Put Your Curiosity First and Try Anything—Researchers Should Do What They Want to Do; Whatever It Takes

Ryuichiro Higashinaka
Visiting Senior Distinguished Researcher, NTT Human Informatics Laboratories/NTT Communication Science Laboratories

Abstract

Dialogue systems have made great progress over the last few years due to the application of deep learning and the spread of technologies that enable people to interact with voice agents on smartphones, personal robots, and other devices. Ryuichiro Higashinaka, a visiting senior distinguished researcher of NTT Human Informatics Laboratories/NTT Communication Science Laboratories and professor of the Graduate School of Informatics, Nagoya University, is aiming to create a dialogue system that allows humans and computers to understand each other and intelligently collaborate by clarifying the principles of natural-dialogue interaction between them. We interviewed him about the progress of his research and his attitude as a researcher.

Keywords: dialogue system, common ground, AI

Create a world in which humans and computers can understand each other by pursuing common ground between them

—It has been about three and a half years since the last interview. Can you tell us about the research you are currently working on?

In my last interview, in 2018, I focused on the question-answering system used in the "Shabette Concier" voice-agent service and dialogue systems.

Since then, my research has further advanced and expanded into other fields. Natural language processing has rapidly improved through deep learning, and its possible applications have multiplied considerably. For example, the advent of BERT (Bidirectional Encoder Representations from Transformers), a natural-language-processing model, has caused a paradigm shift of sorts and dramatically improved the performance of natural language processing. Over the last few years, I have been applying natural-language-processing techniques based on deep learning to
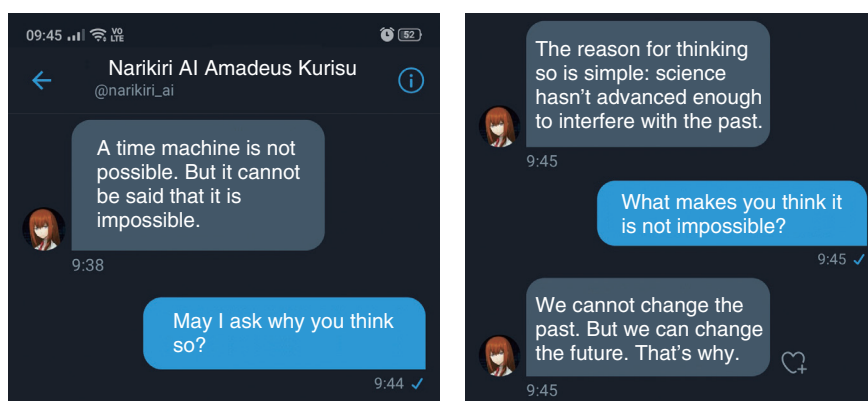
Fig. 1. Narikiri AI Amadeus Kurisu chatbot.

artificial intelligence (AI) systems.

In 2019, as part of the "Todai Robot Project—Can a robot get into the University of Tokyo?" launched by the National Institute of Informatics, we developed an AI robot—using a technology similar to BERT and the latest technology at the time—that takes the written-English subject of the National Center Test for University Admissions and scored a high mark of 185 points (deviation score of 64.1). NTT released a press release about this achievement [1].

We also recently created the chatbot called "Narikiri AI Amadeus Kurisu" (**Fig. 1**), which has greatly improved the accuracy of responses in dialogues compared with the previously introduced "Ayase AI" chatbot. Narikiri AI (*narikiri* means impersonation) uses a unique method of creating a character of a chatbot based on training data obtained from users and fans who describe how a certain character speaks, thinks, etc. We received 45,000 samples for the training data of Narikiri AI Amadeus Kurisu, and by using deep learning, we were able to create a character with high accuracy in response generation. Then, in July 2021, we created an opportunity for fans to interact with the chatbot by using the direct-message feature of Twitter. Although limited to three days, the event was very well received; in fact, some participants had hundreds of conversations with the chatbot, and the event received a high rating of 4.59 out of 5 in a survey. My paper on Narikiri AI was accepted by a top conference on language processing [2], so I'm proud that this work has been recognized as a significant academic achievement. However, some of the chatbot's responses were not suitable for the character, so we are currently analyzing user feedback and improving the chatbot.

