Health Innovation Wessex

Independent evaluation of Healthcare AI Mentor for Life Sciences Hub Wales

Authors:

Amanda Lees
Emily Hunter
Rebecca Player
Julia Wilson

Date of Report:

09/04/2025

Part of the Health Innovation Network

Evaluation Team

Emily Hunter, Senior Programme Manager
Dr Amanda Lees, Evaluation Programme Manager
Dr Rebecca Player, Evaluation Programme Coordinator
Julia Wilson, Evaluation Programme Coordinator

Correspondence

Dr Amanda Lees, Evaluation Programme Manager
Health Innovation Wessex
Innovation Centre, 2 Venture Road, Southampton Science Park, Chilworth, Hampshire SO16 7NP.

Disclaimer

This report presents the findings of an independent evaluation of the Healthcare AI Mentor. The findings of this independent evaluation are those of the authors and do not necessarily represent the views of the wider stakeholders.

Declaration of Interest Statement

Health Innovation Wessex supports innovators to bring their innovations to the NHS as well as provide an evaluation service more broadly to our members and others. On occasion, we evaluate innovations that we have also supported. Whilst these evaluations are independent, for transparency we disclose our dual role where applicable. In this instance, Health Innovation Wessex was only involved in the evaluation.

Acknowledgements

The evaluation of Healthcare AI Mentor was funded by Life Sciences Hub Wales. We would like to thank the project's stakeholders within Health Education and Improvement Wales and Life Sciences Hub Wales for their support and facilitation of this evaluation. We extend our gratitude to the staff of NHS Wales and partner organisations who gave their time to test Healthcare AI Mentor and to share their views and experiences with us. Our thanks also go to Jenny Curtis, Health Innovation Wessex, for her help in developing the visuals in this report.

Health Innovation Wessex

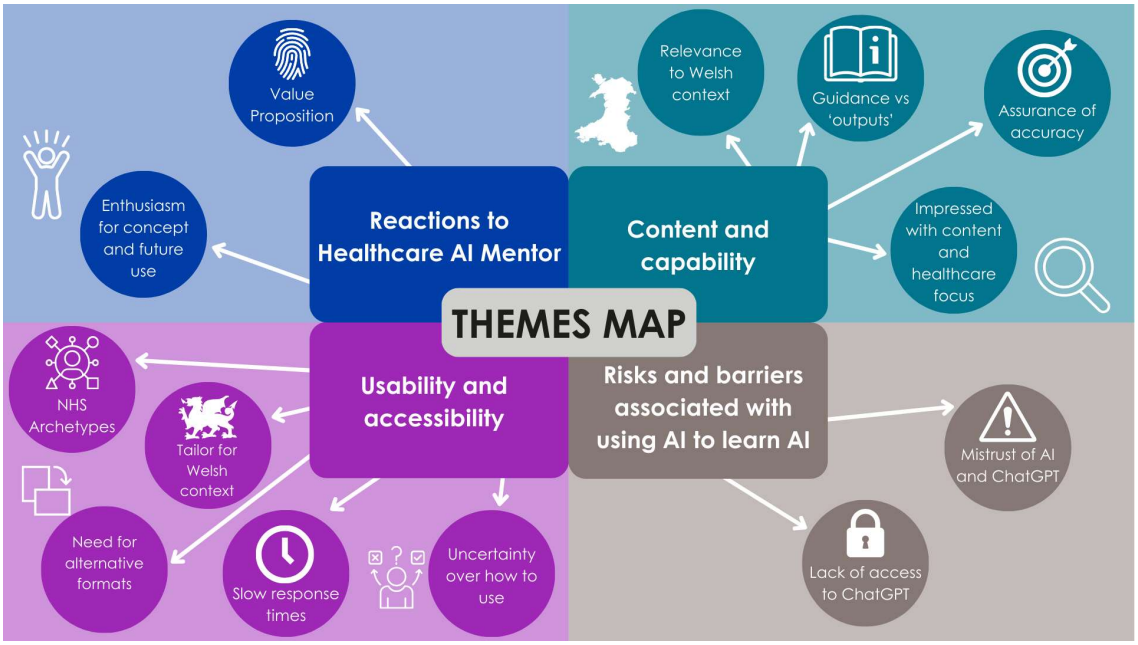
Table of Contents

Executive Summary

Healthcare AI Mentor is a ChatGPT[1] based tool, developed for use within NHS Wales to enhance artificial intelligence (AI) related skills, understanding and operational efficiency within the workforce. Employing the concept of 'using AI to learn AI', it aims to improve levels of AI literacy and upskill the workforce as well as to assist users in identifying AI use cases.

