





RESEARCH ARTICLE

Exploring the role of ChatGPT in rapid intervention text development [version 1; peer review: awaiting peer review]

Hannah Bowers ^{1,2}, Cynthia Ochieng², Sarah E Bennett², Sarah Denford², Milly Johnston ³, Lucy Yardley^{2,4}

¹Primary Care Research Centre, University of Southampton, Southampton, SO15 5ST, UK

²School of Psychological Science, University of Bristol, Bristol, BS8 1TU, UK

³Bristol Medical School, University of Bristol, Bristol, BS8 2BN, UK

⁴School of Psychology, University of Southampton, Southampton, SO17 1BJ, UK

V1 First published: 23 Oct 2023, 12:1395
<https://doi.org/10.12688/f1000research.140708.1>

Latest published: 23 Oct 2023, 12:1395
<https://doi.org/10.12688/f1000research.140708.1>

Abstract

Background

There have been successful applications of AI to answering health-related questions, which suggests a potential role for AI in assisting with development of intervention text. This paper explores how ChatGPT might be used to support the rapid development of intervention text.

Methods

Three case studies are presented. In the first case study, ChatGPT (using GPT-4) was asked to generate sleep advice for adolescents. In case study two, ChatGPT (using GPT-3) was asked to optimise advice for people experiencing homelessness on staying hydrated in extreme heat. Case study three asked ChatGPT using GPT-3 and GPT-4 to optimise an information sheet for participation in a study developing an intervention for maternal blood pressure. Outputs were evaluated by the researchers who developed the text, and in case studies two and three were shown to public and patient contributors for feedback.

Results

ChatGPT was able to generate informative advice about sleep in case study one and was able to accurately summarise information in case studies two and three. In all three cases, elements or aspects were omitted that were included in the researcher-generated text that was

Open Peer Review

Approval Status *AWAITING PEER REVIEW*

Any reports and responses or comments on the article can be found at the end of the article.

based on behaviour change theory, evidence and input from public and patient contributors. However, in case study three, feedback from public contributors suggested ChatGPTs outputs were preferred to the original, although the outputs omitted information and were not at the requested accessible reading level.

Conclusions

ChatGPT was able to accurately generate and summarise health information. However, this information typically excluded core behaviour change techniques and was sometimes inappropriate for the target users. There is likely to be a valuable role for generative AI in the intervention development process, but this will need to be combined with detailed scrutiny and input from researchers and public contributors.

Keywords

ChatGPT, Intervention development, AI, behaviour change

Corresponding author: Hannah Bowers (hannah.bowers@bristol.ac.uk)

Author roles: **Bowers H:** Conceptualization, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Ochieng C:** Conceptualization, Investigation, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Bennett SE:** Conceptualization, Investigation, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Denford S:** Conceptualization, Investigation, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Johnston M:** Conceptualization, Methodology, Writing – Review & Editing; **Yardley L:** Conceptualization, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2023 Bowers H *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Bowers H, Ochieng C, Bennett SE *et al.* **Exploring the role of ChatGPT in rapid intervention text development [version 1; peer review: awaiting peer review]** F1000Research 2023, 12:1395 <https://doi.org/10.12688/f1000research.140708.1>

First published: 23 Oct 2023, 12:1395 <https://doi.org/10.12688/f1000research.140708.1>

Introduction

ChatGPT is an artificial intelligence (AI) chatbot, launched by OpenAI in November 2022. It is capable of generating responses to prompts or questions, input by the user, based on OpenAI's Generative Pre-Trained Transformer (GPT) language model (launched in 2018) (chat.openai.com). Since its launch it has gained significant attention and discussion of the ways in which it might be used across various sectors.

It has been suggested that generative AI, such as ChatGPT, can be used in the generation of content with the purpose of providing information on public health issues, answering questions about health promotion and disease prevention strategies, explaining the role of community health workers and health educators, discussing the impact of social and environmental factors on community health, and providing information about community health programs and services ([Biswas, 2023](#)).

