

Facilitating Discussion on COVID-19 with Autonomous Facilitator: A Case Study on the Comparison of Expert Discussion versus the Public Paradigm Ratio of Reply for Agent Facilitated Post

Jawad Haqbeen, Sofia Sahab, Takayuki Ito

Abstract—An important way to promote large-scale online debate development and improve discussion environment is discussion support platform, with artificial intelligence-based facilitation being its key part. To support such debates, software agents as facilitators need to be developed to facilitate these discussions. In this study, we propose to study this phenomenon using an online-debate system based on facilitation called D-Agree. We aimed to investigate the influence of software entities as autonomous facilitators (AF) on the evaluation of online debate involving cross-class people on COVID-19-related discussion in Afghanistan. This study was conducted with two classes of people: (1) health workers ($n=16$) as experts on COVID-19 related debate, and (2) private citizens ($n=16$) as public and non-experts on COVID-19-related discussions. Initially the health workers were selected using a non-probability sampling technique of convenience sampling survey in collaboration with Afghanistan national public health institute, and the private citizens were selected using convenience sampling, and then we used stratified random sampling to select 16 people from each class. We created 8 online groups, four for each class namely, A–D, and randomly assigned subjects of each class to a group based on a female and three male members ($n=4$; female = 1 and male = 3). The agent will dynamically interact with participants of each group or class of people based on predefined facilitation ratio (1:2 = A & C groups; 1:3 B & D groups). For the sake of experimental evolution, we used discussion annotated datasets that contain human and AF posts, and the number of human posts towards AF posts. According to the results, agents with a facilitation threshold of two people (1:2) had a significant impact on discussion development in terms of both discussion elements and posted characters compared with facilitation threshold of 3 people (1:3). With 1:2 setting, we found that the agent improved the responsiveness of both expert class and public class (A&C groups) than 1:3 setting (B&D groups). That means Afghans engage more-write more characters with 1:2 than 1:3 with agent-based facilitation. Hence, the agent increased the number of identified discussion elements. The output of this research can be used as a precondition on setting agent facilitation for least development countries like Afghanistan.

Index Terms— *Conversational agent, online discussion, public paradigm, expert paradigm, COVID-19, artificial intelligence, online forum*

I. INTRODUCTION

Online forums have become prominent in the pandemic era as main venues for infusing societies with secure participatory elements [1]. It is because Internet forums provide a convenient channel for users concerned about social issues to discuss and share information with each other. However, forums are criticized as being unsuitable to provide a range of services like discussion moderation to lead discussion, insight visualization to raise awareness, and most importantly analyzing discourse data to understand participants' sentiments elements [2]. Furthermore, online forums without support means cannot promote argumentative reasoning and deliberative virtues [3-4], as these forums are

usually based on self-expressive talk and participants do not interact with each other properly and their collaborations are challenging [5]. An emerging way to promote the quality of online discussion development to reach agreement for solving social problems like pandemic is supporting discussion using AI technologies, in which autonomous facilitation is its key part. To support large-scale discussion, software agents have been introduced in the forms of autonomous agents to facilitate discussion among groups of individuals, for example, to engage with debate participants by mimicking their submitted posts while inviting debate participants to engage with each other to solve their common problems. A notable branch of AI, deep learning (DL) and machine learning (ML) has been broadly used to empower a suitable discussion structuring format called Issue-based Information System (IBIS) [19]. These systematic methodologies not only lead discussion by posting facilitated messages, but also

Jawad Haqbeen is with the Department of Computer Science, Nagoya Institute of Technology, Nagoya, 466-8555 Japan, Corresponding author e-mail: jawad.haqbeen@itolab.nitech.ac.jp

Sofia Sahab and Takayuki Ito is with the Department of Social Informatics, Kyoto University, Kyoto, 606-8501 Japan.

uses IBIS as part of its automated categorization of sentences and extracts discussion's structure in real-time. The catalogued discussion trees not only help human participants gain knowledge in online debate, but also, the classified insights of crowd help policy-making institutions.

Drawing on the "autonomous agent" as facilitator role within argumentative reasoning of online discussion. We hypothesize that agents within non-creative online discussion ameliorates discussion on guiding creative discussion and spurring interaction among human participants with different education backgrounds (Please see other hypotheses in section 2). Towards this end, we conducted a cross-people online experiment on the challenges posed by COVID-19 in Afghanistan adopting a suitable discussion structuring format called Issue-based Information System [19] (IBIS) and using an online forum based on facilitation called, D-Agree [3]. Aiming to explore the influence of the agent on the evaluation of online discussion with non-creative topic setting involving two class of human participants with different education background: (1) health workers as experts on COVID-19-related discussion, and (2) private citizens as public and non-experts on COVID-19-related discussion. We pretend that one class has good information regarding thematic area (topic) while others do not. We set two types of facilitation rate where the agent will adaptively reply to human messages based on a predefined ration of 1:2 and 1:3. We compared groups (A&C) of each class with groups (B&D) of same class, and then compared the same groups of a class with another class. The experimental results show that agents as facilitators with higher interaction rates indeed have a valuable effect on enhancing interaction with both classes of people, when debaters discuss non-creative topics. The output of this research can be used as a precondition on setting agent-based facilitation rate in online debate of least developed countries like Afghanistan. To the best of our knowledge, this is the first study comparing discussion structures among classes of people while looking at the effect of agents as facilitators on COVID-19-related discussion development.

