This is a peer-reviewed, accepted author manuscript of the following research article: Morgan, P & Cogan, NA 2024, 'Using artificial intelligence to address mental health inequalities: co-creating machine learning algorithms with key stakeholders and citizen engagement', Journal of Public Mental Health. https://doi.org/10.1108/JPMH-07-2024-0095

Title: Using artificial Intelligence to address mental health inequalities: Co-creating machine learning algorithms with key stakeholders and citizen engagement

Abstract

Purpose: Artificial intelligence (AI) is poised to reshape mental health practices, policies, and research in the coming decade. Simultaneously, mental health inequalities persist globally, imposing considerable costs on individuals, communities, and economies. This study investigates the impact of AI technologies on future citizenship for individuals with mental health challenges (MHCs).

Approach: This research employed a community-based participatory approach, engaging peer-researchers to explore the perspectives of adults with MHCs from a peer-led mental health organisation. The study evaluated potential threats and opportunities presented by AI technologies for future citizenship through a co-created film, depicting a news broadcast set in 2042. Data were gathered via semi-structured interviews and focus groups and were analysed using a reflexive thematic approach.

Findings: The analysis identified four key themes: (1) Who holds the power? (2) The divide, (3) What it means to be human, and (4) Having a voice. The findings indicate that adults with living experiences of MHCs are eager to influence the development of AI technologies that affect their lives. Participants emphasised the importance of activism and co-production, while expressing concerns about further marginalisation.

Originality: This study provides new insights into the intersection of AI, technology, and citizenship, highlighting the critical need for inclusive practices in technological advancement. By incorporating the perspectives of individuals with living experiences, the study advocates for participatory approaches in shaping AI technologies in mental health. This includes the co-creation of machine learning algorithms and fostering citizen engagement to ensure that advancements are equitable and empowering for people with MHCs.

Introduction

Mental health inequalities are pervasive global issues, imposing significant burdens on individuals, communities, societies, and economies (Kirkbride et al., 2024). These disparities are driven by various factors, including socio-economic status (Kivimäki et al., 2020), race (Shim, 2021), gender (Borrescio-Higa & Valenzuela, 2021), geographic location (Hudson, 2012), and access to mental health care (Saeed & Masters, 2021). Stigma and discrimination against individuals with mental health conditions (MHCs) further deepen these inequalities (Sukhera et al., 2022), deterring individuals from seeking help and impacting care quality, leading to poorer mental health outcomes (Corrigan, 2004). These challenges are often more pronounced in marginalised communities, compounding socio-economic factors that worsen MHCs (Macintyre et al., 2018a). Barriers to adequate care, particularly for younger and minority populations, lead to disparities in diagnosis, treatment, and outcomes (Mongelli et al., 2020). With rising mental health challenges in these groups, effective and accessible solutions are increasingly critical (Bommersbach et al., 2023).

Artificial Intelligence (AI) refers to computer systems that simulate human intelligence to perform tasks such as learning, decision-making, and problem-solving. In the context of mental health, AI can analyse vast amounts of data to enhance diagnostic accuracy, create personalised treatment plans, and identify effective interventions (Shah, 2022). For example, machine learning, a subset of AI, enables systems to learn from data and improve performance over time without explicit programming. AI has already made strides in other health domains, such as oncology and radiology (Lin et al., 2023; Najjar, 2023), but its adoption in mental health has been slower due to challenges such as data complexity, ethical concerns, and the absence of clear biomarkers for mental health conditions. Despite these hurdles, AI-driven tools are showing promise in mental health care, where algorithms can predict effective treatments based on individual histories, socio-demographics, and even genetic profiles (Chekroud et al., 2021).

AI's ability to process large-scale data from electronic health records, social media, and wearable devices can help identify patterns of mental health conditions and enable earlier interventions (Lyortsuun et al., 2023). Additionally, AI-driven technologies such as **digital mental health applications, virtual therapy sessions, chatbots, and augmented reality can enhance accessibility,** particularly in underserved areas, by providing remote support (Torous et al., 2021). This could reduce barriers to care, improve mental health outcomes, and support earlier intervention efforts (Srividya et al., 2018). Furthermore, AI systems could contribute to suicide prevention by mapping trends and developing early warning systems (Huang & Hu, 2024).

However, the use of AI in mental health must be approached with caution. There are ethical concerns regarding bias, privacy, and the risk of exacerbating existing inequalities (Balcombe & Leo, 2021). AI systems developed using data from predominantly affluent, white populations may fail to perform well for marginalised groups, potentially reinforcing disparities (Trewin et al., 2019). This highlights the need for global ethical, methodological, and regulatory standards to assess and mitigate bias in AI-driven mental health tools (Gray et al., 2024; Cogan, 2024). It is crucial that AI systems promote equity, transparency, inclusion, and fairness (Timmons et al., 2023; Wang et al., 2024).

Digital exclusion, or the inability to access digital technologies, remains a significant barrier for certain vulnerable groups, particularly those with mental health conditions. While digital exclusion has steadily declined in the UK over the past decade, dropping from over 20% to around 5% (ONS, 2023), those who remain excluded are disproportionately affected. Without targeted efforts to bridge this gap, digital exclusion may worsen health inequalities, preventing access to AI-driven mental health advancements (Van Deursen et al., 2015). The shift to digital services, accelerated by the COVID-19 pandemic, has made this issue even more urgent (Spanakis et al., 2021).

The integration of AI into mental health care must also consider how these technologies might impact human relationships in care. While AI has the potential to supplement the work of mental health professionals, the absence of human interaction in AI-driven interventions could reduce the personalised support essential for individuals with mental health conditions (Bohr & Memarzadeh, 2020). Moreover, the concept of citizenship, particularly for those with MHCs, is changing in the digital era (Volti & Croissant, 2024). AI innovations need to align with the rights, roles, responsibilities, and access to resources that define citizenship today (Rowe, 2015).

To ensure AI technologies are inclusive and equitable, it is essential to involve people with lived experience of mental health conditions in their development. Participatory research methods, which engage those with living experience in the research process, are vital for ensuring that AI tools reflect the needs and values of diverse populations (Skorburg et al., 2024). This approach aligns with the principles of citizenship, giving individuals a voice in decisions that affect their lives (Cogan et al., 2022).