In March 2025, Life Sciences Hub Wales (LSHW) and Health Education and Improvement Wales (HEIW) commissioned the Insight Team at Health Innovation Wessex (HIW) to conduct a rapid, six-week evaluation of the proof-of-concept version of Healthcare AI Mentor (1.2), involving staff from NHS Wales and partner organisations in the initial testing and feedback.

The Insight team conducted six online Think Aloud[2] interviews and four online focus groups, during which participants interacted with the AI Mentor and provided their feedback on it. Rapid Assessment Procedure (RAP) sheets were created to categorise the responses and used as the basis for developing themes.



Reactions to Healthcare AI Mentor ranged from very positive to some questioning of its value proposition. Participants were generally impressed with the AI Mentor's **content and capability** which included the information provided by the AI Mentor, its ability to respond accurately to questions and to provide healthcare relevant responses. Some participants would prefer more referencing of sources to be assured

[1] A large language model chatbot, https://chatgpt.com/
[2] Bennett, S.E., Johnston, M.H., Treneman-Evans, G., Denison-Day, J., Duffy, A., Brigden, A., Kuberka, P., Christoforou, N., Ritterband, L., Koh, J., Meadows, R., Alamoudi, D., Nabney, I., Yardley, L. (2024). Co-creating and optimising an app-based intervention to support better sleep for adolescents in the UK. JMIR Preprints. https://doi.org/10.2196/preprints.63341

of its accuracy and a greater focus on the Welsh context in information returned. A small number requested specific outputs as opposed to advice and information. In terms of **usability and accessibility**, most Think Aloud interviewees expressed some uncertainty over how to interact with the AI Mentor, and the terminology around NHS archetypes tended to be seen as inaccessible. A few respondents suggested including more reference to the Welsh context for example use of the NHS Wales logo, or Welsh greeting. Some suggested including other formats like visuals, videos or podcasts to increase engagement. Participants described some **risks and barriers associated with using AI to learn AI** which focused on a lack of access to ChatGPT at work, and a mistrust of AI and ChatGPT with regards to confidentiality and risks of misinformation.

We noted some differences between those with more involvement and familiarity with AI and digital, and those who described themselves as novices. Those with existing knowledge were more likely to readily appreciate the concept and usefulness of the AI Mentor, with those with less familiarity more likely to be hesitant.

The report describes improvements suggested by participants for further development of the AI Mentor. Guided by these, the report's recommendations centre around reinforcing the value proposition, enhancing accessibility, assurance of accuracy and safe use, further testing and positioning within the training offer.

RECOMMENDATIONS	ACTIONS TO CONSIDER			
Reinforce value proposition	Include further information about aims and how this differs from other AI tools	Complement with communications to position the offer and encourage uptake		
Enhance accessibility	Use of plain English	Provide alternative formats and strategies to keep people engaged with content	Acknowledge Welsh context	
Assurance and safe use	Include AI rules of engagement/ disclaimer	Link to organisational AI policy and SOP	Enhance referencing of sources	Consider balance between guidance and outputs – do parameters need adjusting to prevent certain outputs?
Further testing	Test with different audiences to ensure language keeps people engaged	Include feedback mechanism to flag any inaccuracies		
Place in training offer	To sit as part of a suite of options, rather than mandatory training	Consider where preparatory digital and AI literacy skills may be needed		

1 Introduction and evaluation focus

Healthcare AI Mentor is a ChatGPT[3] based tool, developed for use within NHS Wales to enhance AI literacy which should translate into future improvements in operational efficiency and patient outcomes. It offers tailored guidance for various NHS archetypes[4] (Shaper, Driver, Creator, Embedder, User) and specific job roles, enhancing the learning material to suit each user's needs. It prioritises challenges, risks, and considerations associated with AI use and emphasises ethical and equitable practices. Employing the concept of 'using AI to learn AI', it aims to improve levels of AI literacy and upskill the workforce as well as to assist users in identifying AI use cases.