The rest of the paper is structured as follows. A literature review related COVID-19 online discussion using online forums is shown in Section 2, and its objective and methodology are shown in Section 3. Section 4 describes its experimental procedure and setting, and its results are discussed in Section 5. Finally, Section 6 presents a discussion, future work, and conclusions.

II. RELATED WORK

The previous works are presented in this section.

Recent studies have shown that Internet-based direct democratic discussion platforms are considered as the next-generation social democratic platforms for citizen deliberation to solve sustainability social problems like pandemic and climate change [3-4]. Such platforms could help to collect and then integrate public ideas to inform policy-makers in order to consider the effect of social issues mentioned above on the welfare of society [1]. The online discussion is considered and used as a complement component of operational activities of all known communities during pandemic. For example, social platforms like Twitter and Facebook, regardless of quality and source of data, have been proven to be important in distributing information related to the COVID-19 pandemic. Such data can be

disseminated for short and long-term objectives. In the short term, the communities will discuss, use and disseminate COVID-19 related information; whereas in the long term, the research and public institutes will use it to create sustainable policies. Similar to previous pandemics, the recent emergence of COVID-19 has brought several problems under study. For instance, data accumulated via Twitter has been used to track the public behavior and examine health-seeking and public reactions towards outbreak [20]. Another work has collected data from three social platforms in China to assess public concerns and risk perception, as well as to track public behavior in response to the COVID-19 outbreak [21]. An online discussion platform has been used to fight Covid-19 by collecting and analyzing vast amounts of social data to increase public awareness and for public health policy-making [23].

Several studies are also investigating the effect of online discussion as the primary mode for assessment during pandemic [22]. For instance, the work of Reijo Savolainen focuses on the credibility of Covid-19 vaccine disinformation in online discussion [24] by studying the user generated data posted to a Reddit discussion group. Another work examined Covid-19-related discussions (tweets posted by Twitter users) using ML approach [25]. Another work used automated extraction of COVID-19 related discussions from social media and a natural language process (NLP) method based on topic modeling to uncover various issues related to COVID-19 from public opinions. Moreover, the authors also investigate how to use LSTM recurrent neural networks for sentiment classification of COVID-19 comments [26]. However, with the emergence of AI technologies, a more intelligent debate-support system has emerged with analyzing and learning methodologies. For instance, software entities like conversational agent, argumentative agent, NLP agent, pro-active agent are generally employed within discussion systems to not only interact with human participation but also used for sentiment classification of posted opinions (Shown in Figure 3).

D-Agree [3], is one example of online discussion support platform based on AI facilitation, defined as a computer-based system that supports groups of people engaged in a debate, and provides an interface for users to follow the on-going discussions like their ranking, discussion activities, threads and discussion insights. To improve the efficiency and quality of discussion development, facilitation is introduced into D-Agree. Facilitation is defined as the set of activities that are carried out during debate by software entities to support a group of individuals to achieve their goals during the decision-making process [3]. To this end, D-Agree explored agent-based facilitation models in online debate related to city-planning, climate changes, and discussion of Sustainable Development Goals in countries like Japan [15-17], Myanmar [18], and Afghanistan [12-14], and showed that group discussion improved in all these discussions.

Most of these studies about utilizing online-debate based on facilitation suggest that they only considered predefined facilitation while conducting a mixed groups of people social experiment. However, there is no case study on the comparison of expert discussion versus the public paradigm regarding ratio of agent facilitated posts to investigate how much agents should interact to promote better discussion while considering different classes of people. As a result, it becomes highly important to investigate the influence of software entities as autonomous facilitation on the evaluation

of online debate involving cross class of people settling like expert vs. public paradigm to conceive a conversational agent predefinition setting that can smoothly proceed discussions. The current work uses online discussion where instead of only examining posted discussion from social media, we used our discussion system where two types of participants based on their knowledge background on discussion topics were invited to post their opinions and the discussion reinforced by an agent with two facilitation ratios to interact with human participants of both classes. In this work, we wanted to investigate how agents should interact to promote better discussion while considering different classes of people while assuming that a class of people (expert) have prior knowledge about the discussion theme, and another class (public) don't have the prior knowledge same as expert. We assumed that the more interaction of agent with debate participants, the more persuasive effect of conversational agent will modulate the distributions of the opinions within discussion while reducing the negative opinions (issues and cons) and increasing positive opinions (ideas and pros) among all groups of individuals. Based on the above statement, we propose the following hypothesis.

Hypothesis 1 (H1). *Compared with the public class of individuals (individuals without prior knowledge on discussion topics), the expert class of individuals has a more significant impact on the development discussion with agent-based facilitation.*

Hypothesis 2 (H2). *Compared with facilitation threshold of three people (agent interaction after three human posts), agent-based facilitation threshold of two people (agent interaction after two human posts) has a more significant impact on the development discussion both in terms of modulating the distributions of the IBIS elements and also increasing number of posted characters within cross class of participants debate.*

III. AIMS AND METHODS

Our research's general methodology (Figure 1) conducts a cross class of people control experiments using online discussion support systems based on facilitation, with the objective of verifying discussion structures and investigating the phenomenon of agent-based facilitation ratio in the discussion among groups of people with and without prior knowledge on discussion topics. The hypothesis testing is performed and $p < .01$ is considered statistically significant.

