J SYST SCI SYST ENG
Vol. 30, No. 4, August 2021, pp. 450–464
DOI: https://doi.org/10.1007/s11518-021-5497-1

ISSN: 1004-3756 (paper), 1861-9576 (online)
CN 11-2983/N

Argumentative Conversational Agents for Online Discussions

Rafik Hadfi,[a] Jawad Haqbeen,[b] Sofia Sahab,[a] Takayuki Ito[a]

[a]Department of Social Informatics, Kyoto University, Yoshidahonmachi, Sakyo Ward, Kyoto, Japan 606-8501
rafik.hadfi@i.kyoto-u.ac.jp (✉), sahab.sofia.4h@kyoto-u.ac.jp, ito@i.kyoto-u.ac.jp
[b]Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya, Aichi, Japan 466-8555
jawad.haqbeen@itolab.nitech.ac.jp

Abstract. Artificial Intelligence is revolutionising our communication practices and the ways in which we interact with each other. This revolution does not only impact how we communicate, but it affects the nature of the partners with whom we communicate. Online discussion platforms now allow humans to communicate with artificial agents in the form of socialbots. Such agents have the potential to moderate online discussions and even manipulate and alter public opinions. In this paper, we propose to study this phenomenon using a constructed large-scale agent platform. At the heart of the platform lies an artificial agent that can moderate online discussions using argumentative messages. We investigate the influence of the agent on the evolution of an online debate involving human participants. The agent will dynamically react to their messages by moderating, supporting, or attacking their stances. We conducted two experiments to evaluate the platform while looking at the effects of the conversational agent. The first experiment is a large-scale discussion with 1076 citizens from Afghanistan discussing urban policy-making in the city of Kabul. The goal of the experiment was to increase the citizen involvement in implementing Sustainable Development Goals. The second experiment is a small-scale debate between a group of 16 students about globalisation and taxation in Myanmar. In the first experiment, we found that the agent improved the responsiveness of the participants and increased the number of identified ideas and issues. In the second experiment, we found that the agent polarised the debate by reinforcing the initial stances of the participant.

Keywords: Artificial intelligence, conversational agents, natural language processing, online discussion, computational social science

1. Introduction

The field of Artificial Intelligence (AI) is gradually changing the ways in which humans interact with each other. Most importantly, it is changing the nature of the partners with whom we communicate. It is currently possible to communicate with artificial agents in the forms of virtual assistants, socialbots, or chatbots.

With the increasing sophistication of Natural Language Generation (NLG) techniques, it is not inconceivable for humans to forget that they are actually interacting with a machine, in what is known as the "pretended intimacy" (Weizenbaum 1966). While it is believed that it will take many years until we start seeing the widespread use of such chatbots, their development raises important ethical questions like their deliberative nature, moral accountability, and the extent of their polarising capabilities. Such issues need to be addressed at this early stage, and the design choices made by the developers of such agents need to be carefully scrutinised.

Recent studies have shown that online platforms are vulnerable to deceptive automated activities. For instance, Stella et al. (2018) showed that socialbots deliberately targeted central election hubs with inflammatory content during the 2017 Catalan independence referendum. Similarly, Ferrara (2017) investigates the disinformation campaign that has been coordinated by means of socialbots disguising themselves as legitimate human users during the 2017 French presidential election. Shao et al. (2018) analysed 14 million messages spreading 400 thousand articles on Twitter and found evidence that socialbots played a disproportionate role in spreading articles from low-credibility sources. In an attempt to characterise such bots and their impact on social media, Varol et al. (2017) presented a framework for the detection of socialbots on Twitter. In this light, we propose to investigate how an elaborate conversational agent could gradually influence the evolution of an online debate. We particularly look at the persuasive consequences associated with an agent that uses argumentative cues to influence the opinions of human debaters. The debaters start by taking a stance on a predefined theme and then try to elaborate and defend their positions by posting their ideas and arguments on an online forum. The agent will adaptively reply to their messages by mediating the discussion or by supporting or attacking the users that agree (or disagree) with the main stance. This work has two main contributions.

1. A platform centred around an intelligent conversational agent that uses Natural Language Processing, Natural Language Generation, and argumentative reasoning to interact with humans in online discussions.
2. A study on the effect of polarised and non-polarised conversational agents in online discussions between human participants.

The results suggest that the agent could increase the responsiveness of the participants and their ability to identify ideas and issues. Second, when the debaters have prior knowledge of the issue, their stances do not change under the effect of a bipolarised agent.

The paper is structured as follows. In section 2, we cover the literature relative to conversational agents and their effects in online discussions. In section 3, we employ a methodology based on an artificial agent and a social experiment. In section 4, we present the results of the experiment. Finally, we summarise our major findings and provide future research directions.

2. Related Work

Discussion platforms are considered as the next-generation democratic platforms for citizen deliberation. Such platforms could integrate ideas, opinions, and could lead to enhanced consensuses (Malone 2018, Malone and Klein 2007). For instance, the Collagree platform was employed for opinion gathering and city planning in Japan (Ito et al. 2019 2014 2015). The CoLab platform was used to harnesses the collective intelligence of thousands of people worldwide to address global climate change (Malone and Klein 2007). The Deliberatorium is another platform where people submit ideas by following an argumentation map that frames the ideas within a given discussion structure (Iandoli et al. 2007).

Such platforms are also being used to empower citizens and help implementing sustainable goals (Savaget et al. 2019). For instance, the D-Agree platform was employed to collect opinions and the implementation of Sustainable Development Goals in Afghanistan (Haqbeen et al. 2020a). Another work has recently used the same platform to fight COVID-19 by collecting and analysing vast amounts of social data to increase public awareness and for public health policy-making (Haqbeen et al. 2020b).

