Artificial Ideal Companion Agent

Anushk Mittal & Noorie Dhingra
Georgia Institute of Technology
Atlanta, USA
anushk@gatech.edu
noorie@emory.edu

ABSTRACT
UPDATED—30 November 2019. We propose a bring home conversational agent with the objective to keep the host engaged in intellectual conversations. The proposed solution is (a) an ensemble open context, open domain question answering agent that is capable of answering questions in context and is rooted in knowledge based methods with self-structured memory (b) undertakes supervised learning in the form of *curiosity questions* to fill gaps in knowledge and (c) uses a reinforcement learning mechanism to predict k-future-turns to allow optimizing long term rewards. We sanity check the performance against popular conversational and question answering challenges with future work directed towards empirical studies.

Author Keywords
Open domain; Question Answering; Open Context; companion agent; knowledge-based ML ensemble, agent memory.

CSS Concepts
• **Human-centered computing~Natural language interfaces** • Human-centered computing~Collaborative interaction • Computing methodologies~Reasoning about belief and knowledge • Computing methodologies~Information extraction

AIMS AND OBJECTIVES
The overarching goal of the project is to build a natural language conversational agent that is capable of holding intellectual conversations. This involves:

1) Being able to understand and answer single turn open domain factual questions E.g. "What is the capital of USA?"
2) Being able to understand and answer context-based questions such that the answer to the question relies on additional information gathered from a part of the question. E.g. "What is the next movie of the actor who played Jim in The Office?"
3) Being able to understand and answer questions in multi-turn conversations such that the question involves coreference from previous dialogues. E.g.:
   a. **User:** Who is the president of United States?
      Agent: Donald J. Trump
   b. **User:** And of France?
      Agent: Emmanuel Macron
4) Being able to give data driven responses that are not bald. (based on user preferences or factual knowledge). For e.g.

Old conversation snippet	new conversation
User: I just got a blue Tesla **Agent:** wow, that's great!	**User:** I am going to paint my room. **Agent:** Good luck. Are you painting blue? Like your blue Tesla.
-	**User:** I am going to Kasukabe tonight! **Agent:** ~~Have a great time!~~ You'll love it! Try Onsake, the best in town.

To achieve this behavior, the agent relies on an ensemble of retrieval and learning techniques:

a) Deep Learning based machine reading at scale that involves machine comprehension to *read* documents to find an answer and a data retriever component to *find* relevant documents pertaining to a question.
b) Knowledge Base lookup through popular ontology structures as NELL [1], DBpedia [2], Freebase [3], and WikiData [4] that allows abstracting semantic relationship in questions and potential answers to *find evidence* for best choice.
c) Long term memory (LTM) / Short term memory (STM) structure that allows storing retrieved information as a part of agent's *own memory* along with episodic events for future referencing.

The agent's task can thus be put into three simple steps

Task 0: understanding the question using attentive history, coreference, named entity recognition (NER), parts of speech (POS), and intent classification.

Task 1: finding answer(s) to the question based on knowledge base lookup, machine reading and memory.

Task 2: predicting future turns and pick the best answer that optimizes for reward policy as a reinforcement learning method.

The reward policy is built to ensure that the agent succeeds in keeping the user engaged with its choice of answer.

We hope to advance state-of-the art in the following ways:

- Addressing the issue of conversation agents with no commonsense and ethical knowledge
- A cheaper and more reliable way of finding evidence to answers through knowledge-based reasoning rather than additional nodes in end-to-end deep learning models
- A deployment ready agent that categorically undertakes intellectual conversations rather than mere mimicking.

MOTIVATION

We choose to build this agent because it's a yet unexplored applied field and feeling alone is a major issue with college students [5], yet all companion solutions researched so far are specifically tailored for those of older age [6], [7], [8], [9].

