

# Feasibility and Acceptance of chatbots embedded in healthcare curricula:



Matthew Pears- with CEPEH Team James Henderson,  
Natalia Stathakarou, Klas Karlgren, Panagiotis D Bamidis,  
Iraklis Tsoupouroglou, Eirini Schiza, Constantinos S  
Pattichis, and Stathis Th Konstantinidis

CEPEH report 2023

—

*December*

2023

see [cepeh.eu](http://cepeh.eu) for more information

# Acknowledgements

This work is supported by the ERASMUS+ Strategic Partnership in Higher Education “Chatbot Enhance Personalise European Healthcare Curricula (CEPEH)” ([www.cepeh.eu](http://www.cepeh.eu)) (2019-1-UK01- KA203-062091) project of the European Union.

Thanks to Mohammed Roshan for data processing.

CEPEH Team

# Abstract

This document details the evaluation of each resource in terms of the feasibility and acceptance from the end-users. There was evidence of identifying the feasibility of such resources into formal training and studies exist on the acceptance of such resources, with promising results. However, all these studies defined the need for further research in the area until the use of chatbots in healthcare education became common. Furthermore, the creation process of CEPEH resources was significantly different and had improvements to current methods, due to the co-creation process, and use of low cost but effective technology.

# Contents

<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>viii</b>
<b>List of Abbreviations</b>	<b>ix</b>
<b>Introduction</b>	<b>1</b>
Background . . . . .	2
<b>1 Method</b>	<b>4</b>
1.1 Participants . . . . .	4
1.2 Procedure . . . . .	5
1.3 Design . . . . .	6
<b>2 Results</b>	<b>11</b>
2.1 Participants' Characteristics . . . . .	11
2.2 Chatbot Usability Questionnaire (CUQ) . . . . .	14
2.3 System Usability Scale (SUS) Scores . . . . .	15
2.4 Technology Acceptance Model . . . . .	16
2.5 Other Measures . . . . .	17
2.6 Inferential Statistics . . . . .	22
2.7 Other Findings . . . . .	23
<b>3 CEPEH Focus Group Discussion Analysis</b>	<b>26</b>
3.1 Tokenising . . . . .	26
3.2 Plotting word frequencies - bar graphs . . . . .	27
3.3 Word clouds . . . . .	30
3.4 Sentiment analysis . . . . .	32

## *Contents*

<b>4</b>	<b>Discussion</b>	<b>33</b>
4.1	Quantatative Results . . . . .	34
4.2	Qualatative Results . . . . .	34
4.3	Limitations . . . . .	34
4.4	Conclusions . . . . .	34
<b>5</b>	<b>(Additional Analyses) Training Events</b>	<b>36</b>
5.1	CEPEH Training Event C1 . . . . .	36
5.2	CEPEH Training Event 2 . . . . .	37
5.3	# bibliography: [bibliography/references.bib, bibliography/additional- references.bib] . . . . .	39
	<b>Appendix</b>	<b>40</b>
	<b>Appendices</b>	
<b>A</b>	<b>The First Appendix</b>	<b>42</b>
	<b>References</b>	<b>43</b>

# List of Figures

1.1	Location and Profession of Participants . . . . .	5
2.1	Chatbot Usage History- Pre . . . . .	12
2.2	Chatbots are Useful Opinion- Pre . . . . .	13
2.3	CUQ CEPEH Score . . . . .	14
2.4	CUQ Scatter Plot . . . . .	15
2.5	Improvements in Knowledge . . . . .	18
2.6	Trust Chatbots POST use . . . . .	19
2.7	Ease of Use Comparison . . . . .	21
2.8	Intend to Reuse-Post . . . . .	23
2.9	Easy to Use- Post . . . . .	24

# List of Tables

2.1	Previous Chatbot Usage of Participants . . . . .	12
-----	--------------------------------------------------	----



# List of Abbreviations

<b>CEPEH</b>	. . . . .	Chatbot Enhance Personalised European Healthcare curricula
<b>HELM</b>	. . . . .	Health E-Learning and Media team
<b>RLO</b>	. . . . .	Reusable Learning Object
<b>ASPIRE</b>	. . . .	Aims Storyboarding Production Implementation Release Evaluation
<b>NLP</b>	. . . . .	Natural Language Processing
<b>NLU</b>	. . . . .	Natural Language Understanding
<b>A.I</b>	. . . . .	Artificial Intelligence
<b>TAM</b>	. . . . .	Technology Acceptance Model
<b>SUS</b>	. . . . .	System Usability Scale
<b>CUQ</b>	. . . . .	Chatbot Usability Questionnaire
<b>HIG</b>	. . . . .	HELM is Great

# Introduction

Personalised Healthcare Education is needed to meet growing demand and quality maintenance. There is a growing evidence around chatbots, namely machine conversation systems- these programs have the potential to change the way students learn and search for information.

Chatbots can quiz existing knowledge, enable higher student engagement with a learning task, or support higher-order cognitive activities. In large-scale learning scenarios with a high student-to-lecturer ratio, chatbots can help tackle the issue of individualized student support and facilitate personalised learning. However, limited examples of chatbots in European Healthcare Curricula have been utilised to combine both the continuum of cognitive processes presented in Bloom's taxonomy, with the idea that some repetitive tasks can be done with a chatbot- to provide greater access or to scale faculty time.

Thus, CEPEH strategic partnership has co-created open access chatbots utilising artificial intelligence, promoting innovative practices in digital era, by supporting current curricula and fostering open education.

CEPEH Erasmus+ strategic partnership aimed to co-design and implement new pedagogical approaches and, in particular, chatbots for European medical and nursing schools. CEPEH used participatory design to engage stakeholders (students, healthcare workforce staff, lecturers, clinicians, etc.) in order to co-design effective chatbots and release them as open access resources. Through CEPEH, effective use of digital technologies and open education were be incorporated into healthcare curricula. This enabled students to increase their health and medical related skills through flexible learning.

## *Introduction*

CEPEH expected that students adopted this new digital pedagogy and improve their skills and competences through flexible personalised learning, while the teaching staff enhanced their e-learning tool co-creation competences and make use of co-design best practices and recommendations for use. It is also expected increased cooperation between the partners. Thus, in the long term, CEPEH expects to influence the development of medical and nursing curricula with this digital innovation, foster the quality of the future healthcare workforce and further improve international competitiveness of the partners' healthcare curricula. This document details the evaluation of the resources created by the CEPEH team.

The evaluation specifically explored the feasibility and acceptance from the end-users. These end-users are learners in European healthcare higher education institutions.

There was firstly evidence for the need to identify the feasibility of chatbots and similar resources into formal education and training, with a further need to improve access to these types of learning resources. Of course, studies exist on the acceptance of chatbots, virtual patients, and many other healthcare applications, with promising results. However, through various limitations, we believed there was further research to be completed to accelerate the design, development, implementation, and evaluation processes. These have financial, stakeholder, time, and efficacy benefits. The creation process of CEPEH resources was significantly different to most in the literature, and this report highlights the approach of the CEPEH team towards enhancing personalised healthcare education can be achieved.

## **Background**

The working practices of CEPEH are aimed at maximizing efficacy of these chatbots as learning resources, and provided a sense of shared development and ownership from all stakeholders. The process normally begins with workshops in which the project is scoped and team building occurs. The CEPEH workshops involve the widest possible team of stakeholders including tutors, students, healthcare

## *Introduction*

workers, learning technologists, health service users and carers- depending on the materials being created.

For readers who are interested in using these high quality digital resources please access them for free at [CEPEH.EU](http://CEPEH.EU)

The next section will now present the evaluation of all CEPEH chatbot resources.

# 1

## Method

### Contents

---

<b>1.1</b>	<b>Participants</b>	<b>4</b>
<b>1.2</b>	<b>Procedure</b>	<b>5</b>
<b>1.3</b>	<b>Design</b>	<b>6</b>
1.3.1	Chatbot Usability Questionnaire (CUQ)	7
1.3.2	UTAUT2 (Unified Theory of Acceptance and Use of Technology)	7
1.3.3	System Usability Scale	8
1.3.4	Computer Self-Efficacy Scale Tool	8
1.3.5	Technology Acceptance Model (TAM)	8
1.3.6	Qualitative Measure- Focus Group Discussions	9

---

## 1.1 Participants

This dataset had 14 males and 28 females therefore a total of 42 participants. It was a repeated measure design whereby each participant used the 4 chatbots developed by the CEPEH team. Therefore, there are 42 points of data in the condition before testing, and 126 data points after testing the chatbots- for a total of 168 row of data, 5 per participant. There were 78 questions asked in total, therefore the full dataset has over 4000 cells recorded.

There were 22 females and 7 males from Greece. There were 3 females and 4

## 1. Method



**Figure 1.1:** Location and Profession of Participants

males from Cyprus. There were 2 females and 2 males from Sweden, and there were 2 participants from the United Kingdom (see(1.1)).

The majority 36 participants, were student, with 3 being learning technologists, 2 were lecturers, and 1 was a doctor. Although there could be a difference in these groups, the design was within- groups therefore each participants pre-usage metrics were the comparative control data, and participant differences did not affect the evaluation.

## 1.2 Procedure

For each resource created by the Partners, the same experimental methodology was followed. For each resource created by partners, students performed a study within an online or face to face workshop or course. Student participants joined from Greece, Cyprus, Sweden, and the United Kingdom. A repeated measures

## 1. Method

design was used as the same group measures were taken before and after usage of the chatbots. They were recruited via staff members in the CEPEH group.

Participants were asked prior to the study if they agree to participate, providing them with a PIS form. Participants had the opportunity to discuss with the research team prior to the study and before consent is given. Then, participants used the chatbot resources independently and technical support was provided. Finally, post-intervention measures were recorded.

Some of the participants were invited to participate in Focus Group Discussions (FGD), and each FGD lasted between 15 to 25 minutes, with 5-10 participants. Participants were asked if they would like to be informed of the findings of the study.

### 1.3 Design

The data captured from the participants were their initials and numerical day of birth, used as anonymous identifier for pre-post analysis. Their institution was captured (Aristotle University of Thessaloniki, CYENS Centre of Excellent, Karolinska Institute, and The University of Nottingham), and Sex (Male/Female/Other).

Before any interaction with the learning resources, various perceptions of chatbot such as confidence and easy of use, usefulness, Influence from others, and current learning resources (videos, textbooks, Google, friends etc), were captured. Descriptive data was produced alongside repeated measures t-tests. Repeated measures t-tests were the appropriate test to use as this explores differences between groups, there were no covariates and we did not have several dependant variables. There was one Independent factor being Chatbot use having 2 levels (pre/post). There were 3 chatbots therefore there was option for ANOVA to determine where differences lie if statistical differences were found however this was not wholly appropriate for the data type and not necessary for pre-post comparison.