In practice, sophisticated discussion platforms combine algorithmic methods and machine learning techniques to harness the intelligence of the crowd. In our work, we particularly focus on the use of artificial intelligent agents for their ability to adapts to human behaviour and to the problems at hand. In this case, a conversational agent is defined as a computer program that is designed to interact with users using natural language in ways that mimic human conversation. Most of the existing chatbots utilise algorithms to generate adequate responses. The earlier versions of conversational agents merely created an illusion of intelligence by employing much simpler pattern matching and rule-based models in their interaction with users. However, with the emergence of new technologies, more intelligent systems have emerged with learning methodologies and knowledge-based models. Conversational agents are generally classified into categories based on the knowledge domain, the mode of interaction, and the design aspects. Overall, we distinguish task oriented agents and non-task oriented agents (Chen et al. 2017, Yan et al. 2017). Task-oriented agents are designed for a particular task and are set up to have short conversations, usually within a closed domain such as online shopping, customer support, or medical expertise. When the task is not specific, a non-task oriented agent can simulate a conversation with a person for entertainment purposes in open domains (Hussain et al. 2019). Many approaches could be employed when building task-oriented conversational agents. Such approaches generally use predefined rules, information retrieval, or generative models. Each of these approaches could rely on multiple techniques such as parsing (Weizenbaum 1966), pattern matching (Wallace 2009), ontologies (Al-Zubaide and Issa 2011), and more recently using Artificial Neural Networks (Nuez Ezquerra 2018, Csaky 2019). The approach we are adopting here is a combination of rule-based and generative methodologies. Most importantly, the generated utterances must be constrained by the argumentative nature of the discussion discourse.

Conversational agents are expected to make judgments when interacting with humans or with other artificial agents. Any knowledge that is required for such decisions may be missing, incoherent or conflicting. Formal argumentation is a viable approach for handling conflicting opinions and beliefs. It is the process by which arguments are constructed, compared, and evaluated in order to establish whether any of them are justifiable. Argumentation is inherent to human reasoning and to our native ability to decide collectively about a problem. It is therefore important to conceive of an autonomous conversational agent that can exploit argumentation theories in order to reason about complex problems. Argumentation has been applied in various domains and its applications range from decision making to negotiation (Dung 1995, Fox and Parsons 1997, Amgoud and Parsons 2000). In general, an argumentation process consists in the construction of the arguments, the definition of the interactions between the arguments, the evaluation of each argument, the selection of the acceptable arguments, and the conclusion. An interesting type of interaction between arguments is the case where an argument can defeat or support another argument. These two independent types of information suggest a notion of bipolarity illustrated by the defeat or support relations. Bipolarity has been widely studied in different domains such as knowledge and preference representation, and of course in argumentation frameworks (Cayrol and Lagasquie-Schiex 2005, Amgoud et al. 2008). Herein, we use bipolar arguments in a controllable fashion since they represent the main components encountered in normal discussions and in debates. Using these two relationships, our conversational agent will engage in discussions with humans. Part of our goals is to look at the polarisation effect of these two relationships on discussions.

Several studies have investigated the effect of artificial agents on social media platforms. The work of Ozer et al. (2019) focuses for instance on the polarising effects of bot activities on a political social media network. By studying the retweet network of 3.7 million users during the tragic Stoneman Douglas High School shooting event, the authors found that bot accounts heavily contributed to online polarisation. Bots lead to statistically significant increased polarisation on 65% of the most popular debate related hashtags. In another study by Bail et al. (2018), the authors surveyed a large sample of democrats and republicans who visit Twitter at least three times each week about a range of social policy issues. They found that the republicans who followed a liberal Twitter bot became substantially more conservative while democrats exhibited slight increases in liberal attitudes after following a conservative Twitter bot. The results suggest a phenomenon known as the "echo chambers" (Sunstein 2001) whereby social media sites contribute to political polarisation by creating “echo chambers” that insulate people from opposing views about current events. The current work suggests a similar phenomenon where instead of being isolated in chambers, the positions of the debaters are reinforced by an agent that creates a bipolarisation of the discussion through means of argumentative attacks or supports.

3. Methodology

3.1 Social Experimentation Using AI

Our general methodology is to conduct a discussion between human participants and then introduce a conversational agent. We will distinguish two types of discussions: a non-polarised discussion where the agent is encouraging the participants to provide ideas and raise issues; and a polarised discussion with an agent that is actively supporting some participants while attacking others. The general methodology combines humans and artificial agents to identify generalisable mechanisms that might give rise to emergent properties of hybrid social systems (Keuschnigg et al. 2018). The methodology benefits from computational tools such as agent-based simulations, machine learning, and large-scale web experimentation.

3.1.1 Non-polarised Discussions

There were two objectives behind the first social experiment. First, the Kabul municipality wanted to promote public engagement by collecting insights from the citizens for city-related Sustainable Development Goals. Second, we wanted to verify the effect of our conversational agent in non-polarised discussions by conducting a large-scale social experiment. The call for participation has been posted by Kabul municipality official Facebook and home pages, and the municipal government asked Kabul residents to register for the social experiment. We received 1076 registrants, in which 76 were female. The experiment was conducted from January 20th to March 18th divided to two equal phases. The discussion themes were the following.

1. How Sustainable Development Goals should be adopted effectively in Kabul city?
2. How to promote citizens’ participation and civic engagement in Kabul city?

The first phase was conducted without the artificial agent and the second phase used the artificial agent. We chose not to activate the agent in the first phase to kick start the discussion by the human participants alone and to allow the agent to learn from the mined data.

3.1.2 Polarised Discussions

The second experiment was conducted with 16 Computer Science graduate students over a period of two days. The participants were divided uniformly based on their gender and were assigned anonymous nicknames. The participants discussed the two following themes.

1. Do you agree with the recent governmen-

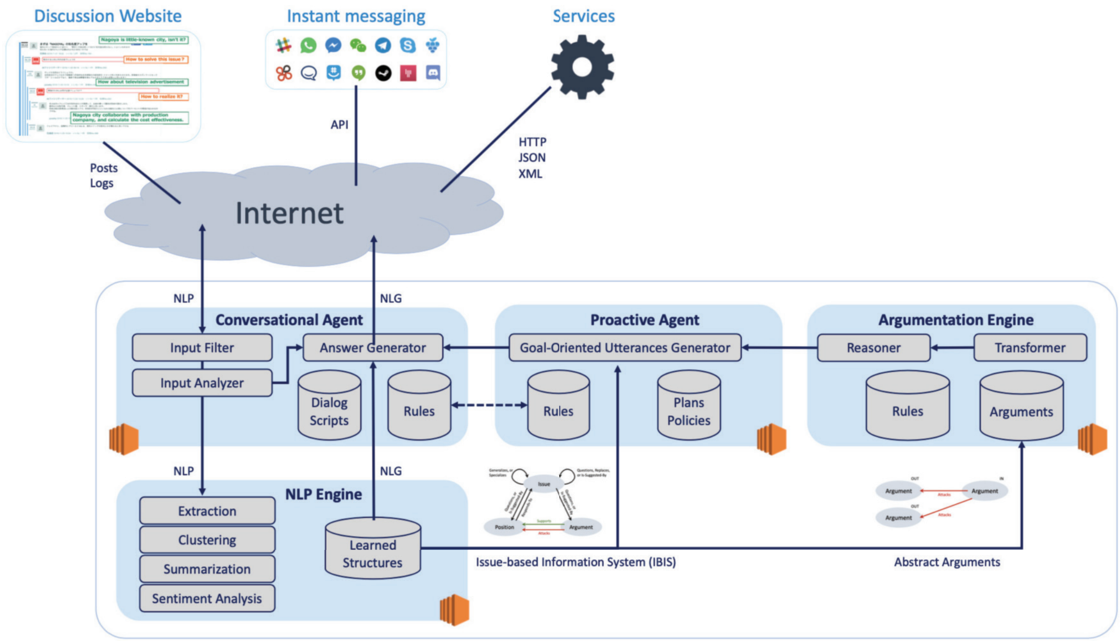


Figure 1 Architecture of the Conversational Agent and the Online Discussion Platform