Over the last few years there has been a significant progress in end-to-end deep learning based conversational agents that essentially learn to *mimic* human behavior. This means that while state-of-the-art agents perform well in answering questions or generating fake conversations, they lie susceptible to gibberish responses when exposed to conversational turns with humans. For example, DialoGPT [10] a state-of-the-art single turn conversational agent on the popular DSTC-7 [11] challenge when queried, only produces sensible responses 1 in 5 times. Such agents are clearly not capable of holding intellectual conversations. Moreover, since these agents do not deeply understand the meaning of what they say, they are highly susceptible to the quality of the large text corpus used for training that raises ethical concerns of agent behavior when let run in the wild [10]. Outputs may reflect gender and other historical biases implicit in the data. Responses generated using such models may exhibit a propensity to express agreement with propositions that are unethical, biased or offensive (or the reverse, disagreeing with otherwise ethical statements). Microsoft's Tay [12] is a popular example of practical ethical issues present in deploying such deep learning agents on a wide scale.

In contrast, agents relying on knowledge bases (KBs) have inherent limitations (incompleteness, fixed schemas) that prohibit such systems to produce complete responses on their own.

Finally, most conversational agents tend to produce bald responses while playing on the safe side that ensures that their responses are grammatically correct and make sense but are uninteresting compared to human responses. For example, when a user says, "I am going to Kasukabe tonight!" end-to-end DL based agents typically respond as "Have a great time!" which is uninteresting compared to a human response of "You'll love it! Try Onsake, the best in town."

BACKGROUND

There has been a significant progress on conversational systems in the past few years. Pertaining to our work on human-like human-level conversational agents, this research can be categorized into: Sequence to sequence models, open domain question answering (QA) based on knowledge base (KB), open domain QA based on machine comprehension, multi-hop question answering and finally, questions and answers in dialogue form.

Natural Language Processing

The current state-of-the-art approaches, sequence to sequence models of various kinds [13], [14], [15], [16] attempt to address some of these skills, but generally suffer from an inability to bring memory and knowledge to bear; as indicated by their name, they involve encoding an input sequence, providing limited reasoning by transforming their hidden state given the input, and then decoding to an output.

Open domain QA based on KB

Open-domain QA was originally defined as finding answers in collections of unstructured documents, following the setting of the annual TREC competitions. With the development of KBs, many recent innovations have occurred in the context of QA from KBs with the creation of resources like WebQuestions [17] and SimpleQuestions [18] based on the Freebase KB [19], or on automatically extracted KBs, e.g., OpenIE triples and NELL [20].

Several commonsense knowledge bases have been constructed during the past decade, such as ConceptNet [21] and SenticNet [22]. The aim of commonsense knowledge representation and reasoning is to give a foundation of real-world knowledge. Typically, a commonsense knowledge base can be seen as a semantic network where concepts are nodes in the graph and relations are edges. Each <concept1, relation, concept2> triple is termed an assertion. Based on the Open Mind Common Sense project [23], ConceptNet not only contains objective facts such as "Paris is the capital of France" that are constantly true, but also captures informal relations between common concepts that are part of everyday knowledge such as "A dog is a pet". This feature of ConceptNet is desirable in our experiments, because the ability to recognize the informal relations

Open domain QA based on machine comprehension

A second motivation to cast a fresh look at this problem is that of machine comprehension of text, i.e., answering questions after reading a short text or story. That subfield has made considerable progress recently thanks to new deep learning architectures like attention-based and memory augmented neural networks [24], [25], [26] and release of new training and evaluation datasets like QuizBowl [27], CNN/Daily Mail based on news articles [28], CBT based on children books [29], or SQuAD [30] and WikiReading [31], both based on Wikipedia and TriviaQA [32]. Chen et al in [33] proposed a two stage approach of retrieving relevant content with the question, then reading the paragraphs returned by the information retrieval (IR) component to arrive at the final answer. This "retrieve and read" approach has since been adopted and extended in various open-domain QA systems [34], [35], but it is inherently limited to answering questions that do not require multi-hop/multi-step reasoning. This is because for many multi-hop questions, not all the relevant context can be obtained in a single retrieval step. The QuAC dataset investigates similar themes, but as a sequence of questions and answers in dialogue form instead [36]. Numerous neural models have been proposed [37], [38], [39] and achieved promising performances on several different MRC data sets, such as SQuAD, NarrativeQA [40] and CoQA [41]. The performance was further boosted after the release of the Bidirectional Encoder Representations from Transformers (BERT) model [42], which has delivered state-of-the-art performance on several RC/QA data sets. Most existing research in machine RC/QA focuses on answering a question given a single document or paragraph. Although the performance on these types of tasks have been improved a lot over the last few years, the models used in these tasks still lack the ability to do reasoning across multiple documents when a single document is not enough to find the correct answer [43].

