Digital Futures in Mind: Reflecting on Technological Experiments in Mental Health & Crisis Support

Jonah Bossewitch, Lydia X. Z. Brown, Piers Gooding, Leah Harris, James Horton, Simon Katterl, Keris Myrick, Kelechi Ubozoh and Alberto Vasquez

DIGITAL FUTURES IN MIND: REFLECTING ON TECHNOLOGICAL EXPERIMENTS IN MENTAL HEALTH & CRISIS SUPPORT

Suggested citation: Jonah Bossewitch, Lydia X. Z. Brown, Piers Gooding, Leah Harris, James Horton, Simon Katterl, Keris Myrick, Kelechi Ubozoh and Alberto Vasquez, *Digital Futures in Mind: Reflecting on Technological Experiments in Mental Health & Crisis Support* (University of Melbourne, 2021) <https://automatingmentalhealth.cc/>

ACKNOWLEDGEMENTS

Ms Ann Njambi and Ms Charity Muturi provided invaluable feedback on the virtual support network in Kenya (page 39).

PG: This work wouldn't have occurred without funding and support from the Mozilla Foundation. Thanks to Jenn Beard and all Mozilla Fellows of 2019-20. Invaluable feedback on an early draft was provided by A/Prof Nev Jones. Stephanie Slack, Timothy Kariotis and David Clifford assisted at different drafting stages. Support from the Melbourne Law School and its Office for Research, as well as the Melbourne Social Equity Institute, was also crucial. The report was partly produced with funding from the Australian Research Council (project no. DE200100483).

KU: I want to acknowledge the co-authors of this paper, and a special acknowledgment for Piers Gooding for modeling that our future technologies should be more than just inclusive of lived experience, but fully integrate the people who will be most impacted by the proposed solutions to co-create and co-develop culturally responsive real mental health solutions that respond to our needs.

Images for this report were generously provided by the Science Gallery Melbourne at the University of Melbourne, drawn from its inaugural exhibition **MENTAL: Head inside** from January – June 2022. Science Gallery Melbourne is part of the global Science Gallery Network pioneered by Trinity College Dublin.

This report was largely written on the unceded lands of the Wurundjeri People of the Kulin Nation.

Table of Contents

Foreword

[TEXT TO BE ADDED]

[TEXT TO BE ADDED]

Your Face is Muted by Tilman Dingler, Zhanna Sarsenbayeva, Eylül Ertay, Hao Huang and Melanie Huang in Science Gallery Melbourne's MENTAL. Photo by Alan Weedon.

Introduction

Urgent public attention is needed to make sense of the expanding use of algorithmic and data-driven technologies in the mental health context. On the one hand, well-designed digital technologies that offer high degrees of public involvement can be used to promote good mental health and crisis support in communities. They can be employed safely, reliably and in a trustworthy way, including to help build relationships, allocate resources, and promote human flourishing.[1]

On the other hand, there is clear potential for harm. The list of 'data harms' in the mental health context is growing longer, in which people are in worse shape than they would be had the activity not occurred.[2] Examples in this report include the hacking of psychotherapeutic records and the extortion of victims, algorithmic hiring programs that discriminate against people with histories of mental healthcare, and criminal justice and border agencies weaponising data concerning mental health against individuals. Issues also come up not where technologies are misused or faulty, but where technologies like biometric monitoring or surveillance work as intended, and where the very process of 'datafying' and digitising individuals' behaviour – observing, recording and logging them to an excessive degree – carry inherent harm.

Public debate is needed to scrutinise these developments. Critical attention must be given to current trends in thought about technology and mental health, including the values such technologies embody, the people driving them, and their diverse visions for the future. Some trends – for example, the idea that 'objective digital biomarkers' in a person's smartphone data can identify 'silent' signs of pathology, or the entry of Big Tech into mental health service provision – have the potential to create major changes not only to health and social services but to the very way human beings experience ourselves and our world. This possibility is also complicated by the spread of 'fake and deeply flawed' or 'snake oil' AI,[3] and the tendency in the technology sector – and indeed in mental health sciences[4] – to over-claim with the promise of a 'silver bullet' and, inevitably,under-deliver.

Meredith Whitaker and colleagues at the *AI Now* research institute observe that disability and mental health have been largely omitted from discussions about AI-bias and algorithmic accountability.[5] This report brings them to the fore. It is written to promote basic standards of algorithmic and technological transparency and auditing, but also takes the opportunity to ask more fundamental questions, such as whether algorithmic and digital systems should be used at all in some circumstances—and if so, who gets to govern them.[6] These issues are particularly important given the COVID-19 pandemic, which has accelerated the digitisation of physical and mental health services worldwide,[7] and driven more of our lives online.

1 Claudia Lang, 'Craving to Be Heard but Not Seen – Chatbots, Care and the Encoded Global Psyche', *Somatosphere* (13 April 2021) <http://somatosphere.net/2021/chatbots.html/>. Lang describes the potential for tech to 'weave together code and poetry, emotions and programming, despair and reconciliation, isolation and relatedness in human-techno worlds.'

2 Joanna Redden, Jessica Brand and Vanesa Terzieva, 'Data Harm Record – Data Justice Lab', *Data Justice Lab* (August 2020) <https://datajusticelab.org/data-harm-record/>.

3 Frederike Kaltheuner (Ed.) *Fake AI* (Meatspace Press, 2021) <https://fakeaibook.com/> (accessed 7/12/2021).

4 Anne Harrington, *Mind Fixers: Psychiatry's Troubled Search for the Biology of Mental Illness* (W. W. Norton & Company, 2019

5 Meredith Whittaker et al, *Disability, Bias, and AI* (AI Now, November 2019) 8.

6 Frank Pasquale, 'The Second Wave of Algorithmic Accountability', *Law and Political Economy* (25 November 2019) <https://lpeblog.org/2019/11/25/the-second-wave-of-algorithmic-accountability/>.

7 John Torous et al, 'Digital Mental Health and COVID-19: Using Technology Today to Accelerate the Curve on Access and Quality Tomorrow' (2020) 7(3) *JMIR Mental Health* e18848.

0.1 Structure

Part 1 charts the rise of algorithmic and data-driven technology in the mental health context. It outlines issues which make mental health unique in legal and policy terms, particularly the significance of involuntary or coercive psychiatric interventions in any analysis of mental health and technology. The section makes a case for elevating the perspective of people with lived experience of profound psychological distress, mental health conditions, psychosocial disabilities, and so on, in all activity concerning mental health and technology.

Part 2 looks at prominent themes of accountability. Eight key themes are discussed – fairness and non-discrimination, human control of technology, professional responsibility, privacy, accountability, safety and security, transparency and explainability, and promotion of public interest. International law, and particularly the Convention on the Rights of Persons with Disabilities, is also discussed as a source of data governance.

Case studies throughout show the diversity of technological developments and draw attention to their real-life implications. Many case studies demonstrate instances of harm. This may seem overly negative to some readers. Yet, there is a lack of readily available resources that list real and potential harms caused by algorithmic and data-driven technologies in the mental health and disability context. In contrast, there is an abundance of public material promoting their benefit. This report seeks to rebalance public deliberation and promote a conversation about public good and harm, and what it would take to govern such technological initiatives responsibly. The case studies also seek to ground discussion in the actual agonies of existing technology rather than speculative worries about technology whose technical feasibility is often exaggerated in misleading and harmful ways (for example, Elon Musk's claim that his 'AI-brain chips will "solve" autism and schizophrenia').[8]

This resource is meant for diverse audiences, including advocates and activists concerned with mental health and disability, service users and those who have experienced mental health interventions and their representative organisations, clinical researchers, technologists, service providers, policymakers, regulators, private sector actors, academics, and journalists.

0.2 How was the Report Written?

This report emerged from a two-year exploration conducted throughout 2020 and 2021. The work was undertaken by the authors, with backgrounds in media studies, policymaking, law, data ethics, and so on, and most of whom have had firsthand encounters with mental health services, distress or disability. The report co-ordinator, Piers Gooding, received funding as a Mozilla Fellow in 2020. With Simon Katterl, Piers led the drafting of Part 1 and 2 of the report, with guidance from the other co-authors. The report recommendations were jointly and equally authored.