#### ## Materials and Measures

The measures used fit within a newly developed Chatbot Evaluation Framework which takes the best measures of 5 previous frameworks. Denecke and Warren

## 1. Method

[2] derived several quality dimensions and attributes from previous chatbot literature. They formed six perspectives from their review of articles and mobile health applications.

These six perspectives were: 1) Task-oriented, 2) Artificial intelligence, 3) System quality perspective, 4) Linguistic perspective, 5) UX Perspective, 6) Healthcare quality perspective.

To capture these perspectives, we used several validated materials that can distinguish these elements of the CEPEH chatbots.

### 1.3.1 Chatbot Usability Questionnaire (CUQ)

The Chatbot Usability Questionnaire (CUQ) [4] is a new questionnaire specifically designed for measuring the usability of chatbots by an interdisciplinary team from the Ulster University. CUQ can be used alongside the prevalent System Usability Scale Score (SUS) [5]. Multiple metrics are more appropriate when measuring usability of chatbots [6] therefore a combination of two scores can provide an all-inclusive overview.

### 1.3.2 UTAUT2 (Unified Theory of Acceptance and Use of Technology)

The underpinning theory of the UTAUT2 is that there are four key constructs to the intentions of using technology based resources: 1) performance expectancy, 2) effort expectancy, 3) social influence, and 4) enabling conditions.

The TAM and the UTAUT2 have cross over in measuring technology acceptance, however the UTAUT2 has more applied probing questions. Few studies exist that use technology acceptance theories for the intention to use products that explicitly incorporate AI. A recent extension of the UTAUT2 model added five (health, convenience comfort, sustainability, safety, security, and personal innovativeness) additional influencing factors to accommodate for AI [7]. This can be used for products in either health, household use, or mobility and can help to explain behavioural intention and use behaviour of chatbots.



## *1. Method*

### **1.3.3 System Usability Scale**

The System Usability Scale (SUS) was used [10] and is a widely used and adopted usability questionnaire. It is popular due to its unbiased and agnostic properties, a non proprietary, and a quick scale of 10 questions. However, as there are the CUQ and parts of the UTAUT2 we have selected only 2 questions which do not cross-over with the other measures. These are important statements however and good indicators of usability when assessed with the other results.

### **1.3.4 Computer Self-Efficacy Scale Tool**

The 10 question CSEST is based on the 32-item questionnaire by Murphy, Coover, and Owen (1989). It can be adapted for any technology and we have selected only a few pertinent questions. Participants are asked to think about using the CEPEH chatbots and answer that they would use the chatbots if I had never used a product like it before; If they could call someone for help if I got stuck, or if someone showed them how to do it first, and other similar usage questions.

### **1.3.5 Technology Acceptance Model (TAM)**

The Technology Acceptance Model (TAM) [1] was specifically developed with the primary aim of identifying the determinants involved in computer acceptance in general; secondly, to examine a variety of information technology usage behaviours; and thirdly, to provide a parsimonious theoretical explanatory model.

TAM suggests that attitude would be a direct predictor of the intention to use technology, which in turn would predict the actual usage of the technology. The only modification to the nine sub-scales of the questionnaire consists of applying the items to the context of chatbots. All the items, except those measuring attitudes, utilize a seven-point Likert scale ranging from “strongly agree” to “strongly disagree” with a middle neutral point [2].

The nine sub-scales of the questionnaire:

## *1. Method*

Ease of use of chatbots   Perceived usefulness of chatbots   Intention of use. Attitude toward usage of chatbots.   Perception of personal efficacy to use a chatbot resource.   Perception of external control toward chatbots.   Anxiety toward chatbot use.   Intrinsic motivation to use chatbot resources.   Perceived costs of chatbots.

### **1.3.6 Qualitative Measure- Focus Group Discussions**

Focus groups are a pervasive means of market research and provides credible acceptance evaluators regarding the penetration that a product or service will have on a target demographic. Focus groups are a form of qualitative research consisting of interviews or structured discussions, in which a group of people are asked about their perceptions, opinions, beliefs, and attitudes towards a product, service, concept, advertisement, idea, or packaging.

Questions are asked in an interactive group setting where participants are free to talk with other group members. During this process, the researcher either takes notes or records the vital points he or she is getting from the group. Researchers select members of the focus group carefully for effective and authoritative responses. Relevant stakeholders, then, can use the information collected through focus groups to receive insights on a specific product, issue, or topic focus [7].

A series of short focus group sessions identified the feasibility of CEPEH resources for formal curricular integration. These sessions, spanning no more than 1-1.5 hours and consisting of no more than 5-7 persons each explored all axes of curricular integration such as accessibility in the classroom, use case scenarios, technology requirements for curricular integration etc. These axes were formalized by the research team, in each evaluation site, to consider the curricular details of each institution.

## 1. Method



Figure 1: Flow diagram of the recruitment process

# 2

## Results

### Contents

---

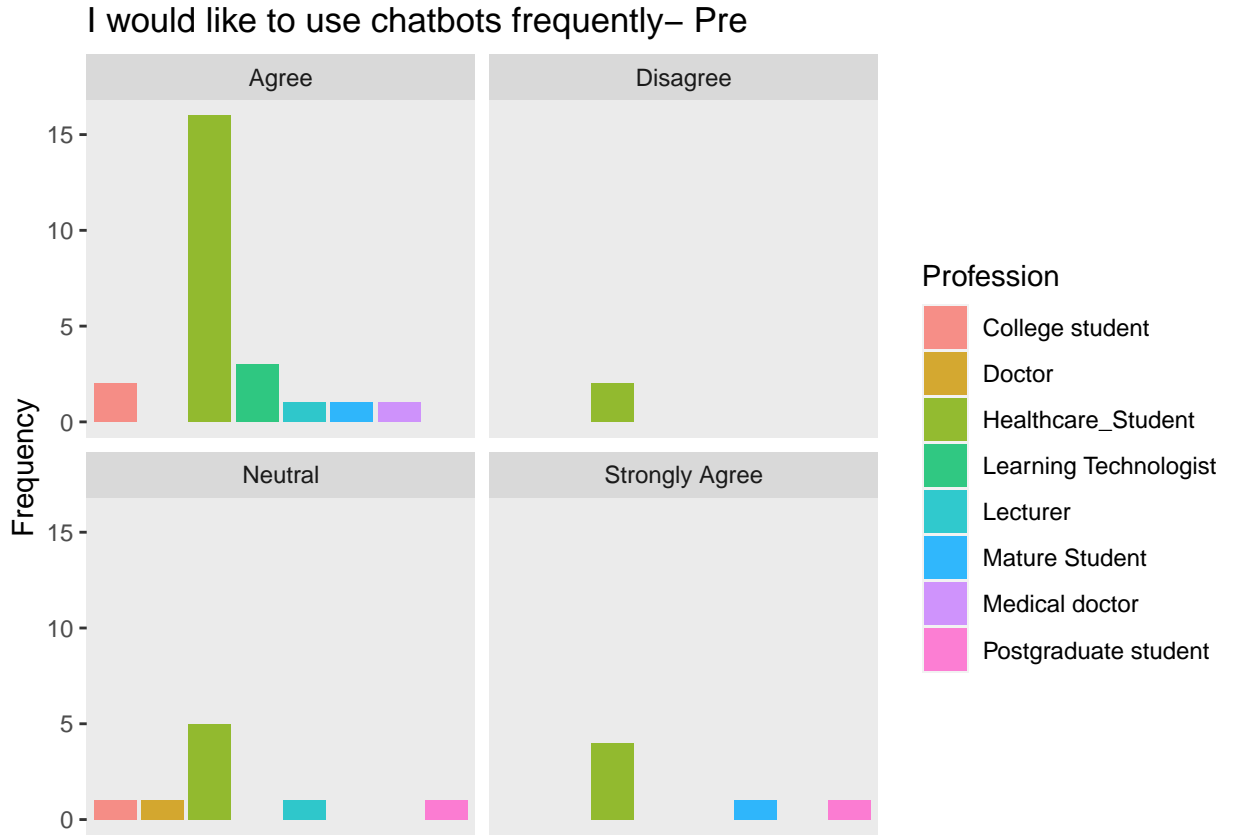
<b>2.1</b>	<b>Participants' Characteristics . . . . .</b>	<b>11</b>
<b>2.2</b>	<b>Chatbot Usability Questionnaire (CUQ) . . . . .</b>	<b>14</b>
2.2.1	CUQ Calculation tool . . . . .	14
<b>2.3</b>	<b>System Usability Scale (SUS) Scores . . . . .</b>	<b>15</b>
<b>2.4</b>	<b>Technology Acceptance Model . . . . .</b>	<b>16</b>
<b>2.5</b>	<b>Other Measeures . . . . .</b>	<b>17</b>
2.5.1	Knowledge of Topics after Use . . . . .	17
2.5.2	Trust in CEPEH chatbots after Use . . . . .	18
2.5.3	Personailty and Interactions . . . . .	20
2.5.4	Ease of Use and Seeking Support . . . . .	21
<b>2.6</b>	<b>Inferential Statistics . . . . .</b>	<b>22</b>
2.6.1	Repeated Measures T-test results . . . . .	22
<b>2.7</b>	<b>Other Findings . . . . .</b>	<b>23</b>

---

## 2.1 Participants' Characteristics

When participants were asked the amount of time they have used a chatbot in any form or subject, 23 stated they had never used a chatbot. Further, 19/42 stated having used a chatbot at least once for between 0-4 hours of use in total. These are likely commercial/website- based assistant chatbots however there are some medical/healthcare resources known to be used in anatomy and/or patient

## 2. Results



**Figure 2.1:** Chatbot Usage History- Pre

interactions. One individual had spent much longer time with usage- this was the mature student.