Multi-hop question answering

More recently, the emergence of multi-hop question answering datasets such as WIKIHOP [44], QAngaroo [45] and HOTPOTQA [46] has led to more interesting work on multi-hop QA. Designed to be more challenging than SQuAD-like datasets, they feature questions that require context of more than one. document to answer, testing QA systems' abilities to infer the answer in the presence of multiple pieces of evidence and to efficiently find the evidence in a large pool of candidate documents. Qi et al [72] recently published Answering Complex Open-domain Questions where rather than relying purely on the original question to retrieve passages, the central innovation is that at each step the model also uses IR results from previous hops of reasoning to generate a new natural language query and retrieve new evidence to answer the original question. These data sets are challenging because they require models to be able to do multi-hop reasoning over multiple documents and under strong distraction. HotpotQA also encourages explainable QA models by providing supporting sentences for the answer, which usually come from several documents (a document is called "gold doc" if it contains the answer or it contains supporting sentences to the answer). To solve the multi-hop multi-document QA task, two research directions have been explored. The first direction focuses on applying or adapting previous techniques that are successful in single-document QA tasks to multi-document QA tasks, for example the studies in [47], [48], [49], [50]. The other direction resorts to Graph Neural Networks (GNN) to realize multi-hop reasoning across multiple documents, and promising performance has been achieved [51], [52], [53], [54], [55]. Select, Answer and Explain (SAE) system described in [56] solves the multi-document RC problem. The system first filters out answer-unrelated documents and thus reduce the amount of distraction information. This is achieved by a document classifier trained with a novel pairwise learning-to-rank loss. The selected answer-related documents are then input to a model to jointly predict the answer and supporting sentences. The model is optimized with a multi-task learning objective on both token level for answer prediction and sentence level for supporting sentences prediction, together with an attentionbased interaction between these two tasks. Evaluated on HotpotQA, a challenging multi-hop RC data set, the proposed SAE system achieves top competitive performance in distractor setting compared to other existing systems on the leaderboard.

However, these implementations for the most part rely on Neural response generation, a subcategory of text-generation that shares the objective of generating natural-looking text (distinct from any training instance) that is relevant to the prompt. Most open-domain neural response generation systems suffer from content or style inconsistency [57], [58], [59], lack of long-term contextual information [60], and blandness [61], [62], [63].

Questions and answers in dialogue form

In the domain of open chit-chat, Open-Subtitles [14], Persona-Chat [64] and Twitter [65] have tested the ability of sequence-to-sequence models that attend over the recent dialogue history, but do not attempt to recall long-term knowledge beyond encoding it directly into the weights of the feed-forward network. Wizards of Wikipedia [71] utilizes Memory Network architectures [66] to retrieve knowledge and read and condition on it, and Transformer architectures [16] to provide state-of-the-art text representations and sequence models for generating outputs.

In the area of non-goal directed dialogue incorporating knowledge. [67] employed Memory Networks to perform dialogue discussing movies in terms of recommendation and open-ended discussion from Reddit, conditioning on a structured knowledge base. [68] also links Reddit to structured knowledge. Both [69] and [70] use unstructured text instead: the former to discuss news articles using Wikipedia summaries as knowledge, and the latter to discuss local businesses in two-turn dialogues using Foursquare tips as knowledge. [70] uses an extended Encoder-Decoder where the decoder is provided with an encoding of the context along with the external knowledge encoding.