8 Isobel Asher Hamilton, 'Elon Musk Said His AI-Brain-Chips Company Could "solve" Autism and Schizophrenia', *Business Insider Australia* (14 November 2019) <https://www.businessinsider.com.au/elon-musk-said-neuralink-could-solve-autism-and-schizophrenia-2019-11>.

0.3 What Recommendations Does the Report Make?

These broad recommendations seem justified based on the discussion in this report.[9]

1. *The well documented negative impacts of algorithmic and data-driven technology on people in extreme distress, persons with psychosocial disability, people with lived experience of mental health issues, and so on, need to be openly acknowledged and rectified by governments, business, national human rights institutions, civil society and people with lived experience of distress and disabilities working together.*[10]

2. *Authentic, active, and ongoing engagement with persons with lived experience of distress and disability and their representative organisations is required at the earliest exploratory stages in the development, procurement and deployment of algorithmic and data-driven technology that directly impacts them. This engagement is required under the Convention on the Rights of Persons with Disabilities, and is key to technology being a force for good in the mental health context. Instead of more technology 'for' or 'about' distressed and disabled people and the collection of vast amounts of data to be fed into opaque processes, these groups themselves should be steering discussions on when and how emerging technologies should be integrated into mental health and crisis responses – if at all. True partnership and engagement with people with lived experience should include compensation for their time and true decision-making power which counteracts tokenisation and minimal involvement.*

3. *'Techno-solutionism'*[11] *must be resisted, in which digital initiatives in the mental health context are presented as self-evidently virtuous and effective, and a simple fix to the complex issues of human distress, anguish and existential pain. Not only must proven and potential harms be squarely acknowledged, but so must unproven benefits. Technology is not neutral. When new technologies are presented as technocratic and apolitical, this overlooks the significant role of human decision-making, power, finance, and social trust, which should be part of public discussion. Fundamental questions must be asked as to whether certain systems should be built at all, whether proposals are technically feasible (or merely unrealistic and over-hyped), and – if they are to be pursued – who should govern them.*

4. *Given the limited (and sometimes highly limited) evidence-base for many algorithmic and data-driven technologies in the mental health context, standards are required that are developed with active involvement of people with lived experience and disability, for use as a mechanism for consensus on scientific integrity standards. This involvement can help limit the many sensational and misleading claims about what AI and other algorithmic technology can achieve and curb their use as cheap alternatives to well-resourced face-to-face support. Government funding for digital initiatives in the mental health and disability context should be dependent upon submissions regarding stringent evidence of safety and efficacy, and in accordance with disability-inclusive public-procurement standards.*

9 An important caveat is that we are not a representative body. The authors are based in high-income countries, as the initial scope of this project looked to regulatory arrangements in the EU and US. The recommendations are not meant to be exhaustive, and nor should they foreclose other strategies and recommendations, particularly by persons with lived experience of mental health crises, disabled people, and their representative organisations.

10 This recommendation draws from the 2021 thematic report on artificial intelligence and disability by the UN Special Rapporteur for the Rights of Persons with Disabilities, Gerard Quinn. Human Rights Council, *Report of the Special Rapporteur on the Rights of Persons with Disabilities* (UN Doc A/HRC/49/52, 28 December 2021) para 73 <https://undocs.org/pdf?symbol=en/A/HRC/49/52>.

11 Evgency Morozov coined this term to describe a pervasive ideology that recasts complex social phenomena like politics, public health, education, and law enforcement as "neatly defined problems with definite, computable solutions or as transparent and self-evident processes that can be easily optimized—if only the right algorithms are in place!" Evgeny Morozov, *To Save Everything, Click Here: Technology, Solutionism, and the Urge to Fix Problems That Don't Exist* (Penguin UK, 2013) 5.

5. *There should be an immediate cessation to all algorithmic and data-driven technological interventions in the mental health context that have a significant impact on individuals' lives that are imposed without the free and informed consent of the person concerned. Regarding algorithmic forms of diagnosis or proxy-diagnosis, the consequences of being diagnosed and pathologised in the mental health context, whether accurately or not, are often profound. Such measures should never be undertaken without the free and informed consent of the person. Among other things, informed consent processes should provide explicit details of data safety and security measures, and clarify who shall monitor compliance.*

6. *Governments, private companies, not-for-profits, and so on must, at a minimum, eliminate forms of mental health- and disability-based bias and discrimination from algorithmic and data-driven systems, particularly in areas such as employment, education and insurance. Such steps should extend to preventing discrimination against people who are marginalised across intersecting lines of race, gender, sexual orientation, class, and so on. Those facing discrimination must have access to an effective and accessible remedy, such as a clear source for complaints and legal review.*

7. *Ethical standards will never be enough. Robust legal and regulatory frameworks are required that acknowledge the risks of employing algorithmic and data-driven technologies in response to distress, mental health crises, disability support needs, and so on. As part of this, a legal and regulatory framework is required that effectively prohibits systems that by their very nature will be used to cause unacceptable individual and social harms and infringe human rights. This could include:*

    a. *mandatory, publicly accessible, and contestable impact assessments for forms of automation and digitisation to determine the appropriate safeguards, including the potential for prohibiting uses that infringe on fundamental rights;*
    b. *proportionality testing of risks against any potential benefits to ensure opportunities to interrogate the objectives, outcomes and inherent trade-offs involved in using algorithmic systems, and doing so in a way that centres the interest of the affected groups, not just the entity using the system such as a healthcare service or technology start-up;*[12]
    c. *strengthening non-discrimination rules concerning mental health and psychosocial disability to prevent harms caused by leaked, stolen, or traded data concerning mental health and disability.*[13]

8. *Public sector accountability needs to be strengthened, including adequately resourcing relevant institutions, which will be vital to addressing the dangers of private sector actors, not-for-profits and government agencies that (mis)use people's data concerning mental health. This includes developing a willing and empowered state-sponsored regulatory framework as well as resources for affected people and civil society organisations to proactively contribute to enforcement. This includes supporting the capacity-building of representative organisations of service users and persons with disabilities to effectively monitor the impact of data driven technology on persons with lived experience of mental health crises or disability. Monitoring could include: advocating for responsible and inclusive data-driven technology, interacting effectively with all key actors including the private sector, and highlighting harmful or discriminatory uses of the technology.*[14]

12 Alexandra Givens, 'Algorithmic Fairness for People With Disabilities: The Legal Framework' (Georgetown Institute for Tech Law & Policy, 27 October 2019) <https://docs.google.com/presentation/d/1EeaaH2RWxmzZUBSxKGQOGrHWom0z7UdQ/present?ueb=true&slide=id.p17&usp=embed_facebook>.

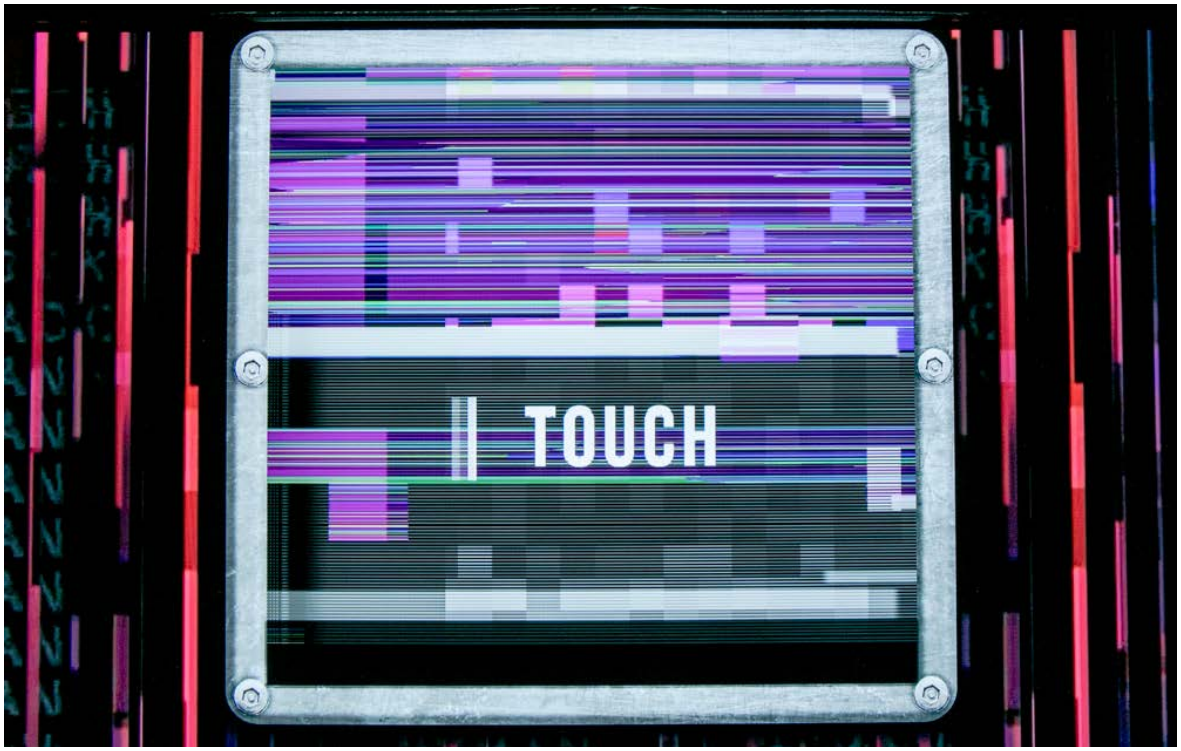
13 Mason Marks, 'Algorithmic Disability Discrimination' in Anita Silvers et al (eds), *Disability, Health, Law, and Bioethics* (Cambridge University Press, 2020) 242

