# Addressing Popular Concerns regarding COVID-19 Vaccination with a Persuasive Chatbot

Lisa A. CHALAGUINE [a] and Anthony HUNTER [a]
[a] *Department of Computer Science, University College London, London, UK*

**Abstract.**
Chatbots have the potential of being used as dialogical argumentation systems for behaviour change applications. They thereby offer a cost-effective and scalable alternative to in-person consultations with health professionals that users could engage in from the comfort of their own home. During events like the global COVID-19 pandemic, it is even more important than usual that people are well informed and make conscious decisions that benefit themselves. Getting a COVID-19 vaccine, once one becomes available, is a prime example of a behaviour that benefits the individual, as well as society as a whole. In this paper, we present a chatbot that engages in dialogues with users who do not want to get vaccinated, with the goal to persuade them to change their stance and get a vaccine. The chatbot is equipped with a small repository of arguments that it uses to counter user arguments on why the user is reluctant to get a vaccine. We evaluate our chatbot in a study with participants.

**Keywords.** chatbots, argumentative persuasion systems, computational persuasion, natural language argumentation, concerns, knowledge base construction

## 1. Introduction

Conversational agents, also known as chatbots, have the potential of being used as dialogical argumentation systems for behaviour change applications and could thereby offer a cost-effective and scalable alternative to in-person consultations with health professionals. During events like the global COVID-19 pandemic, it is even more important than usual that people are well informed and make conscious decisions that benefit themselves and society. One such example is the willingness to get a vaccine, once one is developed. Vaccines have historically proven to be highly successful and cost effective public health tools for disease prevention [13]. But the effectiveness of a vaccine in controlling the spread of COVID-19 depends on the willingness to get vaccinated in the general population. A sufficiently high vaccine coverage may generate herd immunity, which will protect everyone, including those particularly susceptible to the virus [6]. However, a barrier to reaching herd immunity is the prevalence of people who refuse or are hesitant to take vaccines [9,14]. For example, a recent YouGov survey[1] found that one in six

[1] https://docs.cdn.yougov.com/5mkju0kxbj/CCDH_RESULTS_062620_PR%20(002).pdf

UK respondents indicated that they “definitely” or “probably would not” get vaccinated if a COVID-19 vaccine became available, and another 15% do not know whether they would get one. Interviewing all those people who refuse to vaccinate in person and trying to convince them to get the vaccine, would be an impossible task. A chatbot, however, could engage in a persuasive dialogue with people over the internet from the comfort and safety of their own home.

Arguments can be used to provide information and overturn misconceptions [8]. They are an essential part of sensible discussions on controversial and problematic topics. However, different people worry about different things and hence arguments for not getting a COVID-19 vaccine will vary amongst people. One person might be worried about the side effects of a newly developed vaccine, whereas someone else might have lost all trust into the government and might be reluctant to follow the government’s advice to get vaccinated. A chatbot could address those different concerns by providing counterarguments tailored to the different user arguments and during the course of an argumentative dialogue, try to persuade the user to change their stance about getting vaccinated.

In this paper, we present a chatbot that engages in persuasive dialogues with users who are reluctant to get a COVID-19 vaccine, once it becomes available. We show how new concerns can be automatically identified as the number of chats with different users increases. We also demonstrate how to automatically evaluate the persuasiveness of the chatbot’s arguments in order to update the chatbot’s central repository of arguments, also known as knowledge base, and increase the persuasiveness of future dialogues.

The rest of the paper is structured as follows: Section 2 gives some background theory on concerns and the use of argument graphs to construct chatbot knowledge bases; Section 3 gives the aim of the paper and the hypotheses; Section 4 describes the chatbot architecture that was used for the experiments; Section 5 describes the experiments that were conducted with the chatbot, Section 6 presents the results, and in Section 7 we conclude our findings.

## 2. Conceptualising Argumentation

### *2.1. Using Arguments and Concerns for a Persuasive Chatbot*

In order for the chatbot to be able to engage in persuasive dialogues, it needs to be equipped with arguments from both sides. It needs arguments for (not) engaging in a certain behaviour, that could potentially be given by the user, and arguments that attack the user’s arguments (counterarguments). Furthermore, the chatbot should also be able to identify the *concerns* of the user, a concern being a matter of interest or importance to the user. We have shown in a previous study unrelated to healthcare [3], that arguments and counterarguments can be crowdsourced and used as the chatbot’s knowledge base. Further, we have shown that the chatbot can automatically identify the concerns of the user that she raises in her arguments during the chat, in order for the chatbot to present arguments that address the user’s concerns.