In March 2025, Life Sciences Hub Wales (LSHW) and Health Education and Improvement Wales (HEIW) commissioned the Insight Team at Health Innovation Wessex (HIW) to conduct a rapid, six-week evaluation of the proof-of-concept version of Healthcare AI Mentor (1.2), involving staff from NHS Wales and partner organisations in initial testing and feedback. The evaluation investigated staff's reactions to Healthcare AI Mentor (also referred to as 'the AI Mentor') including their perceptions of the usefulness of using AI to learn AI for themselves and the wider workforce, perceived barriers and facilitators to the use of the AI Mentor, and suggestions for improvement and optimising its use.

2 Evaluation approach

2.1 Sampling and recruitment

Brief information about the evaluation was shared by Health Education and Improvement Wales via a range of relevant forums and networks, including HEIW student forums, HEIW social media platforms, the HEIW staff intranet and the Welsh Government AI Commission for Health and Social Care. People wishing to take part were asked to complete a form registering their interest and providing their consent for their details to be shared with, and to be contacted by, the Insight team. The form also asked people to state whether they would prefer to take part in a focus group or a one-to-one interview. The Insight team then contacted those who had expressed an interest in taking part to arrange the preferred appointments. Once appointment times were agreed, we provided an information sheet containing further details about the structure of the sessions, and the requirement to sign up for a ChatGPT account to use the AI Mentor. Adobe consent forms were shared for e-signature.

We aimed to achieve a sample of 20 staff from within NHS Wales or partner organisations, to test and feedback on the AI Mentor. We also aimed to include a mix of job roles including administrative positions. We received an excellent response to our recruitment, with expressions of interest from individuals in a range of job roles and at various level of seniority and experience. We reviewed responses and categorised respondents broadly into role categories (including clinical, business, leadership and administration). As far as possible we grouped people with similar roles and experience levels together in focus groups. The aim of this was to facilitate ease of discussion and comfort for the participants.

[3] A large language model chatbot, https://chatgpt.com/

[4] NHS AI Lab and Health Education England (2022) *Developing workers' confidence in AI* *Understanding healthcare workers' confidence in AI*

2.2 Rapid approach to evaluation

The Insight team employed a rapid evaluation approach (Vindrola-Padros et al. 2020[5]) to enable data collection, analysis and reporting within the agreed six-week evaluation timescale. Rapid evaluations are characterised by team approaches to data collection and analysis and employ a real-time approach to summarising key categories of qualitative information, allowing for the quick generation of insights and outputs. Our team consisted of four evaluators, all of whom contributed to data collection and analysis.

2.3 Data collection methods

Over a two-week period, we conducted six online Think Aloud[6] interviews and four online focus groups. Think Aloud interviews are a method in which respondents experience an intervention (in this case, the AI Mentor) in the presence of the evaluator, allowing the observation of users' reactions, and interactions in real time. Interviews lasted up to one hour. Following a short introduction and sharing of the link to the AI Mentor, interviewees used the AI Mentor for up to 15 minutes, whilst sharing their screen, and simultaneously describing their thoughts and actions out loud to the interviewer (the Think Aloud section). At the end of the Think Aloud section the interviewer asked some general questions concerning views, perceived usefulness of the AI Mentor and suggestions for improvement. At the end of the hour, the participants completed a short online demographic survey, shared via the meeting chat bar.

Each online focus group comprised up to six participants (the final number varied due to some last-minute cancellations). Following a short introduction and sharing of the link to the AI Mentor, focus group participants used the AI Mentor individually for around ten minutes (without sharing screens), followed by a group discussion to elicit their views of the AI Mentor, perceptions of its usefulness and suggestions for improvement. At the end of the hour, the participants completed a short online demographic survey, shared via the meeting chat. The brief surveys collected demographic data from the testers including age, ethnicity, role, length of time in role and self-rated level of familiarity with AI.

2.4 Data analysis methods

We created Rapid Assessment Procedure (RAP) sheets to summarise the main categories of information collected. We created one RAP sheet for the Think Aloud sections of interviews and a second RAP sheet for the remaining interview and focus group questions. RAP sheets were populated as soon as possible after each interview or focus group, guided by transcripts and video recordings where necessary. Once complete, RAP sheets represented a summary of the entire data set, including illustrative quotes, which were then used as the basis for developing themes. The

---

[5] Vindrola-Padros, C., Chisnall, G., Cooper, S., Dowrick, A., Djellouli,N., Mulcahy Symmons, S., Martin, S., Singleton, G., Vanderslott, S., Vera, N., Johnson, G.A. (2020) *Carrying out rapid qualitative research during a pandemic: emerging lessons from covid-19* Qualitative Health Research, 1-13.

