Sociotechnical Considerations for Accessibility and Equity in AI for Healthcare

Adriane Chapman
University of Southampton
Southampton, UK
adriane.chapman@soton.ac.uk

Chloe L Harrison
Adferiad Recovery
Colwyn Bay, UK
chloe.harrison@adferiad.org

Caroline Jones
Swansea University
Swansea, UK
caroline.jones@swansea.ac.uk

James Thornton
Nottingham Trent University
Nottingham, UK
james.thornton@ntu.ac.uk

Rose Worley
Swansea University
Swansea, UK
rose.worley@swansea.ac.uk

Jeremy C Wyatt
University of Southampton
Southampton, UK
J.C.Wyatt@soton.ac.uk

ABSTRACT

As AI systems are built and deployed to support mental health services, it is imperative to fully understand the stakeholder acceptability of such systems so that these concerns can be taken into account in system design. As such, we undertook a consultation with staff (therapists) and service-users at Adferiad Recovery (a large mental health charity). The aim was to capture insights about their understanding of trust, and different trust factors for AI in mental health care. Surveys, interviews and focus groups were conducted with service users and therapists. Key takeaways for computer scientists and the developers of AI systems are presented.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in collaborative and social computing**; • **Security and privacy** → *Human and societal aspects of security and privacy*; • **Computing methodologies** → Artificial intelligence.

KEYWORDS

AI, Mental Health, Accessibility, Equity

ACM Reference Format:
Adriane Chapman, Chloe L Harrison, Caroline Jones, James Thornton, Rose Worley, and Jeremy C Wyatt. 2024. Sociotechnical Considerations for Accessibility and Equity in AI for Healthcare. In *Companion Proceedings of the ACM Web Conference 2024 (WWW '24 Companion), May 13–17, 2024, Singapore, Singapore.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3589335.3651455

1 INTRODUCTION

A large number of AI frameworks have been created to mitigate ethical problems that have been identified as development of AI has fast outpaced regulation. The AI Ethics Guidelines Global Inventory counts 173 frameworks (last update in April 2020) [1]. Several surveys of AI ethics frameworks exist focusing on different aspects, e.g. availability of concrete tools to facilitate development [2], usability of guidelines [8], human-centrism [20], and health [17, 22, 28]. Eight general ethical principles emerge: privacy, accountability, safety and security, transparency and explainability, fairness and non-discrimination, human control of technology, professional responsibility, and promotion of human values [7, 12].

At the same time, AI is being actively applied to mental health. In [25], 128 applications of AI for mental health were reviewed; the bulk of AI systems were used to stratify patients or to evaluate care quality. Other systems attempt to predict mental health problems via AI [26]. Recently, several investigations into providing cognitive behavioural therapy (CBT) have been undertaken [24]. For instance, Woebot has been used in randomized controlled trials to provide CBT, with indications of good uptake [6].

While the ethics of AI in health applications is a widely discussed problem, ranging from general frameworks above to concerns for using ChatGPT [18, 27] and other chatbots [3, 9] within the health context, these efforts rely largely upon literature reviews and thought experiments to identify challenges and pitfalls.

Sociotechnical principles developed and refined through research on the web indicate that society's perception of a technology influence its adoption, which then changes both society and the technology [21]. In 2016, Chopra et al. noted that "Existing approaches for social machines emphasize the technical aspects and inadequately support the meanings of social processes, leaving them informally realized in human interactions" [4]. The same set of problems can be seen today with the use of AI within our health social machines. Frameworks focus on high-level principles; the tools themselves emphasize the technical ability to provide the desired capability. It is imperative to fully understand the social protocols and relationships of the concerned parties as they interact in order to develop a new wave of accessible and equitable AI social machines in healthcare. To this end, we undertook surveys, and a mixture of focus groups and one-to-one interviews, based on the availability and preference of the participants. Themes were then extracted and cross checked by multiple researchers from each of these data.

The contributions of this work include:

(1) Multiple, triangulated qualitative methods including surveys, focus groups and interviews with mental health service users

WWW '24 Companion, May 13–17, 2024, Singapore, Singapore

ACM ISBN 979-8-4007-0172-6/24/05
https://doi.org/10.1145/3589335.3651455

and therapists focused on factors for trust and trustworthiness in AI systems for mental health (Section 2).
(2) A thematic analysis of the results around accessibility and equity for AI within mental health applications (Section 3).
(3) A discussion of these results to provide a "bottom-up" approach to the issues of ethics within AI for health for better AI application design (Section 4).