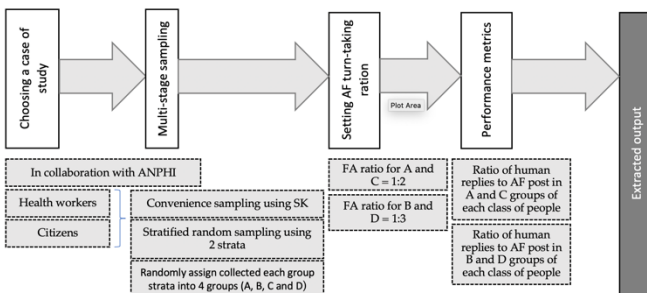


Fig. 1. Summary of sampling process, study instruments and experiment setting.

A. Objective

The objective of this study is to quantitatively assess cross class of people related discussion metrics (number of IBIS and posted characters), where the discussions are reinforced by different ratios of agent-based facilitation. We want to investigate how agent-based facilitated posts affect discussion development across class of people, while considering a group of individuals with prior knowledge on discussion topics as experts; and other groups of individuals without prior knowledge on discussion topics as public.

B. Study Area

Kabul (Figure 2) is Afghanistan's capital and home to the largest share of total urban population, and also, is one of fastest-growing cities [6] which made Kabul the world's 75 largest city [8]. It is also a municipality [7], forming part of the greater Kabul Province, and divided into 22 districts. According to estimates in 2021, the population of Kabul is 4.6 million [9]. All consented subjects living in Kabul city and wanted to debate on COVID-19 crisis management in Afghanistan had taken part in the study. The fifth participant is the agent who can objectively observe conversational situations, and then post facilitation messages accordingly.

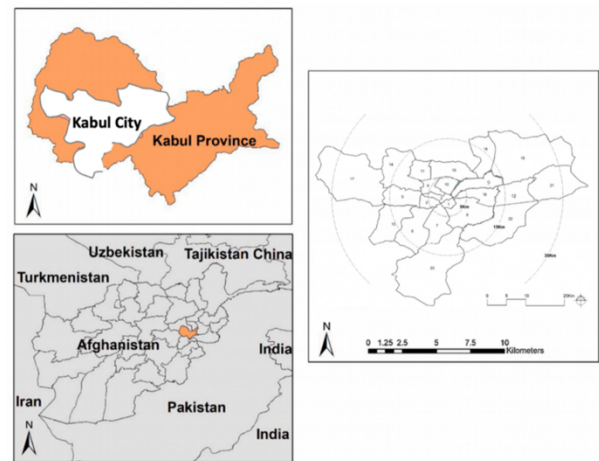


Fig. 2. Location of the study.

C. Sample and sampling method

Since it was difficult for authors to approach health workers. Thus, we recruited them from Afghan-Japan Communicable Disease Hospital (AJCDH) in collaboration with ANPHI using an online survey called, SurveyMonkey of SVMK inc. The survey procedure and methods in this study have gone through multi-stage sampling. In statistics, multistage sampling is the taking of samples in stages using smaller and smaller sampling units at each stage [10]. A nomination questionnaire is used as a survey instrument to collect two samples' responses and support the experimental process. They were chosen based on their availability using convenience sampling, and were recruited through the support of ANPHI as mentioned before. The request for open call participation for the second class, private citizens, was boosted using a Facebook ad. Thus, to whom online survey links reached and anyone who agreed took part in the surveys.

The health workers ($n = 565$), and citizens ($n = 1,085$) responded to our survey and consented to participate in the study. We then used the gathered samples and set the two strata as female and male groups by using stratified random sampling to select 16 candidates from each sample to take part in a control experiment for this study. The authors choose to select four participants from female strata and 12 from male strata for each class of people out of consented subjects. Since we wanted to compare cross-people discussion, authors made a decision to randomly assign 12 male and four female candidates of each sample into four groups, namely A to D, to make a group of four people (a female and three male). We selected four human participants plus AF per group because four is the special number in multiparty conversations [11].

D. Discussion Instrument

D-Agree, a text-only discussion support platform with artificial intelligence-based facilitation is the online discussion instrument for this study. The system was developed by our team and deployed in many countries, such as Afghanistan [12-14], Indonesia, Japan [15-17] and Myanmar [18] through conducting real world social experiments. It is used to host ongoing city planning debates by gathering, facilitating, extracting, and visualizing real-time discussion summaries. The system extracts the submitted opinion based on IBIS, which automatically classifies all submitted opinions based on a combination of four types of elements: issue, idea, pros, and cons. The system has the capability to summarize the submitted opinions. The proactive agent set the behavior of the conversational agent based on predefined consensus policy, where the conversational agent uses IBIS as part of its automated categorization of sentences and then posted facilitated messages to lead the discussion. We set the facilitation threshold where the agent will be introduced to interact with human participants. In this study, we set two types of facilitation ratios for conversational agents: 1:2 and 1:3.