The potential of generative AI to support healthcare has been investigated in a range of topic areas. In a review of 60 papers, there was evidence that ChatGPT could be beneficial in supporting scientific writing, research activity (*e.g.* analysis of large datasets), medical education and healthcare practice ([Sallam, 2023](#)). With regards to writing, there were reported benefits in the speed of conducting literature reviews and improved language to communicate ideas, providing information that is easily accessible and understandable. In another systematic review of 31 studies, ChatGPT was effective at generating information on a range of health topics ([Muftić *et al.*, 2023](#)). However both reviews highlight potential limitations such as generation of inaccurate information, and limited knowledge on topics or research after 2021. In some cases ChatGPT produced overly detailed content, but this may be addressed by using clear and specific prompts from the user about the desired output ([Sallam, 2023](#)).

There is mixed evidence on the quality of health-related information generated by generative AI (with the majority of research using ChatGPT). There is some evidence that AI-generated responses from ChatGPT are outperforming human responses in terms of their quality and empathy. [Ayers *et al.* \(2023\)](#) asked licensed healthcare professionals to evaluate physician responses and AI responses to health questions from a public social media forum. Responses were rated in terms of the quality of the information and the empathy provided. AI responses were preferred 78.6% of the time and were rated significantly better than physician responses with regards to quality and empathy, suggesting AI may be useful in generating public-facing health information. [Grünebaum *et al.* \(2023\)](#) reported a high level of accuracy in answering questions about obstetrics and gynaecology, while [Barat *et al.* \(2023\)](#) report only a 40% accuracy rating for responses to questions about interventional radiology.

Fewer studies have explored the utility in generative AI summarising and optimising text. ChatGPT was effective in summarising text based on palliative critical care scenarios written by human experts ([Almazyad *et al.*, 2023](#)). [Casella *et al.* \(2023\)](#) argue that while ChatGPT can generate and summarise public health text effectively, this is dependent on the information the researcher inputs, and it did not always adhere to strict word limits.

The Coronavirus disease 2019 (COVID-19) pandemic has highlighted the need to extend existing best practice in intervention development so that interventions can be co-developed and evaluated quickly in response to a rapidly-changing public health concern ([Yardley *et al.*, 2023](#)). In an effort to address this issue, the Agile Co-Production and Evaluation (ACE) Framework was proposed ([Yardley *et al.*, 2023](#)). The framework is intended to provide a focus for investigating new ways of rapidly developing effective interventions by combining co-production methods with large-scale testing and evaluation. This framework builds upon the Person-Based Approach (PBA) to intervention development ([Yardley *et al.*, 2015](#)). This approach to intervention development is grounded in the idea that co-production with an intervention's intended users is vital in ensuring the techniques and messaging are appropriate and more likely to be effective.

In this paper, we consider and explore the potential of generative AI (using ChatGPT) in supporting the rapid development and optimisation of interventions. We present three case studies in which ChatGPT was used to support intervention development. This includes using ChatGPT to generate information for a health intervention, in order to evaluate its ability to rapidly produce intervention materials. Following this, we explore the ability of ChatGPT to optimise text that was generated by researchers involved in intervention development. In these two latter case studies, we asked patient and public involvement (PPI) contributors for their views on the outputs generated by ChatGPT.

Methods

In a series of case studies, we asked ChatGPT to generate or simplify text relating to health interventions. Text to be generated or simplified for these case studies was selected from studies being undertaken by the Behavioural Interventions Group. These studies were in the process of developing public-facing intervention content using the PBA or ACE

(Yardley *et al.*, 2015, 2023). This approach to intervention development ensures the developed intervention is grounded in theory, evidence and views of the intended users. It does this through the use of theoretical mapping (*i.e.* mapping techniques and ideas against behaviour change and psychological theories), ‘guiding principles’ for how the intervention should be developed, and a logic model to hypothesise how the intervention may work.