2 METHODOLOGY

After securing approval from the Hillary Rodham Clinton School of Law's Research Ethics Committee, we began with a pilot study consultation with staff (therapists) and service-users at Adferiad Recovery (a large mental health charity). The aim was to capture initial insights about their understanding of trust, and different trust factors (including terminology) around AI in mental health care. The findings of the pilot study fed into the project design, resulting in two surveys (one for service-users, one for therapists), hosted on JISC; and semi-structured interviews and focus groups.

Prospective participants were contacted via gatekeepers at Adferiad Recovery, and were not additionally compensated for their voluntary participation. All participants were provided with a briefing about the study, what consent means, and given a debrief form with details of who to contact should they experience negative effects after taking part. Each participant provided informed consent to participate in the study.

In total, we received 70 survey responses and conducted 12 interviews and focus groups, with 32 participants (10 service-users; 10 peer-mentors; 12 therapists). Interviews and focus groups were recorded and transcribed. Closed question survey responses were analysed quantitatively, using MS Excel. All text responses to the open-ended survey questions and transcripts from the interviews and focus groups were read by multiple members of the research team, and thematically coded and analysed using NVIVO software. We now turn to the main themes identified on accessibility and equity matters.

3 RESULTS

Analysis of the surveys, interviews and focus group data identified key themes on accessibility and equity. On accessibility, these ranged from concerns over possible lack of access to the internet, and literacy issues (including computer literacy), to matters of choice and agency for patients, including the potential to improve support and/or reduce waiting times. Whereas on equity, core concerns included bias, and generalisability versus personalised care. Pervading these two broad areas were concerns over data privacy and confidentiality.

Note on quotes: in this section the text in quotation marks are quotes from the surveys, and interview and focus groups transcripts. All contributions are anonymous. Some quotes have been edited to improve readability (e.g. removing filler words); but no changes have been made that would alter the meaning of the text.

Internet/computer access: There are considerable challenges in ensuring that access to, and the use of AI, will be fair for all potential service users. Without such matters being addressed, the issue of 'choice' will be an illusory one, as illustrated by the following: "if laptop based, the healthcare professional must provide a room for patients to complete courses in. Most clients that I see do not own a laptop, let alone WiFi routers. They rely on the libraries to access the internet, but this isn't a confidential setting" (therapist response). Reliance on AI could exacerbate inequalities for those from disadvantaged backgrounds who may not have the means to engage with digital services, including those relating to healthcare [23].

Literacy: Literacy issues can pose significant challenges for engaging with AI/digital healthcare: "I've got someone who can't read. So if he was struggling, he'd struggle to even type or read back the answer" (therapist response). Similarly, computer literacy cannot be assumed: "Accessibility: many of my clients are not computer literate" (therapist survey response); and "must be easy to access and navigate. Options for computer to read out what it says on the screen for blind persons and large letter options" (therapist survey response). There was a plea to "make sure it is easy to use for people who are not good with computers" (survey response).

Waiting times/better support: One of the positive perceived benefits of AI in healthcare was in the context of reducing waiting times (to see a doctor/therapist): "everything is in a bit of a state at the moment, especially mental health services and I just think that if AI could help with that then great" (therapist interview response). AI was, therefore, perceived as having a potentially valuable role by service-users: "They could greatly reduce waiting times and access to services for those in need". The increased hours of access offered by AI were positively referred to: "You can access it any time, 24/7, it may change some people's lives" (service-user interview response); and "I don't mind the idea of having an AI to talk to about mental health because it's a lot easier to get in contact with than a regular doctor" (service-user interview response). Therapists shared similar sentiments: "Easier for people to talk 24/7. Clients can message as opposed to talk to somebody on the phone. And the answers may be more useful than speaking to untrained staff" (therapist survey response); and "providing 24-hour access: the mind can 'play tricks' at any hour of the day or night" (therapist survey response).

Patient choice: In addition to the benefits of 24 hour access, the fact that some service-users might prefer to engage with AI rather than a human was also an emergent theme: "And for some people that struggle with humans as well, you know, some people might prefer to interact with a computer" (therapist interview response). However, the importance of patient choice was also emphasised: "Also an option for that person to speak to a human, if they're not happy with how it's going with the robot" (therapist interview response); and "So for me, I think the AI would always have to be complementary to what is existing and to not actually take away anything that already exists" (interview response).

