1

Results

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

Table 1.1: 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.
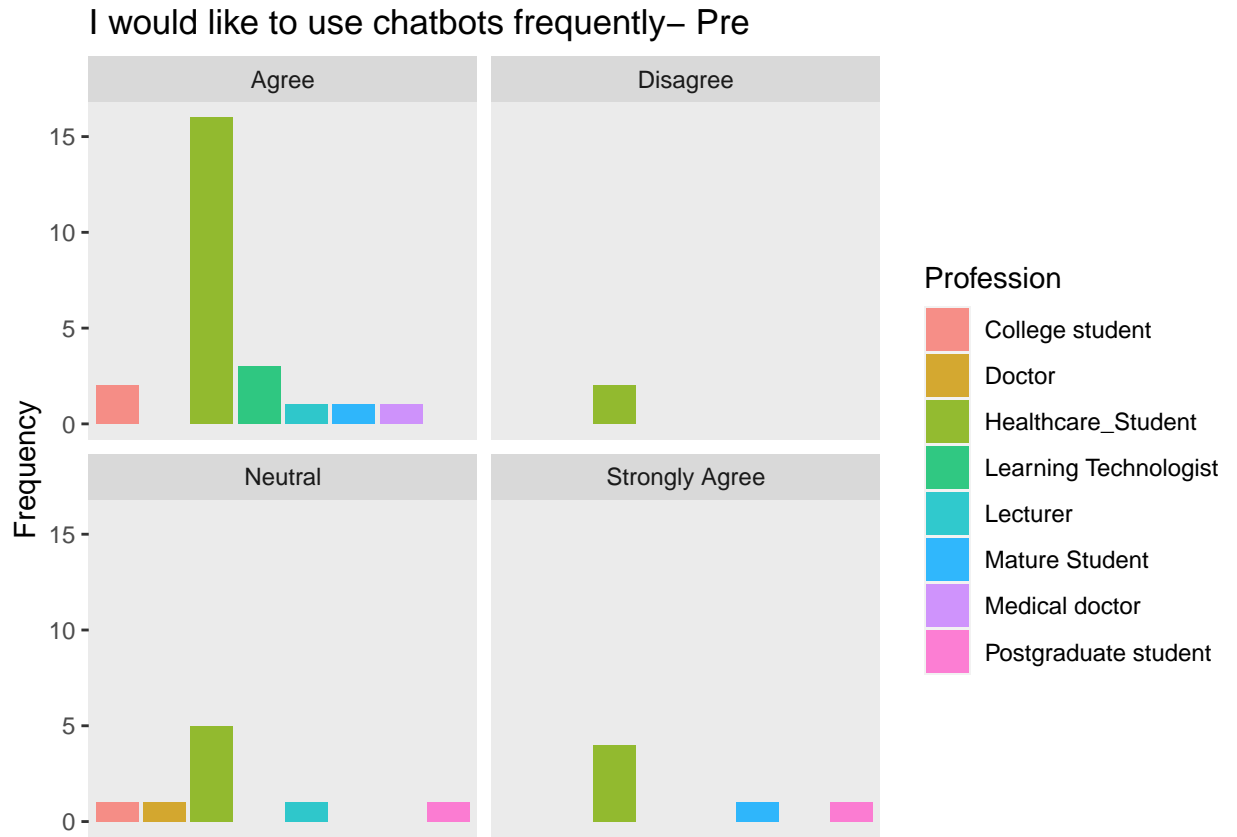
Figure 1.1: Chatbot Usage History- Pre

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 tools as chatbots.

The first boxplot (1.1) shows learners perceptions of easy of use of mobile app and other educational mobile resources