In the first case study, ChatGPT was asked to generate intervention content to improve sleep in teenagers (‘Sleep Solved’). We constructed a prompt for ChatGPT: “How would you improve sleep in teenagers? Write your answer in simple language for someone with a reading age of 9 and use bullet points.” We then evaluated the effects of providing a more elaborate request, to encourage ChatGPT to use the ‘programme theory’ (Skivington *et al.*, 2021) for the intervention to guide the text development in the same way as the researchers were doing. The elements of the programme theory that ChatGPT was additionally asked to follow were the ‘guiding principles’ (Pollet *et al.*, 2020): to avoid being patronising or paternalistic; to use question-answer formats to make the advice accessible and engaging; and to make the advice convincing by giving the scientific explanations for advice, but keeping the scientific explanations very brief and simple. In accordance with the behavioural theory informing the content of Sleep Solved, ChatGPT was also asked to base the advice on Bandura’s Social Cognitive Theory (Bandura, 1977), with the aim of building users’ self-efficacy (confidence) that they can follow the advice, and positive outcome expectations regarding the consequences of following the advice.

In the remaining case studies, ChatGPT was asked to optimise text. In case study two, the text was written by a behavioural scientist from the Behavioural Interventions Group who was involved in the development of an intervention to support people experiencing homelessness to stay safe during extreme weather events. In case study three, the text was from an information sheet used in the recruitment of study participants in research conducted by the Behavioural Interventions Group, aiming to co-produce an intervention for post-partum blood pressure management. This text was produced by a study team outside of the Behavioural Interventions Group.

In order to maintain consistency for comparison across the case studies, ChatGPT was provided with the same wording in all three case studies to request accessible text: “make this text suitable for a reading age of 9 and use bullet points”. Each identically-worded request (with different example text to be simplified) was conducted twice, each output compared, and the best output was selected by the researcher.

OpenAI has recently upgraded ChatGPT by using the GPT-4 model. This requires a paid subscription to ChatGPT Plus. Case study two used ChatGPT using GPT-3, whereas case study one and case study three were conducted in ChatGPT using both GPT-3 and GPT-4. In these two case studies, identical prompts were input into the two versions of ChatGPT (*i.e.* ChatGPT using GPT-3 and ChatGPT using GPT-4) and outputs from each version of ChatGPT were compared.

The outputs from all case studies were evaluated by the researchers developing the intervention text through reflecting on the output compared with the guiding principles, planning tables and text written by the researcher. In case studies two and three feedback was also sought from PPI members of the team and service providers. PPI members were invited to comment on the text produced by ChatGPT as part of their ongoing role within the research teams developing the interventions. Ethical approval is not required for the involvement of public and patient contributors in research.

Results

Case study one: How well can ChatGPT generate intervention text?

As part of a separate study, members of the Behavioural Interventions Group were in the process of developing Sleep Solved, an app designed to improve sleep in teenagers. In the current case study, we explored whether ChatGPT could generate simple advice to be used within this app. Text generated by ChatGPT using GPT-3 in response to our initial, simple request was compared to text generated using ChatGPT using GPT-4 (see ‘Case study one data and review of ChatGPT output 05.09.2023.docx’ (Bowers *et al.*, 2023)) in response to our theory-based request. Both GPT models provided useful advice that mapped onto many elements of the researcher-generated text in the Sleep Solved app. For example, advice to wake up at the same time each day, to create a sleep-friendly environment and avoid phone use before bed were all directly comparable to Sleep Solved (see ‘Case study one data and review of ChatGPT output 05.09.2023.docx’ (Bowers *et al.*, 2023)). The text using ChatGPT using GPT-4 in response to the theory-based request offered more detail, and some reasoning for why the advice was offered. For example, for the ‘sleep-friendly advice’ section, ChatGPT using GPT-3 gave a list of ideas for making the bedroom more calming, whereas ChatGPT on GPT-4 explained why these were helpful.