tal taxation as a way to improve education in Myanmar?

2. Should we promote globalisation in Myanmar?

Each theme of the experiment was allocated a whole day with an intermediary period for resting. The two themes are selected from a list of candidate themes decided beforehand through a face to face brainstorming that involved all of the participants. This is performed in order to create a polarisation similar to the democrat/republication split in Bail et al. (2018). Before and after the online discussion, the participants are surveyed for their agreement or disagreement on the stance of each theme. The users will engage in argumentative discussions to address the issues by providing supportive or oppositional arguments, proposing ideas and positions, or by raising new issues. After conducting the experiment, we measured the level of polarisation of the participants by looking at their stances as well as the ratios of their posted argumentative messages. The 16 participants are divided into two groups: group M with 11 participants with a prior knowledge on the issues, and group J with 5 participants with no knowledge on the issues. The conversational agent will be biased towards group M and unbiased towards group J. In other words, the agent will either attack or support the participants of group M while remaining in a neutral mode with the participants of group J. The neutral behaviour consists in replying to the participants with non argumentative statements such as mediation and facilitation messages.

3.2 The Discussion Platform

The agent platform is composed of a number of modules that interact within the architecture of Figure 1. The architecture is composed of two parts. The bottom part of the figure hosts the agent modules. The top part contains the front-end that interacts with the users through the discussion website or any other messaging service.

The overall agent platform is composed of four agent modules: a conversational agent, a proactive agent, a Natural Language Process-

Table 1 Features of the Sentences Generated by the Conversational Agent

Tag	Description		
Text	Generated sentence that will be posted to the discussion.		
Meaning	Semantics behind the sentence. Useful for the analysis of the discussion.		
Source, Target	Types of the IBIS elements that characterise the sentence and its parent post. Source and target take values in {None, Any, Issue, Idea, Cons, Pros}		
Sentiment	Sentiment of the sentence.		
	Examples	Neutral: "We begin our discussion now."	
		Positive: "Thank you for your valuable contribution"	
		Negative: "I sense some weaknesses to this idea."	
Category	*start_declare*	Declarative sentence at the beginning of a discussion.	
		Example: "Today's discussion theme is about climate change."	
	end_declare	Concluding sentence.	
		Example: "Thank you all for attending today's debate We hope we will see you soon in another discussion".	
	req_child	Sentences set using "target" given "source".	
		Examples	Issue-Idea: Sentence contains an idea related to a parent issue.
			Idea-Cons: Sentence contains a cons related to a parent idea.
			Any-Idea: Sentence contains an idea posted after any message.
	req_thread	Used to solicit or probe for new opinions.	

ing (NLP) engine, and an argumentation engine.

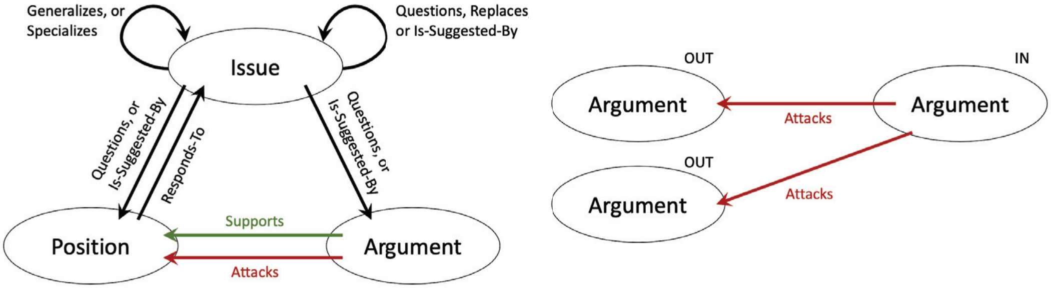
3.2.1 Conversational Agent

The conversational agent interacts directly with the users throughout the website, messaging applications, or external services. This module generates replies based on a set of rules that dynamically combine dialogue utterances depending on the expected behaviour of the agent: mediating, attacking or supporting the posts. A basic dialogue between two humans has a particular structure that usually starts with greetings and declarative sentences, and then dives into the topic of discussion before ending it with concluding sentences. The similar structure is followed by the agent when interacting with the participants. The utterances of the agent are categorised depending on their position in the discussion and whether they contain IBIS elements or not. Table 1 illustrates the types of sentences that the agent generates in a discussion.

3.2.2 Proactive Agent

The proactive agent performs goal-driven actions according to predefined plans and policies that set the behaviour of the agent and guide the discussion towards desired outcomes. There are three main types of policies adopted by the proactive agent.

1. **Consensus policy**. This policy characterises discussions that start from a particular topic, moves to a discussion , and ends with a deliberation. If the discussion is set as a consensus, the proactive agent will provide utterances with types that evolve over time according to the temporal sequence: starting utterances, discussion mediation utterances, and deliberation utterances. Table 1 illustrates these types in the categories of the sentences that the agent could generate.
2. **Brainstorming policy**. This policy characterises a discussion that does not have a specific target or conclusion but seeks an active participation that is often quantifiable by the size of the produced textual content. The proactive agent in this case will probe the participants for more ideas and comments. For instance, a probing message could take the form of "Thank you for the idea. Could you elaborate


Generalizes, or Specializes
Questions, Replaces or Is-Suggested-By
Issue
Questions, or Is-Suggested-By
Responds-To
Questions, or Is-Suggested-By
Position
Supports
Attacks
Argument
OUT
IN
Argument
Attacks
Argument
OUT
Attacks
Argument

(a) The Issue-based Information System (IBIS)

(b) An Abstract Argumentative Graph

Figure 2 Discussion and Argumentation Models

more on the drawbacks of your suggestion?".