(**??**) 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: ## Chatbot Usability
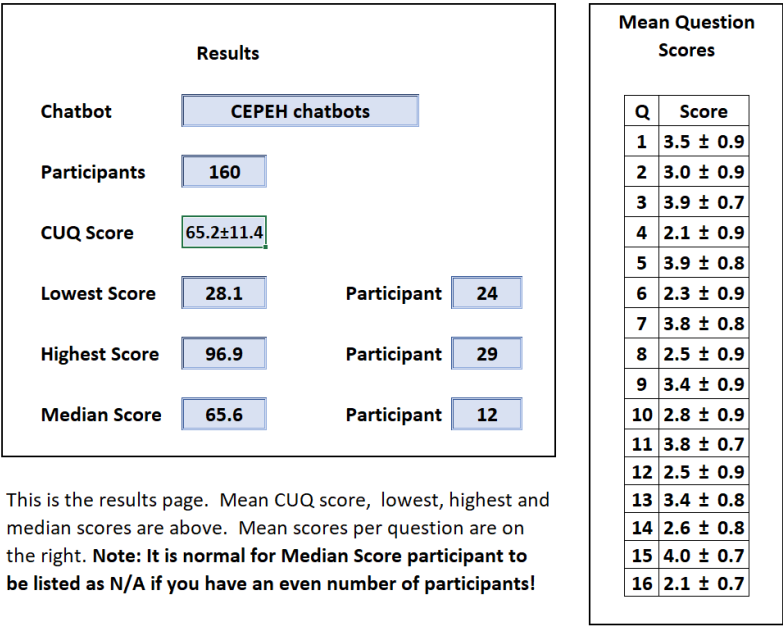
Chatbot Usability Questionnaire Results

	Results			
Chatbot	CEPEH chatbots			
Participants	160			
CUQ Score	65.2±11.4			
Lowest Score	28.1	**Participant**	24	
Highest Score	96.9	**Participant**	29	
Median Score	65.6	**Participant**	12	

This is the results page. Mean CUQ score, lowest, highest and median scores are above. Mean scores per question are on the right. **Note: It is normal for Median Score participant to be listed as N/A if you have an even number of participants!**

Mean Question Scores

Q	Score
1	3.5 ± 0.9
2	3.0 ± 0.9
3	3.9 ± 0.7
4	2.1 ± 0.9
5	3.9 ± 0.8
6	2.3 ± 0.9
7	3.8 ± 0.8
8	2.5 ± 0.9
9	3.4 ± 0.9
10	2.8 ± 0.9
11	3.8 ± 0.7
12	2.5 ± 0.9
13	3.4 ± 0.8
14	2.6 ± 0.8
15	4.0 ± 0.7
16	2.1 ± 0.7

Figure 1.2: CUQ CEPEH Score

(#fig:cuq image)

Questionnaire (CUQ)

1.1.1 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 University of Nottingham CUQ score = 68/100
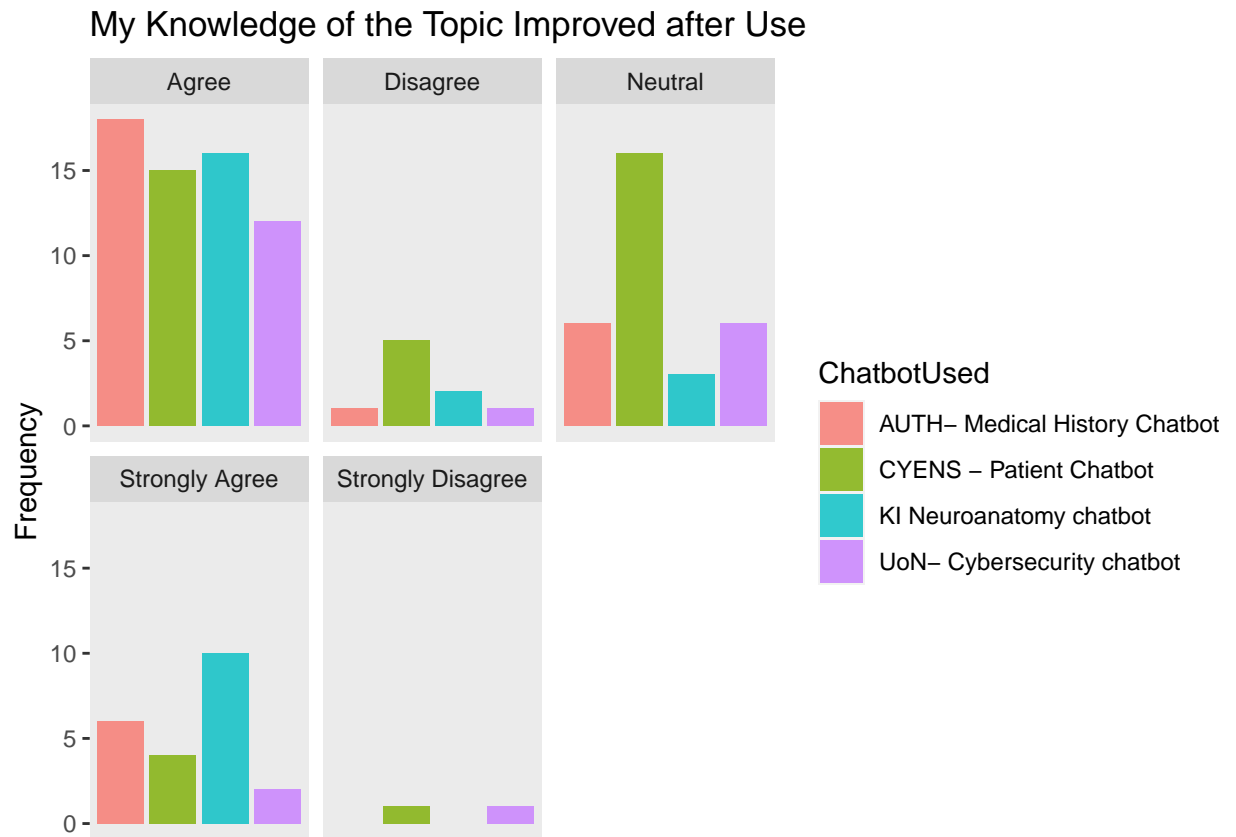
Figure 1.3: Improvements in Knowledge