This paper emphasises the importance of co-creating AI technologies within a public mental health framework to address mental health inequalities. We argue that the involvement of individuals with living experience is crucial to developing machine learning algorithms that meet their needs (Le Glaz et al., 2021). By prioritising inclusion, collaboration among mental

health professionals, individuals with living experience, technology developers, and community organisations can foster a more equitable and inclusive approach to AI in mental health care. This study is the first to explore how people with living experience perceive AI's impact on their sense of citizenship, aiming to identify key areas for developing inclusive, future opportunities for people with MHCs.

Method

We present the findings from an in-depth qualitative study, exploring the views of participants from an English peer-led mental health organisation (PLMHO; n= 10), examining the threats and opportunities for AI technologies and the impact on inclusive future citizenship for people with MHCs. The research team was a mental health professional working alongside peer-researchers (n=3). The project draws upon community based participatory research (CBPR), which is a collaborative research approach that involves community members, organisational representatives, and researchers in all aspects of the research process (MacIntyre *et al.,* 2018b). This method ensures that the perspectives and needs of those affected by mental health inequalities are considered in the development of AI technologies (Carr, 2020). Individuals with living experience of mental health issues bring valuable personal insights into the challenges and needs of those affected (Cogan *et al.,* 2022). Engaging individuals with living experiences of MHCs in the co-creation of AI technologies ensures that the tools developed are relevant, accessible, and effective for diverse populations (Zidaru *et al.*, 2021). Their involvement can also help identify potential risks and ethical considerations, such as privacy concerns and the potential for bias in AI algorithms.

Data Collection

The research project had two phases. The first, consisted of semi-structured interviews and a co-production workshop (participants were from a PLMHO (n = 7) and digital technology company n = 4). The data from this phase was co-analysed and an output of this analysis was a film created by the researchers. The film was a news broadcast set in 2042 exploring the potential impacts of technology on citizenship for people with MHCs. Whilst there are many studies looking at future developments in mental health, the utilisation of the academic approaches of Future Studies (Connolly, 2012) within mental health is a novel approach (Morgan et al, 2020) The use of creative methods, such as film making, within research can provide approaches that express rich social, cultural and contextual factors that can go beyond traditional research methods and highlight complexities and ambiguities (Baumann *et al..* 2020). Film making has 'transformative potential' within participatory research by giving voice to people with living experience, and in this case flattening the power dynamic

within the research team (Kindon, 2003, p143). In the second phase, participants from the PLMHO (n=6) watched the film and then with the research team participated in a focus group. The focus group discussed the film and what could be done to promote more inclusive futures. As is recommended for qualitative research that has an ideographic aim (Malterud *et al.,* 2016), this sample size for the co-production workshop and the focus group were sufficient to capture the voices of the participants yet allowed individuals to have a locatable voice within the reporting of the findings.

Participants

This paper is focused on the views of people with living experience of MHCs and so only the data from the peer-led organisation (n=10) have been included in this paper. All the participants from the PLMHO self-identified as having MHCs. All participants are referred to by pseudonyms (see table 1 for demographic details). Participants were recruited through the mental health practitioner researcher's professional networks, with the aim of fostering long-term co-productive relationships centred on citizenship and technology (Edwards and Brannelly, 2017).

Table 1: Participant demographics

Characteristic	**Number of Participants (n = 10)**
Gender	
Female	3
Male	4
Prefer not to say	3
Ethnicity	
Black British	1
White British	6
Prefer not to say	3
Age	
25-39	1
40-55	5
56-74	1
Prefer not to say	3

Data Analysis

The transcribed data was analysed using a reflexive thematic approach (Braun and Clark 2006; Braun & Clarke, 2021; Jenning *et al.*, 2019) which is an accessible and theoretically flexible interpretative approach to qualitative analysis (Campbell *et al.*, 2021). It facilitates the

identification, analysis, and reporting of data to derive meaning and develop themes (Braun & Clarke, 2023). Given that thematic analysis is a flexible qualitative method, an inductive, ideographic approach was adopted, since this was in keeping with a data-driven, pragmatic analysis (Blandford, 2013).

Due to the cyclical nature of CBPR, the analysis was not conducted in a linear, step-by-step manner but rather as a series of overlapping processes that were frequently revisited. This was especially true during the initial phase of data analysis, aligning with Braun and Clarke's (2021) approach, where analysis is viewed as a recursive process. Initially, the mental health professional on the research team coded the interview data using NVIVO, while peer researchers reviewed sample transcripts. The team met to agree on initial codes and later reconvened to finalise the themes. Although peer researcher involvement was initially limited by time and funding constraints, these issues were resolved, enabling more equitable participation in the later stages.

The peer researchers preferred not to use NVIVO, so the team adopted manual methods for the remainder of the analysis. Themes from the interviews were presented, reviewed, and discussed with participants during a co-production workshop. The transcript of this workshop was then co-analysed by the research team, and this analysis informed the development of the film, which can be seen as a form of reporting, as described by Braun and Clarke (2006).

In the second phase, each researcher familiarised themselves with the data and individually generated initial codes. The team then met to search for themes and share insights, resulting in an initial coding table that included the frequency of themes. A subsequent face-to-face meeting involved manually mapping out and refining the themes and sub-themes, leading to their final agreement.

Reflexive journals were kept, and the research team held reflexive meetings. The meetings were not just to analyse the data but also to critique and review the process of analysis. Alongside these notes, mind maps, and coding tables were kept providing an account of the data analysis providing transparency and rigour (Johnson *et al.,* 2020.) The quality standards for reporting qualitative research were followed in accordance with maintaining an audit trail and transparency in the recording and reporting of data (O'Brien *et al.* 2014).

Ethics

Ethics approvals for this study were given by the University Ethics Committee (ref no. 29315). All participants provided informed consent prior to their participation in the study.