[6] Bennett, S.E., Johnston, M.H., Treneman-Evans, G., Denison-Day, J., Duffy, A., Brigden, A., Kuberka, P., Christoforou, N., Ritterband, L., Koh, J., Meadows, R., Alamoudi, D., Nabney, I., Yardley, L. (2024). Co-creating and optimising an app-based intervention to support better sleep for adolescents in the UK. JMIR Preprints. https://doi.org/10.2196/preprints.63341

Insight team discussed emerging themes and further refined and finalised these during regular project meetings to capture an accurate representation of the data.

3 Participant characteristics

We conducted four focus groups, with 16 participants, and six one-to-one interviews, totalling 22 participants overall. Table 1 below summarises the participant characteristics. Of the staff described, around a third (7/22) had digitally focused roles (for example digital clinician, or digital transformation lead).

Table 1 Participant profile

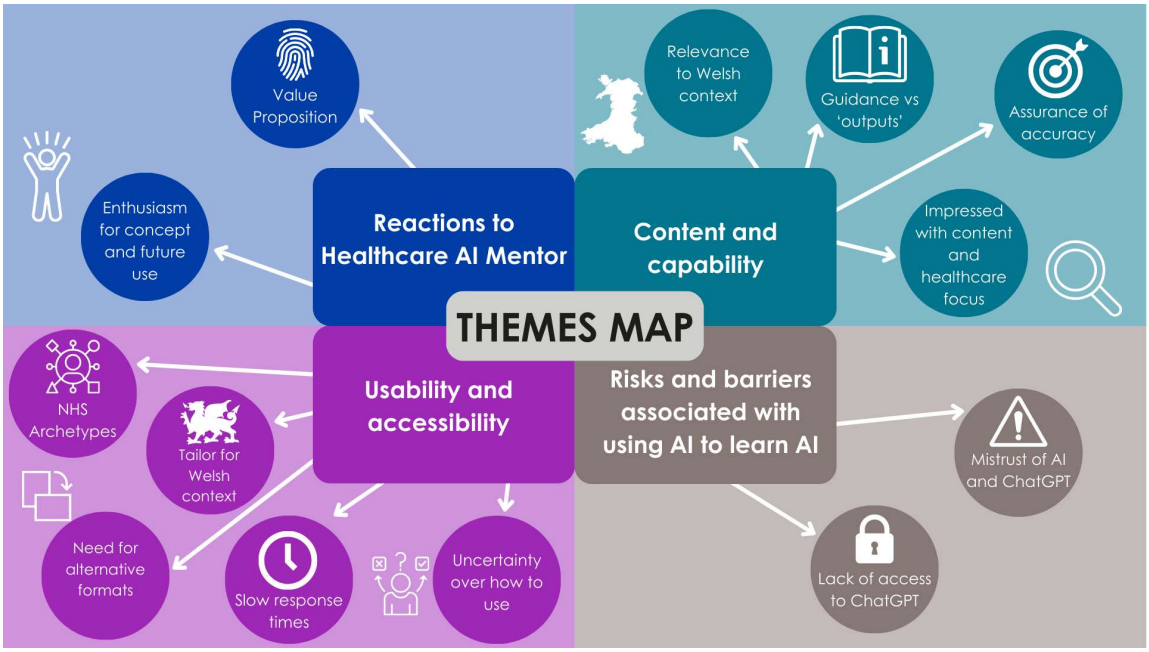
Characteristics	Number
Job category	
Administration/ Business/Leadership	11
Clinical (including clinical educator roles)	11
AI experience (self-rated by participants)	
Advanced	1
Intermediate	14
Novice	6
Never used	1
Focus group or interview	
Focus group	16
Interview	6
Total	**22**

4 Findings

The findings are presented as a theme map in figure 1 below. Themes were grouped into clusters concerning reactions to Healthcare AI Mentor, content and capability, usability and acceptability and risks and barriers associated with using AI to learn AI. This section goes on to describe the themes using illustrative quotes. Please note:

- Each participant was allocated an individual ID number. These relate to the number of expressions of interest received and therefore may exceed 22 (the total number of participants in the evaluation).
- Quotes are allocated using the following conventions: *Interviewee 7* for those taking part in one-to-one interviews, where the number represents the participant ID number and *Participant 6*, for focus group participants, where the number represents the participant ID. The focus group number is also shown.