The score for all 3 chatbots grouped was 65/100, This scoring system was designed to be comparable to SUS and may be freely used alongside it, or in combination with other usability metrics. There has been evidence of correlation of 76% between the CUQ and SUS therefore we expect the SUS scored to be between 48.75 and 81%. We believe the CUQ has the highest validity in the measures used towards the concepts of interest on this study.

Figure **??** 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.

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

I Trust CEPEH Chatbots to Provide me with my Course Information
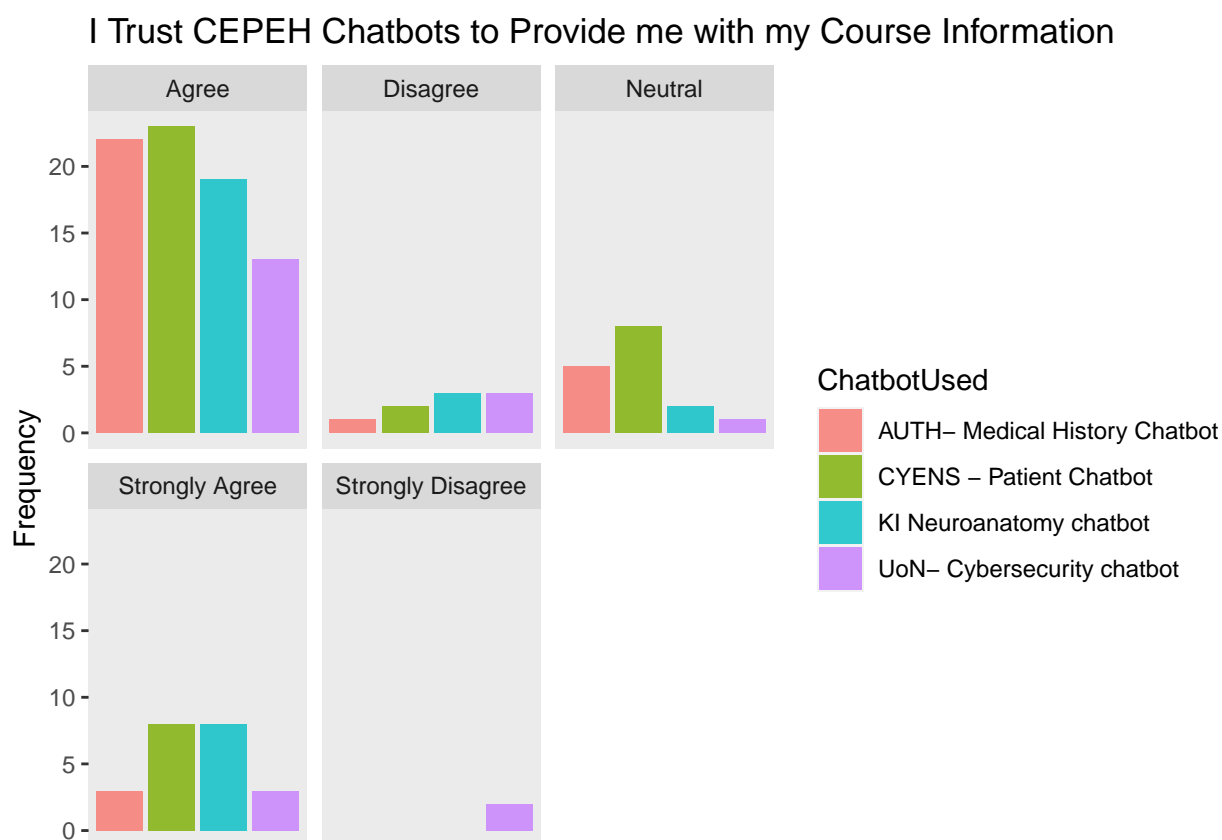
The figure above, (**??**) shows the ratings by participants of the CEPEH Chatbots

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.

