

---

# Alexa Prize Socialbot Grand Challenge: Yet Another Year of Progression in Conversational AI

---

**Dilek Hakkani-Tür**

Amazon Alexa AI  
hakkanit@amazon.com

Building conversational systems that enable natural language interactions with machines has been an attractive research area since the early days of computing, as exemplified by earlier text-based systems such as ELIZA [Weizenbaum, 1966]. The research and development work in this area has been increasing since then, with many publications on both task-oriented [Hosseini-Asl et al, 2020; Papangelis et al, 2020; Vlasov et al, 2019; Zhang et al, 2019, among others] and open-domain [Adiwardana et al, 2020; Roller et al, 2020; Yavuz et al, 2019; among others] dialogue applications. Recent advances in deep learning, and especially Transformer based language models also enabled building large chatbot models, using conversational data mined from the web.

One of the main obstacles in conversational systems research is the scarcity of conversational datasets that include real interactions with users. To advance human-computer interaction through conversations, Amazon has been organizing the Alexa Prize Challenges since 2017. In these challenges, university teams are supported to create socialbots that can converse coherently and engagingly with humans on a range of current events and popular topics such as entertainment, sports, politics, technology, and fashion. In addition to the support for accessing Amazon’s computing infrastructure and tools, Alexa Prize Challenges have been providing a unique opportunity for university teams to connect their systems with millions of real users for spoken interactions. Furthermore, automatically evaluating open domain conversational systems is critical for advancing the quality of them and still remains an open question [Celikyilmaz et al., 2016]. Real user ratings coupled with these conversations provide university teams a large-scale experimentation framework, accelerating the advances in open domain conversational response generation systems and socialbots.

In Alexa Prize Challenge 3, 10 university teams were selected to participate in the Alexa Prize Challenge.<sup>1</sup> The participants were announced in June 2019 and received a research grant, Alexa-enabled devices, free Amazon Web Services (AWS) to support their development efforts, and access to other tools (such as the Conversational Bot, i.e., Cobot, tools as described later in the proceedings), data sources, and Alexa team support.

The semifinals period took place between March 20<sup>th</sup> and April 29<sup>th</sup>. During the semifinals, users interacting with Alexa were connected with the teams’ socialbots by saying “*Alexa,*

---

<sup>1</sup> The announcement is at: <https://developer.amazon.com/blogs/alexa/post/7f75052d-ec3e-46ce-b4f8-0da5b0eb0975/announcing-the-new-class-of-alexa-prize-teams>

#####

*let's chat*" to any Alexa-powered device. Competing socialbots were randomly invoked in response to this utterance. At each turn, user's spoken utterances were transcribed using automatic speech recognition and these utterances along with several other high-level metadata (such as confidence scores) were provided to the socialbots, which in return produced textual responses (possibly with Amazon speech synthesis markup language (SSML) markings to format the prosody of these responses) that were converted to spoken utterances by text-to-speech synthesis. At the end of the interaction, users were prompted to provide a rating from 1 to 5 stars (5 being the highest) on how they felt about speaking with that socialbot. After the rating, the users were also asked if they would like to leave a verbal feedback to the university team that built the socialbot they just interacted with. The top 3 teams automatically advanced to finals based on user ratings (Alquist, Emora, Gunrock), and Amazon also selected two teams (Athena and Chirpy Cardinal) as wildcard finalist teams based on the ratings from Alexa customers, depth and breadth of topics covered, appropriateness and accuracy of responses, and scientific merit as determined by the content of the technical paper.

During the previous challenges, Alexa customers had provided feedback on challenging areas, such as smooth transitioning into new topics, staying and going deep on a topic. Based on the feedback and in order to enable the teams to study these problems, before the start of the 3<sup>rd</sup> challenge, Amazon Alexa AI team collected and released the largest knowledge grounded human-human conversation corpus, Topical Chat dataset [Gopalakrishnan, 2019], publicly. Throughout the challenge, Amazon Alexa Prize team has provided a suite of tools and models to the participants including automatic speech recognition with improved word and entity accuracies, the CoBot toolkit, improved dialogue act and topic detection, sensitive content detection, conversation evaluator models, and finally neural response generators trained on the Topical Chat dataset as a back-off response generator. Additionally, our team invested in novel approaches to generate more appropriate and engaging responses through policy-driven response generation [Hedayatnia, 2020]. Furthermore, in order to help university teams to build systems robust to speech recognition (ASR) noise, Amazon released a version of the Topical Chat dataset enriched with simulated ASR errors [Wang et al, 2020].

