037/xap0000092
12. Dinan, E., Abercrombie, G., Bergman, A., Spruit, S., Hovy, D., Boureau, Y.L., Rieser, V.: SafetyKit: First aid for measuring safety in open-domain conversational systems. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 4113–4133. Association for Computational Linguistics,

Dublin, Ireland (May 2022). https://doi.org/10.18653/v1/2022.acl-long.284, https://aclanthology.org/2022.acl-long.284
13. Dinan, E., Roller, S., Shuster, K., Fan, A., Auli, M., Weston, J.: Wizard of wikipedia: Knowledge-powered conversational agents. Proceedings of the International Conference on Learning Representations (2019)
14. Donatelli, L., Koller, A.: Compositionality in computational linguistics. Annual Review of Linguistics **9** (2023)
15. Ehrlinger, L., Wöß, W.: Towards a definition of knowledge graphs. SEMANTiCS (Posters, Demos, SuCCESS) **48**(1-4), 2 (2016)
16. Eisenberg, J., Sheriff, M.: Automatic extraction of personal events from dialogue. In: Proceedings of the First Joint Workshop on Narrative Understanding, Storylines, and Events. pp. 63–71. Association for Computational Linguistics, Online (Jul 2020). https://doi.org/10.18653/v1/2020.nuse-1.8, https://aclanthology.org/2020.nuse-1.8
17. Eisenberg, J., Sheriff, M.: Automatic extraction of personal events from dialogue. In: Proceedings of the First Joint Workshop on Narrative Understanding, Storylines, and Events. pp. 63–71 (2020)
18. Ethayarajh, K., Choi, Y., Swayamdipta, S.: Understanding dataset difficulty with $\mathcal{V}$-usable information (2021). https://doi.org/10.48550/ARXIV.2110.08420, https://arxiv.org/abs/2110.08420
19. Gilmartin, E., Cowan, B.R., Vogel, C., Campbell, N.: Explorations in multiparty casual social talk and its relevance for social human machine dialogue. Journal on Multimodal User Interfaces **12**(4), 297–308 (Dec 2018). https://doi.org/10.1007/s12193-018-0274-2
20. Gopalakrishnan, K., Hedayatnia, B., Chen, Q., Gottardi, A., Kwatra, S., Venkatesh, A., Gabriel, R., Hakkani-Tür, D.: Topical-Chat: Towards Knowledge-Grounded Open-Domain Conversations. In: Proc. Interspeech 2019. pp. 1891–1895 (2019). https://doi.org/10.21437/Interspeech.2019-3079, http://dx.doi.org/10.21437/Interspeech.2019-3079
21. Grosman, J.S., Furtado, P.H., Rodrigues, A.M., Schardong, G.G., Barbosa, S.D., Lopes, H.C.: Eras: Improving the quality control in the annotation process for natural language processing tasks. Information Systems **93**, 101553 (2020). https://doi.org/https://doi.org/10.1016/j.is.2020.101553, https://www.sciencedirect.com/science/article/pii/S0306437920300521
22. Hildt, E.: What Sort of Robots Do We Want to Interact With? Reflecting on the Human Side of Human-Artificial Intelligence Interaction. Frontiers in Computer Science **3** (2021)
23. Horton, W.S.: Conversational common ground and memory processes in language production. Discourse Processes **40**(1), 1–35 (2005)
24. Horton, W.S.: Theories and approaches to the study of conversation and interactive discourse. The Routledge handbook of discourse processes pp. 22–68 (2017)
25. Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y.J., Madotto, A., Fung, P.: Survey of hallucination in natural language generation. ACM Comput. Surv. **55**(12) (mar 2023). https://doi.org/10.1145/3571730, https://doi.org/10.1145/3571730
26. Kalouli, A.L., Crouch, R.: GKR: the graphical knowledge representation for semantic parsing. In: Proceedings of the Workshop on Computational Semantics beyond Events and Roles. pp. 27–37. Association for Computational Linguistics, New Orleans, Louisiana (Jun 2018). https://doi.org/10.18653/v1/W18-1304, https://aclanthology.org/W18-1304

