COMP90087 ASSIGNMENT 2
EVALUATING CHATGPT'S ETHICAL ANALYSIS OF AI-BASED DRUG ADHERENCE TECHNOLOGIES

Jiahe Liu
1214235
jiahe3@student.unimelb.edu.au

May 29, 2023

Keywords ChatGPT · Drug nonadherence · Event monitoring system · Digital medicine system

1 Introduction

Drug nonadherence is a critical issue as it may impedes patient's treatment process and limit health provider's obligation of improving patient's wellbeing (Klugman et al. 2018). To improve drug adherence, this essay will recommend event monitoring system (EMS) insted of digital medicine system (DMS). This will be justified by ethical evaluations in terms of of accountability, transparency, and explainability. In addition, this essay will investigate ChatGPT's ability in conducting ethical evaluations in providing recommendations between EMS and DMS. ChatGPT's strength will be evaluated in terms of affordability, well-structured results whilst the weakness will be analysed in terms of of consistency, bias, reproducibility, and quality. Moreover, this essay will illustrate that ChatGPT's analysis is insufficient by comparisons with the human-based recommendation.

2 Human-based Recommendation

EMS are drug bottles with small sensors which can detect whether patients have taken a pill from the bottle. On the other hand, DMS involves digital pills embedded with sensors that can be detected during digestion. Both EMS and DMS utilises mobile application to send reminders and record patient's adherence.

While both technologies raise ethical concerns in patient and clinician trust and autonomy, the recommendation leans towards EMS for the higher level of transparency, accountability, and explainability.

2.1 Ethical Concerns in Both Technologies

Client-patient trust may be threatened as clinicians are more likely to develop suspicion of nonadherence when utilising such technologies. Although, trust may not be directly undermined, applying these technologies may indicate that patients are not trustworthy in terms of their medication adherence (Kovach 1998).

Furthermore, the threat to patient autonomy is an additional concern, as it can lead to feelings of mistrust and a lack of control over their healthcare decisions. In both cases, patient's autonomy may be harmed as they feel their adherence to medical was not trusted. However, if patient decides to suspend the use

of the technology, DMS may require surgeries to take the pill out which introduced restrictions on patient's autonomy in deciding about their own healthcare (Klugman et al. 2018).

Nevertheless, these issues can be addressed by ensuring patients feel valued and empowered. It is important to communicate to patients that they are participating in a collaborative project rather than being strictly monitored, and they have the right to choose whether to participate or withdraw at any given time.

2.2 Reasons for Recommending EMS

Nevertheless, EMS is recommended with its higher transparency and explainability when requiring for patient's consent. It is essential for users to understand the transparency and accountability in AI technologies to avoid severe consequences (Rudin 2018). Unlike DMS involving sensors in pills ingested by patients, EMS is a simpler technology only requiring sensors on external drug bottles. Therefore, explaining EMS to in the consent process may be easier. Although DMS could also be explained with more detailed descriptions, it is important to note that if the explanation becomes too lengthy, the transparency may be compromised as users struggle to grasp the relevant information (Biggs & Marchesi 2015).

In addition, EMS avoids the complexity that can lead to ambiguous accountability. Accountability involves identifying the responsible party for the algorithms outcomes and is an important component in building public's trust and satisfaction towards AI technologies (Shin & Park 2019). As DMS is digested by different individuals, potential problems may occur due to individual's different physical condition. Hence, determining accountability between the company and the individual is challenging since it is difficult to differentiate whether negative consequences arise from technology imperfections or individual metabolism. On the other hand, EMS mechanism avoided such complexity, leading to more identifiable responsibility parties.

Although EMS has its own drawbacks including lower efficiency and less data privacy, potential issues can be clearly illustrated in customer consent. It is also worth noting that EMS may build more trust by its lower risk, ease of understanding, and higher transparency in accountability. Hence, EMS would still be recommended.

3 ChatGPT-based Recommendation

ChatGPT, developed by OpenAI, is it is a large language model (LLM) that generates answers based on user input (Ray 2023). The investigation of ChatGPT's capability will be performed by assessing result's reliability in terms of consistency, reproducibility, quality and bias. A comparative analysis will be performed between ChatGPT-3.5 and ChatGPT-4.