Despite the considerable overlap in advice content, there were aspects of ChatGPT’s text that did not map onto the intervention guiding principles or logic model (Yardley *et al.*, 2015) (see ‘Case study one data and review of ChatGPT output 05.09.2023.docx’ (Bowers *et al.*, 2023)). The logic model in this PBA intervention was based on Bandura’s Social

Cognitive Theory (Bandura, 1998). ChatGPT was asked to define this theory, and was able to do this. Despite this, when ChatGPT was asked to draw on this theory, there was no evidence in the output that it had done so. While ChatGPT did give reasons for some of the advice it provided when asked to do so, the rationale it gave was less in-depth than in Sleep Solved. For example, the ChatGPT rationale for maintaining a regular sleep routine was ‘A regular sleep routine helps regulate your body’s internal clock, making it easier to fall asleep and wake up’. Sleep Solved was designed to explain the neuroscience behind the recommended sleep behaviours in an accessible way, and therefore described how cortisol affects sleep and is affected by the sleep cycle, light and caffeine. For example, “*You have a get up and go hormone – cortisol- which helps you feel alert. Getting up at the same time every day trails your brain to release cortisol at the right time.*”, and, “*... If you sleep late, your cortisol levels peak late. This may make you feel tired and sleepy when you have to get up earlier. It might also make you feel hungry, worried or low.*”, “*Try to avoid bright lights such as phone screens. Bright light can make your brain think it is time to wake up!*”. ChatGPT was asked to follow the research team’s guiding principles and logic model by using scientific explanations to convince young people of the positive outcomes of recommended behaviours rather than undermining their autonomy by providing paternalistic advice. ChatGPT often did not follow these instructions, for example simply advising ‘Watch what you eat before bed; Avoid heavy or spicy meals close to bedtime’ – whereas Sleep Solved gave scientific explanations for this advice; “*Eating foods high in sugar close to bedtime causes your brain to release hormones that wake you up, like cortisol and adrenaline. Spicy and fatty foods are harder to digest which can keep you awake too*”. Sleep Solved also provided examples of alternative snacks. In addition, Sleep Solved has several functions to improve users self-efficacy, which do not feature in ChatGPT’s advice (these elements are illustrated further in ‘Case study one data and review of ChatGPT output 05.09.2023.docx’ (Bowers *et al.*, 2023)).

While much of the advice from ChatGPT was relevant to sleep hygiene recommendations for young people, the ChatGPT text did not apply the behaviour change techniques employed in a co-produced intervention. Based on this, the decision was made to explore the ability of ChatGPT to simplify complex language in further case studies, making intervention text simple, accessible and engaging for people from diverse backgrounds.

Case study two: How well can ChatGPT optimize content for an intervention?

Intervention development

People experiencing homelessness (PEH) are considered to be at particularly high risk during extreme weather events, and even more so if they are using drugs and alcohol. Therefore, there is a need to co-create interventions (including advice) to help PEH stay safe during adverse weather events. This case study describes the findings of developing one component of an intervention that focuses on supporting people to keep hydrated during extreme heat.

The initial message was written by a behavioural scientist with expertise in intervention development as part of a separate ongoing study (UK Health Security Agency, 2023) (see ‘Case study two data.docx’ (Bowers *et al.*, 2023)). The content of the message was informed by interviews with PEH, people delivering services to PEH, a review of the literature, and PPI/stakeholder feedback (using the PBA to intervention development (Yardley *et al.*, 2015)). The subsequent message focused on providing information for PEH about the importance of keeping hydrated with non-alcoholic beverages during extreme heat, recognising signs and symptoms of dehydration and heat stroke, encouraging PEH to identify where and when they can access water, and prompting people to carry water bottles to ensure hydration. As many PEH find it difficult to carry equipment such as water bottles, we included content acknowledging this and encouraged people to attempt to carry water for short durations only.

ChatGPT outputs

We asked ChatGPT (using GPT-3) to make the text suitable for a reading age of nine and use bullet points (see ‘Case study two data.docx’ (Bowers *et al.*, 2023)). In response, ChatGPT made the following notable changes. First, the AI-generated text was much shorter (150 words compared with the 300-word original). A definition of dehydration was introduced, but a lot of other detail was removed. In particular, the specific detail about the symptoms of dehydration, and information about symptoms that indicate when medical attention is needed was not included. The sentences encouraging people to find locations in which water is freely available, and carrying bottles only during heatwaves were also removed. In addition to reducing content, ChatGPT modified a key message about reducing alcohol consumption during extreme heat to abstaining from alcohol consumption during extreme weather.