3. **Voting policy**. This policy characterises discussions in which the participants are allowed to rank their contributions by liking the posts. This policy sets the agent behaviour as a function of the votes that the posts receive. For instance, a post that gets fewer votes (Likes) will receive comments from the agent as to incentivise the participants into scrutinising the post.

In our experiments, the agent will adopt a consensus policy since we are interested in specific topics of discussion, in an argumentative setting.

3.2.3 Natural Language Processing Engine

This module extracts the arguments from the discussion for classification, clustering, summarisation, and sentiment analysis. The extraction of the arguments is done by classifying the different nodes into issues (or questions), positions (or ideas), and arguments (pros and cons) according to the issue-based information system (IBIS) (Kunz and Rittel 1970). The IBIS system, illustrated in Figure 2a, is an argumentation-based model that is suitable for wicked or ill-defined problems that involve multiple stakeholders (Kolko 2012). Here, we rely on the IBIS model due to the argumentative nature of the discussions.

In practice, extracting the nodes form a discussion thread requires us to represent it in terms of natural language sequences composed of words. To this end, we use a particular type of Recurrent Neural Network (RNN) called Bidirectional Long Short-Term Memory (Bi-LSTM), which is often used to classify time series (Graves and Schmidhuber 2005). The input is the embedding of each word using fastText (Bojanowski et al. 2017) and the output is a normalised probability. We then consider a sentence as the type of the node that acquires the highest probability amongst four possible labels (issues, ideas, pros, and cons.) More details on the algorithmic and implementation details are found in Suzuki et al. (2019)

3.2.4 Argumentation Engine

This module performs reasoning and inferences on the constructed arguments. That is, the engine transforms the IBIS representation into an argumentation graph (Dung 1995) and reasons about its elements. In practice, the engine needs to decide whether subsets of the arguments are valid or not according to predefined semantics (Baroni and Giacomin 2007). The optimal arguments will later be used by the agent to construct new arguments or to query the debaters for new ones.

Reasoning about the arguments requires an adequate representation of the discussion content. In this case, we need to transform the IBIS representation into a bipolar argumentation framework (Cayrol and Lagasquie-Schiex 2005). For instance, the IBIS graph of Figure 2a will be mapped to an abstract argumentative graph similar to the graph of Figure 2b. Formally, a bipolar argumentation graph is a triple $G = A, \mathcal{R}^-, \mathcal{R}^+$ consisting of a finite set of arguments A and two binary relations $\mathcal{R}^-$ and $\mathcal{R}^+$ on A representing the attacks and supports. Any framework G could be represented as a directed graph called argument graph with nodes A and two types of edges $\mathcal{R}^-$ and $\mathcal{R}^+$. Figure 2b illustrates an example of argument graph with three arguments and two attack relations (Dung 1995).

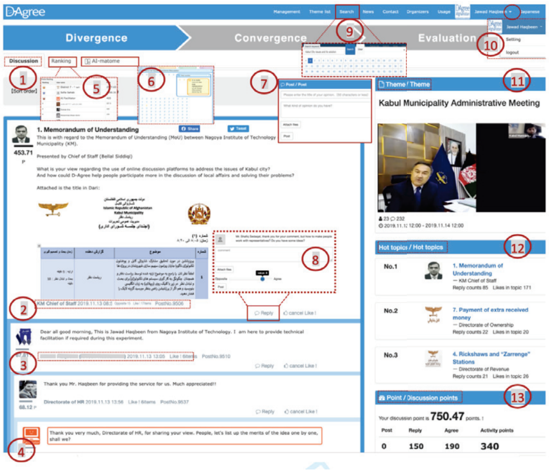
Once the argumentative graph is obtained, the argumentation engine will before a number of operations in order to find the optimal arguments. Following Figure 1, the process starts from the debaters joining an online discussion workspace and engaging in a discussion. The debate starts from one general issue or question that needs to be elaborated by the users. The agent will then operate on the content according to the following algorithm.

1. Collect and restructure the discussion content as a tree.
2. Classify the nodes of the tree based on their IBIS types: issues, positions, pros, or cons. This step uses the output of the NLP Engine.
3. Construct or update the IBIS tree representation of the content.
4. Construct or update the argument graph from the IBIS tree.
5. Set the desired argumentation semantics.
6. Infer the target arguments based on predefined semantics: neutral mediation, attack, or support.
7. Apply Natural Language Generation (NLG) to the the obtained argument(s).
8. Post the generated messages to the discussion at a particular argument point.
9. Go to step 1 and repeat.

In step 5, the agent needs to use specific semantics in order to decide whether some arguments are valid or not in the general course of the discussion. The most relevant semantics we managed to adopt are the admissibility (Nofal et al. 2019), completeness (Bistarelli et al. 2017), and preferred extension (Nofal. et al. 2019). The rules of reply generation are defined as to incentivise the debated to provide either supportive or attacking arguments to an existing argumentation graph. For instance, the agent could prompt debater 1 to provide an argument a_1 to support debater 2's argument a_2, or request debater 3 to attack the argument a_2 with argument a_3. This process, if performed consistently, could increase the completeness of the subsets of the arguments. This will steer the debate by shifting the focus on important issues and strengthening weak argumentative spots. One could imagine an ideal debate to be represented by a quasi complete argument graph, with all arguments being supported in an iterative process involving the conversational agent and the human debaters.

4. Experimental Setting

In the first experiment the agent plays the role of a moderator with supporting messages. In the second experiment, the agent will attack or support users based on one predefined rule: if the agent encounters a message provided by a participant that agrees with the main stance of the discussion, then the agent will reply with a positive message. If the participant disagrees with the stance, then the agent will reply with an attacking message. For example, a supporting message to an issue would look like "{issue} is a good perspective. Who else shares {name} perspective?". An example of an attacking message would look like "{name}, could you provide something more constructive?". Here, the variable {name} is the name of the author of the post, and {issue} is the issue extracted from the message. Other elements such as the pros or cons are also mined from the participants' messages and are exploited in



D-Agree Management Theme list Search News Contact Organizers Usage Jawad Haqbeen Japanese

Divergence Convergence Evaluation

Discussion Ranking AI-matome

[Sort order]

Post / Post

Please enter the title of your opinion. (50 characters or less)

What kind of opinion do you have?