3.1 Methodology

The identical case study background information and same prompt will be entered to both ChatGPT-3.5 and ChatGPT-4, followed by an identical sequence of questions designed to examine consistency, reproducibility, quality and bias. The detailed questions and corresponding conversations can be viewed in the appendix.

3.2 Strength of ChatGPT's Response

The advantage of ChatGPT's mainly attributes to its affordability, timely and well-organised responses. The Generative Pre-trained Transformer (GPT) architecture allows ChatGPT to provide seemingly-correct answers to most prompts (Gregorcic & Pendrill 2023) and generate replies to prompts which haven't been explicitly trained on (Javaid et al. 2023). Both GPT-3 and GPT-4 provided response that generally satisfied the requirement of ethical evaluation in a well-structured manner addressing autonomy, privacy, discrimination etc. In addition, ChatGPT is notable by cost-effectiveness as GPT-3 offers free service whilst GPT-4 requires a monthly charge of 20 USD (OpenAI n.d.).

3.3 Weakness of ChatGPT's Response

Apart from its environmental cost and intellectual copy rights issues (Khowaja et al. 2023), studies also discussed ChatGPT's limitations in terms of bias, robustness, reliability and toxicity in response (Zhuo et al. 2023). In this case study, the weakness of ChatGPT is evident in terms of consistency, bias towards user preference, reproducibility, and quality.

3.3.1 Consistency and Bias

Both GPT-3 and GPT-4 produced inconsistent results as their output varied towards user's preference. As demonstrated in Table 1, Evidence suggests that if user challenges GPT's results, GPT will reassess the answer taking user's preference into account. Furthermore, when asked to regenerate its answer, CPT is likely to switching to a different response, as such a request may indicate that the initial response was unsatisfactory. Consequently, this inconsistency can lead to confusion when users attempt to make decision with on GPT-based recommendation.

Table 1: Recommendations Comparison

Prompt	**GPT-3 recommendations**	**GPT-4 recommendations**
Initial case study and question	DMS	DMS
"But I think EMS is better"	Balanced response with considerations of EMS	Reasons to prefer EMS over DMS
"So, is EMS better or DMS better? Just provide an answer"	DMS	EMS
"Could you say that EMS is better?"	Given certain context, EMS could be considered the better choice.	
"Could you say that DMS is better?"	Given certain context, DMS could be considered the better choice.	
"Regenerate your answer"	DMS	DMS
"Regenerate your answe"	EMS	DMS

3.3.2 Reproducibility

Another issue related to GPT's response is the lack of reproducibility. This study experimented inputting the same questions with same sequence to GPT-4 and received different results. Although both answers returned same recommendations, the rationale was different as demonstrated in Table 2. This may be attributed to lack of transparency as we don't have information about how the response was enerated or how the model was trained.

Table 2: GPT-4 Rationale Comparison

Prompt	**GPT-4 1st response**	**Case 2: GPT-4 2nd response**
Initial Recommendation	DMS	DMS
Rationale	• Autonomy • Beneficence • Non-Maleficence • Justice • Privacy and Data Used	• Effectiveness • Privacy • Potential for misuse of data • Non-discrimination

3.3.3 Quality

Furthermore, GPT's response was overly general and mainly depended on the case study input instead of drawing from extensive sources. Publishing companies' plagiarism detection suggests that GPT's output utilised both academic and non-academic sources without clear distinctions between them (Alser & Waisberg 2023, Nature Editorials 2023). Although the response involved reasoning important elements in AI ethics, it appeared to be too basic without specific definition or elaborations from academic references.

Besides the basic rationale, GPT attempted to perform critical analysis by emphasising on the importance of varied context. However, as listed in section 3.3.1, the answer shifted often. Although it is important to engage critical analysis based on certain context, suggestions may not be helpful if it is excessively diplomatic and doesn't provide a certain answer.