The instructiveness between the participants of groups (A & C); and participants of groups (B & D), and the agent was controlled with two parameters: a period of 1 minute specific to Amazon CloudWatch, and a threshold of 2 and 3 people, respectively. This threshold sets the number of messages that the conversational agent should count human posts before taking part in the discussion. For example, 1:2 means that the conversational agent will wait and count every two human posts before posting her predefined facilitated message. The threshold of 3 people (1:3) means that the conversational agent will wait and count every three human posts before posting her predefined facilitated message.

The system architecture are shown in Figure 3.

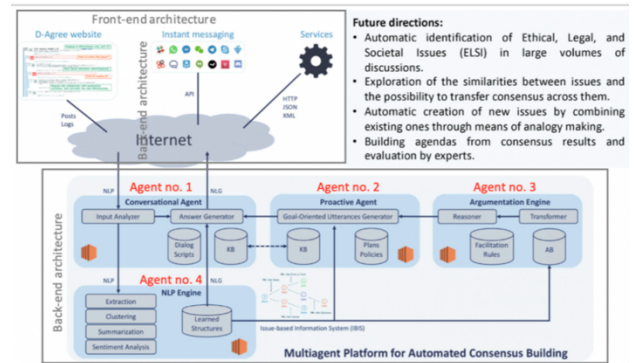


Fig. 3. Outline of system architecture, (top) front-end hosting user interface; and (bottom) back-end hosting four types of agent modules that includes a conversational agent as facilitator that uses IBIS as part of its automated categorization of sentences while posting facilitated messages.

E. Ethics Approval

Our data collection method was compiled with Afghanistan's ministry of public health (MoPH) Institutional Review Board (IRB) committee's ethics board by applying and receiving MoPH exemption letter (IRB no. E.1220.0254). In addition, the ethics committee at Nagoya Institute of Technology (NITech) issued an exemption letter.

IV. EXPERIMENT PROCEDURE AND SETTING

In this section, we introduce the experiment of using the proposed case studies to investigate variance of agent-based facilitation in online discussions by offering predefined facilitation ratio suggestions for cross class of people with (experts) and without prior knowledge (public) regarding theme stance. In this regard, a series of experiments ($n = 8$) are conducted with four groups of each class of people in order to demonstrate the effectiveness of the agent-based facilitation. First, we created eight discussion working spaces, one for each group of people. The invited subjects can join the discussion by entering the discussion code. A request to participate in the group, along with a link directing assigned participants to the D-Agree platform. In addition, a code is required to participate in the group discussions. Then, we sequentially choose facilitation ratio 1:2 for A and B groups and regard the groups (C & D) as the groups based on agent facilitation with 1:3 ratio.

Then, we verify the ability of all participants by conducting a pilot experiment to ensure the functionality of our system, and make sure that subjects can login and post on the system. Finally, we conducted the actual experiment.

All the participants discussed the following discussion topic:

- Challenges of COVID-19 in Afghanistan

We set an issue-giving, or non-creative, topic rather than an issue-solving, because we want to check the effect of our agent facilitation policy which is based on issue-solving-stance.

Note that the agent identity was disclosed as "AI Facilitator" to the participants up until the end of the experiment, and also, the individual's identity was disclosed as their real names. We simultaneously ran separate online experiments for four groups of two classes of people using the same tool, topic and time.

A. Data Collection

We used D-agree to hold online discussions. The actual experiment took place online from December 7 to 28, 2020 for a 20 days period. Expert class consisted of health workers, and the public class consisted of private citizens other than health workers such as students, government employees, etc. The English language was used in these discussions. Each discussion lasted 20 days as mentioned before. We selected this duration because we believe that it is easier for both health workers and private citizens to take their time, and participants were free to generate posts that aim to discuss the discussion topic. The agent-based on two predefined facilitation ratios were set to facilitate the discussion.

The content of the discussion is extracted from used discussion instruments and lightly processed as human-led study. We used discussion annotation files datasets that contained the thread of the discussion; the posts including replies; the IBIS label which contains a combination of four types of elements: issue, idea, pros, and cons; and each submitted posts obtained/evaluated points. In practice, a post is split into sentences while using IBIS as part of its automated categorization of sentences.

In total, 192 posts (1,388 IBIS elements) were generated from the four expert groups discussions (A~D), and 121 posts (961 IBIS elements) were generated from public groups (A~D).

The number of post characters by all expert and social groups and were separately totaled (n= 75,508; n= 50493) (Table 1). The statistics of individual and total post character, total post number, total number of replies for humans, total number of replies for agents, total discussion points, total number of like, and total number of IBIS (Issue, Idea, Pros, and Cons) are shown in Table. 1.