Attach files

Post

1. Memorandum of Understanding

This is with regard to the Memorandum of Understanding (MoU) between Nagoya Institute of Technology Municipality (KM).

Presented by Chief of Staff (Bellal Siddiqi)

What is your view regarding the use of online discussion platforms to address the issues of Kabul city?
And how could D-Agree help people participate more in the discussion of local affairs and solving their problems?

Attached is the title in Dari:

Islamic Republic of Afghanistan
Kabul Municipality

453.71 P

KM Chief of Staff 2019.11.13 08:5 Opposite 15 Like ! 17items PostNo.9506

Mr. Shafiq Sedaqat, thank you for your comment, but how to make people work with representatives? Do you have some ideas?

Attach files

Opposite Agree

Post

Reply cancel Like !

Dear all good morning, This is Jawad Haqbeen from Nagoya Institute of Technology. I am here to provide technical facilitation if required during this experiment.

[redacted] 2019.11.13 13:05 Like ! 6items PostNo.9510

Reply cancel Like !

Thank you Mr. Haqbeen for providing the service for us. Much appreciated!!

68.12 P

Directorate of HR 2019.11.13 13:56 Like ! 6items PostNo.9537

Reply cancel Like !

Thank you very much, Directorate of HR, for sharing your view. People, let's list up the merits of the idea one by one, shall we?

Jawad Haqbeen

Setting

logout

Theme / Theme

Kabul Municipality Administrative Meeting

23 232
2019.11.1: 12:00 - 2019.11.14 12:00

Hot topics / Hot topics

No.1 1. Memorandum of Understanding
— KM Chief of Staff
Reply counts 85 Likes in topic 171

No.2 7. Payment of extra received money
— Directorate of Ownership
Reply counts 22 Likes in topic 20

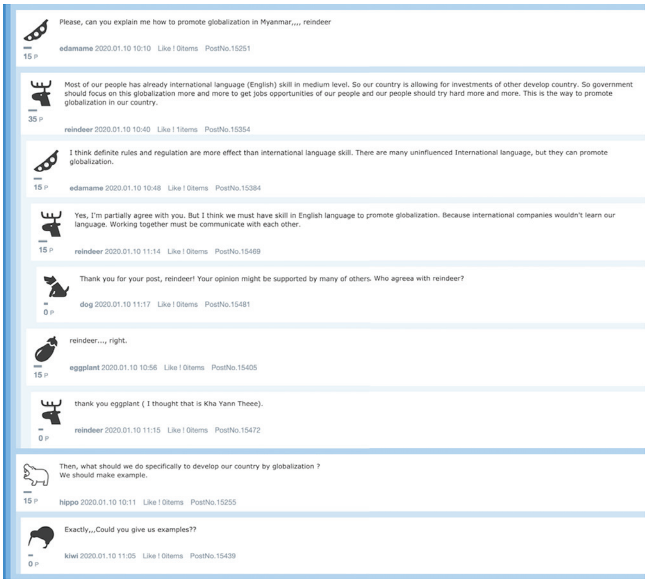
No.3 4. Rickshaws and "Zarrenge" Stations
— Directorate of Revenue
Reply counts 21 Likes in topic 26

Point / Discussion points

Your discussion point is **750.47** points. !

Post	Reply	Agree	Activity points
0	150	190	340

(a) The main interface of the discussion for the Kabul experiment



Please, can you explain me how to promote globalization in Myanmar,,,, reindeer

15 P edamame 2020.01.10 10:10 Like ! 0items PostNo.15251

Most of our people has already international language (English) skill in medium level. So our country is allowing for investments of other develop country. So government should focus on this globalization more and more to get jobs opportunities of our people and our people should try hard more and more. This is the way to promote globalization in our country.

35 P reindeer 2020.01.10 10:40 Like ! 1items PostNo.15354

I think definite rules and regulation are more effect than international language skill. There are many uninfluenced International language, but they can promote globalization.

15 P edamame 2020.01.10 10:48 Like ! 0items PostNo.15384

Yes, I'm partially agree with you. But I think we must have skill in English language to promote globalization. Because international companies wouldn't learn our language. Working together must be communicate with each other.

15 P reindeer 2020.01.10 11:14 Like ! 0items PostNo.15469

Thank you for your post, reindeer! Your opinion might be supported by many of others. Who agreea with reindeer?

0 P dog 2020.01.10 11:17 Like ! 0items PostNo.15481

reindeer..., right.

15 P eggplant 2020.01.10 10:56 Like ! 0items PostNo.15405

thank you eggplant (I thought that is Kha Yann Theee).

0 P reindeer 2020.01.10 11:15 Like ! 0items PostNo.15472

Then, what should we do specifically to develop our country by globalization ?
We should make example.

15 P hippo 2020.01.10 10:11 Like ! 0items PostNo.15255

Exactly,,,Could you give us examples??

0 P kiwi 2020.01.10 11:05 Like ! 0items PostNo.15439

(b) Excerpt from the Myanmar debate[1]

[1] The conversational agent "dog" is not disclosed to the participants. Here, the agent is supporting the participant "reindeer"

Figure 3 Main Interface of the Discussion Website

the same way.

4.1 The Discussion Platform

The agent platform was deployed on Amazon's EC2 infrastructure with each module being allocated to a separate EC2 instance (Amazon 2020). The web interface of the system is shown in Figure 3a. The different components of the interface are described as follows:

1. Discussion phase that includes the discussion thread area.
2. Discussion topic posted by the Kabul city administrator.
3. Human-based facilitated message.
4. Message of the agent.
5. Ranking that includes user performances such as the number of posted items and the activity-based points.

6. Summary of agent activities such as IBIS classification, analysis, and visualisation.
7. Post form used to post discussion topic.
8. Reply function used by users to post opinions.
9. Search is used to refer someone to current and past discussion contents using keywords.
10. Menu bar that includes account setting and logout button.
11. Discussion theme and media. A user can see the total number of discussants, posted items, discussion time information, and live discussion videos if there is any.
12. Ranking of the posted topics.
13. Discussion points earned through participation.

The interactiveness between the participants and the agent was controlled with two parameters: a period of 1 minute specific to Amazon CloudWatch (Wittig et al. 2016), and a threshold of 4 people. This threshold sets the number of messages that the agent should count before taking part in the discussion. For instance, in the excerpt of the discussion in Figure 3b, the agent dog waited for 4 messages from participants edamame and reindeer before posting his message. Note that the agent identity was not disclosed to the participants up until the end of the experiment.