Results

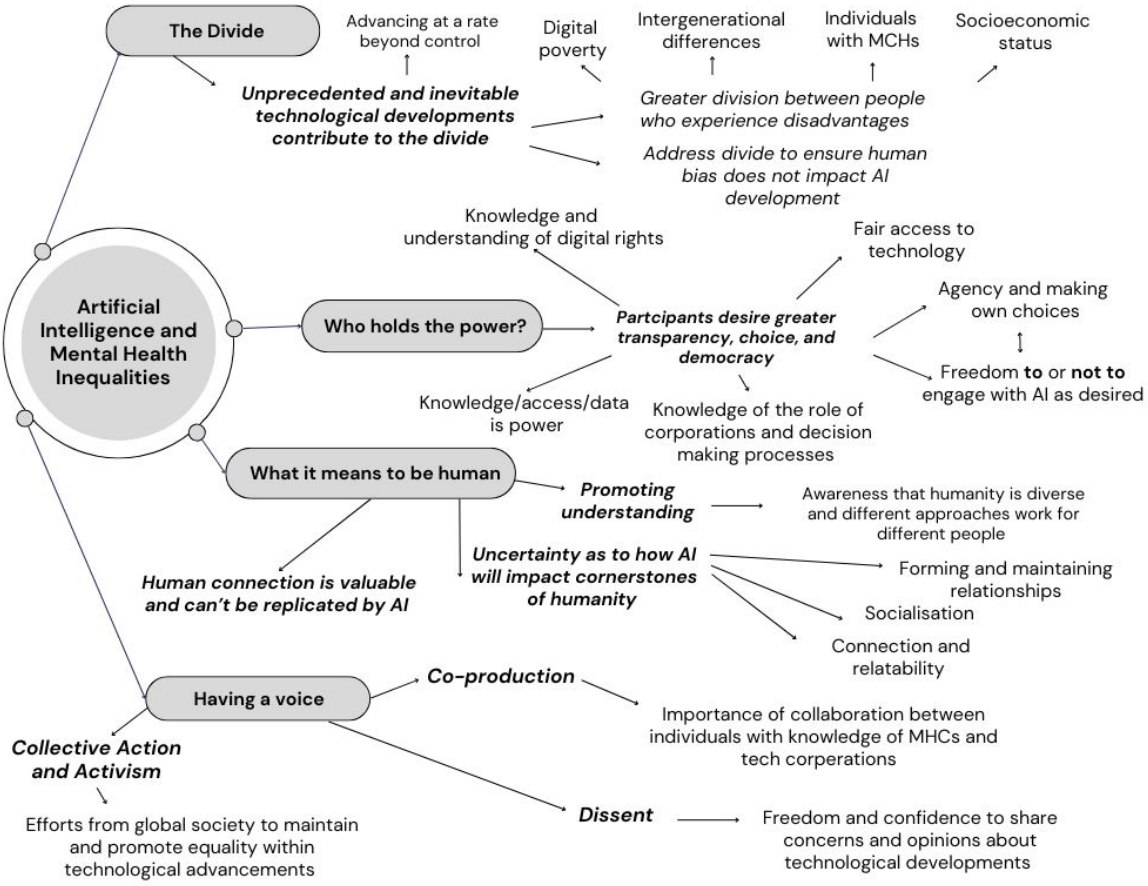
The themes developed were (1) Who holds the power? (2) Divide (3) What it means to be human; and (4) Having a voice (see figure 1). These themes are interconnected, and the findings presented below provide further discourse in relation to the themes with further exemplar quotes detailed in the thematic table (table 2).

Theme 1: Who Holds the Power?: Participants described the importance of knowing where power lies in a technological world, recognising that power did not sit just with governments but also corporations. Knowledge was viewed as essential to gaining power, with democracy playing a crucial role in promoting equality not only for individuals with mental health challenges but for all citizens. There were fears with greater automation there would be a loss of autonomy and reduction in agency. An important part of having personal power was the choice to opt-out of using technology, however participants viewed technological developments as inevitable, happening at a fast pace beyond their control.

> *Feels like it* [technology] *is already moving away from you… for a lot of people, it can whoosh past them* (Maya).

This is not to suggest that all participants were fearful of technology. As Aspienaut noted, it's easy to overlook "the huge positive changes technology has allowed us." While discussions about technology were nuanced, the potential for division remained a significant concern among participants.

Figure 1: Thematic diagram



Theme 2: The divide: This theme describes the way participants saw the impact of technology creating greater division in society. This led to fears about how this would result in further division between people who experience disadvantage, such as people with MHCs and people in poverty, but also intergenerationally. Concerns were raised about how older people may be less likely to engage in the use of technology and therefore be left behind. Participants were worried that younger people were so emersed in technology that they would be subject to additional stressors, such as cyber bullying. Participants raised questions about the role of corporations, who had power, and how decisions were made. They also identified that it was important that people understood their digital rights and had fair access to technology. Participants wanted greater transparency, choice, and democracy.

> *AIs could probably get to the point where they might be able to detect whether somebody is actually going into a crisis from a mental health point of view... I guess it's the choice as to whether you let it in or not. That's always been my thing...choice.* (Pete).

This choice also included not having to adopt technology. Participants questioned without choice and transparency if this could lead to a collapse of agency.

> *How that then looks for people who are having decisions made for them by machines...and that sense of autonomy? (*Laurel).

Theme 3: What it means to be human: Participants questioned whether technology was or had the potential to change the essence of being human. They questioned whether as people increasingly interacted with and via machines this would adversely impact on human-to-human connections. There was concern about the role of touch and what the implications would be for relationships and in particular peer work. Participants questioned whether these changes would impact on our mental health.

> *Is it going to be a massive payoff in 10 years' time? When this AI, suddenly there's, you know, people start becoming more unwell because they are finally realising that human connection is really important for us?* (Colin).

Theme 4: Having a voice: Participants saw having a voice and being able to try to act and make change as a core to their sense of citizenship and to influence the direction of technology. They emphasised the importance of having a greater voice, through activism or by being able co-produce solutions. Co-production was not just seen as important with mental health services but also with digital technology companies.

> *Think that is something that we have a responsibility to do in terms of that kind of activism element of our work… I think it is around just challenging in every context that we're in... How do we keep people connected and keep people having a voice? (*Janet*).*

> *I think for me one of the things that this conversation is really bringing home to me, is the focus around co-producing that kind of future citizenship (*Rose).