V. EXPERIMENTAL RESULTS

The results of the comparison on IBIS catalogued based on: (1) public with 1:2 vs public with 1:3; (2) expert with 1:2 vs expert with 1:3; (3) public with 1:2 vs expert with 1:2; and (4) public with 1:3 vs expert with 1:3 are shown in Table 1. The results of the comparison on average posted characters based on: (1) public with 1:2 vs public with 1:3; (2) expert with 1:2 vs expert with 1:3; (3) public with 1:2 vs expert with 1:2; and (4) public with 1:3 vs expert with 1:3 are shown in Table 2. The results of public all discourse IBIS catalogued are shown in Figure 4; and the results of expert all discourse IBIS catalogued are shown in Figure 5.

A. Intra and Inter class comparisons on average of IBIS development

In Figures 4 and 5, we compare automated categorization of submitted sentences which were created based on IBIS between all public and expert groups (A~D). Then an Independent-samples t-test was conducted to compare their means (Table 1). Firstly, we compared and test public with 1:2 vs public with 1:3, and we found that public groups with threshold of 2 people (M = 53.75, SD = 12.18) had a significant impact on discussion development in terms of average number of IBIS elements compared with public groups with 1:3 (M = 28, SD = 11.80), $t(14) = 2.976, p=0.000 (< .01)$ (two-tailed).

Secondly, we compared and tested expert with 1:2 vs expert with 1:3, and similarly to the above, we found that expert groups with threshold of 2 people (M = 64.75, SD = 11.01) also had a significant impact on discussion development in terms of average number of IBIS elements compared with expert groups with 1:3 (M = 47.25, SD = 3.32), $t(8) = 3.355, p=0.000 (< .01)$ (two-tailed).

Thirdly, we compared and tested the public with 1:2 vs expert with 1:2, and the results were not significant at $p < .01$ (two-tailed). However, we found that expert groups with threshold of 2 people (M = 64.75, SD = 11.01) had a significant impact on discussion development in terms of average number of IBIS elements compared with public groups with 1:2 (M = 53.75, SD = 12.18), $t(14) = 2.976, at p < .05$. Hence, the results obtained through this observation are in a good agreement with H1 and H2 (in terms of increasing IBIS) and confirm the validity of hypothesis 1 and 2.

Finally, we compared and tested public with 1:3 vs expert with 1:3, and we found that expert groups with threshold of 3 people (M = 47.25, SD = 3.32) had a significant impact on discussion development in terms of average number of IBIS elements compared with public groups with 1:3 (M = 28, SD = 11.80), $t(8) = 2.976, at p=0.000 (< .01)$.

This observation is also, in a good agreement with H1 and confirmed the validity of hypothesis 1.

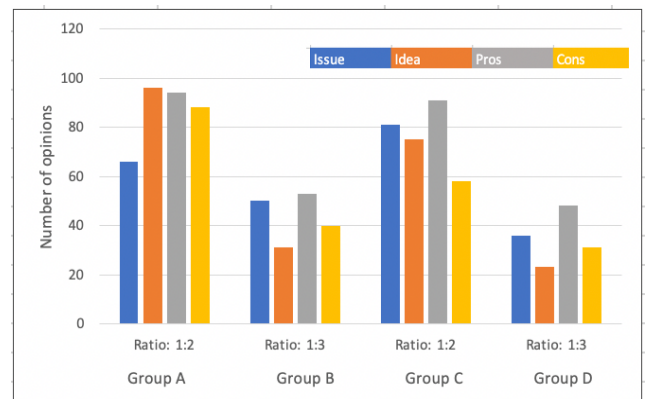


Fig. 4. Number of issues (colored blue), ideas (colored orange), pros (gray) and cons (yellow) coming from the public paradigm (private citizens).

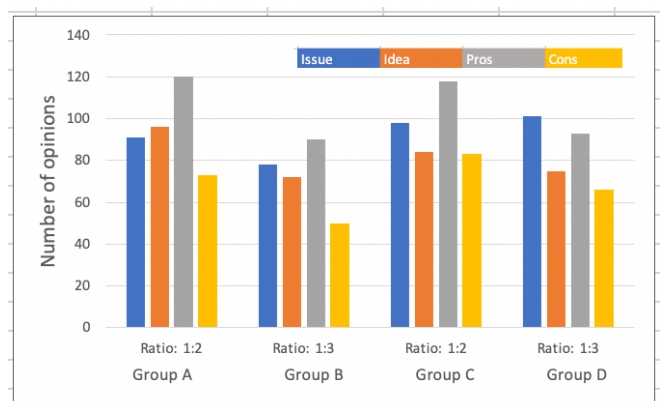


Fig. 5. . Number of issues (colored blue), ideas (colored orange), pros (gray) and cons (yellow) coming from the expert paradigm (health workers).

TABLE I
CHARACTERISTICS OF TESTING ON INTRA AND INTER CLASS COMPARISONS ON AVERAGE IBIS DEVELOPMENT

		Public		Expert	
		1:2	1:3	1:2	1:3
Public	1:2	1.000			
	1:3	0.001***	1.000		
Expert	1:2	0.079*	N.T	1.000	
	1:3	N.T	0.002***	0.003***	1.000

N.T = Not tested for significance

***p<0.01, **p<0.05, *p<0.1 for

B. Intra and Inter class comparisons on average of posted characters

In Table 2, we compare and tested average number of posted characters based on: (1) public with 1:2 vs public with 1:3; (2) expert with 1:2 vs expert with 1:3; (3) public with 1:2 vs expert with 1:2; and (4) public with 1:3 vs expert with 1:3. An Independent-samples t-test was conducted to compare their means.