14 Human Rights Council, 'Report of the Special Rapporteur on the Rights of Persons with Disabilities' (n 10). Para 76(g).

9. *Robust civil society responses are more likely where lived experience groups and disabled people and their representative organisations connect with other activists at the intersections of race, gender, class, and other axes of oppression, rather than viewing algorithmic and data-driven injustices purely through a mental health or disability lens. This could include collectives, nonprofit technology organisations, free and open source projects, philanthropic funders and activists with data practices and skills that help them more fully realise their missions. Those working for economic, social, racial and climate justice can share digital tools, resources and practices to help maximize their effectiveness and impact and, in turn, change the world.*[15]

10. *Interdisciplinary academic input is needed beyond disciplines like medicine, psychology, computer science and engineering, to include researchers from the humanities and social sciences. This will help address the common presentation of algorithmic and data-driven technologies as neutral—as facilitating factual, un-mediated, digital processing. This technocratic framing neglects matters including the significant role of power, the social and economic underpinnings of distress, unjust macroeconomic structures, Big Tech hegemony, and so on.*

11. *Steps must be taken to prevent the undercutting of face-to face encounters of care and support, particularly where private sector interests are expanding into digitised responses to distress or care, and particularly where governments are pursuing digital options as cheap alternatives to well-resourced forms of support. Relations of care and support must be adequately recognised and protected. The over-emphasis of metrics and computational approaches should be resisted in appreciation of the virtues that make for a truly human life.*



Echo by Georgie Pinn in Science Gallery Melbourne's MENTAL. Photo by Alan Weedon.

15 Language for this recommendation is borrowed from 'Aspiration Manifesto | Aspiration' <https://aspiration tech.org/publications/manifesto>. This recommendation comes with a call for caution about the potential misalignment of non-profit organisations with desired aims, and we draw attention to calls to breakaway from the non-profit industrial complex, including turning toward potential alternatives such as grassroots movements and worker self-directed non-profits that aim to improve the accountability and participatory nature of social movements. Jake Goldenfein and Monique Mann, 'Tech Money in Civil Society: Whose Interests Do Digital Rights Organisations Represent?' (2022) 0(0) *Cultural Studies* 1.

0.4 A Note on Terminology

It can be challenging to find clear terminology in an area of rapid technological change.

'Algorithmic and data-driven technologies' will be used to cover diverse technologies that use contemporary computer processing to analyse large amounts of data algorithmically. This includes technology variously described as 'artificial intelligence' (AI), 'machine learning', 'neural networks', predictive analytics, 'deep learning', natural language processing, robotics, speech processing and other forms of automation, that are used for the purposes of making decisions.[16]

Other data-driven technologies that don't explicitly use contemporary algorithmic technology – such as electronic records management software, online-counselling platforms, and even some forms of machine learning – remain relevant, as they form part of a broader communication ecosystem that can generate and transmit data concerning mental health.

We use the term **'communication ecosystem'** to refer to the complex, global networks of information and communication technology. This contemporary communications environment encompasses disparate systems – such as the web, the Internet, and various public and private intranets – which are increasingly converging to create massive, complex, and interconnected flows of data. Other technical terms related to specific technologies will be defined as they arise throughout the report.

The aim of the report is to contribute to public governance. **'Public governance'** includes law and policy but also extends to professional and ethical standards and guidelines, industry norms, civil society advocacy, and cultural expectations around what members of the public consider socially acceptable. Attention to the diverse relationships of power between these mechanisms can help identify the obligations of those employing data-driven technologies, and the rights of those who use and/or are subject to them.

0.5 Minding Language about Mental Health and Technology

Finding appropriate terminology in global discussions about mental health is also challenging. There is no single set of definitions to describe people's experience of mental health. Indeed notions of 'mental health' are contested. Different terms may be preferred according to national and cultural norms, professional conventions, and so on.

'people with lived experience and psychosocial disability' will be used in this report to describe people with firsthand experience of mental health services, mental health crises, extreme distress, psychosis, and so on. We have also sought terminology that conveys our intent to a diverse audience, though we acknowledge that language around mental health is often contested.[17]

People with lived experience can have varying ways of understanding experiences that are often called 'mental health conditions', 'mental health challenges' or 'mental illness'. We acknowledge that mental health can be described using terms such as emotional distress, trauma, mental health crisis and neurodiversity; and that people may describe themselves as 'service users', 'consumers', 'psychiatric survivors', 'disabled', 'ex-patients' and so on, and others may reject designations of their experiences in the terms of psychiatry and psychology.

16 Claude Castelluccia et al, *Understanding Algorithmic Decision-Making: Opportunities and Challenges* (2019) <http://www.europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS_STU(2019)624261_EN.pdf>

17 LD Green and Kelechi Ubozoh, *We've Been Too Patient: Voices from Radical Mental Health--Stories and Research Challenging the Biomedical Model* (North Atlantic Books, 2019).

The term **'people with psychosocial disability'** may not be familiar to readers. It is simply used to refer to people with disability related to mental health. The term has become prominent since the UN Convention on the Rights of Persons with Disabilities ('CRPD') came into force in 2008, and its definition is crucial for this report. The CRPD establishes that 'disability' includes 'mental impairment' (Article 1). Importantly, the CRPD also covers people who experience harms due to *imputed* impairment or disability—that is, where a person is *perceived* by others to be impaired or disabled. This is important here because some algorithmic technologies purportedly function to deduce the inner states of human beings, including *inferring mental health conditions and cognitive impairments*. For example, at the time of writing, Apple is working with multinational biotechnology company Biogen and UCLA to explore using sensor data (such as mobility, sleep patterns, swiping patterns and more) to infer mental health and cognitive decline.[18] Regardless of the accuracy of such predictions there remains the very real possibility of harms against people based on such data, again, *even if those data are false, misleading or inaccurate*.

The wide definition of psychosocial disability also highlights that all persons may interact with systems that generate data concerning their 'mental health'. If you carry a smartphone into a counselling service, visit a depression information website, write about distress on a social media platform, or even simply type and scroll on a mobile device, then various 'digital trails' will be generated that could be used to infer particular mental states—including mental health conditions, distress and cognitive impairment.[19] This is not to endorse or accept such claims but to reiterate that the rise of Big Data and AI has increased the likelihood of *inferences* and *predictions* being drawn from the behaviours, preferences, and private lives of individuals. These data have been described variously as 'emergent health data',[20] or 'indirect, inferred, and invisible health data'.[21] These forms of sensitive personal data create new opportunities for discriminatory, biased, and invasive decision-making about individuals and populations.[22] High risk applications of technologies are likely to have the greatest impact on people who are traditionally marginalised, such as those who are using mental health services or those who are subject to involuntary mental health interventions, and those with intersecting forms of marginalisation, but the range of technologies discussed in this report raise issues that affect everyone.[23]

The term **'data concerning mental health'** is therefore used throughout this report to mean personal data related to the mental health of a person, including the provision of health care services, which reveal information about his, her, their mental health status.[24] Questions remain about what exactly constitutes 'data concerning mental health'. The rising use of indirect, inferred, and 'invisible' health data is blurring the boundary between 'patient', 'service user', 'consumer' and 'citizen', creating multiple issues around the commodification and commercialisation of health, the rise of 'bio-surveillance', and other issues which have profound implications in the mental health context and beyond.