PPI feedback

The two versions of advice were then presented to a PPI team comprising two PEH (one male aged over 50 and one female aged 39, both experiencing street homelessness) and a service provider for PEH. A third version of the message was then

co-produced (see 'Case study two data.docx' (Bowers *et al.*, 2023)). The brevity of the ChatGPT version was viewed positively, but the PPI team also thought that some of the specific detail that had been removed in the shorter version needed to be included. For example, as PEH reported experiencing symptoms similar to dehydration on a regular basis, the inclusion of specific information about symptoms indicating a need for medical attention was considered highly important. The focus on abstinence rather than reduction of alcohol was considered inappropriate and unrealistic. Other feedback focused on information that required modification or was missing from both versions. For example, people thought it important to state that alcohol makes you more dehydrated at the start of the message. The PPI team wanted information about what would happen if they did not get medical attention, and they wanted a map showing the locations where water is available. The PPI team also highlighted the need for an additional, complimentary intervention for service providers, in which detail is provided about how best to support PEH during extreme heat.

Case study three: How well can ChatGPT optimise a participant information sheet for research on intervention development?

Intervention development

High blood pressure during pregnancy can result in elevated cardio-vascular risks over time. Post-pregnancy, blood pressure can change rapidly, often requiring alterations in the patient's medication. In this case study, an intervention was being developed to aid patients to manage their blood pressure post-partum through regular self-checks and pre-planned medication alterations. An information sheet inviting patients and former patients to participate in the development phase of the study was drafted by a team of researchers and refined with other stakeholders including clinicians and two PPI members. This information sheet explained to potential participants that the study had developed an intervention to help patients manage their blood pressure after pregnancy. The information sheet further explained that the patients were being invited because of their experience with high blood pressure in pregnancy and that they would be asked to share their thoughts on the intervention developed. They were also reminded that their participation was voluntary and that if they agreed to participate, they would need to sign a consent form. We considered that the information sheet would require considerable further optimisation to make it more accessible to people with lower levels of literacy, and was therefore a suitable text to ask ChatGPT to simplify.

ChatGPT outputs

Chat GPT was utilised to attempt to optimise the patient information sheet for the study. Initially all the text from the information sheet (four pages) was pasted into ChatGPT using GPT-3. The response it generated was a short paragraph that was too brief to fully explain the study. ChatGPT (using GPT-3) also did not make the language in the sheet simpler and did not produce bullet points in successive iterations of the request (see 'Case study three data.docx' (Bowers *et al.*, 2023)).

The researcher then divided the information sheet into smaller sections (between 94-361 words long) and each section was entered into ChatGPT (using GPT-3) separately with the standard instruction to "make the text suitable for a reading age of 9 and use bullet points". In response to this, ChatGPT organised its responses in bullet points but again the reading level was far more complex than for a nine year old (Flesch-Kincaid grade level 10.9 *i.e.* 16-17 year old).

Since ChatGPT appeared to require short chunks of text, two sections of the information sheet ('Introduction to the Study' and 'Data Protection' sections) were copied into ChatGPT (using GPT-3) separately with the standard instruction *i.e.* to make the text suitable for a reading age of nine and use bullet points. The requests were repeated to compare iterations. When the text retained its original formatting (such as bullet points and line breaks), ChatGPT often did not generate responses with bullet points. However, when the text was pasted as simple unformatted text, it generated bullet points.

The text describing an introduction to the study was then entered into ChatGPT (using GPT-4) with the same prompt ("make the text suitable for a reading age of 9 and use bullet points") to compare the two versions. This was done twice and the best response (as rated by the researcher) used for comparison with the outputs from ChatGPT using GPT-3. GPT-4 had better results than GPT-3 with the first iteration being for a reading age 10-11 (Flesch-Kincaid grade level 5.5) and the output contained all the information. The second iteration on GPT-4 was for a reading age of 9-10 (Flesch-Kincaid grade level 4.9).