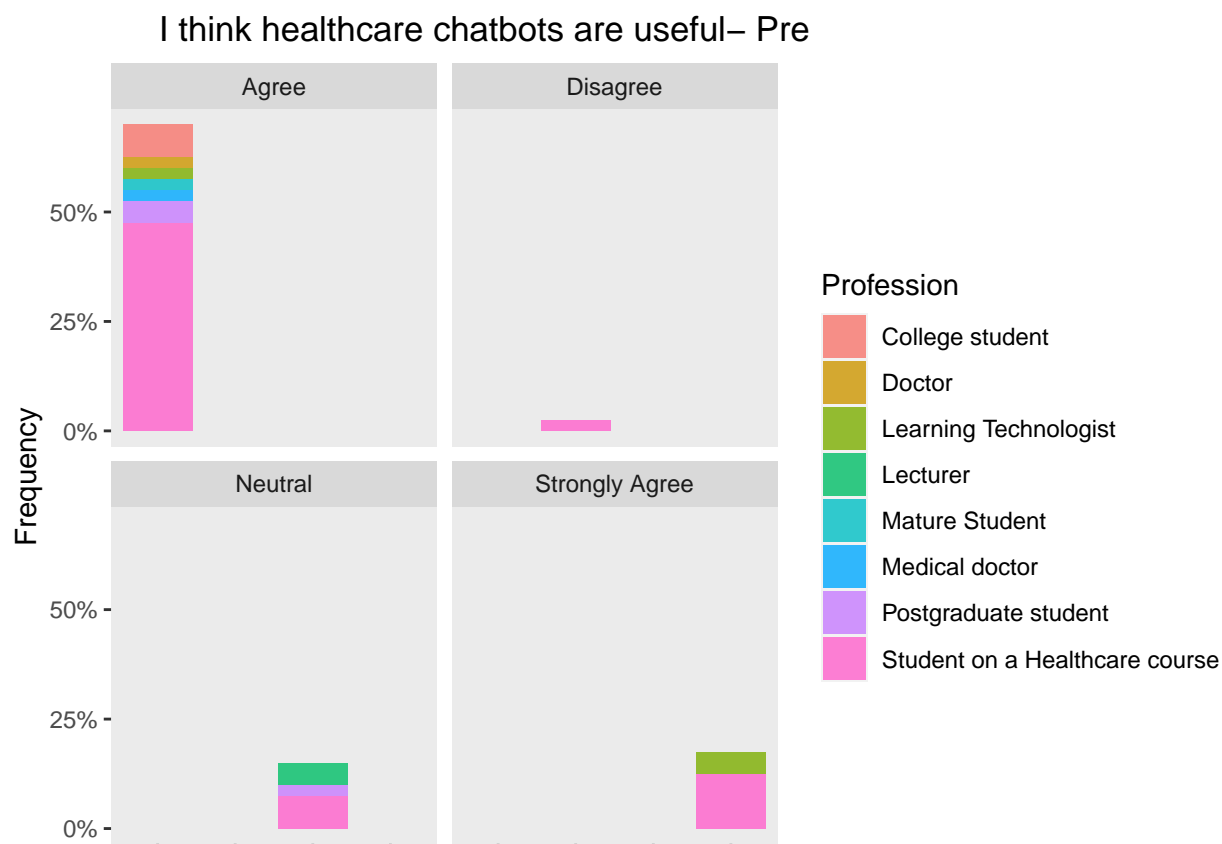
**Table 2.1:** Previous Chatbot Usage of Participants

Previous_Chatbot_Usage	n
1-4 hours	16
10-19 hours	1
5-9 hours	2
Never	23

In short, approximately 50% had never used a chatbot, and 45% had used a chatbot, at some period over the years, for a short period of time.

Most learners use books or online books as resources. They may use multiple sources however they were asked to note the primary source. Only 6 stated their primary sources were *Online videos/interactive materials* which includes such

## 2. Results



**Figure 2.2:** Chatbots are Useful Opinion- Pre

tools as chatbots.

The first boxplot (2.1) shows learners perceptions of easy of use of mobile app and other educational mobile resources

(2.2) shows the opinions of all participants on the usefulness of chatbots. Many had not had experience with them yet had positive rating. This positive opinions of chatbots may be from colleagues, friends, media, tutors, or other social information of the benefits in healthcare education. Around 25% were neutral or disagreed that healthcare chatbots were useful.

*The participants then used the 4 chatbots, and completed the post-usage survey after each chatbot. Pre-and post differences are as followed:*

## 2. Results



**Figure 2.3:** CUQ CEPEH Score

## 2.2 Chatbot Usability Questionnaire (CUQ)

### 2.2.1 CUQ Calculation tool

The CUQ was developed by researchers at Ulster University, [Link](#) and as the calculation can be complex, a dedicated calculation tool has been created.

Please download the CEPEH CUQ calculation tool which has all of the data entered, so you can see the CEPEH CUQ scoring

[Click here to download CUQ calc tool](#)

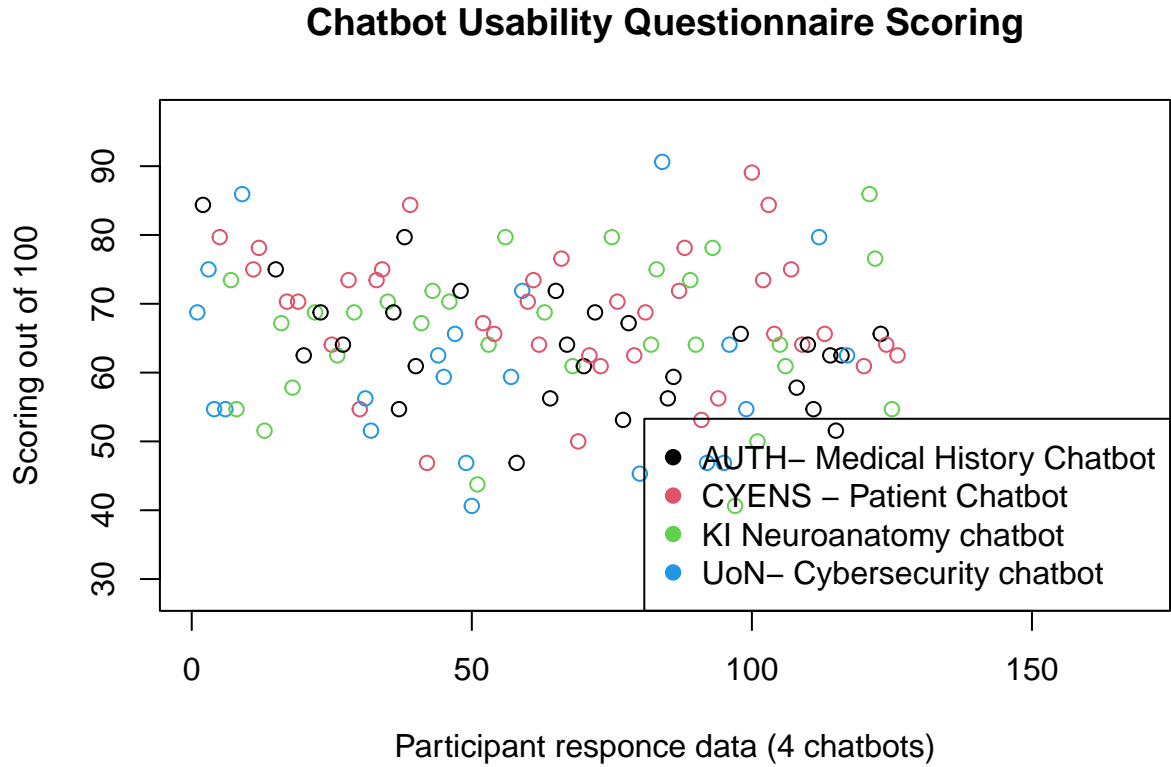
[Click here to download CEPEH CUQ score result](#)

Although the design and development was similar, each chatbot CUQ score was calculated to understand how the topic content may affect usability:

The breakdown of the chatbots was:

- Aristotle University of Thessaloniki CUQ score = 63/100
- CYENS Centre of Excellence CUQ score = 67/100
- Karolinska Institute CUQ score = 63/100

## 2. Results



**Figure 2.4:** CUQ Scatter Plot

- University of Nottingham CUQ score = 68/100

The score for all 3 chatbots grouped was 65/100.

Figure (2.4) shows the CUQ scores as a scatter plot to highlight how there was a moderate distribution of results. Further exploration is required to understand which elements are causing this spread, and if it was due to problems within a small group of learners.

## 2.3 System Usability Scale (SUS) Scores

*Note= The amount of ‘agreement’ is defined as the addition of ‘Agree’ and ‘Strongly agree’ responses.*

The SUS score should consist of 10 items. However, some SUS questions were improved upon by 1 or more CUQ questions, specifically to this Chatbot study.



## 2. Results

The SUS results would be overshadowed by the CUQ scores, expect 2 that did not have cross-over. The two questions were:

- I would like to use the CEPEH chatbot I tested, more frequently (SUS1)(post)
- I felt confident using the CEPEH chatbot (SUS2)(post)

This meant the score of the SUS was not created, however the CUQ score better represented the Learners' perceptions of the CEPEH chatbot in terms of feasibility of reuse and acceptability in healthcare curricula.

I will continue to use CEPEH chatbot(s)	Responses
Agree	71
Disagree	11
Neutral	21
Not Applicable	4
Strongly Agree	19

Confidence using CEPEH Chatbot(s)	Responses
Agree	71
Disagree	11
Neutral	21
Not Applicable	4
Strongly Agree	19

This table

## 2.4 Technology Acceptance Model

The TAM questions were analysed according to their subsets. The subsets were Perceived Usefulness (PU) and Perceived Easy of Use (PEU)

The questions were: Perceived Usefulness (PU)

1. Using CEPEH chatbots would enable me to accomplish tasks more quickly
2. Using CEPEH chatbots would increase performance
3. Using CEPEH chatbots would increase my productivity

## *2. Results*

4. I would find CEPEH chatbots useful on my course

Perceived Easy of Use (PEU)

5. Learning to use CEPEH chatbots would be easy to me
6. It would be easy for me to be skilful at using CEPEH chatbots
7. My interactions with CEPEH chatbots would be clear and understandable
8. I would find CEPEH chatbots easy to use

Results:

The scores as a percentage of agreement, were calculated by averaging the subsets and interpreted as:

Before using the CEPEH chatbots, there was 66% (2.2/5) agreement for the Perceived Usefulness of chatbots in healthcare education, and after 48% (2.6/5) agreed.

Before using the CEPEH chatbots, there was 64% (2.3) agreement for Perceived Ease of Use of chatbots in healthcare education, and after 51% (2.56) agreed.