This report occasionally departs from these key terms when referring to specific data sources, describing research, or quoting an individual or organization, to accurately reflect the views presented in these materials.

18 Rolfe Winkler, 'WSJ News Exclusive | Apple Is Working on IPhone Features to Help Detect Depression, Cognitive Decline', *Wall Street Journal* (online, 21 September 2021) <https://www.wsj.com/articles/apple-wants-iphones-to-help-detect-depression-cognitive-decline-sources-say-11632216601>.

19 Rachel Metz, 'The Smartphone App That Can Tell You're Depressed before You Know It Yourself' (15 October 2018) *MIT Technology Review* <https://www.technologyreview.com/s/612266/the-smartphone-app-that-can-tell-youre-depressed-before-you-know-it-yourself/>; Christophe Olivier Schneble, Bernice Simone Elger and David Martin Shaw, 'All Our Data Will Be Health Data One Day: The Need for Universal Data Protection and Comprehensive Consent' (2020) 22(5) *Journal of medical Internet research* e16879; Rui Wang, Andrew T Campbell and Xia Zhou, 'Using Opportunistic Face Logging from Smartphone to Infer Mental Health: Challenges and Future Directions' in *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers* (Association for Computing Machinery, 2015) 683 <https://doi-org.ezp.lib.unimelb.edu.au/10.1145/2800835.2804391>.

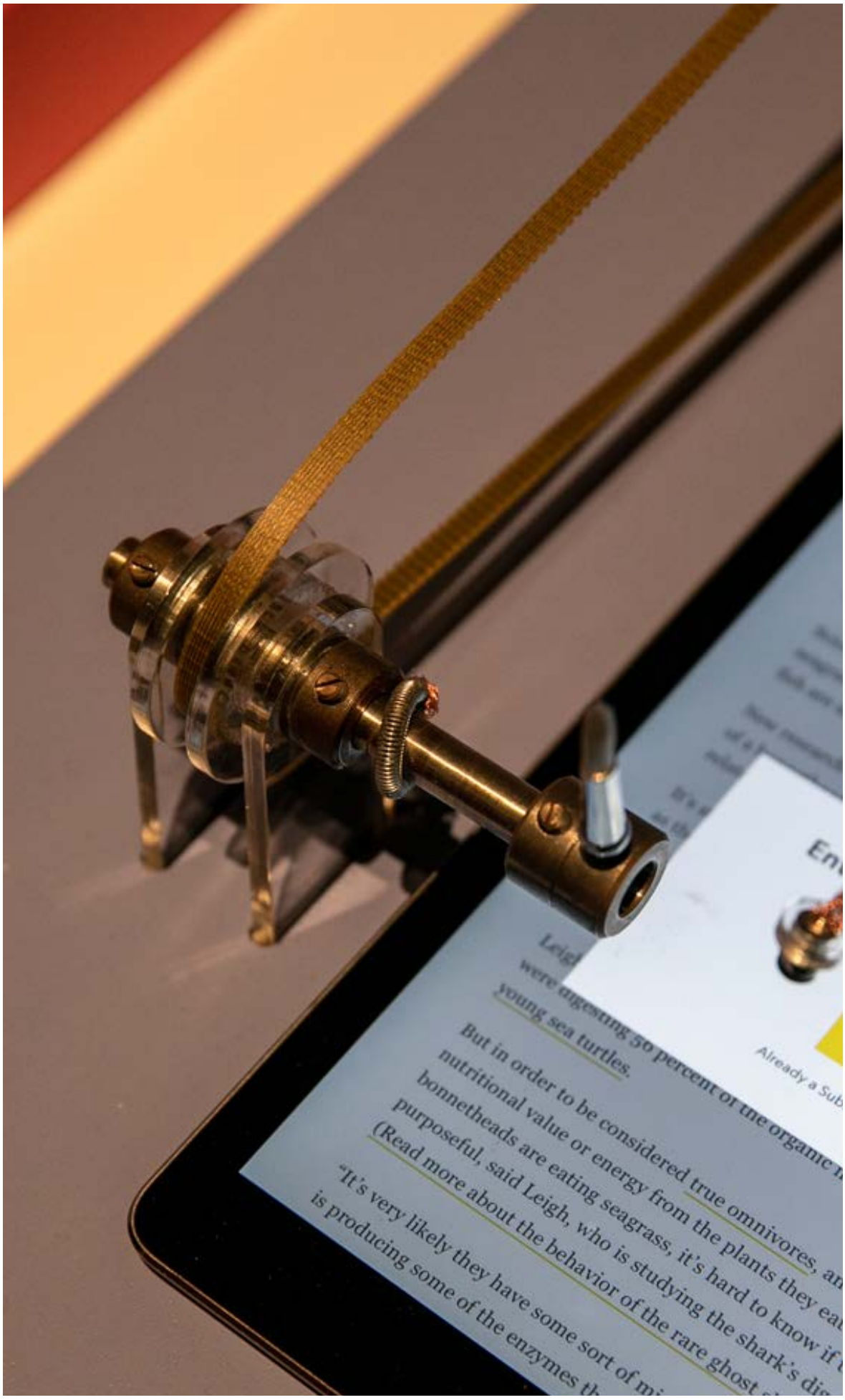
20 Mason Marks, *Emergent Medical Data: Health Information Inferred by Artificial Intelligence* (SSRN Scholarly Paper No ID 3554118, Social Science Research Network, 14 March 2020) <https://papers.ssrn.com/abstract=3554118>

21 Schneble, Elger and Shaw (n 19).

22 Sandra Wachter and Brent Mittelstadt, 'A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI' (2019) 2019(2) *Columbia Business Law Review* 494.

23 This point was borrowed from the excellent report on new technologies of migration management by Petra Molnar. See P Molnar (2019) 'Technological Testing Grounds: Migration Management Experiments and Reflections from the Ground Up' *eDRI*, p.9. <https://edri.org/wp-content/uploads/2020/11/Technological-Testing-Grounds.pdf> (accessed 2/03/2020)

24 This definition draws on the GDPR definition of 'data concerning health', which is defined as 'personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status.' EU GDPR, Article 4 (15).



Stop the Algorithm by Stephanie Kneissl and Max Lackner in Science Gallery Melbourne's MENTAL. Photo by Alan Weedon.

Part 1 - The Rise of Data and Automation in Mental Health

Although 'mental health' is often presented as a purely technical or clinical issue, it is highly political. Controversies abound, including over the language used to describe the issue,[25] the experts who should respond to it,[26] the distribution of resources to help those in need,[27] the use of forced psychiatric intervention by the state and other forced interventions in the name of care,[28] the expansion of psychiatric and psychological ideas to public understandings of human distress and wellbeing,[29] and the socio-political conditions that contribute to profound distress and mental health crises,[30] to name a few contested issues. It is in these political, regulatory, and epistemic struggles that new 'digital mental health technologies' appear.

According to prominent reports, algorithmic and data-driven technology is expanding rapidly in mental health settings. Prominent mental health practitioners and professional associations present algorithmic and data-driven technologies as a way to address the 'global mental health treatment gap'.[31] It can bring about 'radical change', some argue, with the potential for 'scalability' of interventions and unconstrained reach that 'can help reach billions of people.'[32]

Several governments have embraced digital technologies in mental healthcare as a cost-effective, accessible alternative or supplement to face-to-face support. In 2017 in the United Kingdom (UK), for example, former Prime Minister Theresa May announced 'a £67.7million digital mental health package.'[33] In the US between 2009-2015, the National Institute of Mental Health funded $445 million worth of projects concerned with 'technology-enhanced mental health interventions'.[34]

Market interests also play a major role in advancing the proposed digital turn in mental health.

Thoughtforms by Dr Kellyann Geurts and Dr Indae Hwang in Science Gallery Melbourne's MENTAL. Photo by Alan Weedon.

25 Anne Cooke and Peter Kinderman, 'But What about Real Mental Illnesses?" Alternatives to the Disease Model Approach to "schizophrenia' (2018) 58(1) *Journal of humanistic psychology* 47.