Table 2: Themes and exemplar quotes

Themes and associated sub-themes	Exemplar Quote
Who holds the power?: Participants identified that whilst technology and AI is shaping society it is always not clear who is directing these developments. For citizens to increase their own sense of agency and collectively promote equality, they describe importance of understanding where power lies (whether that is with governments or corporations) and the actions they can take.	
Knowledge is power/Data is Power	*The concept of knowledge is power and, actually, that came screaming through… If you don't have access to that knowledge... by default you have less power and less control (*Rose)
Agency	*It's about where we have agency over some of this stuff & agency over our data if data is power, how can we take that back in a way?* (Janet)
Democracy	*How do we, as a society, a global society try and keep those things equal and inclusive and hold people to account on that as well?* (Maya)
Opting-out	*no desire or wish to even partake in a kind of digital community.* (Laurel)
Divide: This theme describes the way participants saw the impact of technology creating greater division in society	
Digital Poverty	*Digital poverty is an issue now, and as the technology advances that will mean you'll need better hardware that would cost more money, and you know what I mean it's like at some point that's going to rule people out.* (Colin)
Generational Divide	*With our young people, predominantly having this presence online I really worry about their safety and their mental health in terms of not knowing what they're facing and not knowing what they're up against and not knowing where criticism or bullying is coming from… My parents-in-law, you know, being as an elderly population as people in their 80s, who have no desire or wish to be part of this at all.* (Laurel)
Addressing the Divide	*Are we actually having intentional conversations about what we do about the inequalities for people? (*Maya)
What it means to be human?: Participants questioned whether technology was or had the potential to change the essence of being human.	
Connection	*When you said that you lose that human touch, you know robots are not going to provide you that human touch. That made me feel really sad and you know, having robots and technology around is great in one respect, but that face-to-face person to person thing is that's like embedded in us. How can you recreate that? You can't (Poppy).*
Uncertainty	*that interaction through technology, will that be fundamentally changed and how we form relationships and how we relate to one another, will that be kind of pulled apart and changed and twisted about and look really different? I find that really, really terrifying in the, you know, the work that we do and supporting each other. (Laurel)*
Promoting Understanding	*I think we all have to be mindful of, that we all relate in different ways and for some people doing that through a*

	*screen is really useful. And how do we think about relational approaches in our digital work if that's the case? (*Janet)
Having a voice: Participants saw having a voice and being able to try to act and make change as a core to their sense of citizenship and to influence the direction of technology.	
Collective Action and Activism	*How do we as a society, a global society, try and keep those things equal and inclusive and hold people to account on that as well?* **(**Maya)
Dissent	*In a world which is around progress and is around future and moving forward, do we lose the opportunity to listen to dissent and to have space to be able to share fears and concerns and worries without feeling that you're not progressive or you're not open to new ways of working and to new challenges?* (Rose)
Co-production	*Thinking about from a co productive process what we do know is this isn't about us having all the expertise and tech companies or corporate organizations not having that [expertise in mental health citizenship]. (*Maya)

Discussion

This paper illustrates the importance of co-creating AI technologies, through presenting findings from a study of a CBPR project conducted with a PLMHO. The participants were particularly concerned about unintended consequences of technology that exacerbated exclusion and distress. They wanted increased agency, choice, transparency and to have a voice in shaping the future. Central to this was access to technology and the promotion of knowledge to understand how to use it and how to protect and promote their rights.

Participants were asked to reflect on both the opportunities and threats posed by technology and AI, with the film presenting both positive and negative impacts. However, perhaps due to the focus on citizenship, concerns centred more on rights and inclusion rather than the potential benefits of AI in the mental health sphere. These concerns align with those raised by practitioners and researchers, particularly regarding bias, privacy, and the importance of human expertise (Olawade et al., 2024).

Inclusion of the voices of people who experience MHCs are often absent in ongoing discourses concerning AI and mental health innovations (Cogan, 2024). However, whilst this is a small-scale study and not a particularly diverse sample there is an appetite for people with living experience to be involved in decisions relating to the implementation of technology. As reported in earlier work (Carr, 2020), our research emphasises the importance of people with living experience collaborating with mental health professionals that have the expertise to guide the development of AI technologies that are clinically sound and ethically responsible. Their involvement ensures that AI tools are designed to complement, rather than replace, humans, and that they adhere to best practices in mental health care, ethics, and treatment (Thieme *et al.*, 2023). There are ongoing questions about

the best ways to ethically involve people with living experience in the development and research of AI (Skorburg et al., 2024; Zidaru et al., 2021). For instance, the tension between the use participatory methods and the use of mass data in AI and machine learning, which often excludes direct involvement from individuals with living experience (Skorburg et al., 2024). This study contributes to the conversation by using Community-Based Participatory Action Research (CBPAR) as a methodological approach, which explicitly seeks to promote inclusion and equality in AI development.

Participants did not just see the impact of AI as solely related to mental health services, but also in the wider context of society. Participants expressed a desire to engage with digital technology companies, either through activism or co-production to ensure their rights and perspectives were included. Participants' perspectives on AI, are very much aligned to critical perspectives on technology, as they do not view digital as automatically positive or even neutral but rather laden with existing power dynamics which therefore requiring engagement and challenge (Choi and Cristol, 2021; Emejulu and McGregor, 2019). This focus on activism aligns with the 'collective citizenship' approaches which promote the involvement of people with living experience in direct participation and political action in social change (Quinn *et al.*, 2020; Reiss *et al.*, 2022). It is important to consider approaches to technology and AI in the context of citizenship and mental health as it provides a means of challenging the structural deficits and inequalities that prevent people with living experience of MHCs of recovering their citizenship (Cogan et al., 2022; MacIntyre *et al.*, 2019).

Community organisations play a crucial role in addressing mental health inequalities by providing support and resources to marginalised populations (Campbell, 2020). Collaborating with these organisations in the development of AI technologies ensures that the tools created are accessible and beneficial to those most in need (Peters *et al.*, 2020). Furthermore, community organisations can help with the dissemination and implementation of AI tools, ensuring they reach the intended populations (Shaw *et al.*, 2019). However, within the UK, community organisations, including those run by people with living experience have been subject to funding cuts (Beresford, 2019). Therefore, consideration and prioritisation of how to best support the empowerment of people with living experience of MHCs and community organisations to fully participate in such co-creation processes in the advancement of AI technologies in mental health is needed. It is essential that mental health services retain a human-centred focus that promotes a person-centred approach enabling personal choice for those that use such services (Ozmen Garibay *et al.*, 2023)

Implications

The findings from the CBPR project underscore the importance of involving all key stakeholders in the development of AI technologies for mental health care and thereby creating AI tools that are relevant, accessible, and effective for diverse populations. This collaborative approach can help mitigate the risks of exacerbating existing inequalities, reduce the likelihood of the AI divide (Sinanan, & McNamara, 2023) and ensure that the benefits of AI are realised for the greater good of all (Feijóo *et al.*, 2020). It is important to acknowledge that participants expressed a desire to be actively involved in decisions about how technology can be used to promote rights. This highlights an area for further study, such as exploring how AI could enhance human interaction or address existing biases within traditional mental health services. It also underscores the value of participatory methods, like those used in this study, in realising the potential for ethical approaches to AI (Skorburg et al., 2024).