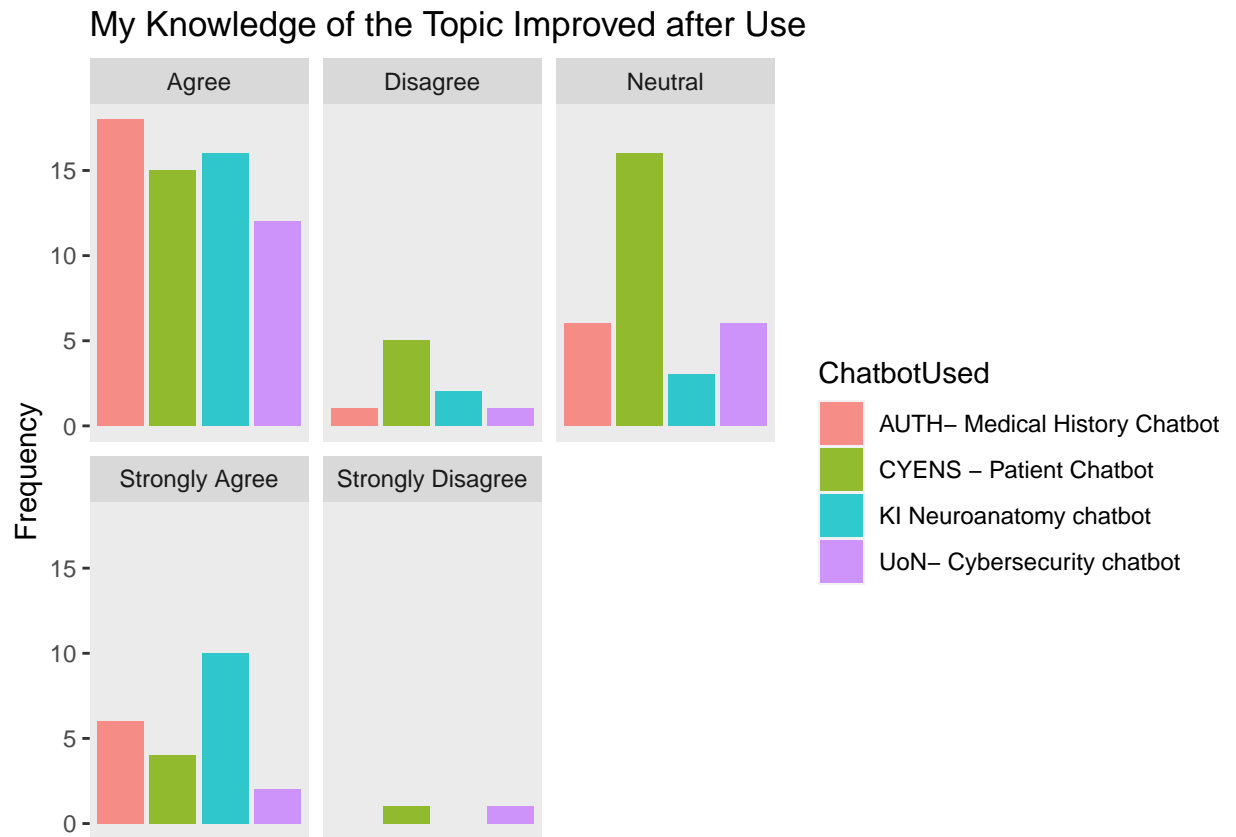
The justification for this may be due to being early versions of applications with limited functionality and functions which can be difficult for user to experience the intended further range of features and learning exercises.

## **2.5 Other Measures**

### **2.5.1 Knowledge of Topics after Use**

CYENS chatbot had around 10 more participants stating that they were neutral on gaining knowledge of the topic

## 2. Results



**Figure 2.5:** Improvements in Knowledge

### 2.5.2 Trust in CEPEH chatbots after Use

The figure above, (2.6) shows the ratings by participants of the CEPEH Chatbots to provide them with the necessary course information. This is a integral element in learners' motivational and educational choices to reuse the learning resources. As previously described, the trust of the information is also a factor in these responses.

## 2. Results

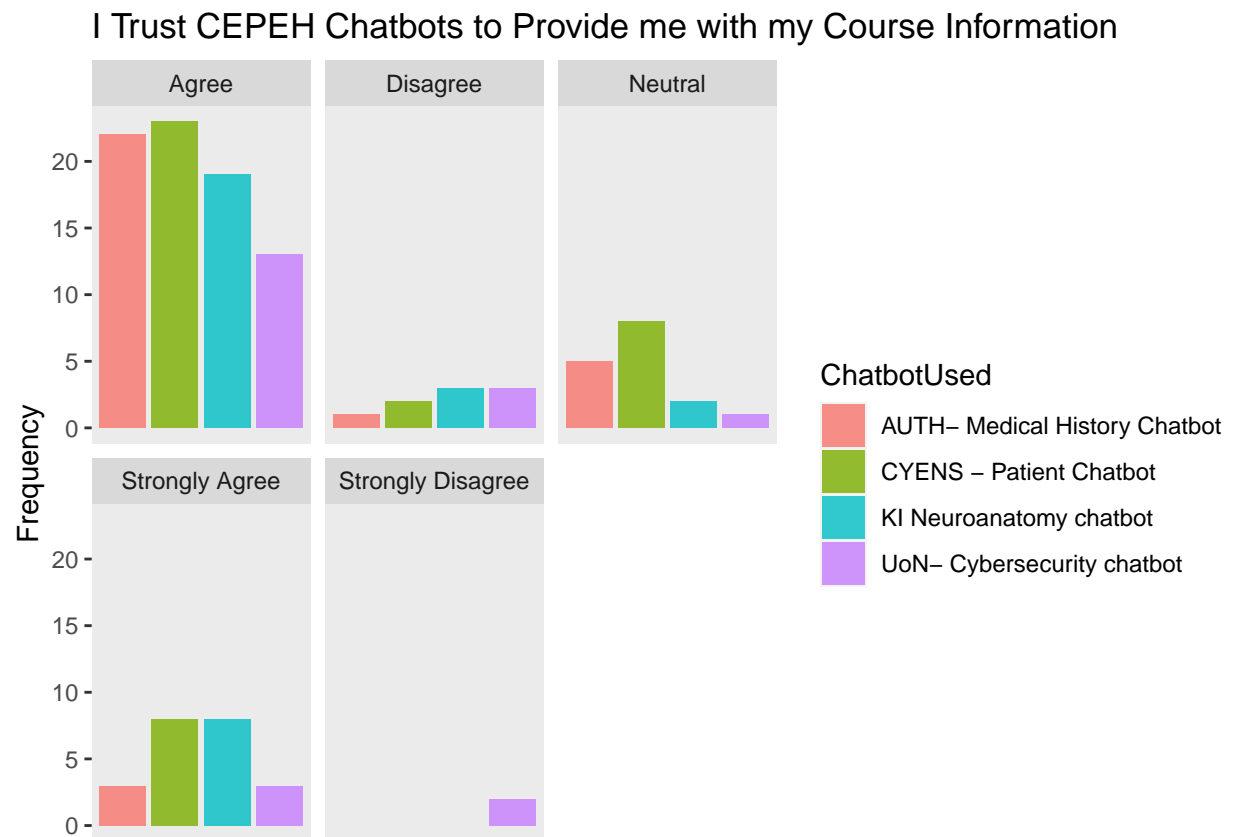
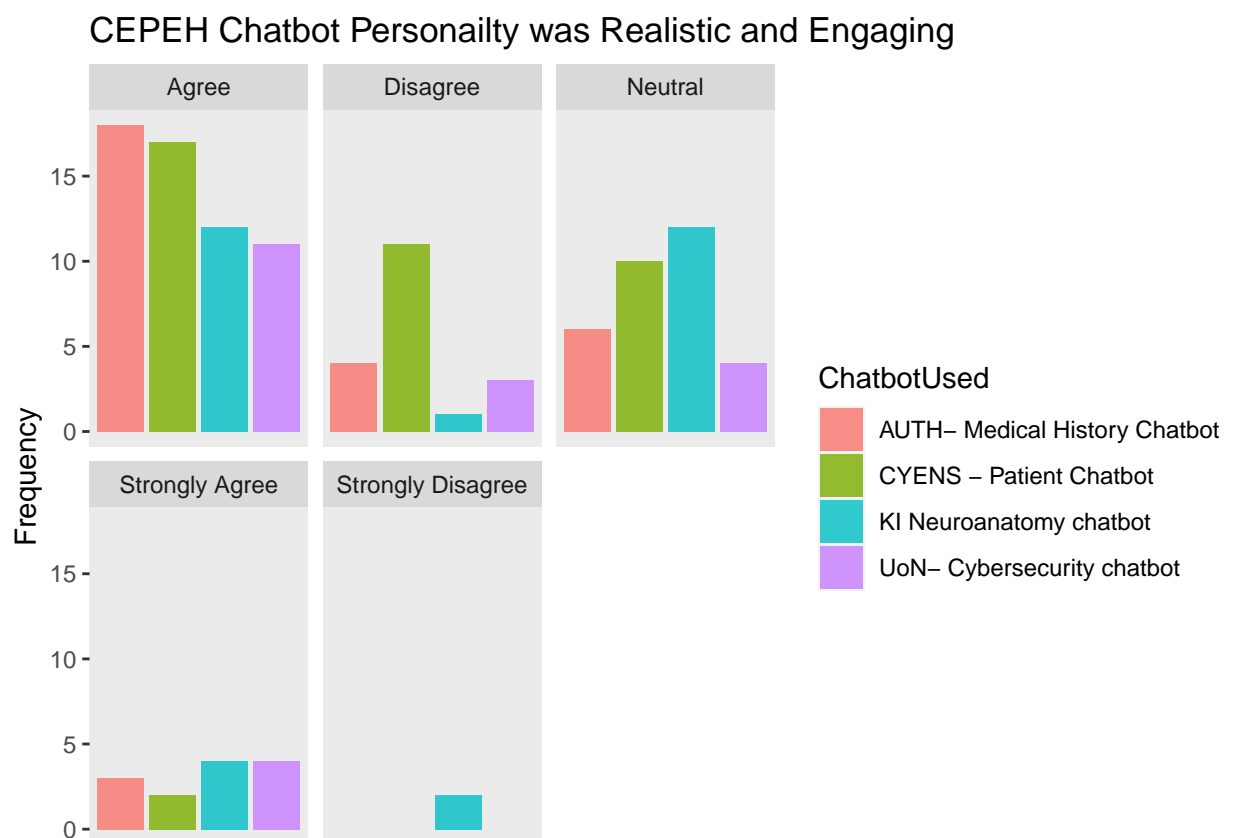