26 Indigo Daya, Bridget Hamilton and Cath Roper, 'Authentic Engagement: A Conceptual Model for Welcoming Diverse and Challenging Consumer and Survivor Views in Mental Health Research, Policy, and Practice' (2020) 29(2) *International journal of mental health nursing* 299.

27 Dainius Pūras and Piers Gooding, 'Mental Health and Human Rights in the 21st Century' (2019) 18(1) *World Psychiatry* 42.

28 Dinah Miller and Annette Hanson, *Committed: The Battle over Involuntary Psychiatric Care* (John Hopkins University Press Baltimore, 2016).

29 Nikolas Rose, *Our Psychiatric Future* (John Wiley & Sons, 2018).

30 Nikolas Rose et al, 'The Social Underpinnings of Mental Distress in the Time of COVID-19 – Time for Urgent Action' (2020) 5 *Wellcome Open Research* <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7411522/>.

31 Vikram Patel et al, 'The Lancet Commission on Global Mental Health and Sustainable Development' (2018) 392(10157) *The Lancet* 1553.

32 Dinesh Bhugra et al, 'The WPA-Lancet Psychiatry Commission on the Future of Psychiatry' (2017) 4(10) *The Lancet Psychiatry* 775, p.803.

33 HM Government, 'Prime Minister Unveils Plans to Transform Mental Health Support', *GOV.UK* (9 January 2017) <https://www.gov.uk/government/news/prime-minister-unveils-plans-to-transform-mental-health-support>

34 National Institute of Mental Health, 'NIMH » Technology and the Future of Mental Health Treatment' (2017)

Global digital health market [US]$118 billion worldwide
A White Paper published by the World Economic Forum states that: The global digital health market has been valued at [US]$118 billion worldwide, with mental health being one of the fastest-growing sectors.

US$1.8 billion in venture-capital funding in 2020
Businesses in "digital behavioural health" reportedly raised $1.8 billion in venture-capital funding in 2020, compared to $609 million in 2019.[35]

Gobal mental health software market: US$4,585m by 2027
According to market speculators, Zion Market Research, the '[g]lobal mental health software market [is] expected to generate revenue of around US$4,585 million by end of 2026.'[36]

Global digital health market expected to reach US$660 by 2026
This growth is mirrored in the 'global digital health market' more broadly, which Statista.com suggests will increase to around US$660 billion dollars by 2026.

10,000+ mental health apps
Over 10,000 apps concerned with mental health are now available for download and use.[37]

Major technology corporations – who happen to be also the largest corporations in the world – have increasingly turned their attention to healthcare activity, with each major firm now appointing chief medical officers and a large staff of physicians and clinicians. This financial activity concerns not just the monetisation of data concerning mental health and the expansion of mental health services online—but also in related areas concerning 'wellness', digitised social and health care, emotion and affect recognition, and so on.

The scale of activity across government and industry is reflected in an expanding body of research, much of which occurs at the intersection of commercial activity and scientific knowledge-making. Lines of accountability across these clinical and commercial domains are not yet clearly defined.[38]

Hence, caution is required in interpreting the global picture of digitised mental healthcare. There at least four reasons for this. First, speculated market value is precisely that—speculated. And those doing the speculating often have vested interests. Examples include technology developers wishing to attract capital, technology vendors seeking to sell products, and corporate services wishing to garner government contracts to build and deliver technological services. Narratives play a strong role in speculative bubbles around new technologies,[39] and there are many who stand to gain by painting a picture of a *rapid* and *inevitable* technological expansion in mental health services and elsewhere. This hype can even be fuelled by humanities scholars who repeat sensational claims about technical feasibility to attract research funding.[40]

35 Molly Fischer, 'The Therapy-App Fantasy', *The Cut* (29 March 2021) <https://www.thecut.com/article/mental-health-therapy-apps.html>; World Economic Forum in collaboration with Accenture, *Empowering 8 Billion Minds: Enabling Better Mental Health for All via the Ethical Adoption of Technologies* (28 October 2019) <https://nam.edu/empowering-8-billion-minds-enabling-better-mental-health-for-all-via-the-ethical-adoption-of-technologies>.

36 Zion Market Research, 'Free Analysis: Mental Health Software Market', *Zion Market Research* (2 January 2019) <https://www.zionmarketresearch.com/market-analysis/mental-health-software-market>.

37 Jennifer Nicholas et al, 'Mobile Apps for Bipolar Disorder: A Systematic Review of Features and Content Quality' (2015) 17(8) *Journal of Medical Internet Research* e198.

38 Nicole Martinez-Martin and Karola Kreitmair, 'Ethical Issues for Direct-to-Consumer Digital Psychotherapy Apps: Addressing Accountability, Data Protection, and Consent' (2018) 5(2) *JMIR Mental Health* e9423.

39 Brent Goldfarb and David A Kirsch, *Bubbles and Crashes: The Boom and Bust of Technological Innovation* (Stanford University Press, 2019).

40 Lee Vinsel, 'You're Doing It Wrong: Notes on Criticism and Technology Hype', *Medium* (1 February 2021) <https://sts-news.medium.com/youre-doing-it-wrong-notes-on-criticism-and-technology-hype-18b08b4307e5>.

Second, governments investing in 'emerging technologies' that are often described in terms of their groundbreaking and revolutionary potential, stand to gain from appearing innovative. Narratives of innovation can be misused in mental health sectors that are commonly painted as broken and crisis-ridden. Innovation-speak may distract from longstanding problems with *existing* mental health policies and practices and the potential need for major investment or restructuring to fix them.[41] Narratives of technological innovation may also detract from broader policies that are toxic to public mental health, such as rising inequality, poor housing, unemployment or employment precarity, pollution and lack of green space. We discuss over-simplified narratives of technological problem-solving later in the report (page 77).

Third, many of the technological claims being made about algorithmic and data-driven technologies in mental healthcare are promissory – that is, they haven't been proven. They lack robust evidence to back them up, particularly to show how they work in applied and real-world settings. In one of the largest surveys of the field, the James Lind Alliance concluded that 'the evidence base for digital mental health interventions, including the demonstration of clinical effectiveness and cost effectiveness in real-world settings, remains inadequate'.[42] Despite this sober finding, and others like it, the flurry of market, government and research activity may falsely suggest an inevitable march of progress toward highly effective and widely adopted digital tools.

Finally, as we will discuss shortly, the very people who are *supposed* to gain from these technological developments – namely, people with lived experience and psychosocial disabilities – are concerningly absent from much of the research and discussion on these topics.[43] Where input from this (diverse) group has been sought for mainstream research or where members of this group have led commentary, the general response appears to be one of ambivalence, with support in some areas through to serious concern in others—though by no means an outright rejection (page 37). As we stress throughout the report, the concerns they raise are not to dismiss the aspirations of those wishing to use technologies in good faith efforts to improve care, and nor is it to uncritically reject technology as necessarily bad or a sure path to a dystopian future. Instead, we aim to express concerns as clearly as possible and promote a sober view of the role of computer technology, with its capacity to simultaneously enable and threaten.