OUTCOMES AND DELIVERABLES

The outcome we expect from this project is a base conversational model that can be replicated and made available to any host user through a texting platform. The agent would be able to hold intellectual conversations out-of-the-box and over time adapt to user preferences such that each replication would respond differently to the satisfaction of their host user.

The deliverable would be the base model and an extended version trained with a host user. We would also demonstrate the model's intellectual performance on a host of question answering challenges.

Finally, there would be a final paper detailing the workings and performance for the proposed agent.

Future directions include:

- Blind Turing Test to assess the human-like behavior of the model.
- Empirical research to measure whether the agent solves the initially proposed problem of keeping the host engaged in intellectual conversations.
- An embodied agent that can bring physical reasoning to its otherwise virtual environment.

REFERENCES

[1] Carlson, Andrew, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R. Hruschka, and Tom M. Mitchell. "Toward an architecture for never-ending language learning." In Twenty-Fourth AAAI Conference on Artificial Intelligence. 2010.

[2] Auer, Sören, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. "Dbpedia: A nucleus for a web of open data." In The semantic web, pp. 722-735. Springer, Berlin, Heidelberg, 2007.

[3] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In Proceedings of the 2008 ACM SIGMOD international conference on Management of data. AcM, pages 1247–1250.

[4] Vrandečić, Denny, and Markus Krötzsch. "Wikidata: a free collaborative knowledge base." (2014).

[5] Center for Collegiate Mental Health. (2018, January). 2017 Annual Report (Publication No. STA 18-166).

[6] Kriglstein, Simone, and Gunter Wallner. "HOMIE: an artificial companion for elderly people." In CHI'05 extended abstracts on Human factors in computing systems, pp. 2094-2098. ACM, 2005.

[7] Sherwood, Timothy, Farilee Mintz, and Miroslava Vomela. "Project VIRGO: creation of a surrogate companion for the elderly." In CHI'05 Extended Abstracts on Human Factors in Computing Systems, pp. 2104-2108. ACM, 2005.

[8] Hollywood, Elaine, Grainne O'Brien, and Sara Lennon. "SIGCHI project: User centered design of a program alleviating loneliness (PAL)." In CHI'05 Extended Abstracts on Human Factors in Computing Systems, pp. 2089-2093. ACM, 2005.

[9] Ring, Lazlo, Barbara Barry, Kathleen Totzke, and Timothy Bickmore. "Addressing loneliness and isolation in older adults: Proactive affective agents provide better support." In 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, pp. 61-66. IEEE, 2013.

[10] Zhang, Yizhe, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and Bill Dolan. "DialoGPT: Large-Scale Generative Pre-training for Conversational Response Generation." arXiv preprint arXiv:1911.00536 (2019).

[11] Michel Galley, Chris Brockett, Xiang Gao, Jianfeng Gao, and Bill Dolan. 2019. Grounded response generation task at dstc7. In AAAI Dialog System Technology Challenges Workshop.

[12] Neff, Gina, and Peter Nagy. "Automation, algorithms, and politics| talking to Bots: Symbiotic agency and the case of Tay." International Journal of Communication 10 (2016): 17.

[13] Sutskever, I., O. Vinyals, and Q. V. Le. "Sequence to sequence learning with neural networks." Advances in NIPS (2014).

[14] Vinyals, Oriol, and Quoc Le. "A neural conversational model." arXiv preprint arXiv:1506.05869 (2015).

[15] Serban, Iulian V., Alessandro Sordoni, Yoshua Bengio, Aaron Courville, and Joelle Pineau. "Building end-to-end dialogue systems using generative hierarchical neural network models." In Thirtieth AAAI Conference on Artificial Intelligence. 2016.

[16] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need." In Advances in neural information processing systems, pp. 5998-6008. 2017.

[17] Berant, Jonathan, Andrew Chou, Roy Frostig, and Percy Liang. "Semantic parsing on freebase from question-answer pairs." In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pp. 1533-1544. 2013.

[18] Bordes, Antoine, Nicolas Usunier, Sumit Chopra, and Jason Weston. "Large-scale simple question answering with memory networks." arXiv preprint arXiv:1506.02075 (2015).