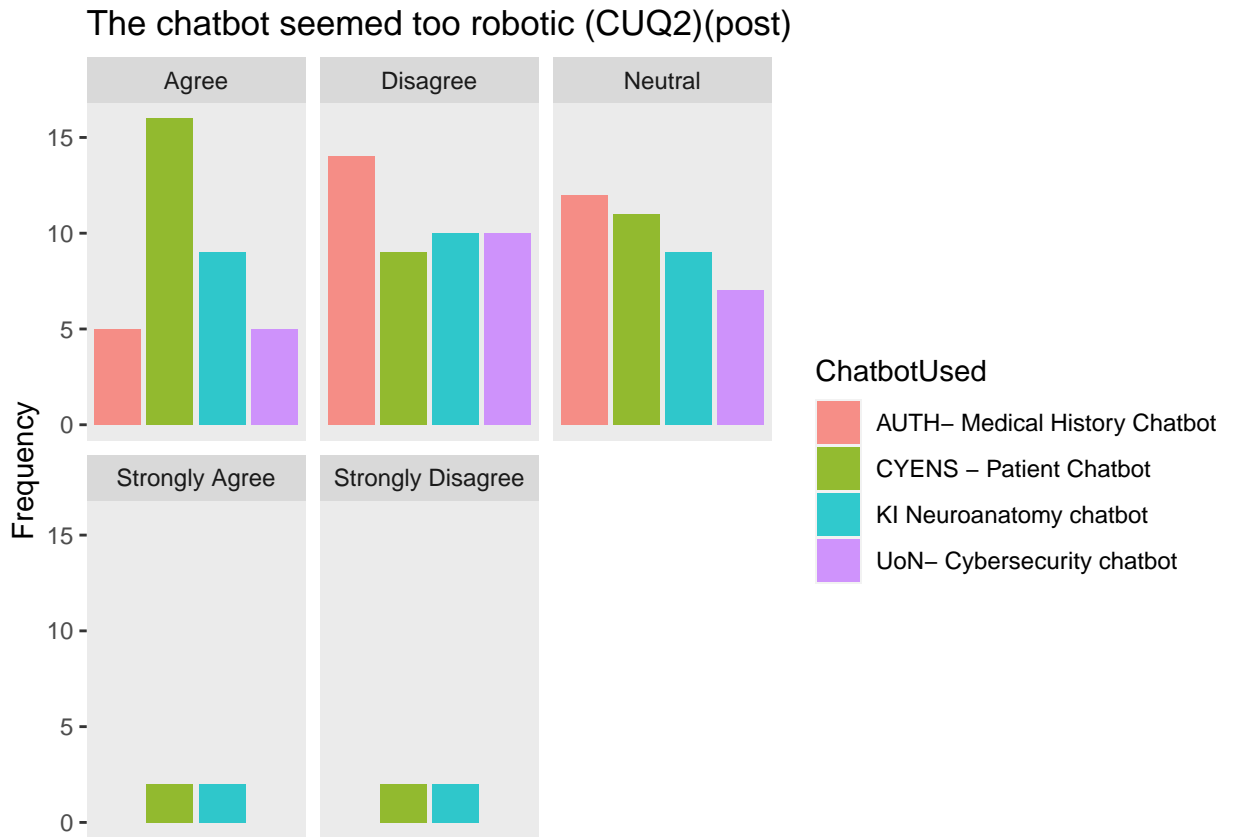
Figure 2.6: Trust Chatbots POST use



## 2. Results

There was mixed results for the chatbot used being realistic and engaging. This question has two descriptive terms however based on the other results we understand that the chatbots' NLP logic, or ability to respond required improvement to be more 'smooth' in replying. The primary limitation was found in the 'robotic' interactions(See Figure 10). This was investigated further in the 'Text Mining' and 'Sentiment Analysis' sections.

### 2.5.3 Personailty and Interactions



*The chatbot seemed too robotic* results had the largest mix of responses, and for all 4 chatbots evaluated. The University of Nottingham Cybersecurity chatbot had more deterministic pathways with exploitation of the NLP modelling to provide illusion of realism. This may explain why there was less agreement. However, Neutrality and/or agreement was not desired.

The CYENS medical patient chatbot had more complex pathways of interactions,

## 2. Results

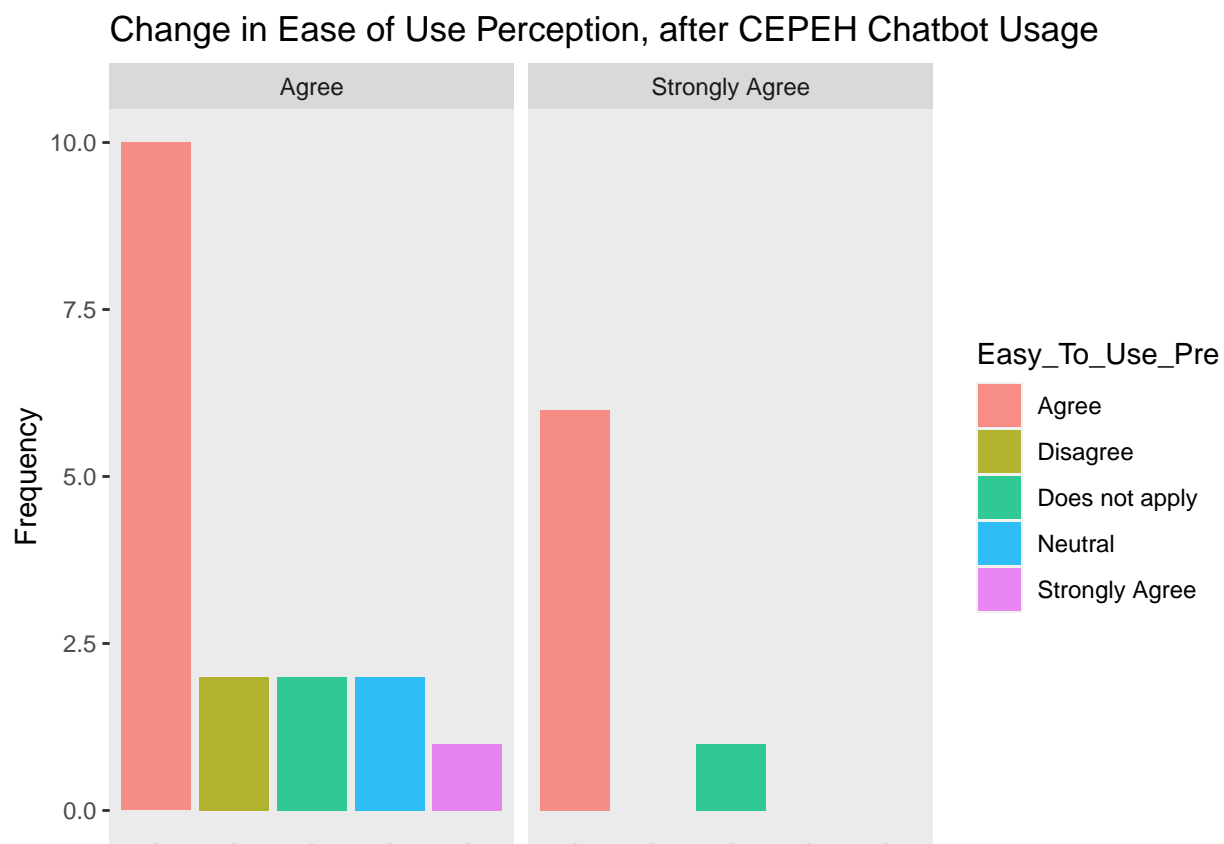
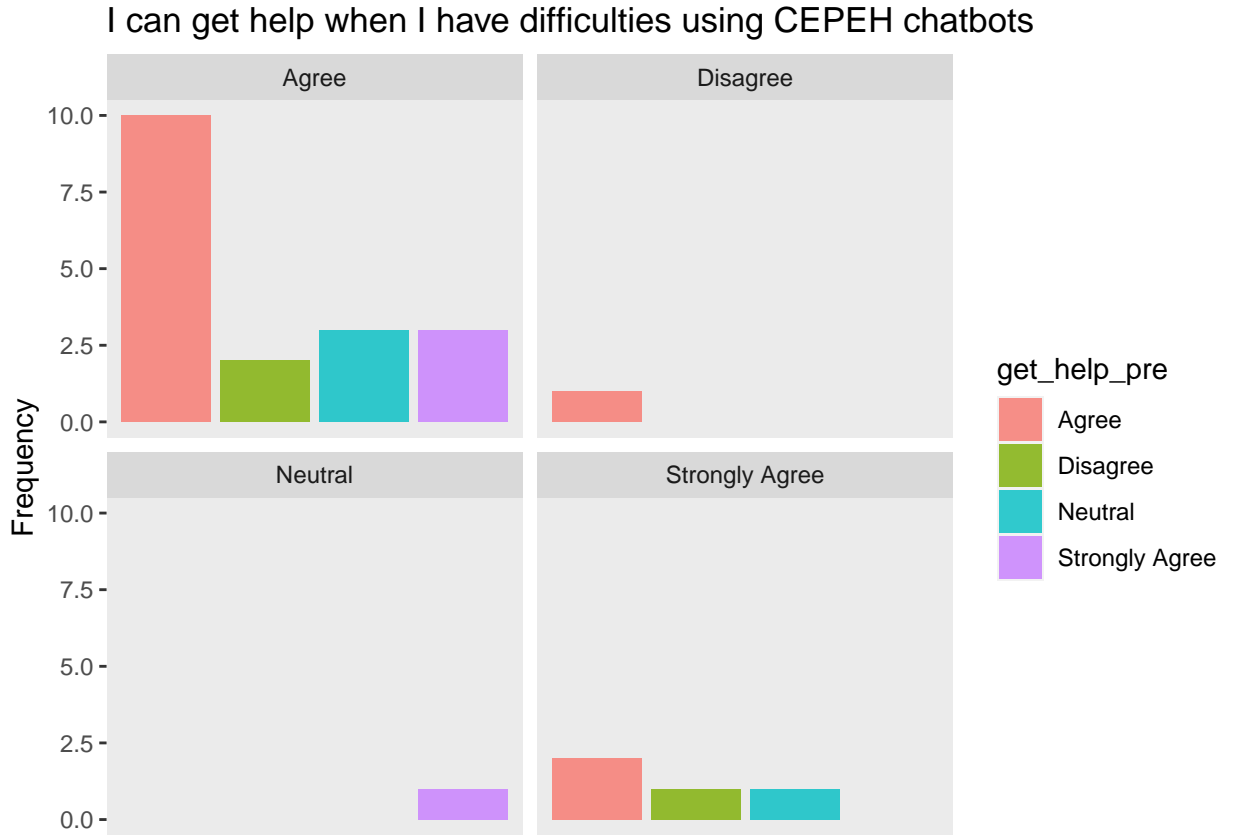


Figure 2.7: Ease of Use Comparison

### 2.5.4 Ease of Use and Seeking Support

After usage, there was only agreement in Ease of Use- as shown in (2.7) as there are no 'Neutral' or disagree columns. Any learners with disagreement before using the CEPEH chatbots, after believed they were easy to use.

## 2. Results



Those who disagreed or were neutral in the pre usage measure, improved their understanding that help was available with the CEPEH chatbots. After usage, 40 participants agreed they could get help if they had difficulty using the resources.