Taking the user’s concerns into account when presenting an argument is important. The chatbot might present a perfectly valid argument that the user might not even disagree with. However, the chatbot’s argument may have no impact on her stance, if the argument does not address her concerns. Whereas, if the chatbot presents an argument that addresses her concern, the user is more likely to be convinced. In order to choose a suitable counterargument in a given dialogue, our chatbot tries to identify the concern of the user argument and counters with an argument that addresses the user's concern.

To illustrate how concerns arise in argumentation, and how they can be harnessed to improve persuasion, consider a person who is reluctant to the idea of getting a COVID-19 vaccine, should one be released, due to the short time it took to be developed compared to other vaccines. The chatbot (that argues for getting a COVID-19 vaccine) could choose one of the following arguments:

- Option 1: A vaccine is the only safe way to create herd immunity which is necessary to stop the virus spreading. This will protect us all from getting the virus, as well as those who, for some reason, cannot get the vaccine.
- Option 2: Funding and research priority is currently directed at developing COVID-19 vaccines around the world. As a result, the process is being sped up, allowing for more clinical tests to take place in a shorter period of time.

Both arguments are perfectly reasonable. However, option 2 addresses the concern regarding the short time frame by acknowledging it and providing an argument on why the user should not worry about it, whereas option 1 does not, even though it provides a valid argument for getting a COVID-19 vaccine.

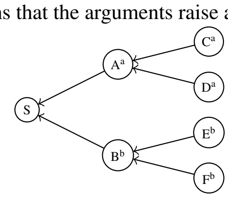
In our previous work we have shown that concerns can automatically be identified by inspecting meaningful, common words of a set of arguments [2]. We used the difference of the users' stance before and after the chat with the chatbot to measure the *persuasiveness* of the dialogues. The scale was from 1-5 (*strongly disagree* to *strongly agree*). For example, if a user changed her stance from *neutral* to *agree* it was considered a positive increase in stance and therefore a persuasive dialogue. We showed that using counterarguments that address the user's concern increases the persuasiveness of a dialogue compared to using other valid arguments that do not address the user's concerns.

### 2.2. Using an Argument Graph as Chatbot Knowledge Base

Argument graphs as proposed by Dung [5] are directed graphs where each node denotes an argument and each arc denotes one argument attacking another. Such graphs provide a useful representation to study attack and support relationships of a given set of arguments. An example of such an argument graph is shown in Figure 1: the chatbot would present the goal argument $S$ (i.e. an argument that we want the persuadee to believe). The user could give an argument, let us assume it is argument $A$ that raises concern $a$. Suppose the chatbot counters it with argument $C$ that addresses concern $a$ and the user replies with another argument, for example, argument $B$, that raises concern $b$ and so on.

In our previous studies we have used chatbots to persuade people to cycle more, instead of driving or using public transport [7]; to reduce meat consumption [4], and to change their stance on UK university fees [3]. The arguments were either compiled and edited by hand by the researchers [7] or crowdsourced [4,3] and stored as directed graphs in the chatbot's knowledge base (KB). In [2] we described a method for automatically acquiring a large argument graph via crowdsourcing. Figure 2 shows a schematic representation of depths in the argument graph. The graph starts with the goal argument (Depth 0). The next level of depth (Depth 1) contains arguments that counter the goal statement. The arguments in Depth 2 counter the arguments in Depth 1 and so on. In

**Figure 1.** Argument graph where child nodes are attacking parent nodes. Each circular node is an argument and each arc denotes an attack. So if there is an arc from X to Y, then X attacks Y, and so X is a counterargument to Y. The superscripts denote the concerns that the arguments raise and the counterarguments address.



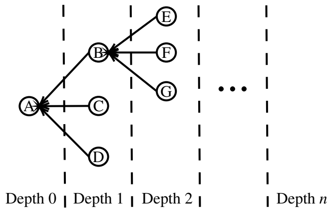

[3] we presented a chatbot that was equipped with the crowdsourced argument graph. The chatbot matched the incoming user argument with a similar argument in the graph (target argument) using cosine similarity of the vector representations of the two arguments (the vectors were created using GloVe word embeddings [12]). The concern of the argument was identified with a multinomial logistic-regression classifier that was trained on the arguments in the graph. If a target argument in the graph was found, the chatbot presented one of the counterarguments of the target argument that addressed the same concern as the user argument in the dialogue. This way the chatbot was able to engage in argumentative dialogues with the users.