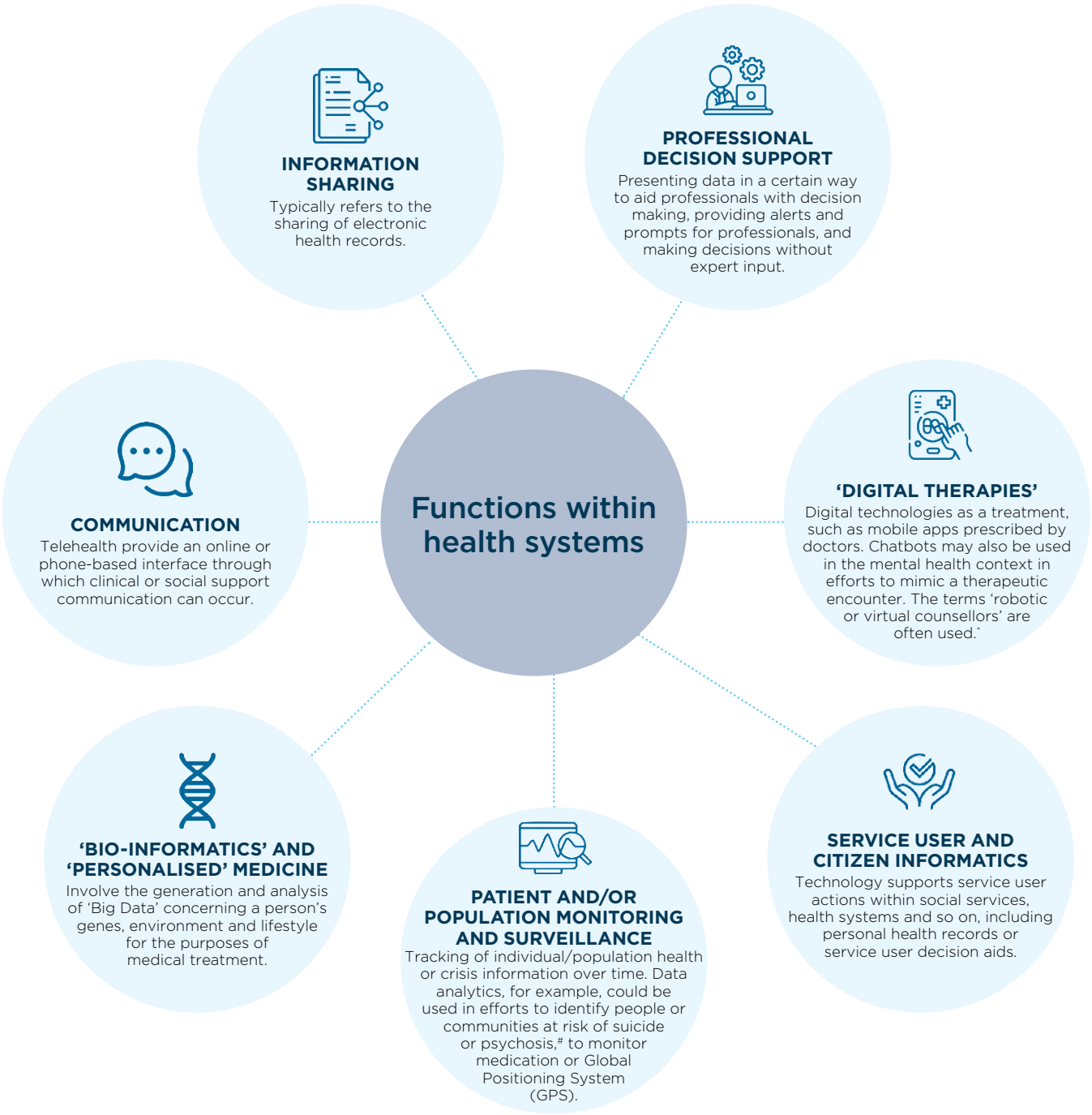
41 Lee Vinsel and Andrew Russell have argued that fetishising innovation can serve to distract from ordinary problems of support infrastructure, including maintenance, repair, and mundane labour. Lee Vinsel and Andrew L Russell, *The Innovation Delusion: How Our Obsession with the New Has Disrupted the Work That Matters Most a Book by Lee Vinsel and Andrew L. Russell* (Currency, 2020).

42 Chris Hollis et al, 'Identifying Research Priorities for Digital Technology in Mental Health Care: Results of the James Lind Alliance Priority Setting Partnership' [2018] *The Lancet Psychiatry* <http://www.sciencedirect.com/science/article/pii/S2215036618302967>; Health Education England likewise raised concerns about 'spurious claims and overhyped technologies that fail to deliver for patients'. Tom Foley and James Woollard, 'The Digital Future of Mental Healthcare and Its Workforce: A Report on a Mental Health Stakeholder Engagement to Inform the Topol Review' (National Health Service (UK), February 2019) p.31.

43 Piers Gooding and Timothy Kariotis, 'A Scoping Review of Algorithmic and Data-Driven Technology in Online Mental Healthcare: What Is Underway and What Place for Ethics and Law?' *Journal of Medical Internet Research - Mental Health.*

1.1 What are the different ways technology is used in crisis support and mental health care?

There are various uses for algorithmic and data-driven technology in the direct provision of mental health care. All are bound up in the contemporary communications eco-system of smartphones, linked devices, and the massive flows of data they enable. *Functions* within health systems include:



* K Oh et al, 'A Chatbot for Psychiatric Counseling in Mental Healthcare Service Based on Emotional Dialogue Analysis and Sentence Generation' in *2017 18th IEEE International Conference on Mobile Data Management (MDM)* (2017) 371.

\# Paolo Corsico, 'The Risks of Risk. Regulating the Use of Machine Learning for Psychosis Prediction' (2019) 66 I*nternational Journal of Law and Psychiatry* 101479; Mason Marks, Artificial Intelligence Based Suicide Prediction, SSRN Scholarly Paper, 29 January 2019 <https://papers.ssrn.com/abstract=3324874>.

These categories are framed in terms of healthcare systems. There may be good reasons to advance other ways of categorising. For example, technologies that analyse data concerning mental health are appearing *outside healthcare services*; for example, in criminal justice agencies, online advertising firms, insurance companies, education settings, employer hiring practices, and so on.[44] Case studies throughout the report will illustrate this expansion.

Some technologies are well-established. Others are exploratory or experimental. Navigating these expanding technologies, including distinguishing which technologies are widely used, which are experimental, which ones are even technically feasible, and which ones are merely sensational and unrealistic, is not always easy. However, certain social, ethical, legal, political and economic themes tend to recur across the range of technology types and the conditions of their usage.

1.2 Benefits Noted in Research

There are several benefits of digital initiatives in the mental health context that are broadly discussed in academic and 'grey' literature:

- Teletherapy, including web-based and other informational communication technology-based forms of support can **break down geographical barriers** and provide effective support to people in distress across large distances, or for those who require or prefer remote support.[45] Hannah Zeavin highlights the way 'care may take unexpected forms through technologies, enabling distanced intimacy and social change that transcends the psychology of the individual'.[46]
- In some cases, online mental health initiatives can **facilitate confidential and anonymous help-seeking** that is a clear social good. This might be extremely important for certain groups, particularly those from small or marginalised communities, for example, people in remote or rural communities, LGBTIQ+ young people, and Indigenous people who are wary of sharing personal information with state-based services,[47] as well as those who may benefit from accessible, digitally facilitated support, including women facing intimate-partner and family violence, or those in sociodemographic groups who may be reluctant to seek traditional forms of care and support.
- There are free web-based programs, some of which may help people to deal with their distress, or identify, name, and better understand their experiences, which can provide a **quick, inexpensive and accessible resource** for those with access to the internet.[48]
- Various kinds of digital technology can help **improve the availability of quality information** to help develop awareness of relevant forms of support. This may include formal services, but also services and organisations outside mental health systems that may be helpful, such as sexual assault services, financial counseling, environmental disaster relief, and informal peer-run support groups for people experiencing distress or addiction. There are examples of community-driven resources, such as online family violence resources and crisis support, created by members of specific cultural communities that are designed to respect their concerns around privacy and cultural respect, while meeting their unique needs.[49]

44 Piers Gooding, 'Mapping the Rise of Digital Mental Health Technologies: Emerging Issues for Law and Society' (2019) 67 *International Journal of Law and Psychiatry* 101498.

45 Bhugra et al (n 32); Hannah Zeavin, *The Distance Cure: A History of Teletherapy* (MIT Press, 2021).

46 Zeavin (n 47).

47 See eg. Mission Australia, 'Accessibility and quality of mental health services in rural and remote Australia Submission' 80, p. 17 https://www.aph.gov.au/DocumentStore.ashx?id=097bdfbe-91ff-44f8-b4ab-ce14217ba1f5&subId=612899 (accessed 9/06/2020); Paul Byron, Digital Media, Friendship and Cultures of Care (Routledge, 2021); Paul Byron, et al. '"You learn from each other": LGBTIQ Young People's Mental Health Help-seeking and the RAD Australia Online Directory' (2016) Western Sydney University Young and Well Cooperative Research Centre, Sydney, p.51; see also https://burndawan.com.au/ (accessed 9/06/2020).