However, to effectively create these collaborations, community organisations and especially peer-led organisations, will need funding to be able to participate on anywhere near equal terms (Shalaby & Agyapong, 2020). There will also need to be action to address the digital divide. Governments and institutions will need to create equal access to digital technologies for all in order to facilitate social inclusion (Deganis et al., 2021) This should include support for those who do not wish to or are not able to access digital technologies to ensure they are not disadvantaged (Lattie et al., 2022).The insights gained from the CBPR project and the broader discussion on co-created AI technologies have several practical implications for addressing mental health inequalities. To develop AI tools that are equitable and effective for diverse populations, it is essential to collect data from a wide range of sources. Further research is required on how mental health inequality affects less studied vulnerable populations, such as ethnic, sexual, and gender marginalised participants, as well as how inequality factors interact to affect mental health in the long term (Gibson *et al.,* 2021). Inclusive data collection ensures that AI algorithms are trained on diverse datasets and can perform well across different populations (Arora *et al.*, 2023). Future work in the field of AI innovations and mental health would benefit from longitudinal evaluation and socio-cultural comparison (Cogan *et al.*, 2022). It would also be valuable to explore how participants actively weigh the risks and opportunities of adopting AI, as well as the potential costs of not integrating AI into mental health care. Additionally, it is crucial to involve people with living experience in discussions about both how AI is introduced by clinicians and policymakers, and what aspects of AI are prioritised for implementation.

Ethical considerations are paramount in the development of AI technologies for mental health care (Fiske *et al.*, 2019). This includes ensuring privacy and confidentiality, addressing potential biases in algorithms, and ensuring that AI tools complement, rather than replace, human care; keeping 'humans in the loop' of AI innovations (Williamson & Prybutok, 2024). As long as there are inequalities in society, there will be inequalities in the technology that humans develop; therefore, it is crucial that these inequalities continue to be addressed (Gichoya *et al.,* 2023). With the deployment of AI in society on the rise, it is important for mental healthcare leaders to determine how they will ethically apply AI within their organisations. It is to the benefit of all stakeholders that the mental health organisations remain mindful of the way technology is deployed throughout the organisation; transparency is key (Davenport & Katyal, 2020). In both the United States and the European Union, draft regulation has been proposed that seeks to address the ethical challenges that arise with widespread use of AI, especially in relation to personal and biometric data (Gerke, *et al.*, 2020). Ensuring that AI systems in mental health are designed and used in ways that are fair, safe, transparent, accountable, non-discriminatory, and aligned with human values will help maximise its benefits while minimising its potential risks and harms (Kalpakos, 2024). At a local level, clinicians should engage the people they serve in conversations about how AI and technology are used, taking into account their preferences, and promoting their rights. NHS Trusts, when developing ethical AI policies, should prioritise co-production by involving people with living experience to ensure that these policies promote inclusivity and transparency within AI systems.

Training and education for mental health professionals and community organisations on the use of AI technologies are crucial (Peters, *et al.*, 2020). This ensures that they are equipped to integrate these tools into their mental health practice and can provide support to individuals using AI-based mental health interventions. Training should also emphasise the importance of maintaining a human-centred approach to care (Tahvanainen *et al.*, 2024). This research would add people with living experience to this group and the findings highlights the importance of their role in co-creating such training, and that there should be a focus on rights, critical perspectives, intersectionality, and inclusive citizenship (Rowe *et al.*, 2015). Policymakers should consider making digital rights and the ethical use of AI core components of the training for registered mental health professionals. Additionally, recovery colleges should incorporate these topics into their curriculums, enabling individuals to understand their digital rights and the role of technology in mental health care.

It is essential that AI systems are designed to promote social inclusion, protect rights, and empower individuals, ensuring that advancements in technology contribute to rather than detract from the citizenship and well-being of all individuals (Cogan *et al.*, 2021).

This will require ongoing monitoring and screening using citizenship orientated measures (Cogan *et al.,* 2022) with data which includes the knowledge from marginalised and politically oppressed communities actively having the power to inform how AI algorithms are constructed. Policymakers and advocates play a critical role in promoting the equitable development and implementation of AI technologies for mental health care (Ogugua, *et al.*, 2023). This includes creating policies that support inclusive data collection, funding research that involves key stakeholders, and advocating for the ethical use of AI in mental health care and treatment (Whittlestone *et al.,* 2019). A key priority for policymakers is to ensure that there are regulations in place to make AI tools accessible and beneficial to marginalised populations. At present, regulatory initiatives are underway yet vary significantly worldwide. Within the European Union (EU) regulations will complement existing privacy regulations, such as General Data protection Regulation (GDPR) or the UK Privacy Act of 2018. The EU have reached political agreement for the implementation of a rights-based AI Act. This will regulate the use of AI within the EU but will also have global ramifications as it applies to any company that extends the use of AI into the EU. The Act seeks to provide a framework that can assess risk, and promote trustworthiness, for example, it seeks to create transparency to be able to determine why an AI system has made a decision and who is advantaged or disadvantaged by this process (EU 2024). However, as Carr (2020) argues, for these regulations to be effective, they need to include people with living experience in their development and delivery.

As well as the tangible implications of policy, training, and involvement, this research also raises a broader question of whether AI and technology is changing what it means to be human and how it impacts public mental health. The data and Society Research Institute recommend using international human rights to guide AI applications to ensure that humans are informed about the impact that AI applications will have on individuals and society (Latonero, 2018). It is important that there are ongoing discussions and research that explore these issues across a wide range of socio-cultural contexts. (Eiroa-Orosa, & Rowe, 2017). It would be valuable to undertake longitudinal research with wider and more diverse groups of people with living experience to understand their views on AI, inclusion, and mental health.