In contrast to our previous works where we covered established domains (e.g. UK university fees), COVID-19 is a novel domain. Due to the novelty of the problem and the lack of information, the domain of potential novel arguments against a COVID-19 vaccine is smaller and less complex than, for example, the domain of arguments for abolishing/maintaining university fees. The number of novel arguments in the chatbot's KB in our previous study[2] was therefore much higher than in this study (how the arguments for the chatbot's KB were crowdsourced and how the KB was constructed will be explained in Section 4.1). It should be noted that the breadth and depth of the argument graph is not a function of the complexity of the topic but rather it is a function of what is in the public domain for ordinary people. Therefore, if media coverage and public engagement in open discussion of the topic increases, so does the argument graph obtained by crowdsourcing.

Taking the concern *student finance* as an example, there are many arguments raising the financial hardship students are facing during and after finishing university. The following three arguments are all distinct, despite all raising the concern *student finance*.

- The amount asked to be paid by the students is very high and most of them might need to take a student loan to continue their studies, placing them in debt even before they start to work.

**Figure 2.** Representation of depths and attack relationships between arguments in our argument graph. Arguments *B, C* and *D* are counterarguments to *A*.



- The student loan interest rates are quite high and students end up repaying much more than they initially borrowed.
- The interest rates are high and puts pressure on the student to find a highly paid job as fast as possible after graduating.

The arguments that address concerns regarding a COVID-19 vaccine, however, are less diverse, as the following three arguments about potential side effects of the vaccine demonstrate:

- I will not get a COVID-19 vaccine because of its potential side effects.
- It's a new vaccine, so we don't yet know what the side effects are.
- The vaccine may have a lot of side effects that could be more dangerous than COVID-19 itself.

The arguments regarding side effects are very similar, and could all be countered with the same argument, for example, that there is high scrutiny over the research on those particular vaccines and nobody would allow giving it to the public if it was unsafe. Whereas, acquiring individual counterarguments would make the graph unnecessarily big and might result in the inclusion of many similar counterarguments.

In the following sections, we outline our hypotheses and describe how we acquired the arguments for the chatbot's KB in order to build a chatbot that can engage in argumentative dialogues to persuade people to get a COVID-19 vaccine, and the experiments conducted with the chatbot.

## 3. Hypotheses

In this paper, we present a chatbot that utilises a set of arguments for taking a COVID-19 vaccine as a KB. The chatbot uses concerns to make strategic choices of moves in order to engage in argumentative dialogues with users to persuade them to get the vaccine.

Given this setting and the nature of the topic, we want to address three questions which were not been tackled in our previous works: Firstly, whether a small and shallow graph as the KB is enough to counter the majority of arguments people might have for not getting the vaccine and thereby create persuasive dialogues. Secondly, whether it is possible by only identifying the concern of a user argument to give a suitable counterargument, without having to use a similarity score and relying on similar target arguments in the graph, as we did in our previous study [3]. And finally, whether the persuasiveness of the chatbot can be increased after chatting to a certain number of people, by updating its KB. We summarise these points in the following three hypotheses:

**H1** Given a novel domain, a relatively small graphical representation of a knowledge base, can be used to represent most of the possible arguments that a set of normal users would know, and can be utilised by a chatbot to create persuasive dialogues, meaning that the stance of the user changes after the chat.

**H2** Given a novel domain, the arguments that address the same concern are sufficiently similar to allow for the provision of suitable counterarguments just by identifying the concern of the arguments.

**H3** The persuasiveness of dialogues can be increased automatically by (1) identifying more (respectively less) convincing arguments and either raising (or respectively lowering) the probability of them being used during the dialogues and by (2) identifying new concerns and adding suitable counterarguments to the chatbot's knowledge base.

In the remainder of this paper, we describe the design of our chatbot that was used for the argumentative dialogues and explain the experiments conducted with the chatbot in order to test our hypotheses.

## 4. Chatbot Design

In this section, we describe the acquisition of arguments used to construct the chatbot's KB, and the concern classifier, used by the chatbot to identify the concerns of the incoming user arguments.

