Good-Bot Bad-Bot: Addressing bias in chatbots and public health information

Natarajan Ganesan[1] and Thanemozhi G. Natarajan[1]

[1]Queromatics

July 26, 2023

Abstract

The proliferation of chatbots in recent months has raised concerns about the potential for bias in these conversations. This is especially true when it comes to public health topics, where accurate and unbiased information is essential. Herein, the focus is on the issue of bias in chatbot conversations using oral contraceptive pills (OCPs) as an example. By raising awareness about this issue and emphasizing the need for critical evaluation, we can empower individuals to navigate the digital landscape with confidence and make informed decisions about their health.

Introduction

The internet has become the go-to source for health information for many individuals, with search engines like Google being the starting point. However, search results often lead to misleading or inaccurate content. To address this issue, conversational agents like chatbots are emerging as an alternative source of health information. Popular chatbots like ChatGPT and Google's Bard promise to provide trustworthy and unbiased information on any topic through natural conversations (Fig 1).

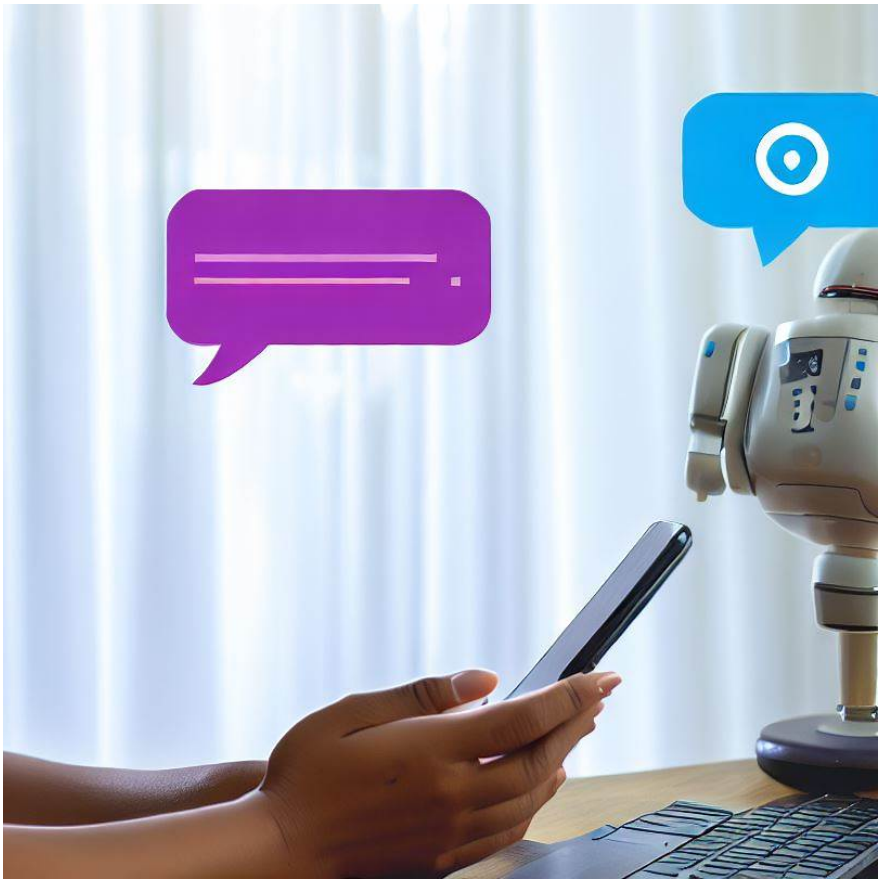

Figure 1: Representative image of a user chatting with a virtual assistant guided by an AI-based chatbot. Image source: Bing image creator

However, these chatbots can exhibit biases that shape the health information they provide. Take oral contraceptives (birth control pills) for example. When asked about the pill, chatbots often provide information that skews towards the benefits like pregnancy prevention, while minimizing discussion of potential side effects. This presents a limited perspective on oral contraceptives. The algorithms driving chatbots are trained on available data, which suffers from reporting and publication biases that accentuate benefits over harms. So chatbots end up perpetuating these biases.

Providing comprehensive, balanced information is vital for truly informed decision-making about health. Biased information from chatbots can steer choices in a particular direction, often aligned with business interests rather than public health goals.

To counter such biases, chatbots can employ oversight from experts to ensure balance in the information provided. Guidelines can be issued urging chatbot creators to minimize biases through training data selection and algorithm tweaking. For example, they can be trained to request the user if they need a more comprehensive view in case a request is pointed in one direction.

Finally, educating people to approach chatbots critically rather than blindly trusting their guidance is key. Just like with human experts, examining chatbot recommendations against alternate credible sources

allows for balanced perspectives.