48 See generally, Productivity Commission, *Mental Health, Draft Report*, Canberra (2019) Ch 6.

49 See eg. https://burndawan.com.au/ (accessed 7/12/2021).

- There is also a positive role for data-driven digital technologies in the **monitoring of services, and collection of vital statistics**, including by civil society monitoring bodies, regulators, health system co-ordinators, managers and advocates. (For examples, page 87)

These are just some of the benefits advanced in the scholarly literature. Some clinically oriented research institutes espouse the benefits of digital forms of mental health care in addressing 'serious access gaps [to mental health-related] education, prevention and treatment services'.[50] We will elaborate on some of these apparent benefits throughout the report, while also attending the risks, challenges, issues, and so on, that may run counter to this optimistic picture of digitally-enabled support.

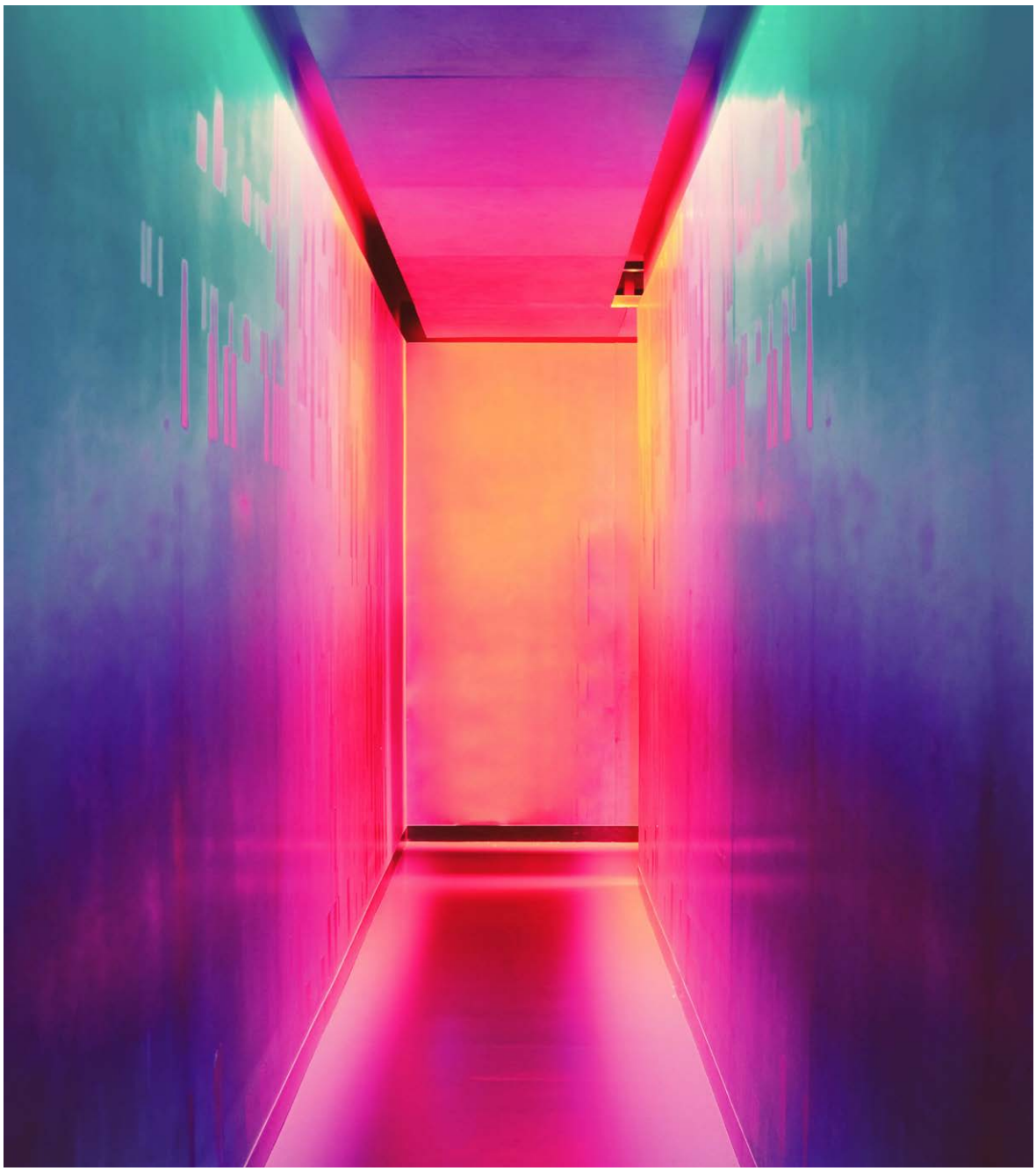


Photo by Efe Kurnaz on Unsplash.

50 Black Dog Institute, 'Saving Lives with Nationally Integrated e-Mental Health Services' https://www.blackdoginstitute.org.au/docs/default-source/research/saving-lives-nationally-integrated-ehealth.pdf?sfvrsn=0 (accessed 7/12/2021)

1.3 Digitising Involuntary Psychiatric Intervention and Other Coercive Measures

> Possibly more than any other group of patients, people with mental health problems can experience particular forms of power and authority in service systems and treatment. They are the only people with long-term conditions who are subject to compulsory treatment under law. The implications of these specific power dynamics as well as potential biases in mental health systems must be considered for the ethical development and implementation of any data-driven technology in mental health.
>
> - Sarah Carr[51]

Unlike data concerning *physical* health, data concerning mental health can be used to initiate state-authorised coercive interventions in certain cases. This possibility adds important legal, social and political dynamics to this discussion. Although the majority of people who access some kind of service for mental health reasons will access those services voluntarily, a small but significant minority of people will be subject to involuntary psychiatric intervention – typically involving the person being detained in hospital and treated against her/his/their wishes – under existing mental health-related legislation.

1.3.1 AI-based Suicide Alerts and Self-harm Surveillance

Government agencies, social media companies, not-for-profits, health services, and others have begun using machine learning and artificial intelligence in suicide prevention, including in efforts to pre-emptively identify people who may self-harm.[52] In some cases, these technologies appear to have been used to activate police powers to detain people for the purposes of involuntary psychiatric intervention.

CASE STUDY: AI-Based Suicide Alerts at Facebook/Meta

In November 2018, a Facebook employee in Texas reportedly alerted police in the Indian state of Maharashtra about a 21-year-old man who had posted a suicide note on his profile. The intervention came after Facebook expanded its pattern recognition software to detect users expressing suicidal intent. Mumbai police reportedly attended the young man's home,[53] for which they have power to authorise involuntary psychiatric intervention under the *Mental Healthcare Act 2017* (India). In 2018, Facebook reported that it had conducted over 1000 'wellness checks' involving the dispatch of first responders.[54]

Facebook/Meta's algorithmic responses also encourage peer-responses from among the person's user-network by drawing their attention to the person's apparent distress.[55] These measures were developed after some form of consultation with suicide attempt survivors and experts on suicide prevention (though few details are available).[56] Facebook/Meta provides some information about the algorithmic process behind the interventions,[57] and has described the ethical issues with which programmers grappled.[58]

51 Sarah Carr, '"AI Gone Mental": Engagement and Ethics in Data-Driven Technology for Mental Health' (2020) 0(0) *Journal of Mental Health* 1.

52 Marks, *Artificial Intelligence Based Suicide Prediction* (n 45).

53 Vijay K Yadav, 'Mumbai Cyber Cops Log into Facebook to Curb Suicides', *Hindustan Times* (online, 5 November 2018) <https://www.hindustantimes.com/mumbai-news/mumbai-cyber-cops-log-into-facebook-to-curb-suicides/story-SMd03alcW0SUBzRJlmdDZJ.html>.

54 Norberto Nuno Gomes de Andrade et al, 'Ethics and Artificial Intelligence: Suicide Prevention on Facebook' (2018) 31(4) *Philosophy & Technology* 669.

55 Catherine Card, 'How Facebook AI Helps Suicide Prevention | Facebook Newsroom' (10 September 2018) <https://newsroom.fb.com/news/2018/09/inside-feed-suicide-prevention-and-ai/>.

56 Gomes de Andrade et al (n 56).

57 Card (n 57).

58 Gomes de Andrade et al (n 56).

However, there remains little information about what precisely is meant by a 'wellness check' (including whether location data are shared with first-responders). Nor is there publicly available research as to the accuracy, scale or effectiveness of the initiative. What Facebook/Meta does with the information following each apparent crisis is also unclear.