4.2 The Discussion Data

The content of the discussion is extracted from the platform and lightly processed. The thread of the discussion, the posts, as well as the IBIS label will have to be extracted from the website and from the classification component. In practice, a discussion thread is split into sentences while preserving the hierarchical links between its constituents (posts) as well as the argumentative types of the underlying sentences. The resulting dataset is described in Table 2.

An entry in the dataset is a sentence "Sentence" that is extracted from a post "Post", identified by "Sentence ID", authored by "User ID", and labelled by "Label".

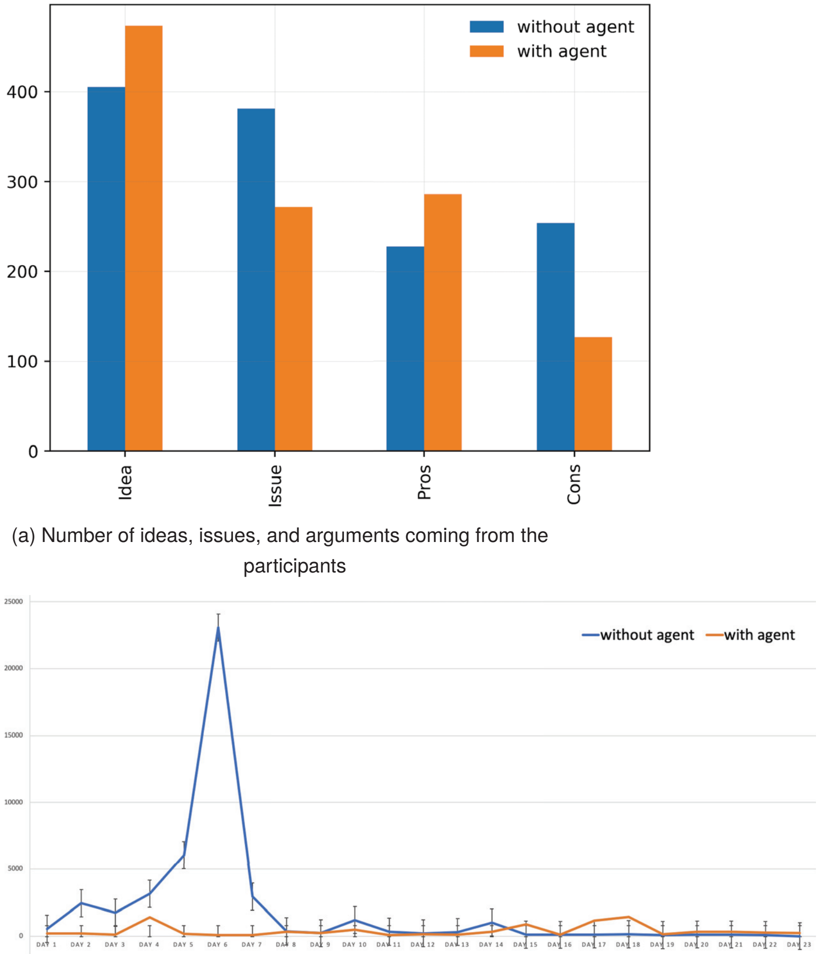
5. Results

The results of the first experiment are shown in Figure 4a. In the first phase of the discussion, the number of issues and cons were higher than when the agent is introduced. Furthermore, the number of ideas and pros increased after the introduction of the conversational agent. Additionally, the average responsiveness time in Figure 4b corresponds to the average waiting time of the same day. If at day d there were n messages posted by the participants and if message i was posted at time t_i then the average responsiveness r_d of day d is computed as $r_d = \frac{1}{n-1}\sum_{i=1}^{n-1}(t_{i+1} - t_i)$. For instance, the peak of the sixth day refers to a day during which the participants' interactions were scattered in the absence of the agent.

The results of the second experiment are shown in Figure 5. For theme 1, only one participant (seahorse) from group M has switched its position from disagreement to agreement. On the other hand, two participants (grape and kiwi) from group J have switched opinions from agreement to disagreement. For theme 2, the participants of group M did not change their opinion while one participant (hippo) from group J switched from disagreement to agreement. In total, 9% of group M and 40% of group J changed their opinions for theme 1. For theme 2, no one from group M had changed his opinion while 20% of group J changed their opinion.

6. Discussion and Conclusion

First, the evolution of IBIS counts in figure 4a suggests that the discussion without the agent is centred around raising issues, and that once the agent is introduced, the discussion will evolve as to find solutions and ideas to the themes. This evolution could be a precondition on how discussions evolve towards a



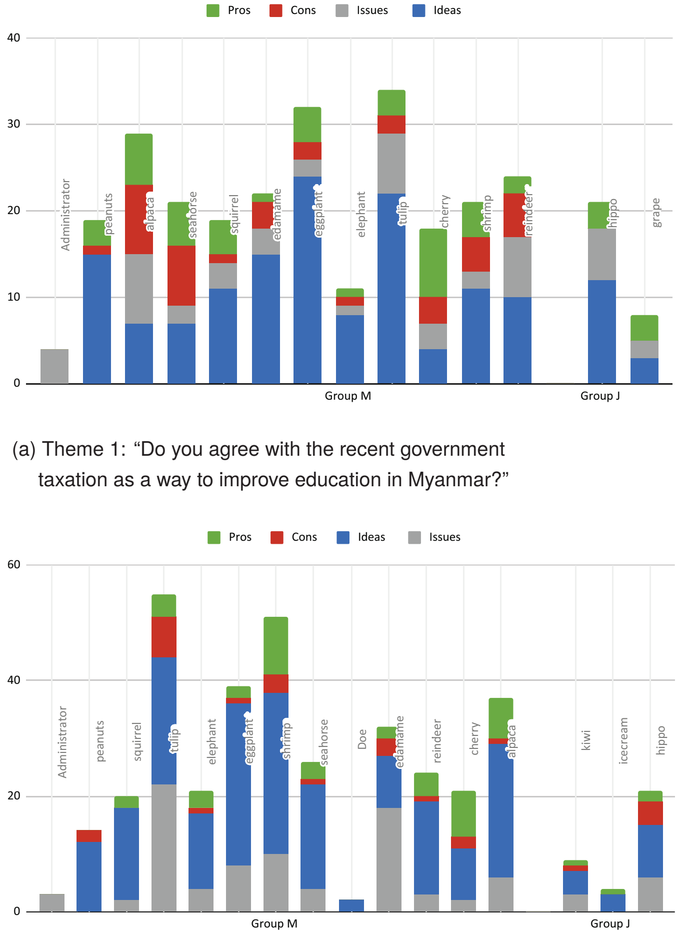
(a) Number of ideas, issues, and arguments coming from the participants