In an evolving digital health landscape, chatbots hold promise in improving access to information. But thoughtfully addressing their limitations is crucial so these tools empower rather than inadvertently mislead people in making health choices aligned with their needs. Openness to oversight and continual learning will allow chatbots to better serve individuals and the public health good.

Methods: Understanding inherent bias in popular chatbots

If Internet search engines utilize complex algorithms to deliver their search results then chatbots like ChatGPT and Bard are even more so and at an advanced level. Concepts like Artificial Neural Networks (ANN) and Natural Language Processing (NLP) are just the surface of it. For example, Fig 2 represents a schematic of an affective conversation where the emotion depends on the context. The health assistant understands the *affective state* of the user in order to generate *effective and empathetic* responses.

To understand a chat was initiated with ChatGPT, a popular chatbot available at https://chat.openai.com/. At this point, one must understand that all conversations are not the same; hence, responses to the same question posed by other users can evoke different answers. Though this approach personalizes the answers to suit each user and their inherent 'intent', the overall objectivity of the answers provided can vary widely. It is entirely possible that responses from AI are in the auto-learning process and keep adapting as the number of users asking the question changes. This makes the process very personalized though not necessarily uniformly objective.

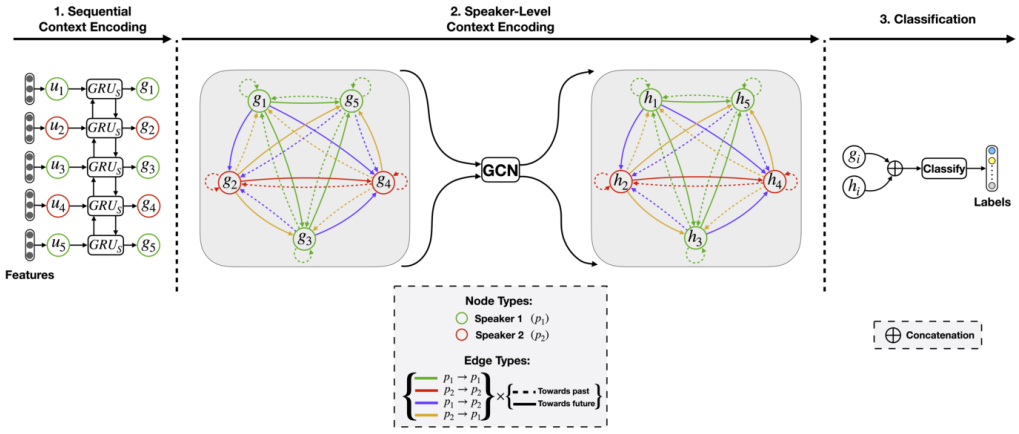

Figure 2: Illustration of an 'affective' conversation where the emotion depends on the context. Health assistant understands 'affective' state of the user in order to generate 'affective' and 'empathetic' responses. Image source: Wikipedia. Original source

Results: Exploring slants in a directed inquiry-based conversation

To further illustrate the biases, let us delve deeper into the example of my recent chat with ChatGPT regarding the FDA approval of an oral contraceptive pill containing Progestin. While major media outlets covered this news in a predictable way, curiosity prompted us to investigate the possible health risks

associated with hormone-based pills. Recognizing the hormonal nature of such contraceptives, it was reasonable to anticipate the existence of risks and seek comprehensive information on the topic. To be fair though, the approach of using oral contraceptives has been in vogue for decades and has proved helpful in supporting women and their reproductive health.

However, as we embarked on a casual chat, we quickly discovered that the information presented was far from comprehensive or unbiased. The chat grew increasingly in favor of the use of the approved pills while I was seeking information, in particular, about the risks associated with its use. Even pointed requests to provide links from Pubmed failed to give satisfactory results. When we countered it by providing it a copy of an abstract text from a very good recent review paper (that itself was a result of many meta-analyses and papers), it evaluated the paper well while still being defensive about what it said earlier. Finally, it had no choice but to accept there are different sides to the issue as well.

The societal implications, politics, and ensuing interests surrounding birth control contribute to an inherent imbalance in the available literature. This imbalance can result in a lack of representation of all sides of the issue, hindering individuals' ability to access a diverse range of viewpoints and evidence.

In this particular case, the push to promote the use of oral contraceptive pills, driven by factors such as gender equality, reproductive rights, and public health initiatives, can influence the information that surfaces in search results. As a consequence, the chat algorithms may prioritize sources that align with the prevailing narrative, emphasizing the benefits and downplaying potential risks associated with hormonal contraceptives. This can inadvertently lead to an incomplete and skewed understanding of the topic, as critical perspectives and studies highlighting the risks may be overshadowed or marginalized.