Firstly, we compared and tested public with 1:2 vs public with 1:3, and we found that the total number of posted characters based on threshold of 2 people (M = 4427.75, SD = 1191.60) had significant compared with the total number of posted characters based on threshold of 3 people (M = 1883.87, SD = 881.41), t(13) = 3.0122, p=0.000 (< .01) (two-tailed).

Hence, the results obtained through this observation are in good agreement with our H2 (increasing number of posted characters) and confirmed the validity of hypothesis 2.

Secondly, we compared and tested experts with 1:2 vs experts with 1:3, and we did not find any significance at p < .01.

Thirdly, we compared and tested the public with 1:2 vs experts with 1:2, and the results were not significant at p < .01.

Finally, we compared and tested public with 1:3 vs expert with 1:3, and we found that the total number of posted characters based on threshold of 3 people in expert groups (M = 4165.12, SD = 641.67) had significant compared with the total number of posted characters based on threshold of 3 people in public groups (M = 1883.87, SD = 881.41), t(13) = 3.0122, p=0.000 (< .01) (two-tailed)

TABLE II
CHARACTERISTICS OF TESTING ON INTRA AND INTER CLASS COMPARISONS ON AVERAGE POSTED CHARACTERS

		Public		Expert	
		1:2	1:3	1:2	1:3
Public	1:2	1.000			
	1:3	0.000***	1.000		
Expert	1:2	0.170	N.T	1.000	
	1:3	NT	0.000***	0.036**	1.000

N.T = Not tested for significance

***p<0.01, **p<0.05, p<0.1 for

VI. DISCUSSION

Specifically, four types of elements (issue, idea, pros, and cons) are fundamental in the discussion structure to clarify discussion mapping (tree). Issue means the common questions that debate participants aim to solve. The issues

generated after other issues could be generalization or specialization of the related issues. In addition, the issues may be questions, sub-issues, or replacements that are generated after other related elements. Idea means possible answers that are generated in response to the related issues. Argument means the opinions that are generated in response to the related ideas. The argument might be in forms of positive opinions which are represented as pros. On the other hand, the argument might be in forms of negative opinions which are represented as cons.

First, the evolution of four elements of discussion: IBIS counts and metrics on contents in Figures 4 and 5 suggests that an agent with threshold of both 2 and 3 people is to solicit people’s positive opinions (ideas and pros). This finding is aligned with our agent consensus policy, where the agents are more willing to demonstrate a higher likelihood of engaging human participants in solving common issues by posting more ideas and cons to ideas.

The total number of ideas and pros were higher for both expert and public groups with both 1:2 and 1:3 setting (n= 591; 381) than the total number of issues and cons for both expert and public groups with both 1:2 and 1:3 setting (n= 387; 273) (Figures 4 and 5).

This finding aligns with the fact that our agent adopts consensus policy, where the agents are more willing to demonstrate a higher likelihood of engaging human participants in solving common issues by posting more ideas and cons to ideas.

This encourages humans to encourage reaching a consensus for solving problems on online communities by mediating and supporting human-generated posts [12]. Hence, the results obtained through this observation are in a good agreement with our agent consensus policy and confirmed the validity of the agent-based facilitation of consensus-building method. This evolution could be very helpful for a complex society like Afghanistan, where people need to focus on solving more issues than raising the issues itself which help to avoid confrontation and help reach agreements. Second, our findings suggest that the more people know about the discussion topic or having prior knowledge about the discussion theme, the higher significance it might have on discussion development. For example, we have collected more IBIS elements (n= 978) from expert groups than public groups (n= 654). In this work, we believe that a class of experts like health workers have prior knowledge about COVID-19 related discussion themes, and another class of public people might don’t have the prior knowledge compared to expert groups.

From the experimental results, we find that the threshold of 2 people-based discussion facilitation can be used to support online discussion facilitation in least developed countries like Afghanistan. From the detailed comparison of the two thresholds, we find that the threshold of two people needs to be considered for conducting discussion facilitation tasks that aim to encourage participants to generate more posts in communities where people are keener to discussion like Afghanistan.

This finding aligns with that of French in Afghanistan [27], which reported that communities with lower socioeconomic levels, such as Afghanistan, are more willing to demonstrate a higher likelihood of engaging in activities and investing more time in discussing common issues. Furthermore, this may be explained by the fact that the AI threshold of 2 people tends to be associated with much more facilitation support

that stimulates people to engage with each other than the threshold of 3 people. This is considered as one reason that leads to higher likelihood of engaging in activities with agent facilitation threshold of 2 people than 3 people.

Second, comparison on evolution of four elements of discussion between expert and public suggests that conversational agents have the ability to increase discussion elements, particularly positive opinions (ideas and cons) with groups of people having prior knowledge about discussion themes than people with non-prior knowledge. This finding is aligned with the fact, health workers are more willing to demonstrate a higher likelihood of knowledge in discussing COVID-19-related issues than public (private citizens).