[19] Bollacker, Kurt, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. "Freebase: a collaboratively created graph database for structuring human knowledge." In Proceedings of the 2008 ACM SIGMOD international conference on Management of data, pp. 1247-1250. AcM, 2008.

[20] Fader, Anthony, Luke Zettlemoyer, and Oren Etzioni. "Open question answering over curated and extracted

knowledge bases." In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 1156-1165. ACM, 2014.

[21] Speer, Robert, and Catherine Havasi. "Representing General Relational Knowledge in ConceptNet 5." In LREC, pp. 3679-3686. 2012.

[22] Cambria, Erik, Soujanya Poria, Rajiv Bajpai, and Björn Schuller. "SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives." In Proceedings of COLING 2016, the 26th international conference on computational linguistics: Technical papers, pp. 2666-2677. 2016.

[23] Singh, Push. "The open mind common sense project." KurzweilAI. net (2002).

[24] Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. "Neural machine translation by jointly learning to align and translate." arXiv preprint arXiv:1409.0473 (2014).

[25] Weston, Jason, Antoine Bordes, Sumit Chopra, Alexander M. Rush, Bart van Merriënboer, Armand Joulin, and Tomas Mikolov. "Towards ai-complete question answering: A set of prerequisite toy tasks." arXiv preprint arXiv:1502.05698 (2015).

[26] Graves, Alex, Greg Wayne, and Ivo Danihelka. "Neural turing machines." arXiv preprint arXiv:1410.5401 (2014).

[27] Iyyer, Mohit, Jordan Boyd-Graber, Leonardo Claudino, Richard Socher, and Hal Daumé III. "A neural network for factoid question answering over paragraphs." In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 633-644. 2014.

[28] Hermann, Karl Moritz, Tomas Kocisky, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. "Teaching machines to read and comprehend." In Advances in neural information processing systems, pp. 1693-1701. 2015.

[29] Hill, Felix, Antoine Bordes, Sumit Chopra, and Jason Weston. "The goldilocks principle: Reading children's books with explicit memory representations." arXiv preprint arXiv:1511.02301 (2015).

[30] Rajpurkar, Pranav, Jian Zhang, Konstantin Lopyrev, and Percy Liang. "Squad: 100,000+ questions for machine comprehension of text." arXiv preprint arXiv:1606.05250 (2016).

[31] Hewlett, Daniel, Alexandre Lacoste, Llion Jones, Illia Polosukhin, Andrew Fandrianto, Jay Han, Matthew Kelcey, and David Berthelot. "Wikireading: A novel large-scale language understanding task over wikipedia." arXiv preprint arXiv:1608.03542 (2016).

[32] Joshi, Mandar, Eunsol Choi, Daniel S. Weld, and Luke Zettlemoyer. "Triviaqa: A large scale distantly supervised challenge dataset for reading comprehension." arXiv preprint arXiv:1705.03551 (2017).

[33] Chen, Danqi, Adam Fisch, Jason Weston, and Antoine Bordes. "Reading wikipedia to answer open-domain questions." arXiv preprint arXiv:1704.00051 (2017).

[34] Nishida, Kyosuke, Itsumi Saito, Atsushi Otsuka, Hisako Asano, and Junji Tomita. "Retrieve-and-read: Multi-task learning of information retrieval and reading comprehension." In Proceedings of the 27th ACM International Conference on Information and Knowledge Management, pp. 647-656. ACM, 2018.

[35] Bernhard Kratzwald and Stefan Feuerriegel. 2018. Adaptive document retrieval for deep question answering. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP).

[36] Eunsol Choi, He He, Mohit Iyyer, Mark Yatskar, Wen tau Yih, Yejin Choi, Percy Liang, and Luke Zettlemoyer. 2018. Quac: Question answering in context. In Empirical Methods in Natural Language Processing.

[37] Seo, M.; Kembhavi, A.; Farhadi, A.; and Hajishirzi, H. 2016. Bidirectional attention flow for machine comprehension. arXiv preprint arXiv:1611.01603.

