Alexa Prize TaskBot Challenge

Eugene Agichtein, Yoelle Maarek, Oleg Rokhlenko
Amazon, Alexa Shopping

As people learn to interact with AI assistants such as Alexa, their needs change and become more complex. Alexa must evolve accordingly, and offer more sophisticated experiences, which require addressing increasingly complex AI research challenges. In the last few years, Amazon has offered university teams a chance to partner with Amazon scientists in order push the boundaries of the state of the art in conversational AI via the Alexa Prize challenge. This competition gives a framework for selected student teams to build agents, or "bots", served by Alexa, which can converse with real Alexa users.

The first Alexa Prize competition was the Social Bot challenge, which has just completed its fourth year [1]. In this challenge, each competing "SocialBot" must converse coherently and engagingly with humans on a range of popular topics. The second competition, the Alexa Prize TaskBot challenge[1], was launched in March 2021, and has a different purpose: the selected academic teams have to build a conversational agent, or "TaskBot", to assist customers in completing tasks requiring multiple steps and decisions.

This challenge was motivated by our north star that Alexa will keep inventing next-generation conversational AI experiences, as our customers' needs change. The TaskBot Challenge is meant to run for three years, with the first year focusing on two domains: cooking and home improvement. One of the unique elements of this challenge is its multi-modal nature, where customers receive both verbal guidance and visual instructions, when a screen is available (e.g., on Echo Show devices).

In order for modern AI assistants to handle a natural conversation around a given task, they should be capable of reasoning, which is a hard AI challenge. As people identify, formulate, and execute real-world tasks, the AI assistant must take into account the real-world context [2] and interpret the user's requests, questions, comments, while providing guidance based on domain knowledge relevant to the task [3]. Accordingly, in this TaskBot challenge. participants had to explore multiple related topics and associated AI assistants capabilities including:

- **Leveraging domain knowledge:** in order to assist Alexa users in their tasks, the agents must identify and represent relevant information about the task, as well as reason about the task steps, outcomes, and goals. Infusing conversations with domain-specific, general knowledge, and common-sense reasoning into the conversation is critical for the success of AI agents[3]. As an example, the challenge participants had to identify and represent the most important information about the task, and break it down into achievable steps, for instance by extracting task-specific knowledge as graphs that augment general knowledge bases.
- **Tracking dialogue state:** in task-oriented conversational systems, dialogue-state tracking refers to the problem of estimating users' goals and requests at each turn of a dialogue. This is a rich research area, as evidenced by numerous studies [4, 5, 6, 7] and associated research challenges (e.g., DSTC[2] now in its tenth instance). Tracking state is even more critical in the context of the TaskBot challenge, as users go to multiple steps as they advance in their task, and in their conversation with Alexa at each turn.

[1] See `https://www.amazon.science/alexa-prize/taskbot-challenge`.

[2] The Dialog System Technology Challenge is "a series of research community challenge tasks for accurately estimating a user's goal in a spoken dialog system", see `https://dstc10.dstc.community/`.

1st Proceedings of Alexa Prize TaskBot (Alexa Prize 2021).

- **Supporting adaptive and robust conversations:** conversational agents should dynamically adapt the dialogue structure to the user's changing needs, be responsive to errors and confusions that might arise during the conversation, as well as potential changes in the physical environment as the tasks progress. The TaskBot challenge allows researchers to test their approaches on diverse users, each with their own needs, communication styles, and personalities, and as such pushes the teams to devise more adaptive and robust conversational agents.
- **Handling multi-modal interactions:** early voice-only AI assistants have now evolved into full multi-modal devices, offering richer experiences to users, who expect the screen to support their conversation. This setting offers the teams a platform to investigate open research questions in multi-modal dialogue systems [8], ranging from foundational capabilities (for instance about the encoding of a multi-modal dialogue state) to user-experience design questions. The latter include, for instance, exploring the appropriate combination of visual and text data to support a fruitful conversation, the interpretation of user touch and voice commands, or the integration of knowledge and visual output into the conversation [9]. These are just some examples of the type of questions TaskBots teams had to address in this first large-scale multi-modal conversation AI challenge.