Figure 2: Theme map



4.1 Reactions to Healthcare AI Mentor

Enthusiasm for concept and future use

Several participants expressed their enthusiasm for the AI Mentor, and a desire to share it with their team and others. Participant 47, Focus Group 4, felt the AI Mentor would be of interest to her team because they are all *"very data and AI literate already."* Focus Group 3 participants identified the tool's usefulness across the multidisciplinary team and corporate functions. Interviewee 7 suggested that if shared via LinkedIn or YouTube, it would "go *viral in the education community*". A few participants stated that the 10-15 minutes testing opportunity was not long enough to fully explore the AI Mentor and that they would have liked the opportunity to trial it for longer.

Some felt that certain conditions needed to be met before using or recommending the AI Mentor. These included improved referencing to be assured of the validity of information returned, and assurance that the tool is compatible with Trust and NHS guidelines. Participant 6, Focus Group 2 explained that they wouldn't want *"everybody to rush in and use it"* until potential risks related to privacy and safety had been mitigated. One participant also noted that ChatGPT would need to be made available at work to allow use of the AI Mentor.

Respondents who expressed enthusiasm for the AI Mentor tended to be already positively disposed towards AI, and digitally and AI literate. Whilst they felt that they may not be the main intended audience for the tool, these participants could see the value in encouraging use amongst colleagues.

"I like the fact that AI is being used to educate about AI…I think that's a nice thing." (Participant 31, Focus Group 3)

"I guess that this audience is already not AI naïve… I think that the greatest gains will be seen in those who have the least exposure." (Participant 31, Focus Group 3).

Value proposition

Several participants questioned how the AI Mentor was different to other chatbots, or the advantages it offered over other ways to learn about AI. Questions about distinctiveness from other forms of AI were shared by those with varying levels of familiarity with AI. Novice users tended to question the benefits that the AI Mentor offered over other sources of information such as a search engine or a reference book. However, as Interviewee 50 indicated, there may be potential to change people's views via a specific *"sales pitch, positioning piece."*

"I can't see the USP... I would have got the same output with Copilot." (Participant 2 Focus Group 1)

"I can't see why you'd want to use an AI tool or a chatbot to do something that feels more like a couple of pages on the intranet." (Interviewee 50)

4.2 Usability and accessibility

Uncertainty over how to use Healthcare AI Mentor

Most Think Aloud interviewees expressed some uncertainty over how to interact with the AI Mentor. Interviewee 50 described the launch page as *"a bit bewildering"* because of a lack of details *"about what it actually does."* Half (3/6) of the Think Aloud interviewees did not answer the initial prompt question on the launch page concerning their job role but instead entered their own query into the first box. It may be that the prompt question had been missed, or that these respondents (who were all familiar with the use of chatbots) just began interacting in the same way as they would normally use AI. Participants sometimes wondered how to answer the questions raised by the AI Mentor; however, users who were more familiar with chatbots exhibited more confidence in how to answer the questions.

"There is nothing on the screen" (Interviewee 38)

"So it's asking me three questions. I don't know, should I answer these all in one go, or should I answer them one at a time?" (Interviewee 50).

NHS archetypes

Participants tended to feel that the concept of 'NHS archetypes' was *"a little bit alien"* (Interviewee 26) and not accessible enough *for* the AI Mentor's target audience. Interviewee 50 was uncertain which archetype they fell into. The one exception was Interviewee 7 who was very impressed with the use of the archetypes to guide content; however, this participant had a highly specialised knowledge of AI and digital.

Need for alternative formats

Some participants felt the AI Mentor, in its current iteration, relied too heavily on reading text. Some suggested including other formats like visuals, videos or podcasts to increase engagement. Interviewees 7 and 26 highlighted the need to

accommodate different learning styles and to give people opportunity to reinforce their learning through 'doing', while one participant suggested 'gamification' as a way to keep users engaged in the content.