Conclusion

Integrating AI in mental health care has the potential to transform mental health practices and improve outcomes, it also poses risks of exacerbating existing inequalities, biases, and digital divides. To address these risks and maximise the benefits of AI, it is essential to involve key stakeholders in the development of AI technologies. It is essential that people

with living experiences of MHCs have access to technology and an equal ‘voice’ in the development of AI technologies in mental health. This is not only important in promoting more inclusive technology but also inclusive citizenship across a wide range of cultural contexts. In promoting inclusive futures for people with MHC, participants in this study identified the importance of and their desire to be involved in the decisions that impact on their citizenship, which include the impact of technology. In the struggle to have an equal voice participants highlighted the importance of activism, co-production, and recognition of the need for transparency about how personal data will be used in AI innovations is a key priority. People with living experience provide valuable insights that could help shape AI tool that are relevant and effective for diverse populations and do not exacerbate existing inequalities and divisions. Co-created AI tools, developed in collaboration with other key stakeholders such as mental health professionals, and community organisations, are crucial for ensuring that AI technologies are equitable and retain a humancentric focus on mental health care and treatment. By adopting a collaborative and inclusive approach, we can harness the power of AI to address mental health inequalities and promote more equitable futures for the greater good of all. Global ethical regulations for AI can ensure the responsible development and implementation of AI technologies and provide a framework that addresses safety, fairness, citizenship, and human rights, while also respecting cultural differences and fostering international collaborations.

References

Alowais, S. A., & Huang, I. W. (2024). Revolutionizing Healthcare: The Role of Artificial Intelligence in Clinical Practice. *Angle Health Law Review*, (94), 95-132.

Arora, A., Alderman, J. E., Palmer, J., Ganapathi, S., Laws, E., Mccradden, M. D., .& Liu, X. (2023). The value of standards for health datasets in artificial intelligence-based applications. *Nature Medicine*, *29*(11), 2929-2938.

Balcombe, L., & De Leo, D. (2021). Digital mental health challenges and the horizon ahead for solutions. *JMIR Mental Health*, *8*(3), e26811.

Baumann, S.E., Lhaki, P. and Burke, J.G., (2020). Collaborative filmmaking: a participatory, visual research method. *Qualitative Health Research*, *30*(14), pp.2248-2264.

Bentley, S. V., Naughtin, C. K., McGrath, M. J., Irons, J. L., & Cooper, P. S. (2024). The digital divide in action: how experiences of digital technology shape future relationships with artificial intelligence. *AI and Ethics*, 1-15.

Beresford, P., (2019). Public participation in health and social care: exploring the Co-production of knowledge. *Frontiers in Sociology*, *3*, p.41.

Blandford, A. E. (2013). Semi-structured qualitative studies. Interaction Design Foundation.

Bohr, A., & Memarzadeh, K. (2020). The rise of artificial intelligence in healthcare applications. In *Artificial Intelligence in Healthcare* (pp. 25-60). Academic Press.

Bond, R. R., Mulvenna, M. D., Potts, C., O'Neill, S., Ennis, E., & Torous, J. (2023). Digital transformation of mental health services. *Npj Mental Health Research*, *2*(1), 13.

Bommersbach, T. J., McKean, A. J., Olfson, M., & Rhee, T. G. (2023). National trends in mental health–related emergency department visits among youth, 2011-2020. *JAMA*, *329*(17), 1469-1477.

Borrescio-Higa, F., & Valenzuela, P. (2021). Gender inequality and mental health during the COVID-19 pandemic. *International Journal of Public Health*, *66*, 1604220.

Braun, V. and Clarke, V., (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101.

Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis?. *Qualitative Research in Psychology*, *18*(3), 328-352.

Braun, V., & Clarke, V. (2023). Toward good practice in thematic analysis: Avoiding common problems and be (com) ing a knowing researcher. *International Journal of Transgender Health*, *24*(1), 1-6.

Campbell, C. (2020). Social capital, social movements and global public health: Fighting for health-enabling contexts in marginalised settings. *Social Science & Medicine*, *257*, 112153.

Carr, E.R. and Ponce, A.N., (2022). Supporting mental health recovery, citizenship, and social justice. *Community Mental Health Journal*,58(1), 1-9.

Carr, S. (2020). 'AI gone mental': engagement and ethics in data-driven technology for mental health. *Journal of Mental Health*, *29*(2), 125-130.

Chekroud, A. M., Bondar, J., Delgadillo, J., Doherty, G., Wasil, A., Fokkema, M., ... & Choi, K. (2021). The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry*, *20*(2), 154-170.

Choi, M. and Cristol, D., (2021). Digital citizenship with intersectionality lens: towards participatory democracy driven digital citizenship education. *Theory into Practice*, *60*(4), pp.361-370.

Cogan, N. (2024). Psychotraumatology and artificial intelligence: a public health approach. *Public Health Open Access*, *8*(1).

Cogan, N., MacIntyre, G., Stewart, A., Harrison-Millan, H., Black, K., Quinn, N., ... & O'Connell, M. (2022). Developing and establishing the psychometric properties of the Strathclyde Citizenship Measure: A new measure for health and social care practice and research. *Health & Social Care in the Community*, *30*(6), e3949-e3965.

Cogan, N. A., MacIntyre, G., Stewart, A., Tofts, A., Quinn, N., Johnston, G., ... & Rowe, M. (2021). "The biggest barrier is to inclusion itself": the experience of citizenship for adults with mental health problems. *Journal of Mental Health*, *30*(3), 358-365.

Connolly, Moira. (2012). "Futurology and mental health services: Are we ready for the demographic transition?" The Psychiatrist 36(5):161-164

Corrigan, P. (2004). How stigma interferes with mental health care. *American Psychologist*, *59*(7), 614.

Davenport, T. H., & Katyal, V. (2020). Every Leader's Guide to the Ethics of AI.

Deganis, I., Haghian, P. Z., Tagashira, M., & Alberti, A. (2021). Leveraging digital technologies for social inclusion. *United Nations Department of Economic and Social Affairs*.

Edwards, R. and Brannelly, T., 2017. Approaches to democratising qualitative research methods. *Qualitative Research*, *17*(3), pp.271-277.

Eiroa-Orosa, F. J., & Rowe, M. (2017). Taking the concept of citizenship in mental health across countries. Reflections on transferring principles and practice to different sociocultural contexts. *Frontiers in Psychology*, *8*, 1020.

Emejulu, A. and McGregor, C., (2019). Towards a radical digital citizenship in digital education. *Critical Studies in Education*, *60*(1), pp.131-147.

European Union (2024, 24th June), AI Act https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

Feijóo, C., Kwon, Y., Bauer, J. M., Bohlin, E., Howell, B., Jain, R., ... & Xia, J. (2020). Harnessing artificial intelligence (AI) to increase wellbeing for all: The case for a new technology diplomacy. *Telecommunications Policy*, *44*(6), 101988.

Fiske, A., Henningsen, P., & Buyx, A. (2019). Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *Journal of Medical Internet Research*, *21*(5), e13216.