We were delighted by the high level of engagement from the academic community in the first year. Ten student teams across the world (covering Asia, Europe and the United States) received a US $250,000 research grant each, an Alexa device, as well as free Amazon Web Services (AWS) computing services to support their efforts. In addition, they all received access to the Alexa Prize CoBot toolkit[3] and to the pre-processed knowledge sources from Whole Foods and WikiHow, and extensive technical, design, and business support from the Alexa Prize team and the broader Alexa and Amazon staff.

The Alexa Prize and the Alexa Shopping organizations collaborated this year to develop the infrastructure and lay the groundwork for a successful challenge, and acquired insights along the way on how to improve the experience for both Alexa customers and challenge participants in the future. One such insight is that customers may not be aware of what the TaskBots can do or how they can be helpful. Unlike the previous SocialBot challenge, the TaskBot experience is most helpful if offered at right time, precisely when the customers' needs arise. Determining when and how Alexa could proactively offer such assistance to customers will be the object of future work. This first year also provided critical data to investigate and validate additional evaluation metrics, both manual and automated, to complement the customer satisfaction ratings, which could be used to improve our evaluation methodology in future challenges.

In this first year of the Alexa Prize TaskBot challenge, both the organizers and the teams made significant progress along the research directions mentioned above, which are described in the participants' reports in these Proceedings. We are excited about the current and future advances to the state of Conversational AI that come out of the Challenge, and look forward to novel research ideas that will enable Alexa to assist and delight customers.

References

[1] Dilek Hakkani-Tür. Alexa prize socialbot grand challenge year iv. In *Alexa Prize SocialBot Grand Challenge 4 Proceedings*, 2021. https://assets.amazon.science/df/7a/b376fb90497a946f3531e5d23b8c/alexa-prize-socialbot-grand-challenge-year-iv.pdf.

[2] Yonatan Bisk, Ari Holtzman, Jesse Thomason, Jacob Andreas, Yoshua Bengio, Joyce Yue Chai, Mirella Lapata, Angeliki Lazaridou, Jonathan May, Aleksandr Nisnevich, Nicolas Pinto, and Joseph P. Turian. Experience grounds language. In *EMNLP*, 2020.

[3] Da Yin, Li Dong, Hao Cheng, Xiaodong Liu, Kai-Wei Chang, Furu Wei, and Jianfeng Gao. A survey of knowledge-intensive nlp with pre-trained language models. *ArXiv*, abs/2202.08772, 2022.

[4] Adarsh Kumar, Peter Ku, Anuj Kumar Goyal, Angeliki Metallinou, and Dilek Z. Hakkani-Tür. Ma-dst: Multi-attention based scalable dialog state tracking. In *AAAI*, 2020.

[3] See the Alexa Prize technical report for a detailed description.

[5] Zhaojiang Lin, Bing Liu, Seungwhan Moon, Paul A. Crook, Zhenpeng Zhou, Zhiguang Wang, Zhou Yu, Andrea Madotto, Eunjoon Cho, and Rajen Subba. Leveraging slot descriptions for zero-shot cross-domain dialogue statetracking. In *NAACL*, 2021.

[6] Yawen Ouyang, Moxin Chen, Xinyu Dai, Yinggong Zhao, Shujian Huang, and Jiajun Chen. Dialogue state tracking with explicit slot connection modeling. In *ACL*, 2020.

[7] Hung Le, Richard Socher, and Steven C. H. Hoi. Non-autoregressive dialog state tracking. *ArXiv*, abs/2002.08024, 2020.

[8] Erkut Erdem, Menekse Kuyu, Semih Yagcioglu, Anette Frank, Letitia Parcalabescu, Barbara Plank, Andrii Babii, Oleksii Turuta, Aykut Erdem, Iacer Calixto, Elena Lloret, Elena Simona Apostol, Ciprian-Octavian Truică, Branislava andrih, Sanda Martinić-Ipić, Gábor Berend, Albert Gatt, and Graina Korvel. Neural natural language generation: A survey on multilinguality, multimodality, controllability and learning. *J. Artif. Intell. Res.*, 73:1131–1207, 2022.

[9] Lizi Liao, Yunshan Ma, Xiangnan He, Richang Hong, and Tat-Seng Chua. Knowledge-aware multimodal dialogue systems. *Proceedings of the 26th ACM international conference on Multimedia*, 2018.