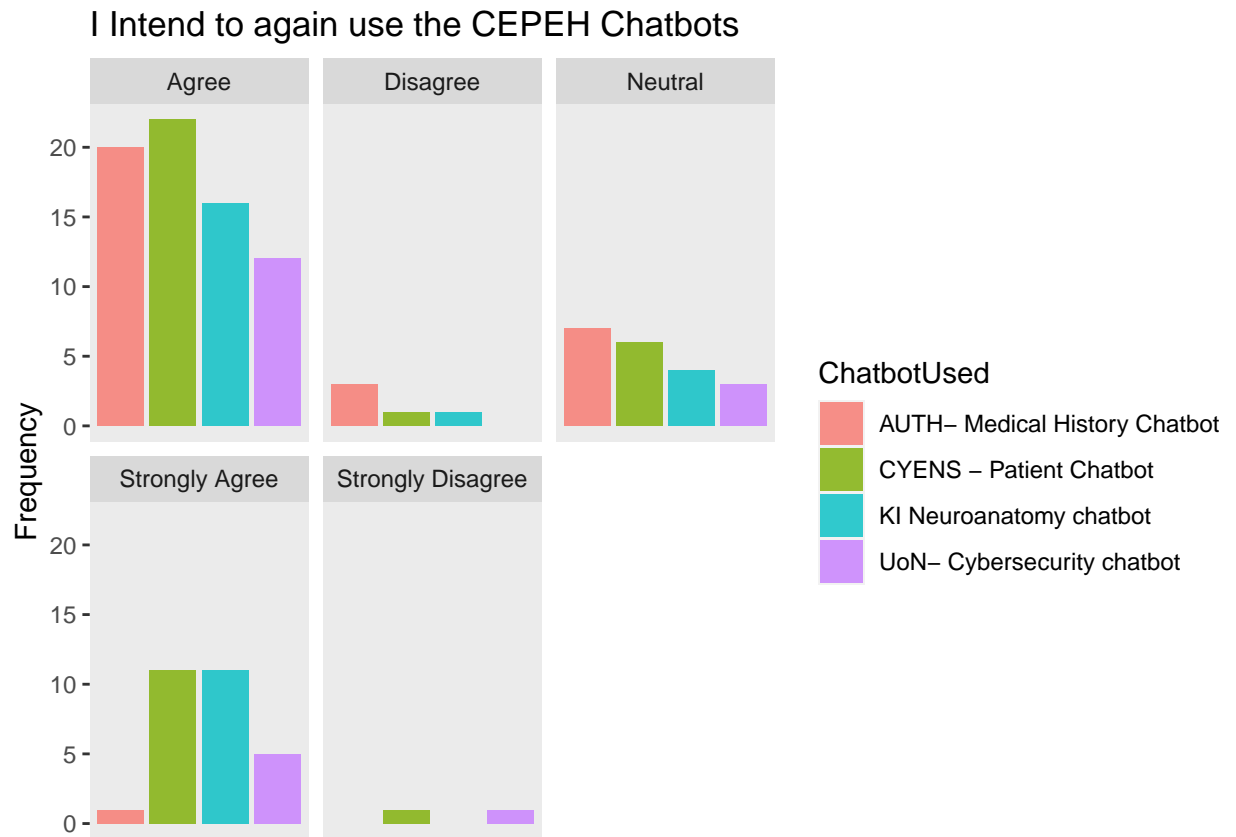
## 2.6 Inferential Statistics

### 2.6.1 Repeated Measures T-test results

After using the CEPEH chatbots, majority of participants stated they would reuse the chatbots. However, there was 6 counts of *disagree* or *strongly disagree* for all 4 chatbots. Further, there were 17 counts of neutral in reuse, which was approximately 4 participants per chatbot (see (2.8)).

For CYENS, even though the knowledge of the topic was not perceived to improve by some participants, this box plot shows how 34/42 stated they would reuse the chatbot developed by CYENS.

## 2. Results



**Figure 2.8:** Intend to Reuse-Post

There was only 1 ‘Strongly Disagree’ response. The agreement options counted for the majority of the data.

## 2.7 Other Findings

Other questions

I intend to continue using chatbots in the future (BI1)

The chatbot provided the information I needed with minimal commands

My knowledge of the topic improved after i had used the Chatbot

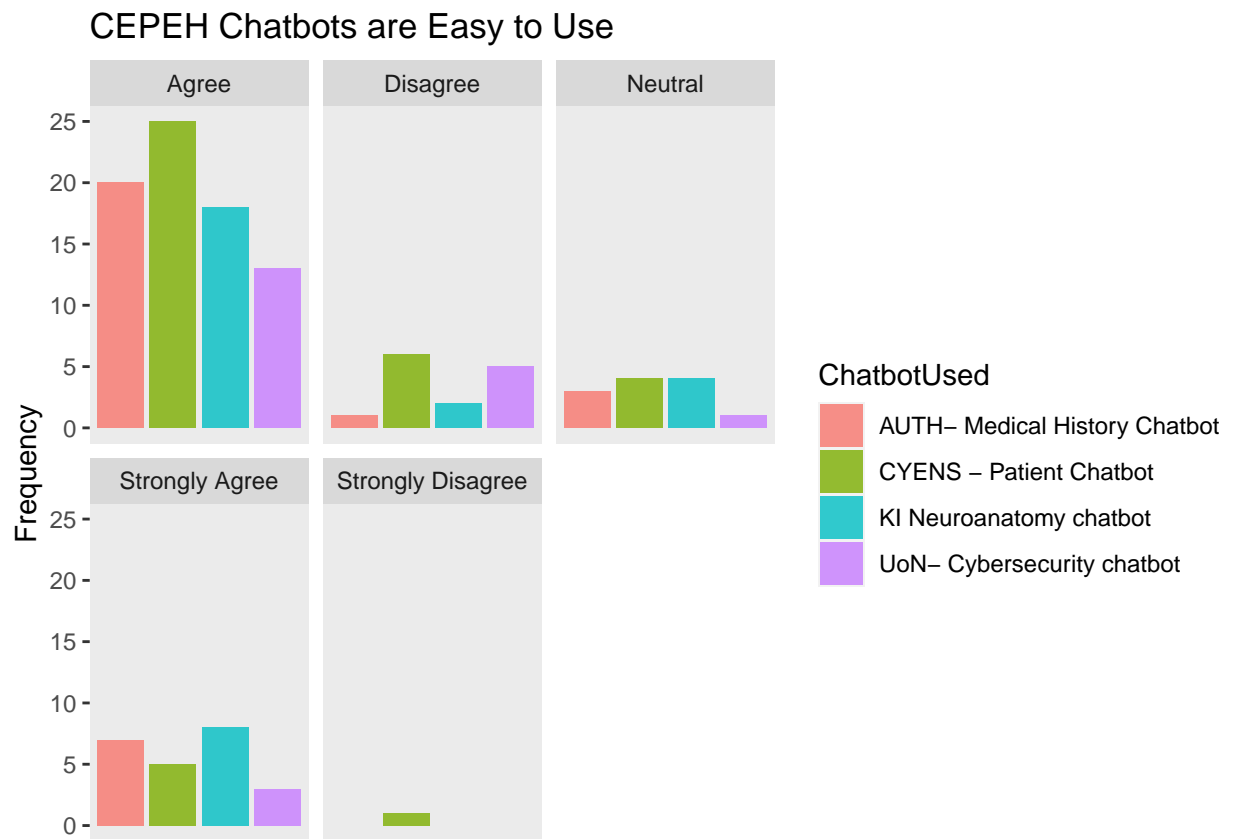
My confidence in understanding the topic improved after I had used the Chatbot

The chatbot provided me with the type of response i expected from asking a tutor/lecturer

The information provided was reliable



## 2. Results



**Figure 2.9:** Easy to Use- Post

The chatbot has a high level of trustworthiness

The duration of conversations to find my answer was too long

The videos/images provided were useful to my questions

The chatbot exceeded my expectation of how it could help me

The chatbot exceeded my expectation of how it could engage with me

I think this learning method could help me to acquire knowledge

I would use this tool again as it has some value to me

I think I will actively use this learning method

I believe I had some choice about learning during chatbot use

I would trust the chatbot to provide me with information for my course

One piece of knowledge I learned from the chatbot was..

Repeated Measures t-test, aka paired t-test (before and after measurements)

## *2. Results*

This t-test compares confident using mobile chatbots before and after CEPEH chatbot usage.

# 3

## CEPEH Focus Group Discussion Analysis

The focus group discussions provided a lot of feedback for how the participants experienced their interactions with the chatbots, and how the CEPEH team can improve them, improve the design and development processes, and improve uptake and sharing.

One method of analysing this data is with use of text mining and data manipulation, creating word clouds, sentiment analysis, and using a model which can distinguish the unique themes in text, and highlights for us what text is used to create these themes.

Therefore, we have created a model to allow efficient and intelligent analysis of this open/free focus group data.

### 3.1 Tokenising

Firstly, we tokenised the words from the FGDs. A Token is “a meaningful unit of text, most often a word, that we are interested in using for further analysis”. For each word we give it a property that we can call upon later.

The data manipulation for this included removing punctuation, converting to lower-case, and setting word type to word (and not such types as “characters”, “ngrams”, “sentences”, “lines” etc)

### 3. CEPEH Focus Group Discussion Analysis

#### 3.1.1 Stop words

The model then removed words with meaningless function. These are called stop words. Words like “the”, “of” and “to” are the most frequent words found, technically, but are of little interest to us.

We also created a custom list of stop words for CEPEH. We know participants may mention other objects, and the list was as followed: found; chatbot; chatbots; presentation.

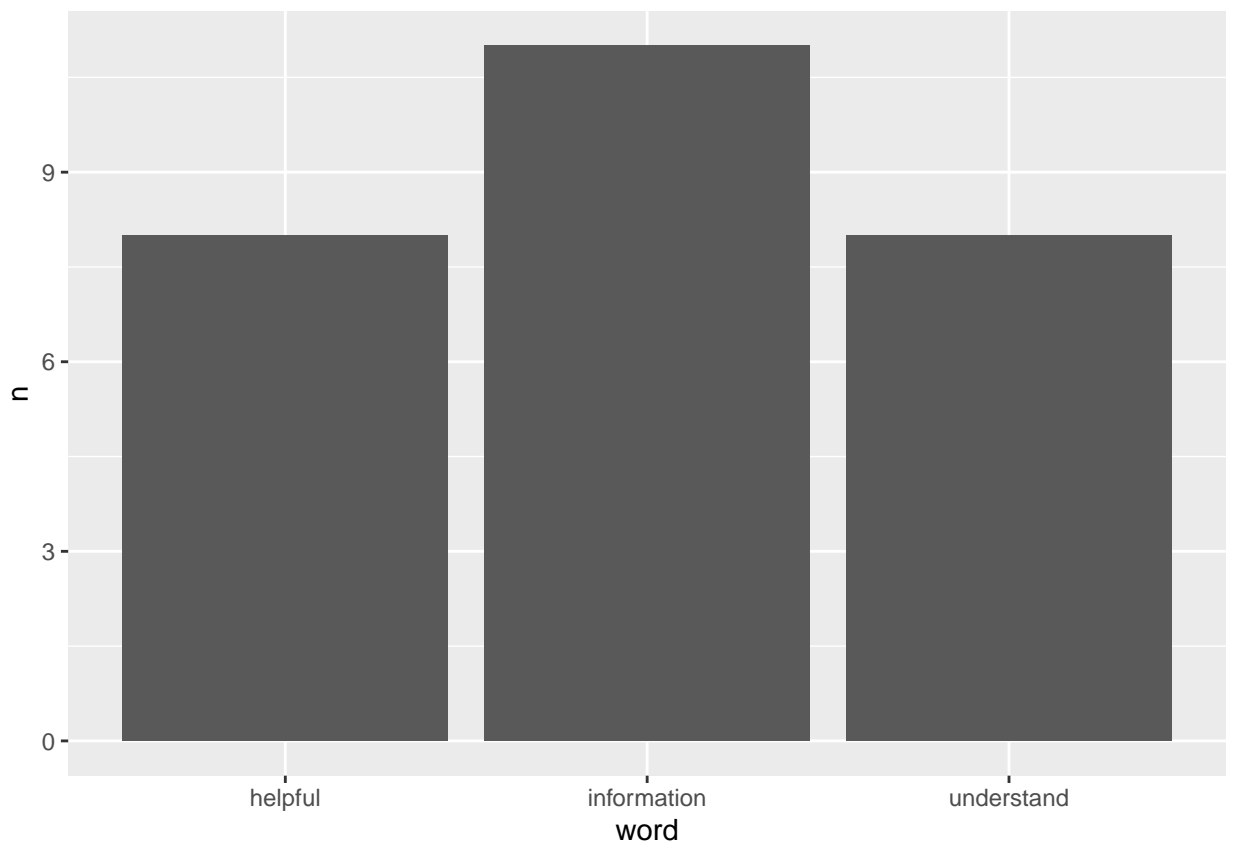
The data was ready for analysis by the model. We ordered it to find the most frequent words.

```
## # A tibble: 384 x 3
## # Groups:   doc_id [1]
##   doc_id word          n
##   <fct> <chr>        <int>
## 1 1      information    11
## 2 1      helpful        8
## 3 1      understand     8
## 4 1      idea           7
## 5 1      ideas           7
## 6 1      workshop       7
## 7 1      beginning      6
## 8 1      dont           6
## 9 1      medical         6
## 10 1     technology      6
## # ... with 374 more rows
```