### *4.1. Knowledge Base Construction*

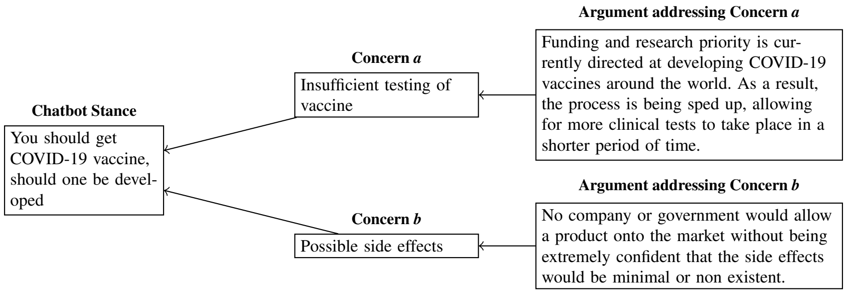
We used a similar method to that described in [2] to acquire an argument graph that was used to construct a KB for the chatbot: for the first level of depth, we recruited 33 participants via Prolific[2], which is an online recruiting platform for scientific research studies, and asked them to provide three arguments against getting a COVID-19 vaccine. We identified 4 concerns that were raised by the majority of the 99 arguments by inspecting the most common, meaningful words, namely: **side effects**, **cost**, **mutation** and **insufficient testing**. Due to the novelty of the topic and the lack of knowledge about the (not yet released) vaccine, the arguments given by users were less diverse and easier to cluster together by the concerns they were addressing. Some examples of arguments that talk about potential side effects of a vaccine are given in Section 2.2. We, therefore, decided to not include individual arguments in the chatbot's KB, but only the concerns, and crowdsource counterarguments that address the concerns, as opposed to countering individual arguments. This makes the resulting graph much smaller and shallower than the traditional argument graph used for the university fees study.

Figure 3 shows a snippet of the graph, where the root node is the goal argument (pro COVID-19 vaccine), the nodes of depth 1 are the concerns that people have, and depth 2 the counterarguments to the user arguments that address that specific concern in depth 1. All concerns, their descriptions and when they were discovered can be found in our github repository [1].

We then collected counterarguments for the 4 identified concerns, which made up the second level of depth of the argument graph. We recruited 20 participants on Prolific and asked them to provide a counterargument for each concern. Participants were presented with the concern (e.g. side effects) and 3 similar arguments that raised the concern in question. They were instructed to give an argument that addressed the concern and countered the given arguments. Out of the 20 acquired counterarguments for each concern, we manually selected 3-5 arguments we considered of high quality based on factors like argument length, grammar, language and validity. We also made sure to

[2]`https://prolific.co`

**Figure 3.** Part of the chatbot's knowledge base with concerns *a* and *b* representing clusters of arguments in level 1 (concerned about insufficient testing of the vaccine and possible side effects respectively) attacking the chatbot's goal argument, and arguments in level 2 addressing the concerns.




only include distinct arguments in the graph, in order to prevent the chatbot from giving similar arguments during the chats.

In order to collect arguments for the third level of depth of our graph, we recruited 20 participants per concern. They were presented with the selected counterarguments for the graph and asked to counter them individually. We noticed that the majority of arguments were very similar to the arguments given in depth 1, raising the same concerns. The arguments that were expressing new concerns, however, could have been easily given in depth 1, meaning that the context of the preceding arguments in depth 1 and depth 2 was not necessary to understand the arguments in depth 3. We, therefore, decided to keep the graph at only two levels of depth: depth 1 being the concerns for not getting a COVID-19 vaccine and depth 2 containing the arguments that address the concerns in depth 1. We identified 4 new concerns during the argument acquisition for depth 3, namely: **long-term side effects**, **mistrust in the government**, **increase in taxes** and **priority in kickstarting the economy**. We acquired counterarguments that address these concerns in the same way as described above for depth 2.

The chatbot also needed default arguments for getting a COVID-19 vaccine that it could use in case the concern of the user argument could not be identified. These arguments are therefore not counterarguments in the traditional sense, as they do not refer to or address the concern of the user argument but instead "change topic" and present a new issue in the debate. We also used crowdsourcing for the acquisition of default arguments and included the best 4 arguments in the chatbot's KB. We also added phrases like *"Ok but", "Have you considered that"* and *"Nevertheless"* to the beginning of the default arguments to indicate that a deviation from the topic occurs. This way the dialogue would resemble argumentation as it would happen between two people: if two human agents engage in an argumentative dialogue, just because one presents an argument the other cannot counter, the dialogue does not necessarily end at that point. The other agent might switch topics and present a new argument he or she believes in, without referencing and directly countering the previous argument.