Bias: The main concern regarding bias was centred on the input data. Participants placed less emphasis on the AI system itself, but rather more on the human responsibility for inputting and creating the datasets, and how their own biases might influence the technology they create. This concern extends to questioning the developers' understanding of mental health issues: "Unless it can learn an AI is only as good as its programmer and it relies on them. There would need to be multiple people involved to avoid bias" (survey response).

Generalisability versus personalised care: During the interviews, therapists highlighted the diversity and complexity of

human beings, and referred to individual differences within mental health: "It's every individual person's mental health. There's no complete cookie cutter sort of mould for it." Therapists were also concerned about how AI would handle people with complex or co-occurring conditions: "if there's lots of things wrong with one person and they fit under multiple categories, then the AI is potentially misdiagnosing or over diagnosing".

Further, they expressed concern about issues arising from generalisability: "I think, in my job [as a therapist], you judge what you say differently to different people. An AI might just give everyone the same kind of response." This was also repeated in the survey responses from therapists: "Content is not personalised to the person, [and] person feeling like there is no one real out there to talk to". This latter comment links to other concerns, expressed by both service-users and therapists, about the loss of the human element: "I just don't like the idea of someone who is in a crisis, really struggling with their mental health, to be not getting that human touch" (therapist response). Similarly, "I think everybody feels that the face-to-face is more valuable and builds a rapport, a relationship. You can't build a relationship with a computer" (service-user response).

Data privacy and confidentiality: As outlined in the introductory paragraph to this section, these themes were also prominent in the interviews, focus groups and survey responses. Examples include: "For me, it just comes down to the security of information"; and "there needs to be transparency between all the parties. It's essential to know how your data is collected and maintained"; and "I don't know where that information is going and who has access to it." Finally: "make that really clear from the offset, where the information is going, in a way that's easy for people to understand. I think a lot of people I work with - if people have psychosis or anything - they are dealing with a lot already and they wouldn't like worrying about where their information will go".

Our results from service users and therapists revealed both benefits and challenges for the use of AI to support mental health services. The identified benefits included reduced waiting times, remote access at any time of day or night and improved patient choice of the channel they use to contact services. Access to an AI via voice or chat may be particularly useful for someone with mild to moderate anxiety or depression for whom a phone call or face-to-face encounter is stressful, as found in an earlier trial of webchat access to therapists [13]. Challenges included service users who cannot communicate on screen due to visual problems or illiteracy, those who lack the necessary hardware or broadband link (for example, rural communities, travellers or homeless people), the risk of bias due to inadequate developer understanding of mental health services (and the lack of current access to representative datasets to train AIs), and the risk that an AI would not deliver responses that are as carefully tailored to user needs and personality as a human therapist, especially in a crisis setting.

4 DISCUSSION

4.1 Strengths and weaknesses

The strengths of our study come from its empirical basis. Rather than reasoning from first principles or speculating about the accessibility or equity challenges posed by AI to support mental health services, we recruited 102 service users and therapists and asked them about their views using a variety of methods. This adds credibility to our findings as these people are embedded in the mental health service context and have a profound understanding of its complexity and subtlety that outstrip any external observer's ability to make sense of this setting.

A complementary weakness is that our service users and therapists had only modest understanding of AI and how it works. We attempted to overcome this by briefing them about AI and giving an example of its use in mental health services before asking for their views. A further weakness is that we assumed that what people said reflects their real views and attitudes. There is some evidence from dietary recall surveys, for example, that social response bias (ie. people say what they think a researcher would approve of) can modify people's reported attitudes and behaviour [16]. Introducing some distance between the participant and researcher through e.g. an online survey may reduce this. A further challenge is that our research was focused on one mental health service provider (albeit covering multiple parts of Wales), so our results may not generalise to other service providers in other countries.

4.2 Comparison to other studies

Our findings are broadly consistent with the limited literature around patient/user acceptance. Davies et al.'s [5] small-scale interview and survey analysis of service user perspectives of a very basic mental health app were generally positive, including (as with our study) in relation to the feeling of not being judged by the app versus speaking to a human being. Likewise, our findings accord with many of the generic research priorities identified by Hollis et al. [11] (following surveys and interviews with some carers, professionals and people with lived experience of mental ill-health). Where our study adds to the literature is in the level of detail and scope of the concerns raised. Given the lack of existing studies on the patient/service-user perspective, our study makes a substantial contribution to the understanding of this issue.