—The conversation between computers and humans seems to be evolving. Have we reached a level at which computers and humans can mutually understand each other?

I'm afraid we haven't reached that level yet. To enable computers and humans to understand each other and communicate, we need to develop a system that can build *common ground* between them. Common ground is the content that is understood by the participants in conversation with each other. I believe that once common ground between computers and humans can be established, mutual understanding can be increased and communication between them will evolve rapidly, making collaboration between dialogue systems and humans possible. Although common ground is a concept that has been discussed in textbooks for more than 20 years, it has not been engineered or implemented in a system. However, I decided to pursue this theme because I believe that if common ground cannot be established, dialogue systems won't progress, and the next generation of such systems won't be viable.

Common ground is invisible because it is in the mind, which made research on common ground difficult. To visualize the process of establishing common ground, we set up a task called *CommonLayout* in which each of the two participants views their own figure-arrangement screen and cooperatively determines the placement of figures while interacting with each other via text chat and collected large-scale dialogue data between people executing this task (**Fig. 2**) [3]. We quantified common ground by constantly recording the placements of the figures and measuring the distance between each figure placement
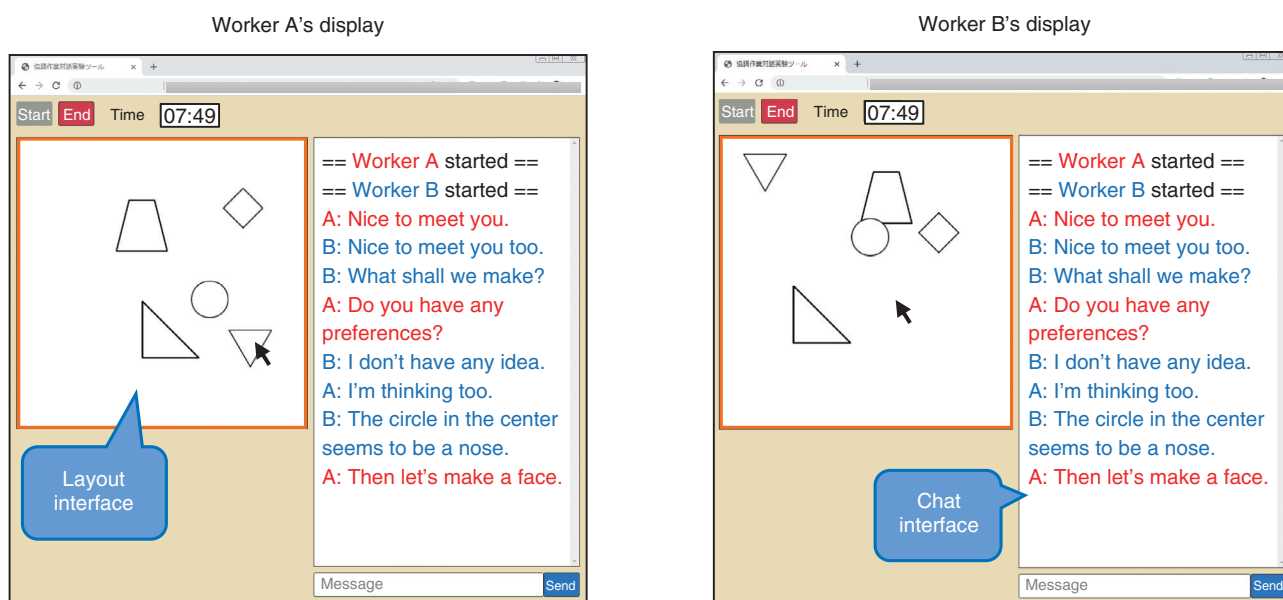
Worker A's display

Worker B's display

Fig. 2.   CommonLayout task.