(b) Daily average responsiveness time of the participants before and after the introduction of the agent

Figure 4 (a) Count of the argumentative elements and (b) Superposed responsiveness of the participants with and without the agent. For instance, day 2 corresponds to January 21st in the first phase of the experiment (without the agent) and to February 18th in the second phase of the experiment (with agent)

Table 2 Data Mined from the Discussion

Element	Description
Post	Original post that contains the sentence
Sentence	Subset of "Post"
Sentence ID	Integer identifying the sentence within the thread
User ID	Integer identifying the user that posted "Post"
Root ID	Identifier of the parent post
Label	IBIS label of the sentence: issue, idea, pros, or cons
Label ID	Unique identifier of "Label" within the thread.
Memo	For general remarks and extended usages



(a) Theme 1: “Do you agree with the recent government taxation as a way to improve education in Myanmar?”

(b) Theme 2: "Should we promote globalisation in Myanmar?"

Figure 5 Distribution of the Issues, Ideas and Arguments for the Two Discussion Themes in the Myanmar Experiment

general consensus with the predomination of ideas and cons, or towards a divergent deliberation with the predomination of issues and cons. Second, the responsiveness rate of figure 4b suggests that the agent increased the reactiveness of the participants to the messages. Here, the total responsiveness time without the agent was 2017 seconds and 381 with the agent.

For the second experiment, the debaters have priori knowledge of the theme (Group M in Figure 5), their stances do not change drastically under the effect of the biased agent and are even reinforced. On the other hand, debaters with no knowledge of the subject matter (Group J) did change their stances more frequently, despite receiving neutral messages from the agent. We suggest that the “echo chambers” effect mentioned in section 2 could be reproduced using radicalised or bipolar biased agents. Notwithstanding the important limitations of our study, the results are similar to Bail et al. (2018) and the use of Twitter chatbots, suggesting that strongly biased agents reinforce existing beliefs and cause a bipolarisation of the debate especially when the users have priorly established views on the subject.

To conclude, the conversational agent had an effect on the argumentative nature of the discussion as well as the interactions between the participants. The persuasive effect modulated the distributions of the IBIS elements in the first example by reducing the issues and cons while increasing ideas and pros. In the second example, the persuasive effect was visible in reinforcing the views of debaters that already had strong prior stances while making the debaters with no knowledge change their stances. Additionally, we found that the agent has the ability to increase participation by finding more ideas and issues.

Our next direction is to apply our polarisation mechanism based on the gender and the socio-cultural context of the discussion. This would allow us to establish and evaluate mediation rules that would empower female participants in group discussions in ways that harness the effectiveness of the whole group in solving wicked problems in developing countries. This form of collective intelligence would rely on Artificial Intelligence as well as cooperative effect linked to the number of women in a group (Woolley et al. 2010).

Acknowledgments

The authors would like to thank the anonymous reviewers for their helpful comments. This work was supported by JST CREST Grant Number JPMJCR15E1, Japan.

References

Al-Zubaide H, Issa AA (2011). Ontbot: Ontology based chatbot. *International Symposium on Innovations in Information and Communications Technology*, IEEE.

Amazon (2020). Amazon.com "Elastic Compute Cloud (EC2)". http://aws.amazon.com/ec2.

Amgoud L , Parsons S , Maudet N (2000). Arguments, dialogue, and negotiation. *European Conference on Artificial Intelligence (ECAI2000)*. Berlin, Germany, August 20-25, 2000.

Amgoud L, Cayrol C, Lagasquie-Schiex MC, Livet P (2008). On bipolarity in argumentation frameworks. *International Journal of Intelligent Systems* 23(10): 1062-1093.

Bail CA, Argyle LP, Brown TW, Bumpus JP, Chen H, Hunzaker MF, Lee J, Mann M, Merhout F, Volfovsky A (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences* 115(37):9216-9221.

Baroni P, Giacomin M (2007). On principle-based evaluation of extension-based argumentation semantics. *Artificial Intelligence* 171(10-15):675-700.

Bistarelli S, Rossi F, Santini F (2017). A conarg-based library for abstract argumentation. *2017 IEEE 29th International Conference on Tools with Artificial Intelligence* (ICTAI). Boston, USA.

Bojanowski P, Grave E, Joulin A, Mikolov T (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics* 5:135-146.

Cayrol C, Lagasquie-Schiex MC (2005). On the acceptability of arguments in bipolar argumentation frameworks. *Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, Springer.

Chen H, Liu X, Yin D, Tang J (2017). A survey on dialogue systems: Recent advances and new frontiers. *Sigkdd Explorations Newsletter* 19(2):25-35.

Csaky R (2019). Deep learning based chatbot models. arXiv preprint arXiv:190808835.

Dung PM (1995). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence* 77(2):321-357.

Ferrara E (2017). Disinformation and social bot operations in the run up to the 2017 French presidential election. arXiv preprint arXiv:170700086.

Fox J, Parsons S (1997). On using arguments for reasoning about actions and values. *Proceedings of the AAAI Spring Symposium on Qualitative Preferences in Deliberation and Practical Reasoning*, Stanford.

Graves A, Schmidhuber J (2005). Framewise phoneme classification with bidirectional lstm and other neural network architectures. *Neural Networks* 18(5-6):602-610.

Haqbeen J, Ito T, Hadfi R, Nishida T, Sahab Z, Sahab S, Roghmal S, Amiryar R (2020a). Promoting discussion with AI-based facilitation: Urban dialogue with Kabul city. *Proceeding of the 8th ACM Collective Intelligence Conference*.

Haqbeen J, Ito T, Sahab S, Hadfi R, Okuhara S, Saba N, Hofiani M, Baregzai U (2020b). A contribution to covid-19 prevention through crowd collaboration using conversational AI & social platforms. *AI for Social Good Workshop*.

Hussain S, Sianaki OA, Ababneh N (2019). A survey on conversational agents/chatbots classification and design techniques. *Workshops of the International Conference on Advanced Information Networking and Applications*, Springer.

Iandoli L, Klein M, Zollo G (2007). Can we exploit collective intelligence for collaborative deliberation? The case of the climate change collaboratorium. *MIT Sloan Research Paper* No. 4675-08.

Ito T, Imi Y, Ito T, Hideshima E (2014). Collagree: A faciliator-mediated large-scale consensus support system. *Collective Intelligence.* 2014.