Gerke, S., Minssen, T., & Cohen, G. (2020). Ethical and legal challenges of artificial intelligence-driven healthcare. In *Artificial intelligence in healthcare* (pp. 295-336). Academic Press.

Gibson, B., Schneider, J., Talamonti, D., & Forshaw, M. (2021). The impact of inequality on mental health outcomes during the COVID-19 pandemic: A systematic review. *Canadian Psychology/Psychologie Canadienne*, *62*(1), 101.

Gichoya, J. W., Thomas, K., Celi, L. A., Safdar, N., Banerjee, I., Banja, J. D., ... & Purkayastha, S. (2023). AI pitfalls and what not to do: mitigating bias in AI. *The British Journal of Radiology*, *96*(1150), 20230023.

Gray, M., Samala, R., Liu, Q., Skiles, D., Xu, J., Tong, W., & Wu, L. (2024). Measurement and mitigation of bias in artificial intelligence: a narrative literature review for regulatory science. *Clinical Pharmacology & Therapeutics*, *115*(4), 687-697.

Huang, Z., & Hu, Q. (2024). Tree hole rescue: an AI approach for suicide risk detection and online suicide intervention. *Health Information Science and Systems*, *12*(1), 45.

Hudson, C. G. (2012). Disparities in the geography of mental health: Implications for social work. *Social Work*, *57*(2), 107-119.

Jardine, J., Nadal, C., Robinson, S., Enrique, A., Hanratty, M., & Doherty, G. (2024). Between rhetoric and reality: Real-world barriers to uptake and early engagement in digital mental health interventions. *ACM Transactions on Computer-Human Interaction*, *31*(2), 1-59.

Jennings, H., Slade, M., Bates, P., Munday, E. and Toney, R., (2018). Best practice framework for Patient and Public Involvement (PPI) in collaborative data analysis of qualitative mental health research: methodology development and refinement. *BMC Psychiatry*, *18*(1), pp.1-11.

Johnson, J. L., Adkins, D., & Chauvin, S. (2020). A Review of the quality indicators of rigor in qualitative research. *American Journal of Pharmaceutical Education*, *84*(1), 7120.

Joseph, A. P., & Babu, A. (2024). The unseen dilemma of AI in mental healthcare. *AI & Society*, 1-3.

Iyortsuun, N. K., Kim, S. H., Jhon, M., Yang, H. J., & Pant, S. (2023, January). A review of machine learning and deep learning approaches on mental health diagnosis. In *Healthcare* (Vol. 11, No. 3, p. 285). MDPI.

Kalpakos, M. E. (2024). Defining the future: The AI Act's potential in equitably safeguarding fundamental rights and promoting AI innovation. *UFITA Archiv für Medienrecht und Medienwissenschaft*, *87*(1), 128-172.

Kirkbride, J. B., Anglin, D. M., Colman, I., Dykxhoorn, J., Jones, P. B., Patalay, P., ... & Griffiths, S. L. (2024). The social determinants of mental health and disorder: evidence, prevention, and recommendations. *World Psychiatry*, *23*(1), 58.

Kivimäki, M., Batty, G. D., Pentti, J., Shipley, M. J., Sipilä, P. N., Nyberg, S. T., ... & Vahtera, J. (2020). Association between socioeconomic status and the development of mental and physical health conditions in adulthood: a multi-cohort study. *The Lancet Public Health*, *5*(3), e140-e149.

Latonero, M. (2018). Governing artificial intelligence: Upholding human rights & dignity. *Data & Society*, *38*.

Lee, C. T., Palacios, J., Richards, D., Hanlon, A. K., Lynch, K., Harty, S., ... & Gillan, C. M. (2023). The Precision in Psychiatry (PIP) study: Testing an internet-based methodology for accelerating research in treatment prediction and personalisation. *BMC Psychiatry*, *23*(1), 25.

Lee, D., & Yoon, S. N. (2021). Application of artificial intelligence-based technologies in the healthcare industry: Opportunities and challenges. *International Journal of Environmental Research and Public Health*, *18*(1), 271.

Le Glaz, A., Haralambous, Y., Kim-Dufor, D. H., Lenca, P., Billot, R., Ryan, T. C., ... & Lemey, C. (2021). Machine learning and natural language processing in mental health: systematic review. *Journal of Medical Internet Research*, *23*(5), e15708.

Lythreatis, S., Singh, S. K., & El-Kassar, A. N. (2022). The digital divide: A review and future research agenda. *Technological Forecasting and Social Change*, *175*, 121359.

Koutsouleris, N., Hauser, T. U., Skvortsova, V., & De Choudhury, M. (2022). From promise to practice: towards the realisation of AI-informed mental health care. *The Lancet Digital Health*, *4*(11), e829-e840.

Lin, B., Tan, Z., Mo, Y., Yang, X., Liu, Y., & Xu, B. (2023). Intelligent oncology: The convergence of artificial intelligence and oncology. *Journal of the National Cancer Center*, *3*(1), 83-91.

Macintyre, A., Ferris, D., Gonçalves, B., & Quinn, N. (2018a). What has economics got to do with it? The impact of socioeconomic factors on mental health and the case for collective action. *Palgrave Communications*, *4*(1), 1-5.

MacIntyre, G., Cogan, N., Stewart, A., Quinn, N., O'Connell, M. and Rowe, M., (2021). Citizens defining citizenship: A model grounded in lived experience and its implications for research, policy, and practice. *Health & Social Care in the Community*, *30*(3), e695-e705.

MacIntyre, G., Cogan, N. A., Stewart, A. E., Quinn, N., Rowe, M., & O'Connell, M. (2019). What's citizenship got to do with mental health? Rationale for inclusion of citizenship as part of a mental health strategy. *Journal of Public Mental Health*, *18*(3), 157-161.

MacIntyre, G., Cogan, N., Stewart, A., Quinn, N., Rowe, M., & O'Connell, M. (2018b). Understanding citizenship within a health and social care context in Scotland using community based participatory research methods. *SAGE Research Methods Cases*.

Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: guided by information power. *Qualitative Health Research*, *26*(13), 1753-1760.

Mongelli, F., Georgakopoulos, P., & Pato, M. T. (2020). Challenges and opportunities to meet the mental health needs of underserved and disenfranchised populations in the United States. *Focus*, *18*(1), 16-24.