"There's a lot of information on the screen." (Interviewee 26)

"If there was a way to deploy this knowledge base and this interface through a kind of a more gamified way then I think I could be quite excited about it." (Participant 31, Focus Group 3).

Slow response times

Participants felt that the AI Mentor was sometimes slow at generating responses – which could make them wonder if they had *"stumped it"* (Interviewee 7) or caused an issue with what they had input. However, in most cases the AI Mentor did provide an answer.

"Ok, that was a very slow start. I think I might have upset it a little bit by not saying I'm a nurse or a doctor." (Interviewee 50).

Tailor for Welsh context

A small number of respondents felt that the AI Mentor should make more use of the Welsh language, such as a greeting, to enhance inclusivity and accessibility. Interviewee 26 suggested that the logo for NHS Wales should be used in place of the logo for NHS England.

4.3 Content and capability

Impressed with content and healthcare focus

Participants were often very impressed with the content generated by the AI Mentor. Participants noted its ability to understand and respond appropriately to questions entered. Interviewee 7 described being *"blown away"* by this. Several participants noted the accuracy of information provided; for example, interviewee 50 stated that content provided on key AI challenges with regards to NHS data *"hit the nail on the head."* Several participants reported that content was relevant and helpful for their roles and included facts that were new to them. The content generated by AI Mentor often appeared to exceed people's expectations with reactions ranging from *"pleasantly surprised"* (Participant 17, Focus Group 3) to *"vastly exceeds"* expectations (Interviewee 7).

Participants were also impressed with the specificity of the content to their healthcare role or context. Participant 47, Focus Group 2 was impressed by the level of detail and healthcare relevance in how the AI Mentor's responded to her query about a programming language. Participants were impressed by its awareness of relevant issues, such as rural healthcare or region-specific challenges. A number were impressed by the utility of the outputs generated such as an AI policy or training plan.

"It's aimed at healthcare, whereas if you go into ChatGPT it looks at everything." (Participant 18, Focus Group 3).

"After a short exchange, AI produced an output, it came up with a structured plan for expanding my AI expertise, which definitely fits my need." (Participant 30, Focus Group 4).

These positive reactions to the content were predominantly made by participants who described themselves as having intermediate experience of AI and working in a range of roles. Less enthusiastic responses to the content were seen within Focus Group 1 where participants described themselves as novices with AI; responses from this group focused more on the need to assure the accuracy of responses, the technical nature of the responses and general concerns about AI.

Relevance to Welsh context

In contrast to the findings above, four respondents suggested that the AI Mentor had signposted them to resources that, whilst in the right subject area, were not relevant for NHS Wales.

"It came up with a load of NHSE stuff to look at and whilst that is sort of correct, it's not pointing me in the Wales direction." (Participant 6, Focus Group 2).

A perceived lack of relevance could result in people being misinformed, *"put people off"* (Participant 2, Focus Group 2) or *"make people a bit sniffy"* (Interviewee 50). On the other hand, participants noted that more specific questioning resulted in responses that were more regionally relevant, and that the AI returned *"depth and quality of information"* (Interviewee 26) regarding regional challenges.

Guidance versus 'outputs'

Some participants asked the AI Mentor to create a specific output. Examples of this included a request for a self-help plan for a patient (which caused a technical issue and required a restart) and a request for an AI policy for Wales (which was provided, and with which the participant was very pleased). In other instances, the AI Mentor generated content based on testers expressing an interest in a particular area, rather than a direct request. Examples of this included the generation of an outline for AI Implementer Training (with which Interviewee 7 was highly impressed) and the generation of advice on how to perform a particular surgery, which indirectly resulted from a tester's interest in whether AI could generate training videos of operations for surgeons. Whilst Interviewee 38 was pleased with the accuracy and 'up to date' nature or the information provided, the provision of a detailed description of a surgical procedure appeared to be outside of expected parameters. The evaluator reflection for this interview noted: *"consider giving instructions for use and setting boundaries in the prompt to avoid the tool answering questions outside its scope"*.

Assurance of accuracy

Several participants expressed the desire for further assurance concerning the accuracy of information returned, and a mechanism by which to highlight any inaccuracies. Two participants in Focus Group 3 emphasised the importance of including references *"to give people some assurance that it's getting this from something that is published"* (Participant 17, Focus Group 3) and where they could cross-check information. Developing this theme, Interviewee 7 suggested that the ability to feed back on any inaccurate content should be incorporated into the next phase of testing.