Ito T, Imi Y, Sato M, Ito T, Hideshima E (2015). Incentive mechanism for managing large-scale internet-based discussions on collagree. *Collective Intelligence.* 2015.

Ito T, Shibata D, Suzuki S, Yamaguchi N, Nishida T, Hiraishi K, Yoshino K (2019). Agent that facilitates crowd discussion.*7th ACM Collective Intelligence.* Pitttsburgh, USA, June 13-14, 2019.

Keuschnigg M, Lovsjö N, Hedström P (2018). Analytical sociology and computational social science. *Journal of Computational Social Science* 1(1):3-14.

Kolko J (2012). *Wicked Problems: Problems Worth Solving.* Ac4d Austin, TX.

Kunz W, Rittel HW (1970). Issues as elements of information systems, vol 131. CiteSeer.

Malone TW (2018). *Superminds: The surprising power of people and computers thinking together.* Little, Brown Spark.

Malone TW, Klein M (2007). Harnessing collective intelligence to address global climate change. *Innovations: Technology, Governance, Globalization* 2(3):15-26.

Nofal S, Atkinson K, Dunne PE (2019). On deciding admissibility in abstract argumentation frameworks. *The 11th International Conference on Knowledge Engineering and Ontology Development.* Vienna, Austria, September 17-19, 2019.

Nofal S, Atkinson K, Dunne PE, Hababeh I (2019). A new labelling algorithm for generating preferred extensions of abstract argumentation frameworks. *Proceedings of the 21st International Conference on Enterprise Information Systems* - Volume 2: ICEIS, INSTICC, SciTePress.

Nuez Ezquerra A (2018). Implementing chatbots using neural machine translation techniques. B.S. thesis, Universitat Politècnica de Catalunya.

Ozer M, Yildirim MY, Davulcu H (2019). Measuring the polarization effects of bot accounts in the us gun control debate on social media. *Proceedings of ACM Conference (Conference'17)*. New York, USA, 2019.

Savaget P, Chiarini T, Evans S (2019). Empowering political participation through artificial intelligence. *Science and Public Policy* 46(3):369-380.

Shao C, Ciampaglia GL, Varol O, Yang KC, Flammini A, Menczer F (2018). The spread of low-credibility content by social bots. *Nature Communications* 9(1):1-9.

Stella M, Ferrara E, De Domenico M (2018). Bots increase exposure to negative and inflammatory content in online social systems. *Proceedings of the National Academy of Sciences* 115(49):12435-12440.

Sunstein CR (2001). *Echo Chambers: Bush v. Gore, Impeachment, and Beyond.* Princeton University Press Princeton, NJ.

Suzuki S, Yamaguchi N, Nishida T, Moustafa A, Shibata D, Yoshino K, Hiraishi K, Ito T (2019). Extraction of online discussion structures for automated facilitation agent. *Annual Conference of the Japanese Society for Artificial Intelligence*, Springer.

Varol O, Ferrara E, Davis CA, Menczer F, Flammini A (2017). Online human-bot interactions: Detection, estimation, and characterization. *Eleventh International AAAI Conference on Web and Social Media*. Montreal, Canada, May 15-18, 2017.

Wallace RS (2009). The anatomy of Alice. *Parsing the Turing Test*, Springer.

Weizenbaum J (1966). Eliza - A computer program for the study of natural language communication between man and machine. *Communications of the ACM* 9(1):36-45.

Wittig M, Wittig A, Whaley B (2016). *Amazon Web Services in Action*. Manning.

Woolley AW, Chabris CF, Pentland A, Hashmi N, Malone TW (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science* 330(6004):686-688.

Yan Z, Duan N, Chen P, Zhou M, Zhou J, Li Z (2017). Building task-oriented dialogue systems for online shopping. *Thirty-First AAAI Conference on Artificial Intelligence*.

Rafik Hadfi is currently an assistant professor in the Department of Social Informatics at Kyoto University. He received his M.Eng. and D.Eng. degrees from Nagoya Institute of Technology in 2012 and 2015. His work spans a broad range of disciplines and R&D activities including automated decision-making and the simulation of smart cities, online debates, and biological organisms. He is currently working on AI-enabled platforms to foster democratic deliberation, sustainable development, and gender equality.

Jawad Haqbeen received the B.S. and M.S. degrees in computer science from Nangarhar University, Afghanistan and Waseda University, Japan, in 2010 and 2013, respectively. He is currently pursuing his Ph.D. degree in artificial intelligence from Nagoya Institute of Technology, Japan. His main research interests include conversational agents, collective intelligence, crowdsourcing and applying artificial intelligence to civic technologies. He was the recipient of the Global Young Scientist Summit award in 2021 and Best International Conference Paper Award in 2020. He is a member of IEEE and IPSJ societies.

Sofia Sahab received the B.S. degree in architectural engineering from Kabul University in 2009, and M.E., and doctor of engineering degrees in urban planning from Nagoya Institute of Technology, Japan, in 2014 and 2017, respectively. She is currently specially appointed researcher at Kyoto University, Japan. She previously worked as assistant professor with Nagoya Institute of Technology, Japan, and Kabul University, Afghanistan. Her current research interests include participative decision support system, participatory urban planning, participatory e-planning, and civic technologies. She has published research articles in journals, such as Journal of Simulation and Gaming (SAGE Publications) and Journal of Architecture and Planning (Transections of Architectural Institute of Japan).

Takayuki Ito is professor of Kyoto University. He received the B.E., M.E, and Doctor of Engineering from the Nagoya Institute of Technology (NIT) in 1995, 1997, and 2000, respectively. From 1999 to 2001, he was a research fellow of the JSPS. From 2000 to 2001, he was a visiting researcher at USC/ISI. From April 2001 to March 2003, he was an associate professor of JAIST. From April 2004 to March 2013, he was an associate professor of NIT. From April 2014 to September 2020, he was a professor of NIT. From October 2020, he is a professor of Kyoto University. From 2005 to 2006, he is a visiting researcher at Division of Engineering and Applied Science, Harvard University and a visiting researcher at the Center for Coordination Science, MIT Sloan School of Management. From 2008 to 2010, he was a visiting researcher at the Center for Collective Intelligence, MIT Sloan School of Management. From 2017 to 2018, he is an invited researcher of Artificial Intelligence Center of AIST, JAPAN. From March 5, 2019, he is the CTO of AgreeBit, inc. as an entrepreneur.