Morgan, P., Brannelly, T. &Eales, S., (2020). Future studies, mental health and the question of citizenship. *Mental Health and Social Inclusion*, *24*(1), 23-32.

Nazir, M. B., & Hussain, I. (2024). Revolutionizing Cardiac Care: AI and Deep Learning in Heart Health. *International Journal of Advanced Engineering Technologies and Innovations*, *1*(4), 189-208.

Ogugua, J. O., Okongwu, C. C., Akomolafe, O. O., Anyanwu, E. C., & Daraojimba, O. D. (2024). Mental health and digital technology: a public health review of current trends and responses. *International Medical Science Research Journal*, *4*(2), 108-125.

Olawade, D. B., Wada, O. Z., Odetayo, A., David-Olawade, A. C., Asaolu, F., & Eberhardt, J. (2024). Enhancing mental health with Artificial Intelligence: Current trends and future prospects. *Journal of medicine, surgery, and public health*, 100099.

Ozmen Garibay, O., Winslow, B., Andolina, S., Antona, M., Bodenschatz, A., Coursaris, C., ... & Xu, W. (2023). Six human-centered artificial intelligence grand challenges. *International Journal of Human–Computer Interaction*, *39*(3), 391-437.

Peters, D., Vold, K., Robinson, D., & Calvo, R. A. (2020). Responsible AI—two frameworks for ethical design practice. *IEEE Transactions on Technology and Society*, *1*(1), 34-47.

Quinn, N., Bromage, B. & Rowe, M., (2020). Collective citizenship: from citizenship and mental health to citizenship and solidarity. *Social Policy & Administration*, *54*(3), 361-374.

Reis, G., Bromage, B., Rowe, M., Restrepo-Toro, M.E., Bellamy, C., Costa, M. and Davidson, L., (2022). Citizenship, social justice and collective empowerment: living outside mental illness. *Psychiatric Quarterly*, *93*(2), pp.537-546.

Roosan, D. (2024). Integrating Artificial Intelligence with Mixed Reality to Optimize Health Care in the Metaverse. In *Augmented and Virtual Reality in the Metaverse* (pp. 247-264). Cham: Springer Nature Switzerland.

Rowe, M. (2015). *Citizenship and mental health*. Oxford University Press.

Saeed, S. A., & Masters, R. M. (2021). Disparities in health care and the digital divide. *Current Psychiatry Reports*, *23*, 1-6.

Shah, V. (2022). AI in Mental Health: Predictive Analytics and Intervention Strategies. *Journal Environmental Sciences and Technology*, *1*(2), 55-74.

Shalaby, R. A. H., & Agyapong, V. I. (2020). Peer support in mental health: literature review. *JMIR Mental Health*, *7*(6), e15572.

Shaw, J., Rudzicz, F., Jamieson, T., & Goldfarb, A. (2019). Artificial intelligence and the implementation challenge. *Journal of Medical Internet Research*, *21*(7), e13659.

Shim, R. S. (2021). Dismantling structural racism in psychiatry: a path to mental health equity. *American Journal of Psychiatry*, *178*(7), 592-598.

Sinanan, J., & McNamara, T. (2023). Great AI divides? Automated decision-making technologies and dreams of development. In *The Dark Social* (pp. 87-100). Routledge.

Skorburg, J. A., O'Doherty, K., & Friesen, P. (2024). Persons or data points? Ethics, artificial intelligence, and the participatory turn in mental health research. *American Psychologist*, *79*(1), 137.

Spanakis, P., Peckham, E., Mathers, A., Shiers, D., & Gilbody, S. (2021). The digital divide: amplifying health inequalities for people with severe mental illness in the time of COVID-19. *The British Journal of Psychiatry*, *219*(4), 529-531.

Sukhera, J., Knaak, S., Ungar, T., & Rehman, M. (2022). Dismantling structural stigma related to mental health and substance use: an educational framework. *Academic Medicine*, *97*(2), 175-181.

Tahvanainen, L., Tetri, B., & Ahonen, O. (2024, May). Exploring and Extending Human-Centered Design to Develop AI-Enabled Wellbeing Technology in Healthcare. In *Nordic Conference on Digital Health and Wireless Solutions* (pp. 288-306). Cham: Springer Nature Switzerland.

Thieme, A., Hanratty, M., Lyons, M., Palacios, J., Marques, R. F., Morrison, C., & Doherty, G. (2023). Designing human-centered AI for mental health: Developing clinically relevant applications for online CBT treatment. *ACM Transactions on Computer-Human Interaction*, *30*(2), 1-50.

Timmons, A. C., Duong, J. B., Simo Fiallo, N., Lee, T., Vo, H. P. Q., Ahle, M. W., ... & Chaspari, T. (2023). A call to action on assessing and mitigating bias in artificial intelligence applications for mental health. *Perspectives on Psychological Science*, *18*(5), 1062-1096.

Torous, J., Bucci, S., Bell, I. H., Kessing, L. V., Faurholt-Jepsen, M., Whelan, P., ... & Firth, J. (2021). The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry*, *20*(3), 318-335.

Trewin, S., Basson, S., Muller, M., Branham, S., Treviranus, J., Gruen, D., ... & Manser, E. (2019). Considerations for AI fairness for people with disabilities. *AI Matters*, *5*(3), 40-63.

Rowe, M. (2015). *Citizenship and mental health*. Oxford University Press.

Van Deursen, A. J., & Helsper, E. J. (2015). The third-level digital divide: Who benefits most from being online?. In *Communication and information technologies annual* (Vol. 10, pp. 29-52). Emerald Group Publishing Limited.

Volti, R., & Croissant, J. (2024). *Society and technological change*. Waveland Press.

Wang, C., Boerman, S. C., Kroon, A. C., Möller, J., & H de Vreese, C. (2024). The artificial intelligence divide: Who is the most vulnerable?. N*ew Media & Society*, 14614448241232345.

Whittlestone, J., Nyrup, R., Alexandrova, A., Dihal, K., & Cave, S. (2019). Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research. *London: Nuffield Foundation*.

Williamson, S. M., & Prybutok, V. (2024). Balancing privacy and progress: a review of privacy challenges, systemic oversight, and patient perceptions in AI-driven healthcare. *Applied Sciences*, *14*(2), 675.

Zidaru, T., Morrow, E. M., & Stockley, R. (2021). Ensuring patient and public involvement in the transition to AI-assisted mental health care: A systematic scoping review and agenda for design justice. *Health Expectations*, *24*(4), 1072-1124.