27. Kann, K., Ebrahimi, A., Koh, J., Dudy, S., Roncone, A.: Open-domain dialogue generation: What we can do, cannot do, and should do next. In: Proceedings of the 4th Workshop on NLP for Conversational AI. pp. 148–165. Association for Computational Linguistics, Dublin, Ireland (May 2022). https://doi.org/10.18653/v1/2022.nlp4convai-1.13, https://aclanthology.org/2022.nlp4convai-1.13
28. Kaushik, D., Khilji, A.F.U.R., Sinha, U., Pakray, P.: CNLP-NITS @ LongSumm 2021: TextRank variant for generating long summaries. In: Proceedings of the Second Workshop on Scholarly Document Processing. pp. 103–109. Association for Computational Linguistics, Online (Jun 2021). https://doi.org/10.18653/v1/2021.sdp-1.13, https://aclanthology.org/2021.sdp-1.13
29. Kensinger, E.A., Schacter, D.L.: Memory and emotion. Handbook of emotions **3**, 601–617 (2008)
30. Koller, A., Oepen, S., Sun, W.: Graph-based meaning representations: Design and processing. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts. pp. 6–11 (2019)
31. Li, Y., Su, H., Shen, X., Li, W., Cao, Z., Niu, S.: Dailydialog: A manually labelled multi-turn dialogue dataset. In: Proceedings of The 8th International Joint Conference on Natural Language Processing (IJCNLP 2017) (2017)
32. Louvan, S., Magnini, B.: Recent neural methods on slot filling and intent classification for task-oriented dialogue systems: A survey. In: Proceedings of the 28th International Conference on Computational Linguistics. pp. 480–496. International Committee on Computational Linguistics, Barcelona, Spain (Online) (Dec 2020). https://doi.org/10.18653/v1/2020.coling-main.42, https://aclanthology.org/2020.coling-main.42
33. Mehndiratta, A., Asawa, K.: Non-goal oriented dialogue agents: state of the art, dataset, and evaluation. Artificial Intelligence Review **54**, 329–357 (2021)
34. Moon, S., Shah, P., Kumar, A., Subba, R.: OpenDialKG: Explainable conversational reasoning with attention-based walks over knowledge graphs. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. pp. 845–854. Association for Computational Linguistics, Florence, Italy (Jul 2019). https://doi.org/10.18653/v1/P19-1081, https://aclanthology.org/P19-1081
35. Oepen, S., Abend, O., Abzianidze, L., Bos, J., Hajič, J., Hershcovich, D., Li, B., O'Gorman, T., Xue, N., et al.: Mrp 2020: The second shared task on cross-framework and cross-lingual meaning representation parsing. Proceedings of the CoNLL 2020 Shared Task: Cross-Framework Meaning Representation Parsing pp. 1–22 (2020)
36. Pramanik, S., Alabi, J., Roy, R.S., Weikum, G.: UNIQORN: unified question answering over RDF knowledge graphs and natural language text. CoRR **abs/2108.08614** (2021), https://arxiv.org/abs/2108.08614
37. Repar, A., Shumakov, A.: Aligning Estonian and Russian news industry keywords with the help of subtitle translations and an environmental thesaurus. In: Proceedings of the EACL Hackashop on News Media Content Analysis and Automated Report Generation. pp. 71–75. Association for Computational Linguistics, Online (Apr 2021), https://aclanthology.org/2021.hackashop-1.10
38. Saurí, R., Pustejovsky, J.: Factbank: a corpus annotated with event factuality. Language resources and evaluation **43**, 227–268 (2009)
39. Searle, J.R.: Speech acts: An essay in the philosophy of language, vol. 626. Cambridge university press (1969)