The first article in the proceedings from the Alexa Prize team presents these advances, in order to support the university teams. These are in addition to the scientific contributions from the previous challenges that have been summarized in [Ram et al., 2018] and [Khatri et al, 2020]. In the following chapters, the Alexa Prize Challenge proceedings are organized to include overviews and contributions of the participating university systems. Many thanks to the Amazon Alexa Prize team and participating university teams for making Amazon Alexa Prize Challenge established and influential, and congratulations to the finalists!

## References

Adiwardana D, Luong MT, So DR, Hall J, Fiedel N, Thoppilan R, Yang Z, Kulshreshtha A, Nemade G, Lu Y, Le QV. Towards a human-like open-domain chatbot. arXiv preprint arXiv:2001.09977. 2020 Jan 27.

Celikyilmaz A, Clark E., Gao J. Evaluation of Text Generation: A Survey. <https://arxiv.org/pdf/2006.14799.pdf>. June 2020.

Gopalakrishnan K, Hedayatnia B, Chen Q, Gottardi A, Kwatra S, Venkatesh A, Gabriel R, Hakkani-Tür D, AI AA. Topical-Chat: Towards Knowledge-Grounded Open-Domain Conversations. In INTERSPEECH 2019 (pp. 1891-1895).

Hedayatnia B, Kim S, Liu Y, Gopalakrishnan K, Eric M, Hakkani-Tur D. Policy-Driven Neural Response Generation for Knowledge-Grounded Dialogue Systems. arXiv preprint arXiv:2005.12529. 2020 May 26.

Hosseini-Asl E, McCann B, Wu CS, Yavuz S, Socher R. A simple language model for task-oriented dialogue. arXiv preprint arXiv:2005.00796. 2020 May 2.

Khatri C, Hedayatnia B, Venkatesh A, Nunn J, Pan Y, Liu Q, Song H, Gottardi A, Kwatra S, Pancholi S, Cheng M, Chen Q, Stubel L, Gopalakrishnan K, Bland K, Gabriel R, Mandal A, Hakkani-Tur D, Hwang G, Michel N, King E, Prasad R. Advancing the state of the art in open domain dialog systems through the alexa prize. arXiv preprint arXiv:1812.10757. Alexa Prize Proceedings,

Papangelis A, Namazifar M, Khatri C, Wang YC, Molino P, Tur G. Plato dialogue system: A flexible conversational ai research platform. arXiv preprint arXiv:2001.06463. 2020 Jan 17.

Ram A, Prasad R, Khatri C, Venkatesh A, Gabriel R, Liu Q, Nunn J, Hedayatnia B, Cheng M, Nagar A, King E, Bland K, Wartick A, Pan Y, Song H, Jayadevan S, Hwang G, Pettigru A. Conversational AI: The science behind the Alexa prize. arXiv preprint arXiv:1801.03604.

Roller S, Dinan E, Goyal N, Ju D, Williamson M, Liu Y, Xu J, Ott M, Shuster K, Smith EM, Boureau YL. Recipes for building an open-domain chatbot. arXiv preprint arXiv:2004.13637. 2020 Apr 28.

Vlasov V, Mosig JE, Nichol A. Dialogue transformers. arXiv preprint arXiv:1910.00486. 2019 Oct 1.

Weizenbaum, J. Eliza – A computer program for the study of natural language communication between man and machine. Communications of the ACM, 9(1):36–45, Jan. 1966. ISSN 0001-0782. <https://web.stanford.edu/class/linguist238/p36-weizenbaum.pdf>

Yavuz S, Rastogi A, Chao GL, Hakkani-Tur D. DeepCopy: Grounded response generation with hierarchical pointer networks. arXiv preprint arXiv:1908.10731. 2019 Aug 28.

Zhang Y, Ou Z, Yu Z. Task-Oriented Dialog Systems that Consider Multiple Appropriate Responses under the Same Context. arXiv preprint arXiv:1911.10484. 2019 Nov 24.