CEPEH Chatbot Personailty was Realistic and Engaging
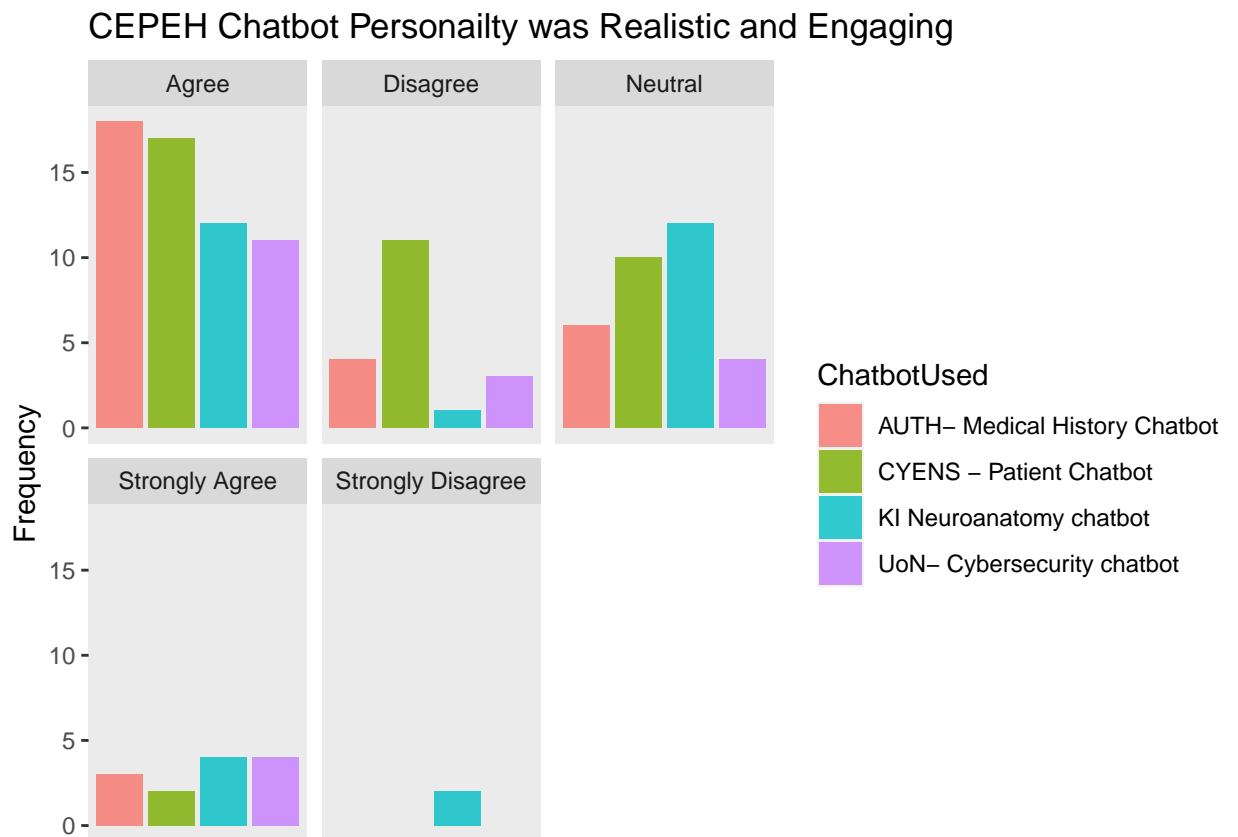
There was mixed results for the chatbot used being realistic and engaging.