40. See, A., Manning, C.: Understanding and predicting user dissatisfaction in a neural generative chatbot. In: Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue. pp. 1–12. Association for Computational Linguistics, Singapore and Online (Jul 2021), https://aclanthology.org/2021.sigdial-1.1
41. Skantze, G., Doğruöz, A.S.: The open-domain paradox for chatbots: Common ground as the basis for human-like dialogue. In: Proceedings of the 24th Meeting of the Special Interest Group on Discourse and Dialogue. pp. 605–614. Association for Computational Linguistics, Prague, Czechia (Sep 2023), https://aclanthology.org/2023.sigdial-1.57
42. Smith, E.M., Williamson, M., Shuster, K., Weston, J., Boureau, Y.L.: Can you put it all together: Evaluating conversational agents' ability to blend skills. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 2021–2030. Association for Computational Linguistics, Online (Jul 2020). https://doi.org/10.18653/v1/2020.acl-main.183, https://aclanthology.org/2020.acl-main.183
43. van Son, C., Morante, R., Vossen, P.: Natural language processing tasks for the extraction of perspectives. Creating a More Transparent Internet: The Perspective Web p. 173 (2022)
44. Soricut, R., Brill, E.: Automatic question answering using the web: Beyond the factoid. Information Retrieval **9**, 191–206 (2006)
45. Sun, H., Dhingra, B., Zaheer, M., Mazaitis, K., Salakhutdinov, R., Cohen, W.: Open domain question answering using early fusion of knowledge bases and text. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. pp. 4231–4242. Association for Computational Linguistics, Brussels, Belgium (Oct-Nov 2018). https://doi.org/10.18653/v1/D18-1455, https://aclanthology.org/D18-1455
46. Sun, K., Moon, S., Crook, P., Roller, S., Silvert, B., Liu, B., Wang, Z., Liu, H., Cho, E., Cardie, C.: Adding chit-chat to enhance task-oriented dialogues. Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies p. 1570–1583 (2021)
47. Tiddi, I., Schlobach, S.: Knowledge graphs as tools for explainable machine learning: A survey. Artificial Intelligence **302**, 103627 (2022)
48. Wang, J., Liu, J., Bi, W., Liu, X., He, K., Xu, R., Yang, M.: Improving knowledge-aware dialogue generation via knowledge base question answering. In: Proceedings of the AAAI conference on artificial intelligence. vol. 34, pp. 9169–9176 (2020)
49. Wen, Z., Cao, J., Yang, R., Liu, S., Shen, J.: Automatically select emotion for response via personality-affected emotion transition. In: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. pp. 5010–5020. Association for Computational Linguistics, Online (Aug 2021). https://doi.org/10.18653/v1/2021.findings-acl.444, https://aclanthology.org/2021.findings-acl.444
50. Wu, J., Zhou, H.: Augmenting topic aware knowledge-grounded conversations with dynamic built knowledge graphs. In: Proceedings of Deep Learning Inside Out (DeeLIO): The 2nd Workshop on Knowledge Extraction and Integration for Deep Learning Architectures. pp. 31–39. Association for Computational Linguistics, Online (Jun 2021). https://doi.org/10.18653/v1/2021.deelio-1.4, https://aclanthology.org/2021.deelio-1.4
51. Yang, S., Zhang, R., Erfani, S.: GraphDialog: Integrating graph knowledge into end-to-end task-oriented dialogue systems. In: Proceedings of

the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 1878–1888. Association for Computational Linguistics, Online (Nov 2020). https://doi.org/10.18653/v1/2020.emnlp-main.147, https://aclanthology.org/2020.emnlp-main.147

52. Yu, D., Sun, K., Cardie, C., Yu, D.: Dialogue-based relation extraction. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 4927–4940. Association for Computational Linguistics, Online (Jul 2020). https://doi.org/10.18653/v1/2020.acl-main.444, https://aclanthology.org/2020.acl-main.444
53. Yu, D., Yu, Z.: MIDAS: A dialog act annotation scheme for open domain human machine spoken conversations. CoRR **abs/1908.10023** (2019), http://arxiv.org/abs/1908.10023
54. Yu, W., Zhu, C., Li, Z., Hu, Z., Wang, Q., Ji, H., Jiang, M.: A survey of knowledge-enhanced text generation. ACM Comput. Surv. **54**(11s) (nov 2022). https://doi.org/10.1145/3512467, https://doi.org/10.1145/3512467
55. Zang, X., Rastogi, A., Sunkara, S., Gupta, R., Zhang, J., Chen, J.: Multiwoz 2.2: A dialogue dataset with additional annotation corrections and state tracking baselines. In: Proceedings of the 2nd Workshop on Natural Language Processing for Conversational AI, ACL 2020. pp. 109–117 (2020)
56. Zhang, S., Dinan, E., Urbanek, J., Szlam, A., Kiela, D., Weston, J.: Personalizing dialogue agents: I have a dog, do you have pets too? In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 2204–2213. Association for Computational Linguistics, Melbourne, Australia (Jul 2018). https://doi.org/10.18653/v1/P18-1205, https://aclanthology.org/P18-1205
57. Zhong, H., Dou, Z., Zhu, Y., Qian, H., Wen, J.R.: Less is more: Learning to refine dialogue history for personalized dialogue generation. In: Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. pp. 5808–5820. Association for Computational Linguistics, Seattle, United States (Jul 2022). https://doi.org/10.18653/v1/2022.naacl-main.426, https://aclanthology.org/2022.naacl-main.426