After our initial argument acquisition, we ended up with 8 concerns and 3-5 counterarguments that addressed each concern (31 arguments in total). This could be represented as a shallow graph where depth 1 has 8 nodes that are all attacking the root statement, which is to get a COVID-19 vaccine, and each of these 8 nodes has 3-5 child nodes in depth 2 which represent counterarguments that are attacking the concern nodes.

*4.2. Understanding the user input*

The chatbot identified the concern of the user argument using a classifier and chose a counterargument from the graph that addressed the same concern, without using a similarity score and relying on similar arguments in the KB, as in our previous study. For the initial classifier, we took the arguments we acquired during the initial argument acquisition process as the training set, using the Python Scikit-learn library[3]. The classifier used multinomial logistic regression and a binary feature representation of the arguments (one-hot encoded vectors) in order to identify the concern of the incoming user argument. We extracted the top two concern predictions. If the top prediction was over 50% in confidence, the argument was labelled with one concern, if the top prediction was below 50% but both of the top 2 predictions were above 30%, the argument was labelled with two concerns, otherwise with none. If a concern was identified, the chatbot chose a counterargument from the graph that addressed the identified concern. If a user argument was labelled with two concerns, an attacker was chosen that addressed the concern with the higher predicted value. If all arguments that addressed the given concern have been used, the chatbot selected an argument that addressed the second concern (if there was one) or selected a default argument. When all arguments had been used up, the chat ended.

## 5. Experiments

The purpose of the chatbot was to test all three of our hypotheses. The chatbots were deployed on Facebook via the Messenger Send/Receive API[4].

Prior to recruiting participants for the study, we ran a survey where we asked 450 people to choose from a scale of 1-5 whether they would get a COVID-19 vaccine, should one be developed. The options were *very unlikely, unlikely, I don't know, likely* and *very likely*. We recruited 150 participants in three batches of 50 from those that chose *very unlikely, unlikely* and *I don't know*, i.e. those that had a negative stance.

Before the chat, the users were directed to a Microsoft Form and asked again how likely they would get a COVID-19 vaccine. After submitting their answer they were redirected to the Facebook page where they could begin the chat. The chatbot started the chat by instructing the user that they could end the chat anytime by sending the word *"quit"* and then asking why the user would not get a COVID-19 vaccine, should one be developed. The user then presented their first argument. The chatbot replied with either a counterargument from the argument graph or a default argument, depending on whether it could identify the concern of the user argument. The counterarguments were stored in a Python dictionary with the concerns as the keys and the list of counterarguments that addressed that concern as the values. If the concern could be identified, the chatbot replied with the first counterargument in the list. If the message of the user was less than 7 words in length and contained a negation, the chatbot queried *Why?* or *Why not?* to force the user to expand. This process was repeated with each subsequent argument given by the user. The chatbot would end the chat as soon as all default arguments were used up and no concern could be identified, or all arguments that addressed the concern

[3]`https://scikit-learn.org`

[4]Since then we host our chatbot on a separate website

were also used up. At the end of the chat the chatbot presented the user with a link that redirected them to another Microsoft Form where they were asked a series of questions:

1. Did you feel understood by the chatbot? (Yes/No/Sometimes)
2. Did you feel that the chatbot's arguments were relevant? (Yes/No/Some of them)
3. Do you feel like all your points were addressed? (Yes/No/Some of them)
4. How likely would you get a COVID-19 vaccine, should one be developed? (Very unlikely - very likely)
5. If the chatbot was successful in changing your mind - tell us why. If the chatbot was not successful, tell us why not. (Open-ended)

Questions 1-3 were used to test our second hypothesis and judge the relevance, length and quality of the chats, and question 4 was to test our first hypothesis and compare the stances of the participants before and after the chat with the chatbot in order to judge persuasiveness.

In order to test our third hypothesis, we analysed the chats after every 50 participants in order to (1) identify new concerns and update the training set for the concern classifier with samples of arguments that raised the newly identified concern, and (2) to identify more and less convincing arguments. After analysing the first 50 chats we discovered that the word *healthy* came up a number of times and inspected the arguments that contained this word. Several participants indeed claimed that they were young and healthy and would be fine coping with COVID-19. We, therefore, included **healthy** into our set of concerns and added the user arguments into the training set of arguments for the concern classifier, so that the classifier was able to identify this concern in the next iteration and provide suitable counterarguments. We also crowdsourced counterarguments for that concern and added the best into the chatbot's KB.