"*...it all comes down to how can you be confident that when the chatbot points you to something, it's pointing you to the right thing, which means things that don't cause harm, but also, things that are comprehensive enough...*" (Interviewee 7)

" *How will we know when we're given the wrong information? What there isn't in this is any way to say, actually I don't agree with that…I'd love some sort of feedback mechanism.*" (Interviewee 7).

4.4 Risks and barriers associated with using AI to learn AI

Lack of access to ChatGPT

One immediate barrier or risk of using AI to learn AI was restriction of access to ChatGPT, which was routinely blocked by several organisations known to participants. A small number of our participants needed to use personal devices to interact with the AI Mentor during focus groups and one participant was unable to access the AI Mentor themselves because of these restrictions.

Mistrust of AI and ChatGPT (confidentiality and quality of information)

Some participants suggested that a generalised mistrust towards ChatGPT may also deter people from using the AI Mentor. Concerns focused on data governance and confidentiality (though these did not tend to be expanded upon) and risk of misinformation. Participant 2 suggested that addressing these concerns should be "*person-led*" and reassuring to build confidence in AI use through discussion.

Participant 52 in Focus Group 1 worried that users might take information generated by AI as "*gospel*" and Interviewee 26 signalled their own concerns about the risks of rapid onward sharing of "*significant misinformation*" generated by AI, with Participant 6 in Focus Group 2 stating that such cases are starting to be the focus of disciplinary hearings.

"*A lot of people are very sceptical about It* (ChatGPT). *For data governance, confidentiality, they don't want to touch it.*" (Participant 6, Focus Group 2)

"*It can be really dangerous, and people are relying on it really heavily now.*" (Participant 52, Focus Group 1)

4.5 Suggestions and next steps

Table 2 below describes suggestions, and next steps, for the further development of the AI Mentor as defined by participants in response to specific questioning on what improvements could be made, as well as points raised during general discussions.

Table 2 Suggestions for improvement

Suggestion	Details
Inclusion of disclaimer and framework for AI use	Include a disclaimer statement and guidance concerning Do's and Don'ts, framework for use, AI etiquette, how data will be used. Linked to a broader organisational framework for AI or standard operating procedure. *"As much as we hate NHS red tape… I think it needs to be wrapped in some 'beautifully comms designed red tape so that people know... not to get their fingers burned."*
Improve orientation to tool on the launch page	Inclusion of further explanation on the AI Mentor and how to use it on the launch page. *" You need an idea of where to go."* Consider reformatting questioning regarding role and archetype. *"Start off conversationally so rather than saying…'What is your job title?' have something like Welcome to..."*
Use accessible language and explain unfamiliar concepts	Do not assume pre-existing knowledge (e.g. concerning what AI is, or what NHS AI archetypes are), consider if terms may be overly technical *"I think it needs to be more understandable"* and include explanations *"I think there needs to be something that says, you know…what artificial intelligence is."*
Consider different learning styles/needs	Consider including additional forms of content and learning, moving away from reliance on text. *"We know that people don't learn best by reading stuff. We know they learn best by doing stuff…So there's something about marrying the signposting value of this with the educationalist. How best to deliver education? And I'd love to explore that."*
Fit to Welsh context	For example, by including some Welsh language, NHS Wales logo, linking to key regional frameworks, and/or include an explanation that Welsh sources may not always be available and explaining the implications of this. *"We like to dragonise our own things."*
Improve referencing/assurance of accuracy	Inclusion of source references and ensure/explain that responses are linked to relevant organisational and NHS frameworks. *"In this tool there is no referencing system...if they can give this reference about that…it will be better."*
Improve access from work computers, or consider alternative interface	There was some mention of the need to ensure that people can access the AI Mentor (and thus ChatGPT) from their work computers, and that it would be good if *"you didn't have to sign up for a ChatGPT login."*
Further testing required	Two participants suggested further testing by providing responses from different perspectives (no AI knowledge, expert user etc), and identified the need for wider testing after this initial phase. *"I'm assuming that after this study, there'll be a wider study, you know, if it's successful in a small group in a closed and safe environment, then you know, exposing it in the wild. And getting some feedback in the wild might be a good next step."*
Unsuited to mandatory training	Consensus that the AI Mentor is not suited to being part of the mandatory training offer. Rather it should be considered as *"another tool in the toolbox"* to offer, alongside other strategies to build AI and digital literacy skills including supportive conversations or a video learning package. It was also suggested that the AI Mentor could signpost people to other relevant learning modules.