# A Appendix


```
[{"speaker": "user",
 "utterance": "Do you like Iron Man?",
 "graph": [{ "dialogue_act": "yes_no_question",
            "text": "Do you like Iron Man?",
            "triple": {
              "id": "AssistantLikeIronMan",
              "subject": {"id": "assistant", "type": "person", "ref": "assistant"
                  , "offset": [3,5]},
              "predicate": {"id": "like", "type": "like", "ref": "like", "offset"
                  : [7,10]},
              "object": {"id": "IronMan", "type": "character", "ref": "Iron Man
                  ", "offset": [12,19]}}}]}
,{"speaker": "assistant",
 "utterance": "Sure do! Robert Downey Jr. is a favorite.",
 "graph": [{ "dialogue_act": "pos_answer",
            "text": "Sure do!",
            "triple": {
              "id": "assistant_affirm_AssistantLikeIronMan",
              "subject": {"id": "assistant", "type": "person", "ref": "", "offset
                  ": []},
              "predicate": {"id": "affirm", "type": "affirm", "ref": "Sure do!",
                  "offset": [0,7]},
              "object": {"id": "AssistantLikeIronMan", "type": "preference", "ref
                  ": "", "offset": []}},
            "perspective": {
              "holder": "assistant", "target_id": "AssistantLikeIronMan", "
                  sentiment": 1, "polarity": 1, "certainty": 1}
         },{ "dialogue_act": "statement",
            "text": "Robert Downey Jr. is a favorite.",
            "triple": {
              "id": "assistant_favorite_RobertDowneyJr",
              "subject": {"id": "assistant", "type": "person", "ref": "", "offset
                  ": []},
              "predicate": {"id": "favorite", "type": "like", "ref": "favorite",
                  "offset": [23,30]},
              "object": {"id": "RobertDowneyJr", "type": "person", "ref": "
                  Robert Downey Jr.", "offset": [0,17]}},
            "perspective": {
              "holder": "assistant", "target_id": "RobertDowneyJr", "sentiment
                  ": 1, "polarity": 1, "certainty": 1}}]}
,{"speaker": "user",
 "utterance": "Yes, I think I like him too. Did you know he will not be in
       Antman?",
 "graph": [{ "dialogue_act": "statement",
            "text": "Yes, I think I like him too.",
            "triple": {
              "id": "user_like_RobertDowneyJr",
```

```
            "subject": {"id": "user", "type": "person", "ref": "I", "offset": [
                13,14]},
            "predicate": {"id": "like", "type": "like", "ref": "like", "offset"
                : [15,18]},
            "object": {"id": "RobertDowneyJr", "type": "person", "ref": "him"
                , "offset": [20,22]}},
          "perspective": {
            "holder": "user", "target_id": "RobertDowneyJr", "sentiment": 1,
                "polarity": 1, "certainty": 0.25}
      },{ "dialogue_act": "yes_no_question",
          "text": "Did you know he will not be in Antman?",
          "triple": {
            "id": "assistant_know_AntmanNotHasActorRobertDowneyJr",
            "subject": {"id": "assistant", "type": "person", "ref": "assistant"
                , "offset": [4,6]},
            "predicate": {"id": "know", "type": "know", "ref": "know", "offset"
                : [8,11]},
            "object": {"id": "AntmanNotHasActorRobertDowneyJr", "type": "
                factoid", "ref": "will not be in Antman", "offset": [13,33]},
          "perspective": {
            "holder": "user", "target_id": "
                AntmanNotHasActorRobertDowneyJr", "sentiment": 0, "
                polarity": 0, "certainty": 1}}}]}
,{"speaker": "assistant",
  "utterance": "I heard that.",
  "graph": [{ "dialogue_act": "pos_answer",
          "text": "I heard that.",
          "triple": {
            "id": "assistant_affirm_AntmanNotHasActorRobertDowneyJr",
            "subject": {"id": "assistant", "type": "person", "ref": "I", "
                offset": [0,1]},
            "predicate": {"id": "affirm", "type": "affirm", "ref": "I heard
                that.", "offset": [0,7]},
            "object": {"id": "AntmanNotHasActorRobertDowneyJr", "type": "
                factoid", "ref": "that", "offset": [8,11]}},
          "perspective": {
            "holder": "assistant", "target_id": "
                AntmanNotHasActorRobertDowneyJr", "sentiment": 0, "
                polarity": 1, "certainty": 0.75}}]}]
```


*Annotation example 1:* Altered excerpt from OpenDialog [34]. Triples, which serve as a basic unit, are assigned an identifier and can be recursively referred to in order to express more complex meanings. *Nodes* -subjects and objects in triples- represent entities (mostly grammatically identified as nouns) annotated at token level. *Edges* represent predicates (mostly identified as relations between entities, or properties of entities) annotated at knowledge level. Predicates are semantically meaningful, and entities' co-references are resolved. Triple elements without explicit referent are represented as `"ref": ""`. Sentiment values may be

-1 (negative), 1 (positive), or 0 (neutral). Polarity values may be -1 (negation), 0 (neutral or question), or 1 (affirmation). Certainty values range between 0 (uncertain) and 1 (certain).