We further analysed the reactions people gave to the most commonly used arguments by the chatbot in order to identify more and less convincing ones. If the majority of people disagreed with an argument by stating '*I disagree, That's a lie* or *This is not true*, we deleted that argument form the chatbot's KB. If only some participants disagreed with an argument, it was moved further down the list and vice versa - if people agreed with an argument it was moved up the list to increase the probability of the chatbot using it in the chat.

After analysing another 50 chats (100 in total) the order of argument's was not changed because very few participants explicitly disagreed with an argument so that it could be automatically detected, and the arguments that participants agreed with were already at the top of the list. However, we discovered two new concerns when analysing 100 chats, namely **natural herd immunity** and **alternative protection**. After another 50 chats (150 in total) a further concern was discovered - that only **vulnerable** people should get the vaccine. These concerns also came up in the first iteration, however, only after collecting 100-150 chats they could be identified via automatic means by analysing common word patterns. Our final chatbot's KB contained 38 arguments to address 12 concerns.

## 6. Evaluation of the Chatbot

Table 1 shows how many participants changed their stance for each chatbot after engaging in an argumentative dialogue with that chatbot. We divided the change in stance into

**Table 1.** Percentage of participants that changed their stance for each chatbot after the knowledge base was updated. The last row also includes the participants that did not change their stance but gave positive feedback for Question 5.

| Change of stance | Chatbot 1 | Chatbot 2 | Chatbot 3 |
|---|---|---|---|
| **Negative to Neutral** | 8% | 26% | 26% |
| **Neutral to Positive** | 8% | 10% | 14% |
| **Negative to Positive** | 2% | 2% | 2% |
| **Total** | 18% | 38% | 42% |
| **Including Q5** | 24% | 58% | 54% |

3 categories: a change from negative to neutral (from *very unlikely/likely* to *I don't know*); a change from neutral to positive (from *I don't know* to *likely/very likely*); and a change from negative to positive (from *very unlikely/unlikely* to *likely/very likely*). We do not consider a change from *very unlikely* to *unlikely*. One can see that even the first chatbot has an impressive persuasion rate of 18%. This verifies our first hypothesis - that a small, shallow argument graph can be utilised by a chatbot to create persuasive dialogues.

We also asked participants to give a short explanation of why the chatbot was either successful or unsuccessful in changing their stance. Overall 45% of the participants gave positive feedback. The positive feedback of those participants that were not convinced was mainly that despite the chatbot making good points, they were still too worried about something (e.g. long term side effects). 42% of participants stated that they were sticking to their point of view, however, without complaining about the performance of the chatbot. Only 13% criticised the chatbot for not addressing their concerns and not feeling understood. The last row in Table 1 includes the participants who did not change their stance after chatting with the chatbot, but gave positive feedback, either by stating that the chatbot gave good arguments or that the chatbot "made them think".

One can see that there was a noteworthy increase after the first update of the KB, where unconvincing arguments were deleted and convincing arguments pushed forward in the list. This is a statistically significant increase with a p-value of 0.025 using Chi-Square and therefore verifies our third hypothesis, that the persuasiveness of dialogues can be increased by automatically identifying more and less convincing arguments and either increasing or lowering the probability of them being used in the chat. The fact that the persuasion rate only increased by 4% after identifying two new concerns and adding counterarguments to address those to the chatbot's KB, can be explained by the small sample size. The arguments about **natural herd immunity** and **alternative protection** were only identified after recruiting 33 participants for the initial argument acquisition *and* after analysing further 100 chats where participants presented 11 arguments on average. Therefore, expecting a significant increase in the persuasion rate after another 50 chats could not be expected.

Table 2 shows the results for the first three questions for all the chatbots combined (150 participants). We did not differentiate between the different chatbots because the numbers were approximately the same. One can see that the majority of the participants

**Table 2.** Answers to the first three questions for each of the three chatbots.

| **Understood (Q1)** | | | **Relevant Args (Q2)** | | | **Points addressed (Q3)** | | |
|---|---|---|---|---|---|---|---|---|
| Yes | No | Sometimes | Yes | No | Some | Yes | No | Some |
| 35% | 13% | 52% | 61% | 6% | 33% | 30% | 22% | 47% |

**Figure 4.** Example chat between a participant and the last iteration of the chatbot (Chatbot 3). Chatbot arguments are in the dark boxes and user arguments in the light boxes. Default arguments are indicated with an *. The chat begins in the left column and continues in the right one.

So tell me, why would you not consider getting a COVID-19 vaccine, should one be developed?