5 Limitations

Respondents were given a short time (10-15 minutes) with which to interact with a proof-of-concept version of the AI Mentor, meaning that these findings are based on a snapshot view of a tool which is still in development. Further testing would be necessary to enable longer interaction and/or to interact with future iterations. A small number of participants were not able to access ChatGPT directly during the sessions and therefore were only able to interact with it via an evaluator sharing their screen and asking for their instructions on what to enter. This may have affected their experiences and views of the AI Mentor.

We received an excellent response to our initial invitation for expressions of interest to participate in this evaluation. The limited timeframe for the evaluation (six weeks in total, with two weeks for data collection) meant that we were unable to include all these volunteers in our evaluation, limiting the range of responses.

It is also important to note that the sample was self-selecting (suggesting their interest in the evaluation's subject matter) and contained a proportion of staff who were in digitally focused roles. Their reactions to the AI Mentor therefore may not be representative of the wider body of NHS Wales staff.

6 Discussion and conclusions

Table 2 provides a useful roadmap of where participants felt improvements could be made to the AI Mentor in its next iteration, as well as where they saw it fitting within a training offer.

We noted some differences in views between those with more involvement and familiarity with AI and digital, and those who described themselves as novice. Those with existing knowledge were more likely to readily appreciate the concept and see the usefulness of the AI Mentor, with some very positive views expressed. On the other hand, those who were less familiar questioned the benefit of using the AI Mentor over other ways of learning (e.g. intranet, books, podcasts) and doubted that they, or their colleagues, would be likely to use it. Both novice and more experienced users questioned the AI Mentor's distinctiveness from ChatGPT and other forms of AI.

Both findings speak to the need to clarify the value proposition for the AI Mentor, particularly for those with less exposure and enthusiasm for AI to 'win hearts and minds.' As Interviewee 50 suggested this may take the form of 'sales and positioning' information, and several signalled the need for further context and information within the tool itself. Some highlighted the need for the AI Mentor to sit within a suite of approaches designed to raise AI and digital literacy overall, using a combination of approaches. Participants felt that building digital and AI literacy would need to be a prerequisite for certain groups of staff. Linked to this, Think Aloud interviews revealed that it was not immediately clear to participants how to begin interacting with the tool, or what its purpose was, suggesting further help to orientate staff to the aims and process for use.

The issue of accessibility was important, especially for those with less familiarity or confidence with AI. Participants suggested that terms related to NHS archetypes may

be too technical and that simplifying questions and reducing information regarding these may make engagement easier. Others felt that offering information in different formats would be important.

Life Sciences Hub Wales may also need to consider whether additional limits (and instructions) are needed around the type of information that can be returned (or requested) by the AI Mentor (e.g. whether it is appropriate for the AI Mentor to return information concerning a medical procedure or a treatment plan, as was observed in some interviews). It may also be worth considering the role of AI Mentor in generating outputs for participants (e.g. policies or training plans); participants tended to be impressed by what the AI Mentor returned, and one participant suggested that using the AI Mentor might help people to develop skills in refining prompts and questioning to achieve required outputs.

7 Recommendations

Based on the findings above, and the observations of the evaluation team, the following actions are recommended to consider for the next iterations of Healthcare AI Mentor.

RECOMMENDATIONS	ACTIONS TO CONSIDER			
Reinforce value proposition	Include further information about aims and how this differs from other AI tools	Complement with communications to position the offer and encourage uptake		
Enhance accessibility	Use of plain English	Provide alternative formats and strategies to keep people engaged with content	Acknowledge Welsh context	
Assurance and safe use	Include AI rules of engagement/ disclaimer	Link to organisational AI policy and SOP	Enhance referencing of sources	Consider balance between guidance and outputs – do parameters need adjusting to prevent certain outputs?
Further testing	Test with different audiences to ensure language keeps people engaged	Include feedback mechanism to flag any inaccuracies		
Place in training offer	To sit as part of a suite of options, rather than mandatory training	Consider where preparatory digital and AI literacy skills may be needed		