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'
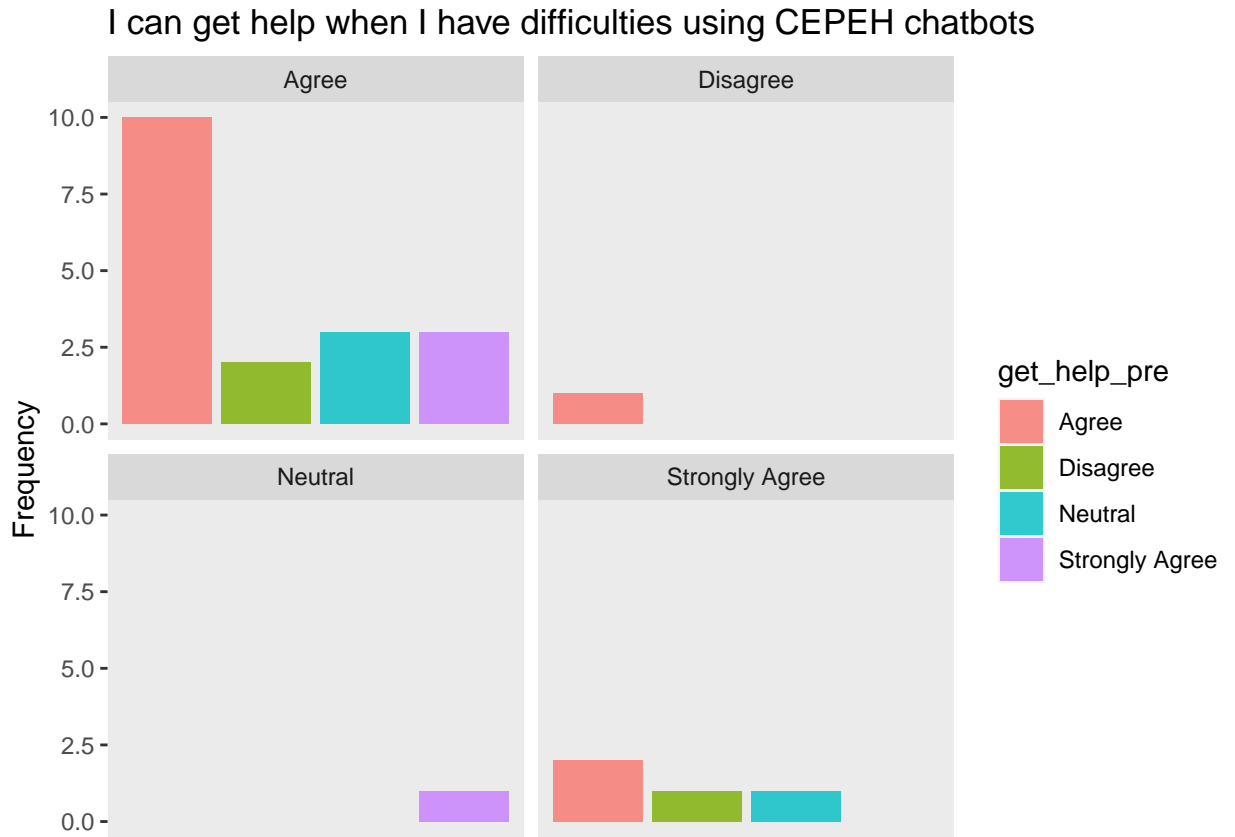
and 'Sentiment Analysis' sections.