I worry that normal vaccine take decades to be approved and a COVID-19 one may be rushed and therefore ineffective and potentially dangerous

Funding and research priority is currently directed at developing COVID-19 vaccines around the world. As a result, the process is being sped up, allowing for more clinical tests to take place in a shorter period of time

This may mean missing crucial steps in the process stage

*Nonetheless, a vaccine is the only safe way to create herd immunity which is necessary to stop the virus spreading. This will protect us all from getting the virus, as well as those who, for some reason, cannot get the vaccine (e.g. allergy to the vaccine)

I believe a vaccine will be necessary at some point but I wouldn't feel safe getting a rushed one. Instead I hope that medication can be found to fight the symptoms

Just because the vaccine has been developed quickly it doesn't mean it is any less safe than one that has taken longer to be developed. However lots of treatments come with possible side effects, It's about weighing the possible side effects against the effects of getting COVID

I agree that it is about weighing up the side effects. However, I would be sceptical

There is high scrutiny over the research on those particular vaccines and nobody would allow it to give it to the public if it is unsafe. A high number of places independently developing vaccines in different parts of the world makes the outcome of trials much more trustworthy on different populations

I understand that it will not be given to the public if it is unsafe but it is hard to know what impact the vaccine could have in the future

with any medicine, there can be severe side effects but these only occur in a small number of cases. The benefits of a vaccine far outweigh the risk

But we do not know the side effects yet

Vaccines are well researched before releasing, they need to go through multiple stages of trails. Risk of COVID is higher than the risk of side effects

I understand they are well researched but I would not want to rush into obtaining one

No company or government would allow a product onto the market without being extremely confident that the side effects would be minimal or non-existent.

A company may if it means profiting from the vaccine

*Ok, but lockdowns are crippling to the economy, as well as your mental health and well-being. Getting a vaccine will make you feel safe when you go to crowded places and allow you to return to your normal life

I agree, but I think it is a risk I'm willing to take. I do not expect other people to feel the same as me

*Despite that, a vaccine will protect you from getting sick and reduce the risk of getting COVID-19 and avoid potentially fatal complications that COVID-19 can cause.

I agree. But I will also try to maintain social distancing. Eventually, I will get the vaccine

We cannot maintain social distancing and other restrictions forever. Eventually, we all want to go back to the gym, theatre and restaurants without having to wait in massive queues due to limited spaces, and having to wear face masks

I understand, but that is other people's choices, not mine

You cannot always maintain a save distance to someone and you might get less cautious over time. Getting a vaccine is easier and saver

I agree

*I'm happy you agree. Don't you also think that getting a vaccine will protect not just you but also reduce the risk of you spreading the disease and potentially infecting vulnerable people

I agree with that. I would get a vaccine if I was sure it was safe. Until then, I will social distance to minimise the risk of spreading

I'll stop here. It has been nice talking with you. I hope you think about my points and do consider taking the vaccine if it becomes available

considered the chatbot's arguments as relevant in most cases and answered the first three questions with either *yes* or *sometimes*. These results are similar to the ones we got in our previous study for a strategic chatbot that took the concerns of the users into account, compared to a baseline chatbot that did not[5]

Given that the chatbot did not use natural language generation and was not able to address novel arguments or expand on existing ones by giving more information, and

[5]Points addressed has a higher satisfaction rate than the strategic chatbot in our previous study (20%), however, the sample size of the previous study was only 50 participants, compared to 150 in this study.

solely relied on correct concern classification, the given results are promising. The length of the chats were on average 21 alternating turns (disregarding the chatbot querying the user on short negated statements). This means that the chatbot, on average, gave 10 arguments, 4 of which were default arguments and 6 from the graph. This supports our second hypothesis that only by identifying the concern of an argument, suitable counterarguments can be presented and that the resulting argumentation dialogues are of satisfactory length and quality. An example of a chat can be seen in Figure 4. All chatlogs and the code for the chatbot can be found on github [1].

Side effects are by far the most popular concern, with 43% of user arguments given during the chats (where a concern could be identified) raising it. This is coherent with previous studies which analysed vaccine hesitancy in France [15], the US [10] and the EU [11]. The second most prevalent concern was about potential insufficient testing of the vaccine (23%). Interestingly, certain concerns that were prominent during the initial argument acquisition, like the potential cost of the vaccine and the possibility of mutation of the virus, were not as prominent during the chats with the chatbot with only 3% and 1% of the user arguments raising them respectively.