ChatGPT appeared to shorten text consistently, which could be useful for generating key points - particularly for more general subjects like data protection in research. ChatGPT using GPT-3 was not successful at making the text simple enough for a reading age of 9; the study introduction text was usually generated for a reading age of 15-17 while the data protection text generated was simpler at reading age 13-14 (Kincaid Flesch grade level 8.8). Additionally, it was not able

Table 1. Demographic characteristics for PPI contributors in Case Study Three.

	Sex	Ethnicity	Age	Socioeconomic deprivation	Employment	Education	Religion	Has a disability
PPI1	F	Black African	40	Living in area of high deprivation	Unemployed	GCSE	Christian	Yes
PPI2	F	Asian	39	Living in area of high deprivation	Unemployed	University degree	Muslim	No
PPI3	F	Asian	Unknown	Living in area of high deprivation	Unemployed	Unknown	Muslim	No
PPI4	F	Black British Caribbean	56	Living in area of high deprivation	Part-time employed	Bachelor's degree	Christian	No

to handle longer documents and the formatting had to be removed. ChatGPT using GPT-4 had better results with regards to the reading level and the information included.

PPI feedback

The original text from a participant information sheet and two simplified versions (one from ChatGPT using GPT-3 and one from ChatGPT using GPT-4) were shown to the study's PPI members (demographic characteristics are presented in Table 1). They found a clear difference between the original wording of the information and the simplified ChatGPT texts. All preferred the output from ChatGPT using GPT-3. Two PPI members sought clarification on an acronym that had been used to name the study regardless of whether the text had explained it. One PPI member said she preferred the output from ChatGPT using GPT-3, which she found to be the easiest to understand. Another PPI member found this same output easiest to read. She said she did not like having too many words on a leaflet and ChatGPT on GPT-3 offered the least words. All the PPI members agreed that the original text had too many words and technical terms and that ChatGPT using GPT-4 was better in that regard. Despite agreeing with the others, one PPI member explained that she preferred the simple headings (in question form) in the version from ChatGPT using GPT-4, but that overall she was happy with the version from ChatGPT using GPT-3.

Some other stakeholders (comprising a general physician (GP), obstetrician fellow, midwife and two researchers) were also shown the original text and asked to compare it with the ChatGPT versions. They all preferred the simplified versions from both ChatGPT using GPT-3 and ChatGPT using GPT-4. The bullet points in the ChatGPT versions were thought to make the text easier to read. The team also noticed that the language on the ChatGPT versions was simpler and hence more accessible than in the original text.

Discussion

Across three case studies, we have explored how ChatGPT might be used to support rapid intervention development. In our first case study, when prompted to produce advice on sleep for adolescents, ChatGPT's output undoubtedly had face validity and value; the advice it gave reflected scientific consensus and was succinct and accessible. Indeed, a fourth case study that was planned for this paper was not pursued because the intervention was being developed alongside rather than before this study, and the developer found the text suggested by ChatGPT so helpful that she was unable to create a version that was not influenced by it. However, the ChatGPT text did not closely follow the 'guiding principles' for how to make the text engaging that it was asked to apply, and did not include any of the key behaviour change techniques that were developed as part of the co-production process using the PBA. In our second case study, key pieces of advice that were generated as a result of co-production with intended users of the intervention were omitted. In this case study, members of the public (specifically, people experiencing homelessness) provided feedback on ChatGPT's output and highlighted that these omissions needed to be included in the advice. Our third case study showed that, when prompted to optimise an entire participant information sheet, ChatGPT (using GPT-3) omitted information that would be necessary to include and the reading level of the text was higher than requested by the prompt. However, when text was entered in smaller chunks without bullet points, the output was improved and the text was considered superior to the original by both PPI and the research team.

We have learned from our case studies that despite the speed with which ChatGPT can provide text, using ChatGPT can be surprisingly time-consuming. Texts requiring optimisation may need preparation prior to being input into ChatGPT (using GPT-4), for example by dividing into short simple sections and simplifying technical terms. ChatGPT using

Table 2. Learning points for future use of generative AI to assist with rapid intervention development.