[38] Xiong, C.; Zhong, V.; and Socher, R. 2016. Dynamic coattention networks for question answering. arXiv preprint arXiv:1611.01604.

[39] Tay, Y.; Luu, A. T.; Hui, S. C.; and Su, J. 2018. Densely connected attention propagation for reading comprehension. In Advances in Neural Information Processing Systems, 4911–4922.

[40] Kocisk ˇ y, T.; Schwarz, J.; Blunsom, P.; Dyer, ` C.; Hermann, K. M.; Melis, G.; and Grefenstette, E. 2018. The narrativeqa reading comprehension challenge. Transactions of the Association of Computational Linguistics 6:317–328.

[41] Reddy, S.; Chen, D.; and Manning, C. D. 2019. Coqa: A conversational question answering challenge. Transactions of the Association for Computational Linguistics 7:249–266.

[42] Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 4171–4186.

[43] Chen, J., and Durrett, G. 2019. Understanding dataset design choices for multi-hop reasoning. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics:

Human Language Technologies, Volume 1 (Long and Short Papers), 4026–4032.

[44] Welbl, J.; Stenetorp, P.; and Riedel, S. 2018. Constructing datasets for multi-hop reading comprehension across documents. Transactions of the Association of Computational Linguistics 6:287–302

[45] Welbl, Johannes, Pontus Stenetorp, and Sebastian Riedel. "Constructing datasets for multi-hop reading comprehension across documents." Transactions of the Association for Computational Linguistics 6 (2018): 287-302.

[46] Yang, Zhilin, Peng Qi, Saizheng Zhang, Yoshua Bengio, William W. Cohen, Ruslan Salakhutdinov, and Christopher D. Manning. "Hotpotqa: A dataset for diverse, explainable multi-hop question answering." arXiv preprint arXiv:1809.09600 (2018).

[47] Dhingra, B.; Jin, Q.; Yang, Z.; Cohen, W.; and Salakhutdinov, R. 2018. Neural models for reasoning over multiple mentions using coreference. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers), volume 2, 42–48

[48] Zhong, V.; Xiong, C.; Keskar, N. S.; and Socher, R. 2019. Coarse-grain fine-grain coattention network for multi-evidence question answering. arXiv preprint arXiv:1901.00603.

[49] Yang, Z.; Qi, P.; Zhang, S.; Bengio, Y.; Cohen, W.; Salakhutdinov, R.; and Manning, C. D. 2018. Hotpotqa: A dataset for diverse, explainable multi-hop question answering. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2369–2380

[50] Nishida, K.; Nishida, K.; Nagata, M.; Otsuka, A.; Saito, I.; Asano, H.; and Tomita, J. 2019. Answering while summarizing: Multi-task learning for multi-hop qa with evidence extraction. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2335–2345.

[51] Song, L.; Wang, Z.; Yu, M.; Zhang, Y.; Florian, R.; and Gildea, D. 2018. Exploring graph-structured passage representation for multi-hop reading comprehension with graph neural networks. arXiv preprint arXiv:1809.02040.

[52] De Cao, N.; Aziz, W.; and Titov, I. 2019. Question answering by reasoning across documents with graph convolutional networks. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 2306–2317.

[53] Cao, Y.; Fang, M.; and Tao, D. 2019. Bag: Bi-directional attention entity graph convolutional network for multi-hop reasoning question answering. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 357–362.

[54] Tu, M.; Wang, G.; Huang, J.; Tang, Y.; He, X.; and Zhou, B. 2019. Multi-hop reading comprehension across multiple documents by reasoning over heterogeneous graphs. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2704–2713.

[55] Xiao, Y.; Qu, Y.; Qiu, L.; Zhou, H.; Li, L.; Zhang, W.; and Yu, Y. 2019. Dynamically fused graph network for multihop reasoning. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 6140–6150.