These biases can be further compounded by political and allied interests that seek to shape the discourse surrounding birth control. Various stakeholders could attempt to manipulate search results, either directly or indirectly, to direct the users toward their agendas. As a result, individuals increasingly relying on chatbot conversations may struggle to access well-rounded and unbiased information about the potential health risks associated with oral contraceptive pills.

This imbalance in the available literature underscores the importance of critically evaluating information obtained through these chatbots. It highlights the need for individuals to be aware of the biases that can be inherent in search results and to actively seek out diverse sources of information. By consulting reputable scientific journals, academic research databases, and trusted healthcare resources, individuals can obtain a more comprehensive understanding of the risks and benefits associated with oral contraceptive pills.

Moreover, this example demonstrates the limitations of relying solely on internet searches for accessing nuanced information on public health topics. It emphasizes the significance of seeking guidance from healthcare professionals who possess the expertise to navigate and interpret scientific literature objectively. Engaging in open and informed discussions with healthcare providers allows individuals to receive personalized advice, address specific concerns, and obtain a more holistic view of the risks and benefits of oral contraceptive pills.

Discussion

Chatbots operate based on sophisticated algorithms that analyze user queries and generate responses. However, these algorithms are not immune to biases, as they are likely to be designed to prioritize certain

information sources and viewpoints. For example, in the case of oral contraceptives, the algorithm may favor sources that emphasize the benefits while downplaying or omitting information about potential risks associated with their use, especially when their use may be desired by public health agencies for the betterment of women and reproductive health. For example, looking at the increased risk of developing cancer over use for a long period of time especially in vulnerable demographics (certain ethnicity). This is based on a recent user experience. Each user experience may surely vary. Consequently, the chatbot may provide incomplete or skewed information, hindering individuals' ability to obtain a comprehensive understanding of the topic.

Moreover, the bias encountered in chatbot responses can hinder the retrieval of scientific papers and systematic reviews that delve into the potential risks associated with oral contraceptives. These studies may present nuanced findings, highlighting adverse effects, contraindications, or specific populations for whom caution is advised. However, due to the bias toward promoting contraceptive use, the chatbot may overlook or underrepresent such studies, limiting individuals' exposure to critical information. Various factors such as optimization techniques, sponsored content, and commercial interests can influence the visibility and ranking of information, potentially skewing the presentation of viewpoints. In the context of public health, biases can significantly impact the availability and accessibility of information related to oral contraceptive pills and their associated health risks.

In navigating information gleaned from such chatbots, it is imperative for individuals to exercise critical thinking and evaluate search results meticulously. A casual conversation with a chatbot may not always yield a balanced view, as the algorithms are likely to prioritize certain sources and perspectives. Furthermore, the "best interests" of the public, as determined by the algorithms (in turn determined by interests that be), may not align with providing comprehensive and unbiased information. It is essential to be aware of these limitations and actively seek out diverse sources of information.

Policy recommendations

Enhanced transparency and disclosure

- Advocate for makers of Chatbots like ChatGPT, Bard, etc. to provide more transparency regarding the factors influencing 'fact presentation' and visibility of information.
- Encourage Chatbot makers to disclose potential conflicts of interest, sponsorships, or biases that may impact search results.

Promoting Critical Health Literacy

- Advocate for the integration of critical evaluation and information literacy skills into chatbot interactions, empowering individuals to critically assess online health information.
- Collaborate in the development of educational campaigns and resources that educate the public about biases in chatbot responses and strategies for effectively navigating and evaluating the information provided by chatbots.

Collaboration between Public Health Experts and Tech Companies

- Foster partnerships between public health experts and technology companies to ensure the development of search algorithms that prioritize the presentation of balanced, evidence-based information.

- Engage in ongoing dialogue to address concerns related to search result biases and work towards optimizing the retrieval of reliable health information.

Conclusions

Bias in public health-related conversations with chatbots such as ChatGPT or Bard or any emerging ones poses significant challenges to individuals seeking accurate and comprehensive information. By acknowledging the existence of biases and actively addressing them, we can foster a digital landscape that enables individuals to make informed decisions about their health. Through policy recommendations such as enhanced transparency, promoting critical health literacy, and collaboration between public health experts and tech companies, we can mitigate the impact of bias and ensure equitable access to reliable information. Empowered by critical evaluation skills, individuals can navigate public health internet searches with confidence, unveiling hidden truths and making informed choices that contribute to their overall well-being.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used grammar-checking tools and generative AI in order to improve the readability and organization of the content. After using this tool/service, the authors reviewed and edited the content as needed, and take full responsibility for the content presented.

References

Deepanway Ghosal, Navonil Majumder, Soujanya Poria, Niyati Chhaya, and Alexander Gelbukh. DialogueGCN: A Graph Convolutional Neural Network for Emotion Recognition in Conversation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics, 2019. doi: 10.18653/v1/d19-1015. URL https://doi.org/10.18653%2Fv1%2Fd19-1015.