## 7. Conclusion

Our contribution in this paper is threefold. Firstly, we have shown that for a new domain, where public awareness is still limited, a small argument graph can be used to represent most of the possible arguments in this domain and utilised by a chatbot to create persuasive dialogues and presented a method how to acquire and structure such a graph. Secondly, we have demonstrated that no sophisticated natural language understanding of the user arguments is needed in order to provide suitable counterarguments that address the concerns of the users. And thirdly, we have shown that the persuasiveness of the dialogues can be increased automatically by identifying more and less convincing arguments and updating the chatbot's KB accordingly. We have also validated that new concerns can be identified as more data comes in, which has the potential to increase the persuasiveness further by making the chatbot able to address more concerns in the future by presenting suitable counterarguments.

The advantage of using a chatbot for such a task is that a chatbot is able to address millions of people at the same time in the comfort of their own home and collect a vast amount of data in a very short time. Our method of analysing the incoming user arguments scales easily and allows obtaining many arguments from different people. The more data comes in, the easier it gets to identify patterns, discover new concerns, acquire arguments that address these concerns and update the KB of the chatbot accordingly. This allows us to identify common misconceptions, address the lack of information and potentially even fake news. The belief that COVID-19 is a "Chinese hoax" and that Bill Gates wants to use the vaccine to insert micro chips into us all, are prevalent in certain groups of the population and with more data a classifier could easily identify those fake news and address them with suitable counterarguments.


## Acknowledgements

The first author is funded by a PhD studentship from the EPSRC.

## References


[1] https://github.com/lisanka93/covid-19_vacc_bot.

[2] L. A. Chalaguine and A. Hunter. Knowledge acquisition and corpus for argumentation-based chatbots. In *Proc. of the 3rd Workshop on Advances In Argumentation In Artificial Intelligence*, pages 1–14. CEUR, 2019.

[3] L. A. Chalaguine and A. Hunter. A persuasive chatbot using a crowd-sourced argument graph and concerns. In *Proc of Computational Models of Argument (COMMA)*, pages 9–20. IOS Press, 2020.

[4] L. A. Chalaguine, A. Hunter, F. L. Hamilton, and H. W. W. Potts. Impact of argument type and concerns in argumentation with a chatbot. In *Proc. of the 31st International Conference on Tools with Artificial Intelligence*, pages 1557–1562. IEEE, 2019.

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

[6] P. Fines, K. Eames, and D. Heymann. "Herd immunity": a rough guide. *Clinical Infectious Diseases*, 52(7):911–916, 2011.

[7] E. Hadoux and A. Hunter. Comfort or safety? Gathering and using the concerns of a participant for better persuasion. *Argument & Computation*, 10(2):113–147, 2019.

[8] A. Hunter. Towards a framework for computational persuasion with applications in behaviour change. *Argument and Computation*, 1:15–40, 2018.

[9] N. MacDonald. Vaccine hesitancy: Definition, scope and determinants. *Vaccine*, 33(34):4161–4164, 2015.

[10] J. E. Marco-Franco, N. Guadalajara-Olmeda, S. G. de Julián, and D. Vivas-Consuelo. Covid-19 healthcare planning: Predicting mortality and the role of the herd immunity barrier in the general population. *Sustainability*, 12(13):5228, 2020.

[11] S. Neumann-Böhme, N. V. NE, I. Sabat, P. Barros, W. Brouwer, J. van Exel, J. Schreyögg, and T. Stargardt. Once we have it, will we use it? a european survey on willingness to be vaccinated against covid-19. *The European Journal of Health Economics*, 21:977–982, 2020.

[12] J. Pennington, R. Socher, and C. D. Manning. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, page 1532–1543. ACL, 2014.

[13] V. Rémy, N. Largeron, S. Quilici, and S. Carroll. The economic value of vaccination: why prevention is wealth. *Journal of Market Access & Health Policy*, 3(1), 2015.

[14] S. M. Sherman, L. E. Smith, J. Sim, R. Amlot, M. Cutts, H. Dash, G. J. Rubin, and N. Sevdalis. Covid-19 vaccination intention in the uk: Results from the covid-19 vaccination acceptability study (covaccs), a nationally representative cross-sectional survey. *medRxiv*, 2020.

[15] J. K. Ward, C. Alleaume, and P. Peretti-Watel. The french public's attitudes to a future covid-19 vaccine: the politicization of a public health issue. *https://doi.org/10.31235/osf.io/xphe9*, 2020.