VII. CONCLUSION

In this paper, we investigated and compared the influence of conversational agents as facilitators on the evaluation of COVID-19-related online debate involving two classes of people: expert versus public paradigm. To this end, we compared the effect of agents on guiding each group of people by setting two types of thresholds of facilitation (threshold of 2 people versus threshold of 3 people). The conversational agent with 1:2 facilitation ratio (threshold of 2 people) had an effect on the discussion development in terms of average number of IBIS elements as well as average number of posted characters. The persuasive effect of conversational agents modulated the distributions of the IBIS elements in the first example by reducing the issues and cons while increasing ideas and pros among all groups of people. Additionally, we found that the conversational agent has the ability to increase discussion elements, particularly positive opinions (ideas and cons) with groups of people having prior knowledge about discussion themes.

In our future work, we plan to further investigate the agent-based facilitation thresholds of n -size people by considering the different group sizes ranging from four to 100 group members and group composition based on genders in order to analyze the efficient facilitation rate for different online group sizes and compositions.

FUNDING AND ACKNOWLEDGMENT

This work has been supported by JST CREST project Grant No. JPMJCR20D1. We would like to thank and express our deepest appreciation to all people who participated in this study.

REFERENCES

- [1] Haqbeen, J., Ito, T., Sahab, S., Sato, T., Okuhara, S., Hofiani, M.: Insights from a large-scale discussion on COVID-19 in collective intelligence. In: IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT20), (2020)
- [2] Wyss, D., Beste, S.: Artificial facilitation: promoting collective reasoning within asynchronous discussions. *Journal of Information Technology & Politics*, 14,3, 214–231 (2017)
- [3] Ito, T., Shibata, D., Suzuki, S., Yamaguchi, N., Nishida, T., Hiraishi, K., Yoshino, K.: Agent that facilitates crowd discussion. In: *ACM Collective Intelligence Conference*, (2019)
- [4] Haqbeen, J., Ito, T., Hadfi, R., Nishida, T., Sahab, Z., Sahab, S., Roghmal, S., Amiryar, R.: Promoting discussion with AI-based facilitation: Urban dialogue with Kabul City. In: *ACM Collective Intelligence Conference* (2020)
- [5] Tavanapour, N., Poser, M., Bittner, E.A.C.: Supporting the Idea Generation Process in Citizen Participation - toward an Interactive System with a Conversational Agent as Facilitator. In: 27th European Conference on Information Systems, Stockholm and Uppsala. (2019).
- [6] UNHABITAT <https://unhabitat.org/sites/default/files/download-manager-files/State%20of%20Afghanistan%20Cities%202015%20Volume%201.pdf>, last accessed 2021/09/1
- [7] Kabul Municipality Official website, accessed at: <https://km.gov.af>, last accessed 2021/09/1
- [8] CMS. The Largest Cities in the World and Their Mayors. World Mayors (City Mayors Statistics), <http://www.citymayors.com/statistics/largest-cities-mayors-1.html>, last accessed 2021/09/1
- [9] National Statistic and Information Authority (NSIA). April 2021. Accessed at: <https://www.nsia.gov.af:8080/wp-content/uploads/2021/06/Estimated-Population-of-Afghanistan-1-1400.pdf>, last accessed 2021/09/1
- [10] Thompson, S.K.: Sampling. third edition, John Wiley & Sons, Inc., chapter 13 (2012), 171–175, (2012)
- [11] Matsuyama, Y., Akiba, I., Fujie, S., Kobayashi, T.: Four-participant group conversation: A facilitation robot controlling engagement density as the fourth participant, *Computer Speech & Language*, 33.1, 1–24 (2015).
- [12] Haqbeen, Jawad; Sahab, S.; Ito, T.; and Rizzi, P. Using decision support system to enable crowd identify neighborhood issues and its solutions for policy makers: An online experiment at Kabul municipal level. *Sustainability*, 13(10), 5453 (2021).
- [13] Haqbeen, Jawad; Sahab, S.; Ito, T. AI-based mediation improves opinion solicitation in a large-scale online discussion: Experimental evidence from Kabul Municipality, *IJCAI Workshop on AI for Social Good*, 2021.
- [14] Haqbeen, J.; Ito, T.; Hadfi, R.; Sahab, Z.; Sahab, S.; Amiryar, R.; Nishida, T. Usage & Application of AI-based Discussion Facilitation System for Urban Renewal in Selected Districts of Kabul City: Afghanistan Experimental View. In *Proceedings of the 34th Annual Conference of the Japanese Society for Artificial Intelligence*, Kumamoto, Japan, 9–12 June 2020; pp. 1–4.
- [15] Ito, T. Towards agent-based large-scale decision support system: The effect of facilitator, *Proceedings of the 51st Hawaii International Conference on System Sciences*, 2018.
- [16] Ito, T.; Suzuki, S.; Yamaguchi, N.; Nishida, T.; Hiraishi, K.; and Yoshino, K. Crowd Discussion for the Future of Your Hometown: An Agent-based Crowd Discussion Support System, *Proceedings of the 2019 Workshop on Artificial Intelligence and United Nations Sustainable Development Goals (AI4SDGs)*, Macao, 2019.
- [17] Ito, T.; Hadfi, R.; Haqbeen, J.; Suzuki, S.; Sakai, A.; Kawamura, N.; Yamaguchi, N. Agent-Based Crowd Discussion Support System and Its Societal Experiments. In *Advances in Practical Applications of Agents, Multi-Agent Systems, and Trustworthiness*, 1st ed.; Demazeau, Y., Holvoet, T., Corchado, J., Costantini, S., Eds.; Springer: L'Aquila, Italy, 2020; pp. 430–433.
- [18] Hadfi, R.; Haqbeen, J.; Sahab, S.; and Ito, T. Argumentative Conversational Agents for Online Discussions, *Journal of Systems Science and Systems Engineering*, 1-15 (2021).
- [19] Kunz, W.; Rittler, H.W. Issues as elements of information systems. *Tech. Rep. working paper no. 131* 1970.
- [20] Ganasegeran, K.; Abdulrahman, S.A. Artificial Application in tracking health behaviors during disease epidemics. In *proceeding Human Bavaieur Analysis using Intelligent Systems*, Cham Springer, pages 141-155, 2020.
- [21] Hou,Z.; Du, F.; Jiang, H.; Zhou, X.; Lin, L. Assessment of public attention, risk perception, emotiannal and behavioural responses to the COVID-19 outbreak: social medial surveillance in China. *medRxiv preprint medRxiv*: doi: <https://doi.org/10.1101/2020.03.14.20035956>
- [22] Ackerman, D.S.; Gross, B.L. Synchronous online discussion board as a primary mode of delivering marketing education: responding to the Covid-19 pandemic and beyond. *Marketing Education Review*, DOI: 10.1080/10528008.2021.1893752
- [23] Haqbeen, J.; Ito, T.; Sahab, S.; Hadfi, R.; Okuhara, S.; Saba, N.; Hofiani, M.; Baregzai, U. A contribution to covid-19 prevention through crowd collaboration using conversational AI & social platforms. *AI for Social Good Workshop*.
- [24] Savolainen, R. Assessing the credibility of Covid-19 vaccine disinformation in online discussion. *Journal of Information Science*, 2021.
- [25] Xue J, Chen J, Hu R, Chen C, Zheng C, Su Y, Zhu T. Twitter Discussions and Emotions About the COVID-19 Pandemic: Machine Learning Approach. *Journal of Medical Internet Research*, 2020;22(11):e20550.
- [26] Jelodar, H.; Wang, y.; Orji, R.; Huang, H. Deep sentiment classification and topic discover on novel coronavirus or covid-19 online discussions:

NLP using lstm recurrent neural network approach. IEEE Journal of Biomedical and Health Informatics, 24.8, 2733-2742 (2020).

- [27] French, M.; Popal, A.; Rahimi, H.; Popure, S.; Turkstra, J. Institutionalizing Participatory Slum Upgrading: A Case Study of Urban Co-production from Afghanistan, 2002–2016. Environ. Urban. 2018, 31, 209–230.



JAWAD HAQBEEN received M.S. degree in Computer Science from Waseda University, Japan, in 2013. 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 is recipient of the Global

Young Scientist Summit award in 2021, received the best presentation award at the 16th International Conference on Knowledge, Information and Creativity Support Systems (2021), the IBM Award of Scientific Excellence (2020), the best paper award at the 15th International Conference on Knowledge, Information and Creativity Support Systems (2020), and the best paper presentation award of the IEEE Nagoya Branch (2018). He is currently research member of JST CREST project.



SOFIA SAHAB is a specially appointed researcher at Kyoto University, Japan. She previously worked as assistant professor with Nagoya Institute of Technology, Japan, and Kabul University, Afghanistan. In 2017, she received a Ph.D. in urban planning from Nagoya Institute of Technology, Japan. Her current research interests include developing participative online decision support systems for supporting city planning processes in collective intelligence. She has

published research articles in journals, such as Journal of Simulation and Gaming (SAGE Publications) and Journal of Architecture and Planning (Transactions of Architectural Institute of Japan). She is one of the recipients of the Japan Science and Technology Agency AIP program. She is currently research member of JST CREST project.



TAKAYUKI ITO is a professor of Kyoto University. He received the B.E., M.E., and D.-Ing. from Nagoya Institute of Technology in 1995, 1997, and 2000, respectively.

He was JSPS research fellow, an associate professor with JAIST, visiting scholar at USC/ISI, Harvard University and MIT, and professor with Nagoya Institute of Technology, Japan. He was board member of IFAAMAS, PC-chair of PRIMA09 and AAMAS13, and general-chair of PRIMA14, ICA16, PRIMA20 and local arrangement chair of IJCAI20 international conferences. He was recipient of the JSIA achievement, JSPS prize, JSSST, MEXT, MEXT young scientist, IPSJ Awards and AAMS06 Best Paper Award. He is the author/co-author of more than 213 conference proceedings articles, and more than 62 journal articles. His research interests include artificial intelligence, crowd-based discussion support systems, collective intelligence and multi-agent systems. He is senior member of ACM, AAAI, JSAI, IEICE, JSST, SICE, JEA, and IPSJ. He is currently principal investigator of JST CREST project.