## 3.2 Plotting word frequencies - bar graphs

With this information a Bar graph of top words from the participants in the FGD can be rendered.

### 3. CEPEH Focus Group Discussion Analysis

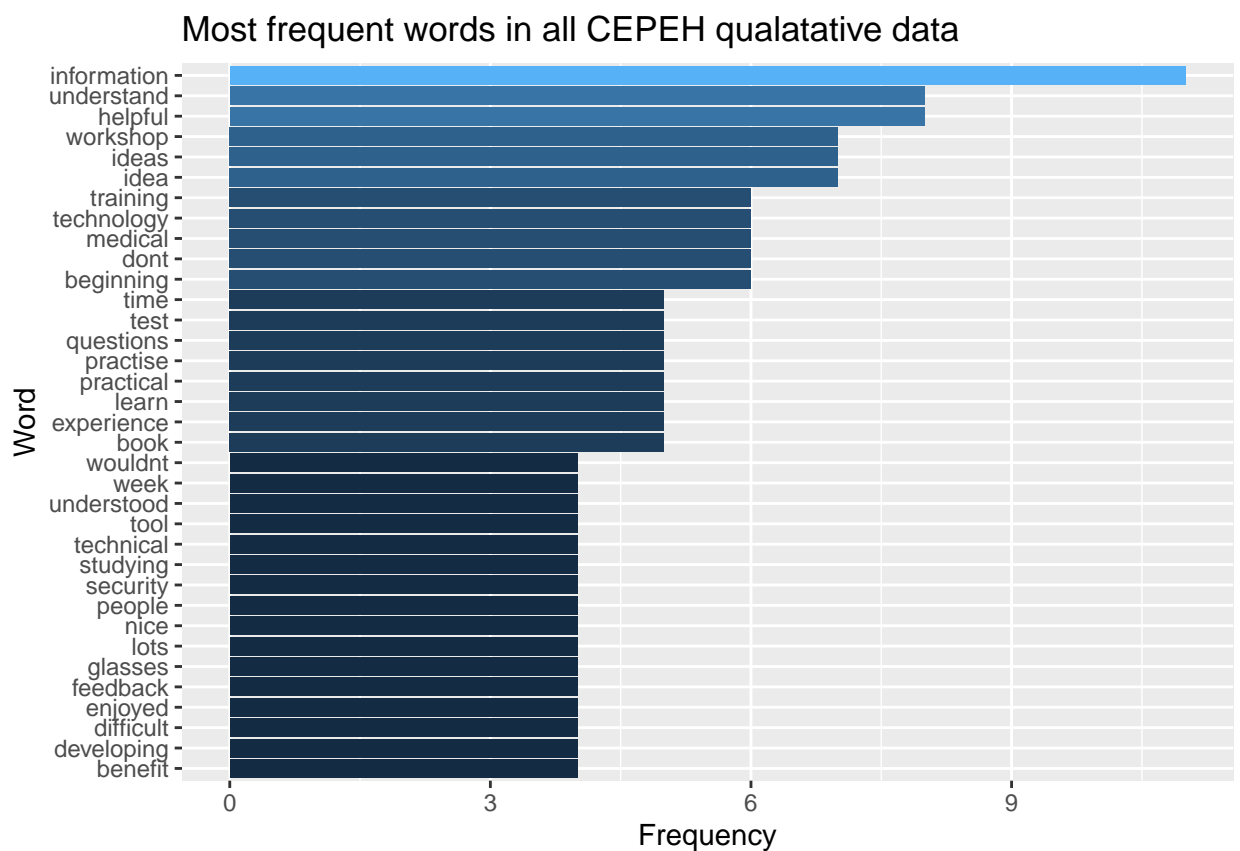


and after some modifications, a graph of the top 35 words is produced, with

better aesthetics. The most frequent words present in focus group discussions after

using the 4 chatbots, are in the Figure below.

### 3. CEPEH Focus Group Discussion Analysis



Although the frequency is not high for each word, we are able to get a general picture of the sentiments, intensities, and concerns which would be immediately occurring when plotted.

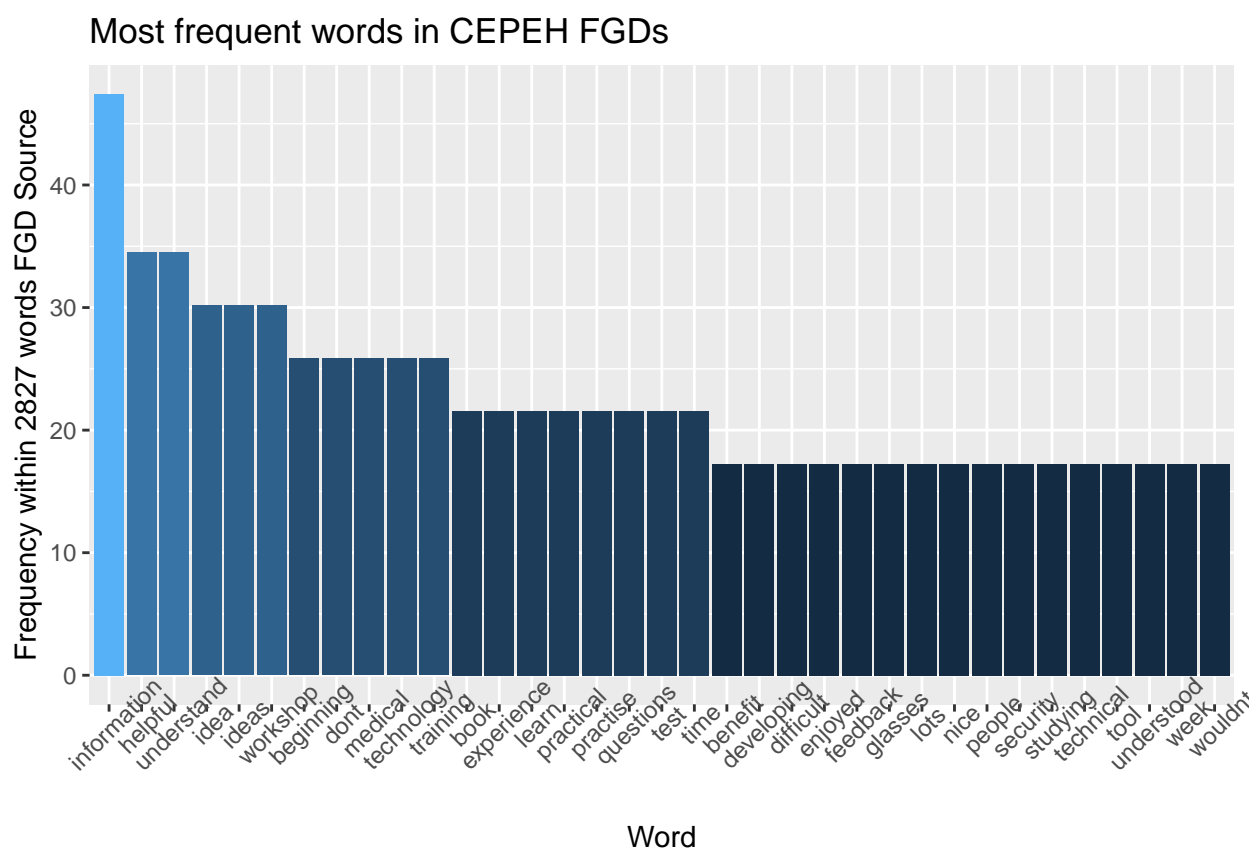
#### 3.2.1 Normalised frequency

A better way to understand this data is to normalise the frequency of occurrences in accordance with the source text. The raw text had 2827 words in total. Therefore we can mutate the ratios to reflect this.

#### 3.2.2 Plotting normalised frequency

Now we can plot, for example, the 20 most frequent words when normalised by the source text.

### 3. CEPEH Focus Group Discussion Analysis



In summary, this understanding of frequent words can help to understand common concurrences and extrapolate to a larger audience. If scope and impact of CEPEH chatbots increased we can understand the type of themes and trends may occur, based on such FGD analysis.

### 3.3 Word clouds

To visualise the most frequent words in another format, below is a word cloud which presents the word size to indicate the frequency- words that occur more often being displayed in a larger font size. This has a normalised data frequency in accordance to the FGD source document analysed.