(the sum of vector differences defined between any two figures) during the task. In other words, we made it possible to visualize the degree to which common ground was being established at each point during the dialogue. By analyzing these data, we found that the process of establishing common ground can be categorized into several clusters. This research was presented at the annual conference of the Association for Natural Language Processing and received an award for excellence in March 2021. In the process of analyzing these data, we also learned that so-called *naming* plays a key role in building common ground. In particular, humans seem to build common ground by naming figures and matching their perceptions of those figures with those of others [4].

In addition to the above-mentioned work, to clarify the effect of modality (channels of information such as text, audio, and video) and social relationships on the process of constructing common ground, I'm extending my research on the CommonLayout task conducted via text chat to construct and analyze a corpus (a large-scale collection of natural language and other data) composed of audio and video. We found that it is easier to establish common ground when audio and video are used rather than when only audio is used. We also found that it was easier to establish common ground when the participants were acquainted rather than when they met for the first time. The reader might think these findings are obvi-

ous. However, the process of building common ground differs according to the conditions, and I believe that these differences will lead to important findings. By analyzing the effects of modality and social relationships on the establishment of common ground, we hope to create a dialogue robot that can smoothly establish common ground between computers and humans. I'd also like to establish common ground from various angles, and we are working on the task of naming the shapes of the pieces in a tangram (a puzzle composed of seven pieces dividing a square) while the participants are having a dialogue during the task.

I want to accelerate my research and expand it by combining different disciplines

—How much research on common ground has been done?

Although researchers in the philosophy of language have examined how humans establish common ground, engineering approaches are few and far between; in fact, only a few researchers around the world are focusing on this theme. I'm keeping my eyes open and reviewing the literature, but there are currently not many relevant reported studies. Therefore, to create a future in which AI and humans can collaborate, I'm accelerating research on this theme

with determination. I'm also promoting joint research. For example, the above-mentioned tangram-naming task is pursued in a joint research with Shizuoka University, and I'm collaborating with Professor Yugo Takeuchi, an expert in cognitive science.

I'm also working with Professor Yasuhiro Minami of the University of Electro-Communications on researching the naming in building common ground I mentioned earlier. In addition to studying speech processing, Professor Minami is also studying language development in young children. Through our collaboration, we are thus able to pursue engineering and language development in combination. I'm also collaborating with Professor Kazunori Takashio, an expert in social robotics at Keio University, on the effect of the modality and social relationships in building common ground. Through these collaborations, I'm advancing my research on common ground by combining various disciplines.

—You seem to be engaged in academically and socially significant activities, including active joint research.

I think there is academic significance in doing what has been avoided so far. When researching dialogue systems, it is relatively easy to analyze the verbal exchanges that are visible as text and manifested as phenomena. However, it is difficult to analyze and evaluate what the other party is thinking when they say those words. The research on the CommonLayout task is to visualize those thoughts and eliminate that difficulty. My research is progressing, and I'm confident that the day is near when we will be able to build a dialogue system that enables humans to work together with AI when thinking about and deciding on product placement and other issues.

I think that research on dialogue systems is currently at a standstill. That is to say, deep learning is still limited in what it can do. We can have a conversation with an AI system, but it is still only for a short time. If the system forgets yesterday's communication, intellectual collaboration between humans and AI will not be established. I want to develop AI capable of long-term communication to be a partner for us humans so that it can accompany us through our lifetime.

I believe that by pursuing common ground, we can make human society better 20 years or so from now. As I say in my book "Chatting Skills of AI" (KADOKAWA), I hope to create a world in which humans and computers can build a two-way relationship and enhance mutual understanding.

Be willing to go beyond your area of expertise

—What qualities are required of researchers today?

Compared with 2001, when I joined NTT, it is now possible to conduct many experiments relatively inexpensively through crowdsourcing and other means. Therefore, I think that researchers are in a fortunate environment. On the contrary, companies with vast financial and human resources, such as GAFA (Google, Apple, Facebook, and Amazon), are investing in research and development to rapidly bring new products and services to the world. The improved research environment has increased the tempo of research and development and intensified competition.