The chatbot seemed too robotic (CUQ2)(post)
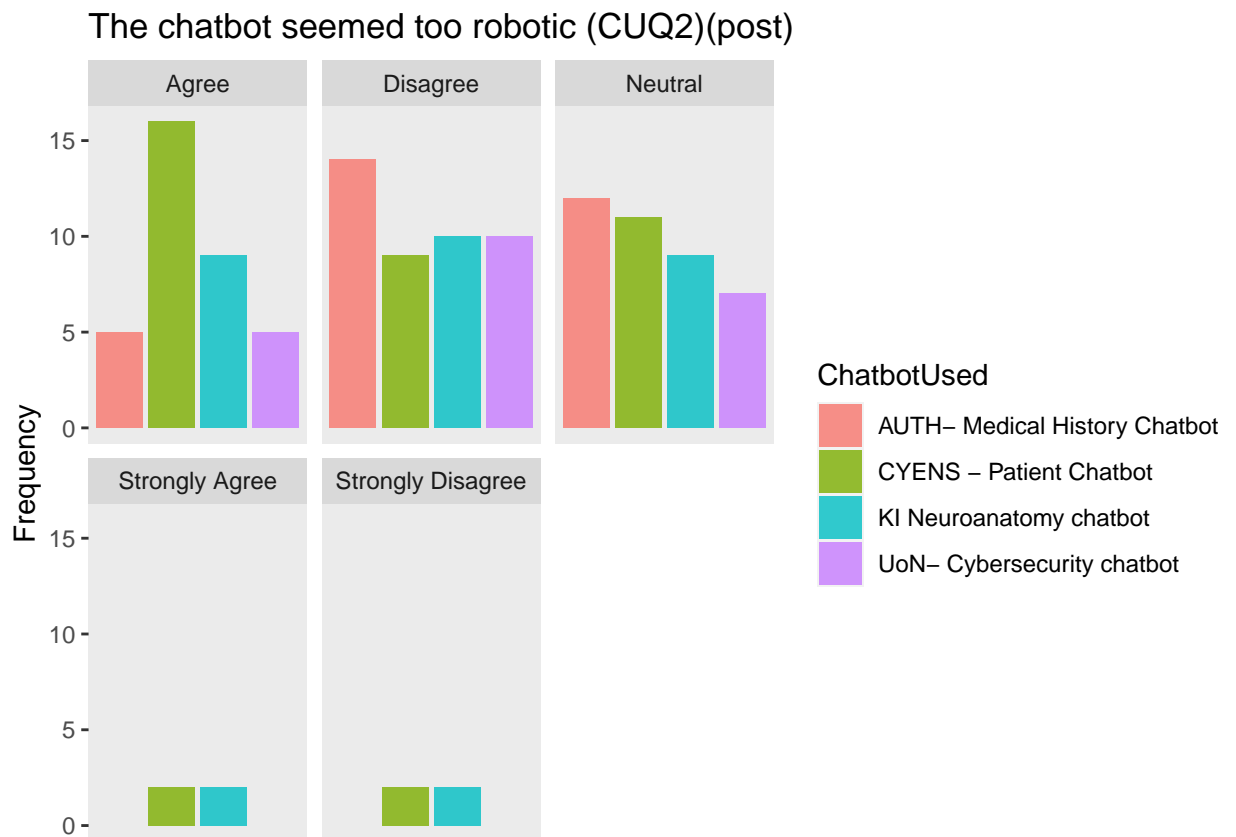
The chatbot seemed too robotic results had the largest mix of responses, and for all

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,

Change in Ease of Use Perception, after CEPEH Chatbot Usage
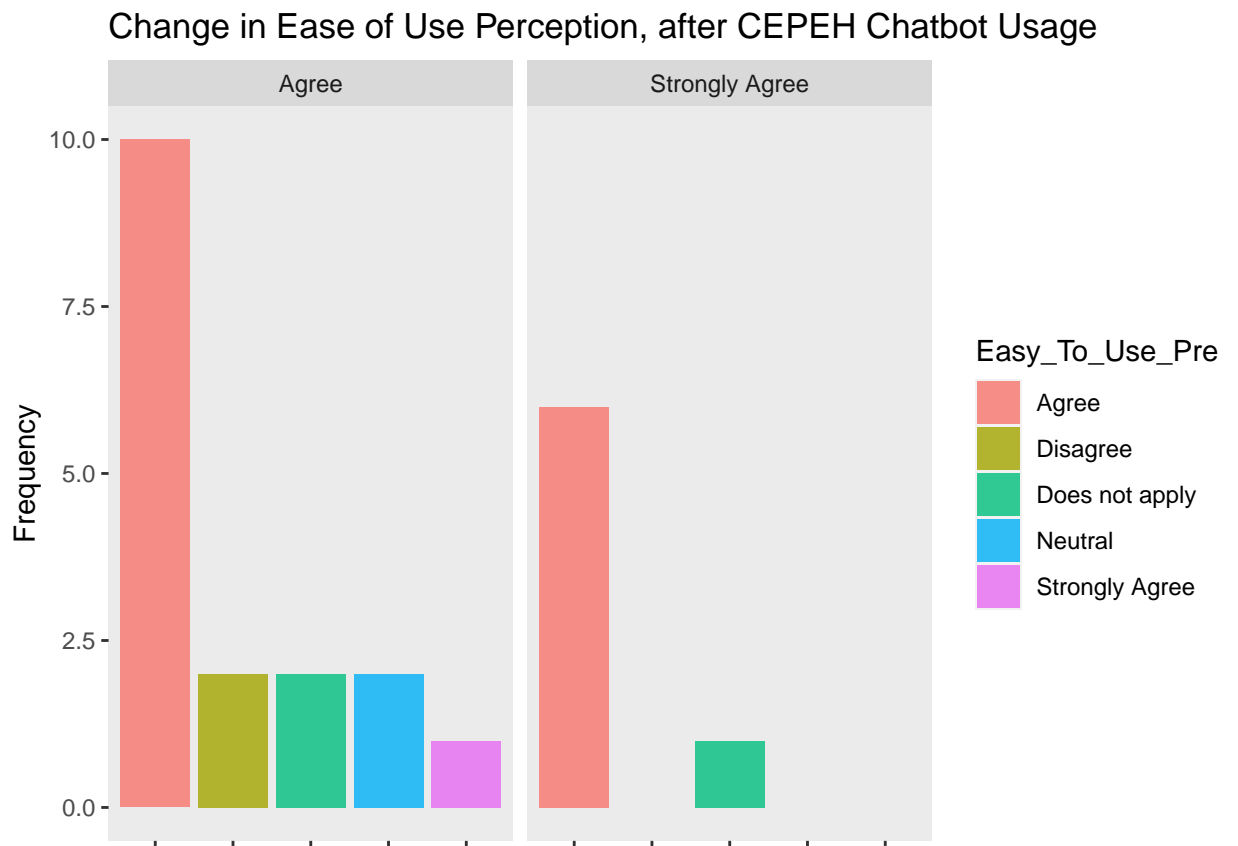
After usage, there was only agreement in Ease of Use- as shown in (**??**).