Benefits	Limitations
AI can be used to very rapidly generate consensus-based advice that summarises current thinking on a topic	<ul style="list-style-type: none"> • All AI-generated advice must be checked by topic experts to ensure that it is correct • If the advice is intended to produce behaviour change then input from behavioural scientists may be needed to add behaviour change techniques to support implementation of the advice • Because AI-generated advice is based on widely available text, additional design input may be needed to generate text and images that is sufficiently novel and interesting to engage users • Because AI learns from widely available text, it may struggle with rarer research or topics
AI can be useful for rapidly optimising text for different target reader groups, such as those with a lower reading age or who need translations	<ul style="list-style-type: none"> • AI-generated text may not be sufficiently sensitive to the contexts of seldom heard groups. Co-production with all intended user groups is therefore essential to ensure that text is appropriate for them
AI may be helpful to assist less experienced intervention text developers to generate accessible, readable content	<ul style="list-style-type: none"> • Both the AI and the researchers are likely to require training and significant input of time to create appropriate, engaging content • Time is also required to check the accuracy and validity of AI generated text, which may not be accurate
AI is able to provide 'plain English' definitions for complex terms or theories	<ul style="list-style-type: none"> • AI may struggle with theoretical concepts and abstract principles. While it can provide a definition, the algorithm is not currently advanced enough to incorporate theory into the content it produces

GPT-4 seemed to perform better, and outputs could be improved further by responding to ChatGPT's outputs with further requests for changes and by giving examples of the required tone and suitable outputs. For instance, generative AI (including ChatGPT) used in health intervention messaging could be prompted to simplify commonly used terms and explain common concepts (e.g. randomisation, anonymisation or informed consent). We faced a steep learning curve in our use of ChatGPT and it is likely that in future researchers will also need to be trained to use generative AI more effectively. Although beyond the scope of this study, it is likely that generative AI can contribute to public health intervention development in other ways. For example, it could be used to rapidly summarise and respond to public reactions to an emergency on social media, and to translate intervention content into multiple languages.

Despite the potential of generative AI to support rapid response during public health emergencies, in its current form, there are significant limitations (see Table 2). Importantly, there is evidence that ChatGPT has problems with accuracy in generating public health information (Biswas, 2023; Jungwirth & Haluza, 2023; Muftić *et al.*, 2023; Sallam, 2023). Furthermore, the use of ChatGPT for arising public health issues is limited due to its language model being trained using a dataset with a cut-off date in 2021 (Jungwirth & Haluza, 2023; Muftić *et al.*, 2023; Sallam, 2023) meaning it is not able to provide up-to-date information about novel public health concerns. In contrast to previous studies (e.g. Muftić *et al.*, 2023; Sallam, 2023), in our three case studies there were no examples of inaccuracies in the text generated by ChatGPT. However the quality of the output (with regards to formatting and language used) was related to the quality of the prompt (Casella *et al.*, 2023).

Conclusions

ChatGPT (particularly ChatGPT using GPT-4) can produce succinct summaries of text, when this text is prepared by a researcher and entered into ChatGPT in the appropriate format with appropriate context and requests. However, ChatGPT tends to miss key elements that address subtle but important barriers to behaviour change identified through collaboration with intended users and grounded in theory and evidence. While ChatGPT is able to rapidly summarise text to be informative and succinct, it may omit key content grounded in behaviour change theory, which may affect its ability to change behaviour (Yardley *et al.*, 2015). A further problem with AI-generated text is that interventions need to offer users advice and support that seems somewhat novel or unique in order to promote engagement. Because interventions produced using generative AI are based on widely available texts they are unlikely to appear innovative and uniquely informative. Moreover, since the outputs from generative AI are inevitably generalist rather than context-specific there is a risk that they will not be well-adapted to the context of under-represented groups. The outputs of generative AI may thus be able to assist with 'agile' content generation but the 'co-production' element of the ACE framework will remain crucial to ensure that content is 'relatable' and appropriate for the intended users (Yardley *et al.*, 2023).

In summary, generative AI is a novel tool that may be useful in rapid intervention development. There is however a learning curve and as we are learning how to use it; it too is learning how best to respond.

Data availability

Underlying data

Repository name: FigShare.