### 3. CEPEH Focus Group Discussion Analysis



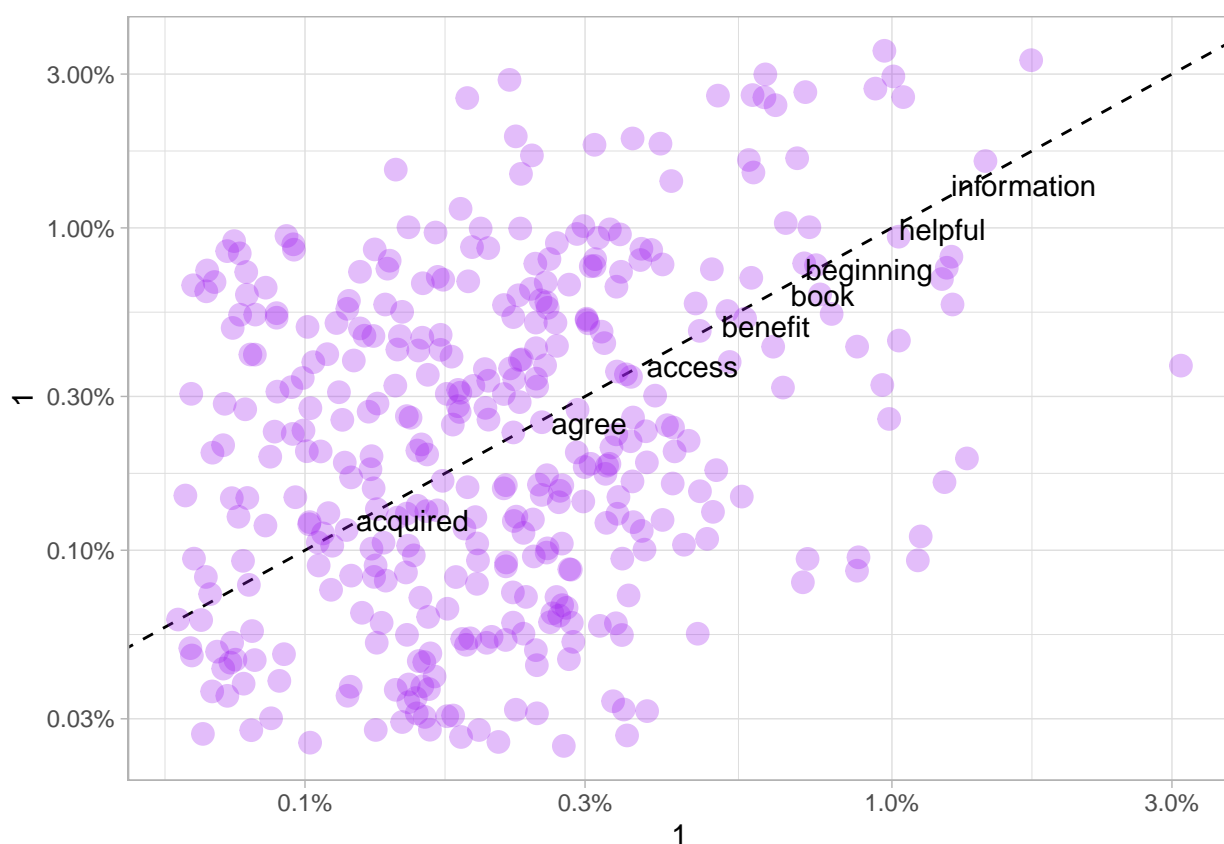
We understand the context has been reduced for each word. However, in general there can be categorised positive/negative words from the word cloud: Positive words are- benefit, practical, nice, helpful, learn, ideas, and enjoyed Negative words are- difficult, test (who likes a test?), don't, and 'lot' may be negative if there is a 'lot' of information.

#### 3.3.1 The vocabulary of Texts

Here is a graph that has plotted the words in places depending on the word frequencies. Additionally, colour hotspots shows how different the frequencies are - darker items are more similar in terms of their frequencies, lighter-coloured ones more frequent in one text compared to the other.



### 3. CEPEH Focus Group Discussion Analysis



## 3.4 Sentiment analysis

What is the sentiment of all participants? What is types of emotional words are being used? The preparation of these words has some use in understanding the frequencies, but their emotional valence are not compared. The table above has the word *'helpful'* which has a positive connotation, however there are 386 words, with many having several occurrences.

```
## max(total_score) min(total_score)
```

```
## 1 38 38
```

negative	positive	total_score
24	62	38

# 4

## Discussion

### Contents

---

<b>4.1</b>	<b>Quantatative Results . . . . .</b>	<b>34</b>
4.1.1	CUQ . . . . .	34
<b>4.2</b>	<b>Qualatative Results . . . . .</b>	<b>34</b>
<b>4.3</b>	<b>Limitations . . . . .</b>	<b>34</b>
<b>4.4</b>	<b>Conclusions . . . . .</b>	<b>34</b>

---

### ##Summary of Findings

The Chatbots were beneficial.

Learners have lots of other choices such as YouTube, but there is a certain need for personalised information gathering , this can save time and prevent learning incorrect information.

This was one reason why they were rated positive as they are able to streamline data finding for learners in a format that is understandable and easy to them.

## *4. Discussion*

### **4.1 Quantatative Results**

#### **4.1.1 CUQ**

Holmes et al. [<https://dl.acm.org/doi/10.1145/3335082.3335094>] designed the CUQ to be comparable with the system usability scale (SUS).

We have calculated both these scores out of 100 to allow the same benchmark, which is 68. A score of 68 is at the centre of the range is thought of as “C”. The average benchmark for CUQ is 68, and in initial/pilot studies 68 may be considered higher than expected when considering technical issues, less developed user interfaces etc.

Previous studies evaluating chatbots have had similar score. For example, in 2022 [Link](#) found a physical activity promotion chatbot received 64.5/100, with lowest score at 40.6

### **4.2 Qualatative Results**

### **4.3 Limitations**

### **4.4 Conclusions**

Those results can be interpreted that the learning objectives of the training event was chosen appropriately for the diverse audience including clinicians, academics, researchers, and learning technologists/IT specialist resulting to a successful training event that enable participants to take the acquired knowledge back to their organisations in order to co-design and implement. As it was expected and can be depicted from self-confidence statements that some participants being very confident before the event, not all the objectives expected to be reached by everyone, since the training was targeting both technical and non-technical participants. However, on both average and individual matched responses participants self-statements showed that they improved their knowledge and understanding in

#### *4. Discussion*

using co-creation approaches to develop digital education resources and in designing and developing chatbots as educational resources.

```
output: #bookdown::html_document2: default #bookdown::word_document2:
default bookdown::pdf_document2: template: templates/template.tex document-
class: book
```

# 5

## (Additional Analyses) Training Events

### Contents

---

<b>5.1</b>	<b>CEPEH Training Event C1</b>	<b>36</b>
<b>5.2</b>	<b>CEPEH Training Event 2</b>	<b>37</b>
<b>5.3</b>	<b># bibliography: [bibliography/references.bib, bibliography/additional-references.bib]</b>	<b>39</b>

---

### 5.1 CEPEH Training Event C1

The CEPEH training event C1 held at the premises of University of Nottingham aiming to prepare participants for the practical elements of co-creation and implementation of chatbots as an educational resource. It combined both theoretical and hands-on training. 15 participants were from RISE, AUTH, UoN.

Project managers of partners signposted the person involved, and relevant announcements were made through social media channels to the wider public. External to the project speakers were from University of Leeds, and Computer Science Department of University of Nottingham. It included academics, medical doctors, and researchers with focus both on clinical research and digital innovations in

## *5. Citations and cross-refs*

healthcare education and IT specialist/learning technologists 11.18 years of experiences (SD=7.2). A balance between male and female participants achieved.

Participants were asked to highlight what they liked for each day and how each day can be improved. Findings are described below per day of the training event

### Day 1

The participants comment that they liked the design method for educational resources presented using a co-creation approach, they liked the interactions with other groups, and they liked the overview of existing chatbot resources of the partners. On the areas that can be improved, more media material were requested.

Day 2 Participants enjoyed the presentation from the invited speaker from another faculty of the University of Nottingham, the CEPEH resources presented and the storyboarding process. Participants highlighted that the participation of more clinicians in the event would be an added value in regards with the storyboarding process.

Day3 Participants liked the hands-on activities of the day also enjoyed the creativity of the groups on the online chatbot development tool. As an area of improvement, participants wanted more time on hands on sections.

## **5.2 CEPEH Training Event 2**

### **Pre-Training Event survey May 9th-13th 2022 Thessaloniki, Greece**

Twenty-six participants attended the Training Event, along with approximately 10 staff members. There were 21 undergraduate students and 5 postgraduate students, who completed the survey for a total of 26 responses. There were 86% of participants who stated they had not been to a similar event like the training event CEPEH facilitated. There were 90% of students who found the event schedule very organised, and 70% agreed most of the planned sessions were relevant to that interest with the remaining 30% not having enough experience to understand the context to determine if they are interested in the training event. There were 95%

## *5. Citations and cross-refs*

of students agreeing or strongly agreeing the training event location is great, the remaining person did not leave additional comments.

Table 1 suggested attendees had minimal intention to share their own ideas due to lack of previous experience of attending such events, or due to lack of knowledge on the area. However, most were interested in listening to other groups and hearing contextual cases in healthcare.

There were 77% of participants stated they were novices in experience with chatbots in healthcare and were attending to learn more. The remaining 23% (7 students) stated they were competent and had limited experience with chatbots in healthcare.

One day had several events regarding cybersecurity in healthcare. When asked before these events, 83% stated they were neutral or disagreed that they felt confident about their cybersecurity knowledge in healthcare. In addition, 80% stated they when neutral or disagreed that they felt they had strong cybersecurity safety in healthcare. Table 2 shows the main pre and post results suggesting a positive experience for more than 75% of attendees on all measures.

There were 90% (23) of students who heard about the event through a lecturer or a professor, the CEPEH newsletter (2), and 1 person was informed through the anatomy tutoring system at Karolinska Institute. Additionally, 60% suggested the training event to somebody else before the course started.

There were six individuals who stated neutral or disagree when asked if having issues on registration or finding the information for the event. This may have been due to being dependent on emails to receive the information, instead of a dedicated website where the information is available anytime.

As this was face-to-face, participants were asked about sufficient Covid-19 precautions in place at the facility, 94% agreed with sufficient precautions, two individuals stated no but did not give further information in the additional input box provided. In summary, most participants were undergraduate students with novice experience, happy with the training event location, felt the sessions were relevant to them, and most shared the event with their colleagues. The values of co-creation,

## *5. Citations and cross-refs*

chatbots in healthcare, and taking patient history were bestowed to students in an engaging and well-received manner. Notably, the highest ratings were for staff friendliness which is key to engagement and consistent interaction throughout the intense and long 5-day duration. The sessions were recorded there for the online recordings may be viewed with higher numbers over the subsequent weeks.

### **5.3 # bibliography: [bibliography/references.bib, bibliography/additional-references.bib]**



# Appendix

# Appendices



## The First Appendix

This first appendix includes an R chunk that was hidden in the document (using `echo = FALSE`) to help with readability:

**In `02-rmd-basics-code.Rmd`**

**And here's another one from the same chapter, i.e. Chapter ??:**

## References

- Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life*. John Murray.
- Goethe, J. W. von. (1829). *Wilhelm Meisters Wanderjahre oder die Entsagen-den*. Cotta.