The clinician perspective is comparatively better covered in research literature. An integrative review of AI decision support systems in mental health found that barriers to adoption primarily arose from clinicians' uncertainty regarding 'trust and confidence, end-user acceptance and transparency'. This study included four distinct systems that were described between 2016 and 2020, and focused on decision support systems [10]. Examples outside mental health include Liberati et al. on implementation of Computerized Decision Support Systems in Italian Hospitals [15], Petkus et al. on British physicians' concerns on AI [19] and likewise Lai et al. on French practitioners' concerns [14]. These studies tend to raise issues in relation to professional autonomy/control of introduction and use of AI tools. Whilst many of our own therapist participants concurred generally, we consider that their concerns in relation to equity and accessibility are nevertheless a significant and novel addition to the literature.

4.3 Recommendations for research and practice

We feel that the number and variety of insights revealed by our methods suggests that researchers into the ethics of AI should be more willing to contact and consult those likely to use or benefit

from the AI, using standard qualitative and quantitative methods, as we did. This will require extra work and ethical approval but is likely to be very productive. Our recommendations for practice are that the developers of AI tools to support mental health services – and probably other kinds of health and care support services – need to take action to reassure everyone using their tools that AI:

(1) Is suitable for all levels of literacy and can be used by those with visual impairments.
(2) Can be accessed by people with minimal hardware, eg. a modest smartphone and limited data allowance.
(3) Is as unbiased as possible, by inclusion of training data from a fully representative sample of service users across age, gender, ethnic group and deprivation categories. Assembling and providing anonymised access to such a dataset in sensitive areas such as mental health will be challenging.
(4) Is able to provide responses that are as carefully tailored to user needs and personality as a human therapist in most cases. This may require the developer to exclude users who are in crisis from using their AI.

5 CONCLUSIONS

Our study engaged with both service users and therapists to identify the benefits and challenges of the use of AI to support mental health services. Key takeaways for computer scientists and the developers of AI systems include:

- bias due to inadequate developer understanding of mental health services
- the lack of current access to representative datasets to train AIs
- delivery of responses that are as carefully tailored to user needs and personality as a human therapist, especially in a crisis setting.

Additional benefits identified from both groups included reduced waiting times, remote access at any time of day or night and improved patient choice of the channel they use to contact services, as well as peace of mind for those who struggle with social interactions. The challenges identified an impact on service users with visual problems, illiteracy, and lack of IT resources. Technologists creating social machines to facilitate mental health must take these concerns into account.

ACKNOWLEDGMENTS

This research was generously supported by the British Academy and Leverhulme Trust (SG2122\210037) and the NIHR Southampton Biomedical Research Centre (NIHR203319). We are particularly indebted to our research participants and to our project partner Adferiad Recovery.