Project title: Input and output for Exploring the role of ChatGPT in rapid intervention text development, DOI <https://doi.org/10.6084/m9.figshare.24084399.v1> (Bowers *et al.*, 2023)

This project contains the following underlying data:

- Case study one data and review of ChatGPT output 05.09.2023.docx (ChatGPT outputs from Case Study One)
- Case study two data.docx (ChatGPT outputs from Case Study Two)
- Case study three data.docx (ChatGPT outputs from Case Study Three)

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/) (CC-BY 4.0).

Acknowledgements

We would like to thank Mr Joe Shervell, Sabroso Ltd., for providing access to ChatGPT Plus and advising on the use of ChatGPT. We would also like to thank the PPI contributors for their comments in case studies two and three. LY is an NIHR Senior Investigator and her research programme is partly supported by NIHR Applied Research Collaboration (ARC)-West and NIHR Health Protection Research Unit (HPRU) for Behavioural Science and Evaluation.

References

- Almazayd M, Aljofan F, Abouammoh NA, *et al.*: **Enhancing Expert Panel Discussions in Pediatric Palliative Care: Innovative Scenario Development and Summarization With ChatGPT-4.** *Cureus.* 2023; **15**: e38249.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ayers JW, Poliak A, Dredze M, *et al.*: **Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum.** *JAMA Intern. Med.* 2023; **183**: 589–596.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bandura A: **Self-efficacy: Toward a unifying theory of behavioral change.** *Psychol. Rev.* 1977; **84**(2): 191–215.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Bandura A: **Health Promotion from the Perspective of Social Cognitive Theory.** 1998.
- Barat M, Soyer P, Dohan A: **Appropriateness of Recommendations Provided by ChatGPT to Interventional Radiologists.** *Can. Assoc. Radiol. J.* 2023; 084653712311701.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Biswas SS: **Role of Chat GPT in Public Health.** *Annals of Biomedical Engineering.* Springer; 2023.
[Publisher Full Text](#)
- Bowers H, Ochieng C, Bennett SE, *et al.*: **Input and output for Exploring the role of ChatGPT in rapid intervention text development.** figshare. Dataset. 2023.
[Publisher Full Text](#)
- Cascella M, Montomoli J, Bellini V, *et al.*: **Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios.** *J. Med. Syst.* 2023; **47**(1): 33.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Grünebaum A, Chervenak J, Pollet SL, *et al.*: **The Exciting Potential for ChatGPT in Obstetrics and Gynecology.** *Am. J. Obstet. Gynecol.* 2023; **228**: 696–705.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Jungwirth D, Haluza D: **Artificial Intelligence and Public Health: An Exploratory Study.** *Int. J. Environ. Res. Public Health.* 2023; **20**(5).
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Muftić F, Kadunić M, Mušibegović A, *et al.*: **Southeast Europe Journal of Soft Computing Exploring Medical Breakthroughs: A Systematic Review of ChatGPT Applications in Healthcare.** 2023; **12**(1).
- Pollet S, Denison-Day J, Bradbury K, *et al.*: **A Qualitative Exploration of Perceptions of a Digital Intervention to Promote Physical Activity in Older Adults.** *J. Aging Phys. Act.* 2020; **29**(3): 442–454.
[Publisher Full Text](#)
- Sallam M: **ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns.** *Healthcare (Basel, Switzerland).* 2023; **11**(6).
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Skivington K, Matthews L, Simpson SA, *et al.*: **A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance.** *BMJ.* 2021; **374**.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- UK Health Security Agency: **Supporting vulnerable people before and during hot weather: healthcare professionals - GOV.UK.** 2023.
[Reference Source](#)
- Yardley L, Denford S, Kamal A, *et al.*: **The Agile Co-production and Evaluation (ACE) framework for developing public health interventions, messaging and guidance.** 2023.
[Reference Source](#)
- Yardley L, Morrison L, Bradbury K, *et al.*: **The person-based approach to intervention development: Application to digital health-related behavior change interventions.** *J. Med. Internet Res.* 2015; **17**(1): e30.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research