[56] Tu, Ming, Kevin Huang, Guangtao Wang, Jing Huang, Xiaodong He, and Bowen Zhou. "Select, Answer and Explain: Interpretable Multi-hop Reading Comprehension over Multiple Documents." arXiv preprint arXiv:1911.00484 (2019).

[57] Jiwei Li, Michel Galley, Chris Brockett, Georgios P Spithourakis, Jianfeng Gao, and Bill Dolan. 2016. A persona-based neural conversation model. In ACL.

[58] Yizhe Zhang, Xiang Gao, Sungjin Lee, Chris Brockett, Michel Galley, Jianfeng Gao, and Bill Dolan. 2019. Consistent dialogue generation with self-supervised feature learning. arXiv preprint arXiv:1903.05759.

[59] Xiang Gao, Yizhe Zhang, Sungjin Lee, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. 2019. Structuring latent spaces for stylized response generation. arXiv preprint arXiv:1909.05361.

[60] Iulian Vlad Serban, Alessandro Sordoni, Ryan Lowe, Laurent Charlin, Joelle Pineau, Aaron Courville, and Yoshua Bengio. 2017. A hierarchical latent variable encoder-decoder model for generating dialogues. In AAAI.

[61] Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. 2016. A diversity-promoting objective function for neural conversation models. In NAACL.

[62] Yizhe Zhang, Michel Galley, Jianfeng Gao, Zhe Gan, Xiujun Li, Chris Brockett, and Bill Dolan. 2018. Generating informative and diverse conversational responses via adversarial information maximization. In NeurIPS.

[63] Lianhui Qin, Michel Galley, Chris Brockett, Xiaodong Liu, Xiang Gao, Bill Dolan, Yejin Choi, and Jianfeng Gao. 2019. Conversing by reading: Contentful neural conversation with on-demand machine reading. arXiv preprint arXiv:1906.02738.

[64] Zhang, Saizheng, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. "Personalizing Dialogue Agents: I have a dog, do you have pets too?." arXiv preprint arXiv:1801.07243 (2018).

[65] Alessandro Sordoni, Michel Galley, Michael Auli, Chris Brockett, Yangfeng Ji, Meg Mitchell, Jian-Yun Nie, Jianfeng Gao, and Bill Dolan. 2015. A neural network approach to context-sensitive generation of conversational responses. In Proc. of NAACL-HLT.

[66] Sukhbaatar, Sainbayar, Jason Weston, and Rob Fergus. "End-to-end memory networks." In Advances in neural information processing systems, pp. 2440-2448. 2015.

[67] Jesse Dodge, Andreea Gane, Xiang Zhang, Antoine Bordes, Sumit Chopra, Alexander Miller, Arthur Szlam, and Jason Weston. Evaluating prerequisite qualities for learning end-to-end dialog systems. In Proceedings of the International Conference on Learning Representations (ICLR), 2016.

[68] Hao Zhou, Tom Young, Minlie Huang, Haizhou Zhao, Jingfang Xu, and Xiaoyan Zhu. Commonsense knowledge aware conversation generation with graph attention. In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI), pp. 4623–4629. International Joint Conferences on Artificial Intelligence Organization, 7 2018.

[69] Prasanna Parthasarathi and Joelle Pineau. Extending neural generative conversational model using external knowledge sources. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 2018.

[70] Marjan Ghazvininejad, Chris Brockett, Ming-Wei Chang, Bill Dolan, Jianfeng Gao, Wen-tau Yih, and Michel Galley. A knowledge-grounded neural conversation model. In Proceedings of the Conference on Association for the Advancement of Artificial Intelligence (AAAI), 2018.

[71] Dinan, Emily, Stephen Roller, Kurt Shuster, Angela Fan, Michael Auli, and Jason Weston. "Wizard of wikipedia: Knowledge-powered conversational agents." arXiv preprint arXiv:1811.01241 (2018).

[72] Qi, Peng, Xiaowen Lin, Leo Mehr, Zijian Wang, and Christopher D. Manning. "Answering Complex Open-domain Questions Through Iterative Query Generation." arXiv preprint arXiv:1910.07000 (2019).