REFERENCES

[1] Algorithm Watch. 2020. AI Ethics Global Inventory. https://inventory.algorithmwatch.org/ Accessed 14 December 2021.
[2] Ayling and Chapman. 2021. Putting AI ethics to work: are the tools fit for purpose? *AI and Ethics* (2021), 1–36. https://doi.org/10.1007/s43681-021-00084-x
[3] Cabrera, Loyola, Magaña, and Rojas. 2023. Ethical dilemmas, mental health, artificial intelligence, and llm-based chatbots. In *International Work-Conference on Bioinformatics and Biomedical Engineering*. Springer, 313–326.
[4] Chopra, and Singh. 2016. From Social Machines to Social Protocols: Software Engineering Foundations for Sociotechnical Systems. In *WWW'16*. 903–914. https://doi.org/10.1145/2872427.2883018
[5] Davies, Craven, Martin and Simons. 2017. Proportionate methods for evaluating a simple digital mental health tool. *Evid Based Ment Health* (2017). https://doi.org/10.1136/eb-2017-102755
[6] Fitzpatrick, Darcy, and Vierhile. 2017. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Ment Health* 4, 2 (06 Jun 2017), e19. https://doi.org/10.2196/mental.7785
[7] Fjeld, Achten, Hilligoss, Nagy, and Srikumar. 2020. Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. *Berkman Klein Center Research Publication* 2020-1 (2020).
[8] Hagendorff. 2020. The ethics of AI ethics: An evaluation of guidelines. *Minds and machines* 30, 1 (2020), 99–120.
[9] Hamdoun, Monteleone, Bookman, and Michael. 2023. AI-based and digital mental health apps: Balancing need and risk. *IEEE Technology and Society Magazine* 42, 1 (2023), 25–36.
[10] Higgins, Short, Chalup, and Wilson. 2023. Artificial intelligence (AI) and machine learning (ML) based decision support systems in mental health: An integrative review. *International Journal of Mental Health Nursing* (2023). https://doi.org/10.1111/inm.13114
[11] Hollis, Sampson, Simons et al. 2018. Identifying research priorities for digital technology in mental healthcare: results of the James Lind Alliance Priority Setting Partnership. *Lancet Psychiatry* (2018). https://doi.org/10.1016/S2215-0366%2818%2930296-7
[12] Jobin, Marcello, Vayena. 2019. The global landscape of AI ethics guidelines. *Nature machine intelligence* 1, 9 (2019), 389–399.
[13] Kessler, Lewis, Kaur et al. 2009. Therapist-delivered Internet psychotherapy for depression in primary care: a randomised controlled trial. *Lancet* (2009). https://doi.org/10.1016/S0140-6736(09)61257-5
[14] Lai et al. 2020. Perceptions of artificial intelligence in healthcare: findings from a qualitative survey study among actors in France. *J. Trans. Med.* (2020). https://doi.org/10.1186/s12967-019-02204-y
[15] Liberati, Ruggiero, Galuppo et al. 2017. What Hinders the Uptake of Computerized Decision Support Systems in Hospitals? A Qualitative Study and Framework for Implementation. *Implement Sci.* (2017). https://doi.org/10.1186/s13012-017-0644-2
[16] Miller, Abdel-Maksound, Crane et al. 2008. Effects of social approval bias on self-reported fruit and vegetable consumption: a randomized controlled trial. *Nutrition J.* (2008). https://doi.org/10.1186/1475-2891-7-18
[17] Morley et al. 2020. The ethics of AI in health care: A mapping review. *Social Science and Medicine* 260 (2020), 113172. https://doi.org/10.1016/j.socscimed.2020.113172
[18] Parray, Inam, Mahfuza et al. 2023. ChatGPT and global public health: applications, challenges, ethical considerations and mitigation strategies. , 50-54 pages. https://doi.org/10.1016/j.glt.2023.05.001
[19] Petkus, Hoogewerf and Wyatt. 2020. What do senior physicians think about AI and clinical decision support systems: Quantitative and qualitative analysis of data from specialty societies. *Clin Med (Lond)* (2020). https://doi.org/10.7861/clinmed.2019-0317
[20] Rigley, Chapman, Evers, and McNeill. 2023. Anthropocentrism and Environmental Wellbeing in AI Ethics Standards: A Scoping Review and Discussion. *AI* 4, 4 (2023), 844–874. https://doi.org/10.3390/ai4040043
[21] Shadbolt, Smith et al. 2013. Towards a Classification Framework for Social Machines. In *Proceedings of the 22nd International Conference on World Wide Web* (Rio de Janeiro, Brazil) *(WWW '13 Companion)*. Association for Computing Machinery, New York, NY, USA, 905–912. https://doi.org/10.1145/2487788.2488078
[22] Solanki, Grundy, and Hussain. 2023. Operationalising ethics in artificial intelligence for healthcare: A framework for AI developers. *AI and Ethics* 3, 1 (2023), 223–240.
[23] Studman. 2023. Access denied? Socioeconomic inequalities in digital health services. In *Report, Health data and COVID-19 tech*. Ada Lovelace Institute. https://www.adalovelaceinstitute.org/report/healthcare-access-denied/
[24] Thieme, Hanratty, Lyons, et al. . 2023. Designing human-centered AI for mental health: Developing clinically relevant applications for online CBT treatment. *ACM Transactions on Computer-Human Interaction* 30, 2 (2023), 1–50.
[25] Tornero-Costa, Martinez-Millana, Azzopardi-Muscat, Lazeri, Traver, and Novillo-Ortiz and others. 2023. Methodological and quality flaws in the use of artificial intelligence in mental health research: systematic review. *JMIR Mental Health* 10, 1 (2023), e42045.
[26] Tutun, Johnson, Ahmed et al. 2023. An AI-based decision support system for predicting mental health disorders. *Information Systems Frontiers* 25, 3 (2023), 1261–1276.
[27] Wang et al. 2023. Ethical considerations of using ChatGPT in health care. *Journal of Medical Internet Research* 25 (2023), e48009.
[28] Zhang and Zhang. 2023. Ethics and governance of trustworthy medical artificial intelligence. *BMC Medical Informatics and Decision Making* 23, 1 (2023), 7.