I can get help when I have difficulties using CEPEH chatbots

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.

1.2 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. 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)	Responces
Agree	71
Disagree	11
Neutral	21
Not Applicable	4
Strongly Agree	19

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

This table

1.3 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

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.

1.4 Inferential Statistics

1.4.1 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 (1.4).

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.

There was only 1 'Strongly Disagree' response. The agreement options counted for the majority of the data.
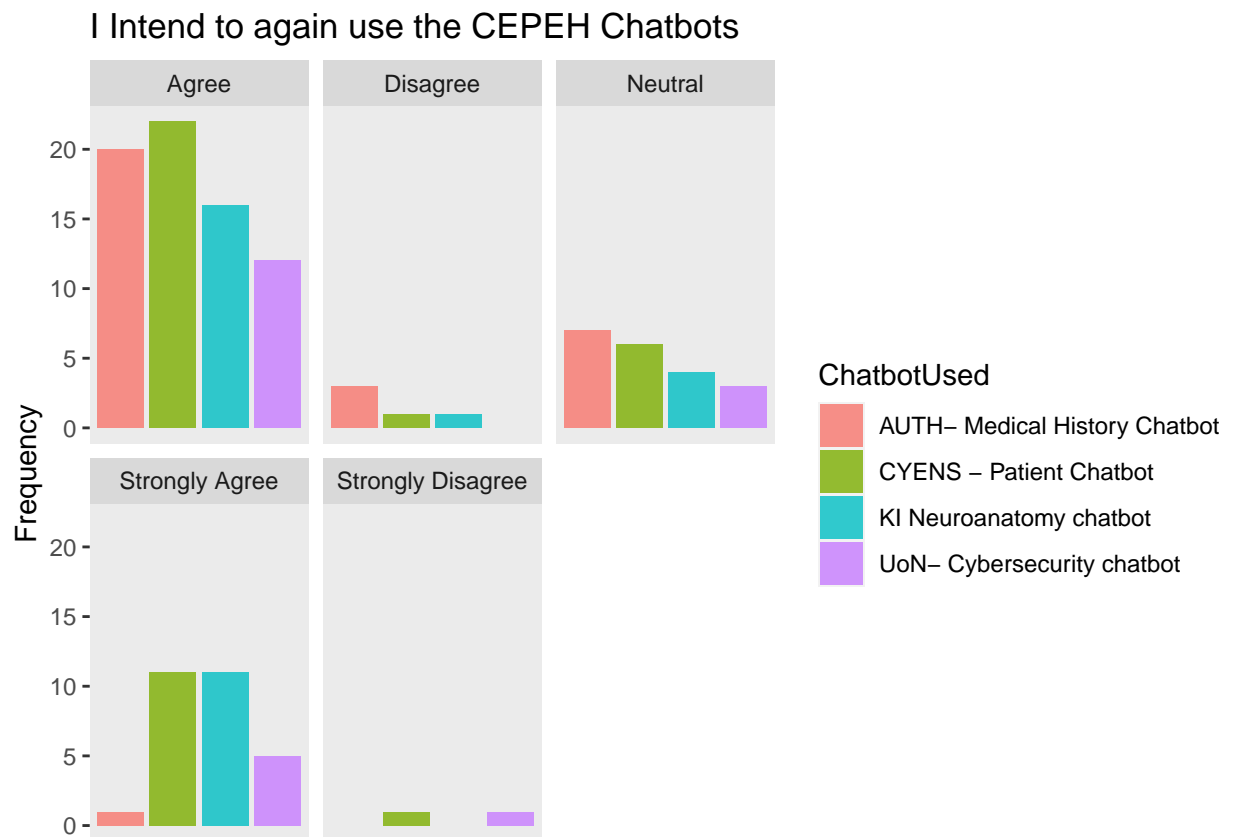
Figure 1.4: Intend to Reuse-Post