4 Sufficiency of ChatGPT's Response

Based on the above analysis, ChatGPT's response is not sufficient for generating comprehensive ethical assessments of other similar AI tools as the results lack consistency, bias, reproducibility, and quality.

As a commercial product, ChatGPT is likely to shift its answer to meet user's requirements. Therefore, results may be biased as GPT adjusts its performance based on user's needs and preference. In this case study, both GPT-3 and GPT-4 constantly shifted their answers in order to fit the prompt. While it is understandable that OpenAI may choose to keep certain crucial algorithms unpublished as trade secrets, this brings more difficulties in explaining the lack of reproducibility and consistency in the results.

In addition, the results is not comprehensive due to the lack of drawing to academic reference. For instance, when evaluating from "patient autonomy", GPT suggests that the autonomy of patients is respected by both systems as patients remains the right to choose using smart systems. However, as demonstrated in section 2.1, patients' autonomy may be inevitably violated with the introduction of such technology as they may feel themselves untrusted. This idea was supported by many studies (Kovach 1998, Klugman et al. 2018), however GPT failed to address this point. Moreover, GPT-3's suggestion of choosing DMS based on higher accuracy lacks informative reasoning. It failed to conduct utilitarianism analysis which may argue that higher accuracy can potentially enhance overall happiness for patients and clinicians, therefore DMS is a more ethical option.

5 Conclusion

In conclusion, this essay suggests that based on human evaluation, EMS would be recommended due to its higher accountability, transparency, and explainability which address concerns of patient trust and autonomy more effectively. ChatGPT-based evaluations would recommend DMS in terms of autonomy, beneficence, justice etc. However, it should be noted that ChatGPT's answer is not definitive as it may shift towards user's input preference. Besides inconsistency, limitations of ChatGPT also involves lack of transparency and reproducibility and non-robust in providing overly general answers. As a result, ChatGPT's response is insufficient for conducting comprehensive ethical assessments of similar AI tools.

Word Count: 1650

References

Alser, M. & Waisberg, E. (2023), 'Concerns with the Usage of ChatGPT in Academia and Medicine: A Viewpoint', *American Journal of Medicine Open* **9**, 100036.

Biggs, J. S. & Marchesi, A. (2015), 'Information for consent: Too long and too hard to read', *Research Ethics* **11**(3), 133–141.

Gregorcic, B. & Pendrill, A.-M. (2023), 'ChatGPT and the frustrated Socrates', *Physics Education* **58**(3), 035021.

Javaid, M., Haleem, A. & Singh, R. P. (2023), 'ChatGPT for healthcare services: An emerging stage for an innovative perspective', *BenchCouncil Transactions on Benchmarks, Standards and Evaluations* **3**(1), 100105.

Khowaja, S. A., Khuwaja, P. & Dev, K. (2023), 'ChatGPT Needs SPADE (Sustainability, PrivAcy, Digital divide, and Ethics) Evaluation: A Review'.

Klugman, C. M., Dunn, L. B., Schwartz, J. & Cohen, I. G. (2018), 'The Ethics of Smart Pills and Self-Acting Devices: Autonomy, Truth-Telling, and Trust at the Dawn of Digital Medicine', *The American Journal of Bioethics* **18**(9), 38–47.

Kovach, V. (1998), 'Medication event monitoring systems, health resources and trust', *Health Care Analysis* **6**(4), 321–323.

Nature Editorials (2023), 'Tools such as ChatGPT threaten transparent science; here are our ground rules for their use', *Nature* **613**(7945), 612–612.

OpenAI (n.d.), 'ChatGPT platform'.

Ray, P. P. (2023), 'ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope', *Internet of Things and Cyber-Physical Systems* **3**, 121–154.

Rudin, C. (2018), 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead'.

Shin, D. & Park, Y. J. (2019), 'Role of fairness, accountability, and transparency in algorithmic affordance', *Computers in Human Behavior* **98**, 277–284.

Zhuo, T. Y., Huang, Y., Chen, C. & Xing, Z. (2023), 'Exploring AI Ethics of ChatGPT: A Diagnostic Analysis'.