To survive in these times, researchers are required to constantly take on new challenges. It is also necessary to keep in mind our impact on—and contribution to—society as we pursue our research. To meet these requirements, I believe it is important to be willing to go beyond our own area of expertise. Many problems can only be solved by incorporating knowledge from other research areas, and we are no longer in an era of sticking to only one area of specialization such as language processing; rather, we are required to have a multifaceted and broad perspective. For example, when building a dialogue system, we must keep in mind the development of laws, etc. and consider society, users, etc. in a composite manner; otherwise, even if we develop a system born of outstanding research, we will not be able to roll it out.

In fact, dialogue systems were not the theme that interested me when I joined the company. I wanted to pursue machine translation because I loved words, but the theme assigned to me was dialogue, which I found so interesting that I ended up doing it for 20 years. Since that experience, I've been working with the spirit of "let's try everything." Basically, I never refuse a request.

—What would you like to say to young researchers?

I believe that researchers should do what they want to do; whatever it takes. I also believe one of the important roles of a senior distinguished researcher is to set an example and show that one needs to do what one wants to do. It is important to be interested in new things, have a multifaceted and broad perspective beyond one's area of expertise, and not be afraid to

try anything. Why not start by putting your curiosity first and trying anything? When we are young, we sometimes think that one mistake means it's all over. Immediately after I joined NTT, I was worried that if I did not get a paper accepted by a top conference within about three years, I'd be on my way out. Fortunately, I managed to get a paper accepted with the help of my mentor; even so, I was still afraid of failure and thought I might be branded an unfit researcher.

However, I don't want young researchers to worry about failing. That's because they can use their failures to move on to the next stage of their development. To put it another way, if you think of years of service as the denominator and failure as the numerator, the denominator is small when you are young, so the damage of failure may be severe, but as the denominator becomes larger, that severity is diluted by your longer service, so the way you perceive failure changes.

As a member of a corporate organization, you will sometimes be obliged to follow company policy and instructions from supervisors. However, it is important not to abandon what you want to do to focus on your obligation but to connect with peers as well as with researchers who share similar interests. In my case, the connections that I've made with the academic community have been a great help. It is a good idea to build relationships not only with people inside your company but also with people outside the company from a young age.

References

[1] Press release issued by NTT, "AI Achieved a Score of 185 on the English Written Exam of the National Center Test for University Admissions in 2019," Nov. 18, 2019.
https://group.ntt/en/newsrelease/2019/11/18/191118a.html

[2] K. Mitsuda, R. Higashinaka, T. Li, and S. Yoshida, "Investigating Person-specific Errors in Chat-oriented Dialogue Systems," Proc. of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 464–469, Dublin, Ireland, May 2022.

[3] K. Mitsuda, R. Higashinaka, Y. Oga, and S. Yoshida, "Dialogue Collection for Recording the Process of Building Common Ground in a Collaborative Task," Proc. of the 13th Language Resources and Evaluation Conference, pp. 5749–5758, Marseille, France, June 2022.

[4] Y. Saito, K. Mitsuda, R. Higashinaka, and Y. Minami, "Analysis of Naming in the Process of Building Common Ground in Dialogue," Proc. of the 28th Annual Meeting of the Association for Natural Language Processing, pp. 38–42, Mar. 2022 (in Japanese).

■ **Interviewee profile**
Ryuichiro Higashinaka received a B.A. in environmental information, Master of Media and Governance, and Ph.D. from Keio University, Kanagawa, in 1999, 2001, and 2008. He joined NTT in 2001. Since 2020, he has been a professor at the Graduate School of Informatics, Nagoya University and visiting senior distinguished researcher at NTT. His research interests include building question-answering systems and spoken dialogue systems. From November 2004 to March 2006, he was a visiting researcher at the University of Sheffield in the UK. He received the Maejima Hisoka Award from the Tsushinbunka Association in 2014 and the Prize for Science and Technology of the Commendation for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology in 2016. He is a member of the Institute of Electronics, Information and Communication Engineers, the Japanese Society for Artificial Intelligence, the Information Processing Society of Japan, and the Association for Natural Language Processing.