1.5   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

The chatbot has a high level of trustworthiness

The duration of conversations to find my answer was too long
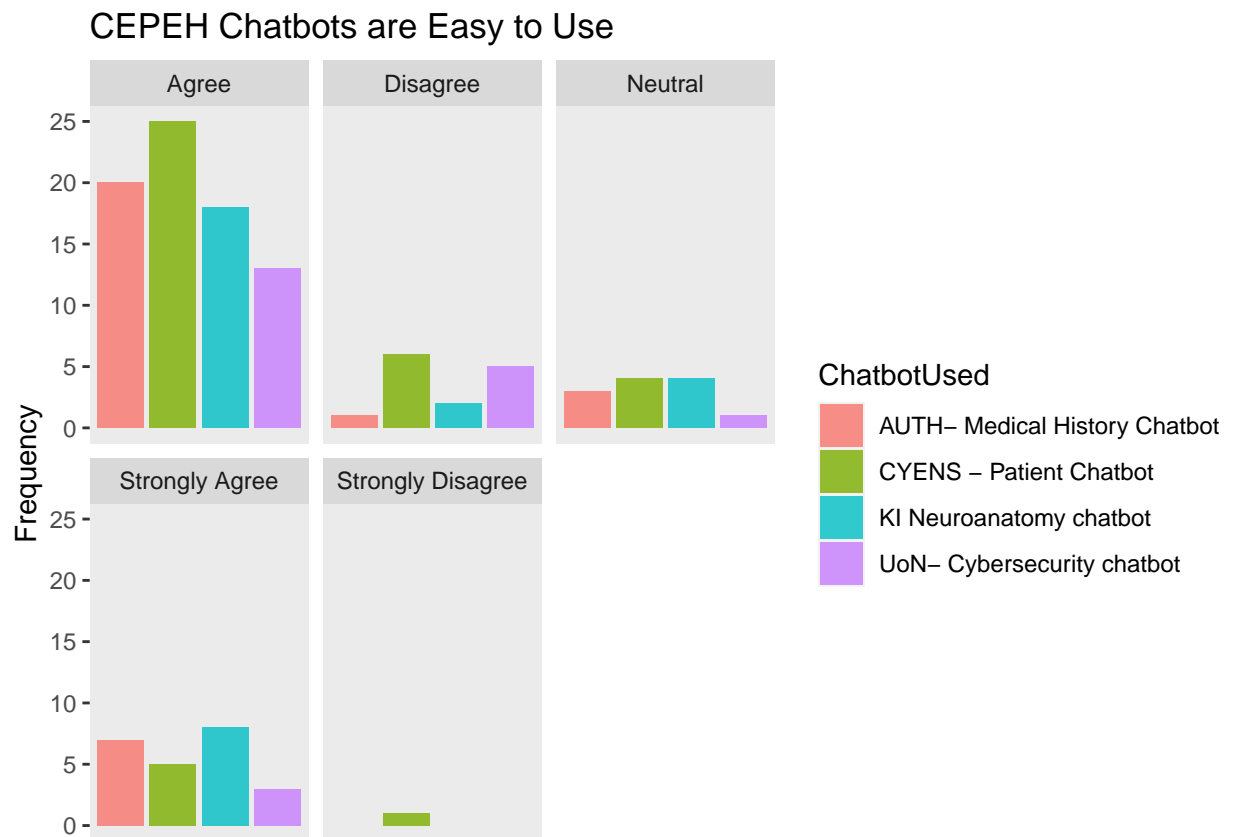
The videos/images provided were useful to my questions

Figure 1.5: Easy to Use- Post

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)

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