Police appear to be the first-responders undertaking 'wellness checks'. Facebook/Meta has therefore drawn criticism for failing to grapple with the reality of anti-Black racism in the US and the prevalence of police violence in their encounters with distressed individuals, particularly Black, Indigenous, people of colour. For example, Joshua Skorburg and Phoebe Friesen write:

> While [Facebook/Meta's wellness checks] may seem like a positive contribution to public health on Facebook's behalf, it is becoming increasingly clear that police wellness checks can do more harm than good. Between 2015 and August 5, 2020, 1,362 people who were experiencing mental health issues were killed by police in the United States. This remarkable number constitutes 23 percent of police fatalities in that time.[59]

The US is by no means alone on such patterns of police violence.[60]

From a legal and regulatory perspective, suicide prediction in medical systems is governed by health information laws, medical practice and clinical governance regimes, and research regulations that require transparency and peer review. Flawed as these frameworks may be, AI-based suicide prediction on social media platforms, as Mason Marks points out, 'typically occurs outside the healthcare system where it is almost completely unregulated, and corporations often maintain their prediction methods as proprietary trade secrets'.[61] To remedy this, Marks recommends several steps to improve people's safety, privacy and autonomy, including:[62]

- making prediction methods more transparent, and giving users unambiguous opportunities to opt-out and delete prediction information;
- protecting consumer privacy and minimising the risk of exploitation, by ensuring suicide predictions cannot be used for advertising or be shared with third parties (such as insurance companies, employers or immigration authorities); and
- the monitoring of ongoing prediction programs by independent data monitoring committees for safety and efficacy.

The use of individual and population monitoring in efforts to prevent suicide or efforts to promote its use, are likely to increase in coming years. In December 2020, the US National Suicide Prevention Lifeline administrator recommended that the US Government authorise geo-location systems to pin-point the exact location of all callers by 2022.[63] Leah Harris has criticised this recommendation, warning that 'Mad and disabled advocates who have experienced mental health crisis intervention, and even some crisis service providers, worry that geolocation would serve to further entrench coercion in mental health and crisis response systems, replicating problematic aspects of [the US emergency services line] 911'.[64] The impact of automated surveillance of callers on rates of involuntary psychiatric interventions, police involvement in crises, citizens' willingness to report to such services, and so on, remains unknown.

59 Joshua August Skorburg and Phoebe Friesen, 'Mind the Gaps: Ethical and Epistemic Issues in the Digital Mental Health Response to Covid-19' (2021) 51(6) *Hastings Center Report* 23.

60 See eg. Piers Gooding, '"The government is the cause of the disease and we are stuck with the symptoms": deinstitutionalisation, mental health advocacy and police shootings in 1990s Victoria' in G Goggin, L Steele, and R Cadwallader (Eds.) *Normality and Disability: Intersections among Norms, Law, and Culture* (Routledge, 2018) 100-110; Anthony J O'Brien et al, 'The Nature of Police Shootings in New Zealand: A Comparison of Mental Health and Non-Mental Health Events' (2021) 74 *International Journal of Law and Psychiatry* 101648.

61 Marks, *Artificial Intelligence Based Suicide Prediction* (n 45).

62 Ibid.

63 Vibrant Emotional Health, *988 Serviceable Populations and Contact Volume Projections* (Vibrant, December 2020) <https://www.vibrant.org/wp-content/uploads/2020/12/Vibrant-988-Projections-Report.pdf>.

64 L Harris, 'The New National Mental Health Crisis Line Wants to Track Your Location', *Disability Visibility Project* (19 April 2021) <https://disabilityvisibilityproject.com/2021/04/19/the-new-national-mental-health-crisis-line-wants-to-track-your-location/>.

1.3.2 'Digitising mental health law'

Some governments have sought to digitise processes of involuntary psychiatric intervention.[65]

CASE STUDY: Electronic Forms and Mobile Technology in Involuntary Psychiatric Interventions

In the UK in 2020, regulations were amended to speed up applications for compulsory psychiatric ntervention orders by providing an online communication platform between mental health professionals involved in involuntary interventions. This web-based interface allows social workers, nurses, psychologists and others who are interacting with a person in crisis to locate and communicate with medical practitioners via videocall who may assess the person and authorise involuntary intervention. One online platform to emerge with government support is reportedly used by over 70% of National Health Service Trusts at the time of writing.[66] David Bradley, the Chief Executive of South London & Maudsley NHS, strongly endorses the practice, describing it as '[t]he Uber of finding doctors for the health service'.[67] Relevant doctors can enter their availability on a personal calendar and 'build a profile containing their location, specialities and languages spoken, and monitor their activity via a dashboard'.[68]

Proponents suggest the electronic forms and digital platforms will improve access to care, reduce errors, and improve information sharing, which ultimately reduces the distress of the individuals and prevents delays in the provision of healthcare.[69] However, some mental health services users have raised concerns about the unknown impact of the digitised process on people subject to orders, which are potentially serious and warrant closer attention.[70]

In recent years, the processing of data about those subject to involuntary psychiatric intervention through electronic records systems has harmed people with lived experience in some cases. Concerns about police agencies sharing data concerning self-harm were raised in Canada, where municipal police collated non-criminal information about individuals who had self-injured or attempted suicide.[71] The information was then circulated to US border authorities, who used the information to deny several Canadians entry into the US. (This example will be discussed at page 52). The ease with which people's sensitive data concerning involuntary treatment can be accessed by various government departments, has raised concerns about the potentially unlawful uses of that data.

65 In the UK, for example, a largescale government review of mental health legislation recommended the 'digitising of the Mental Health Act' HM Government, *Modernising the Mental Health Act: Increasing Choice, Reducing Compulsion Final Report of the Independent Review of the Mental Health Act 1983* (Crown, December 2018) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/778897/Modernising_the_Mental_Health_Act_-_increasing_choice__reducing_compulsion.pdf>.

66 This platform emerged from a public private partnership, with funding from the NHS Innovation Accelerator (NIA), NHS England's Innovation and Technology Payment Evidence Generation Fund, NHS England's Clinical Entrepreneur programme and DigitalHealth.London's Accelerator. S12 Solutions Website, www.s12solutions.com [accessed 3/3/2021])

67 S12 Solutions, 'What is S12 Solutions?' Twitter (21 Jan 2020) https://twitter.com/S12Solutions/status/1219262300667961349 [accessed 19/05/2021])

68 Doctors may also use the app to register assessments that were undertaken, provide supporting evidence for professional development, and complete payment claims for their assessment. The app also generates data about the Act assessment process.

69 Thalamos, 'Mental Health Act Forms: The Benefits of Going Digital', *Thalamos.co.uk* (10 November 2020) <https://www.thalamos.co.uk/2020/11/10/mental-health-act-forms-the-benefits-of-going-digital/>. Small pilot evaluations appear to support this view. S12 Solutions (2017d). Pilot Evaluation. NHS Innovations Accelerator. Available from: https://nhsaccelerator.com/wp-content/uploads/2019/05/S12-Solutions-pilot-evaluation1.pdf (accessed 13/07/2021).

70 M. Stevens, et al. The availability of section 12 doctors for Mental Health Act assessments - a scoping review of the literature. NIHR Policy Research Unit in Health and Social Care Workforce, The Policy Institute, King's College London, p.16; Mental Elf, *Digitising the Mental Health Act: A Public Debate #DigitalMHA* (26 June 2020) <https://www.youtube.com/watch?v=Yuzkctpv1dA>.

71 Office of the Privacy Commissioner of Canada, 'Disclosure of Information about Complainant's Attempted Suicide to US Customs and Border Protection Not Authorized under the Privacy Act' (21 September 2017) <https://www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-federal-institutions/2016-17/pa_20170419_rcmp/>.

CASE STUDY: 'Serenity Integrated Monitoring' – Sharing Sensitive Information and Flagging 'High Intensity Users' of Mental Health Services

The 'Serenity Integrated Monitoring' (or SIM) was a program run in England by police and public mental health services in relation to 'high intensity users'—individuals who have been frequently detained under section 136 of the *Mental Health Act 1983* (England and Wales).[72] This practice does not use algorithmic technology but it highlights sensitive data-sharing issues in current communication ecosystems.

The program involved police officers, described as 'High Intensity Officers', regularly contacting the person to dissuade them from 'unnecessary' interactions with emergency health services, and to instead arrange more 'appropriate' support.[73] Major concerns with the program were reported in May 2021:

> [w]hen tagged under the system, patients can be denied care, prevented from seeing doctors or psychiatrists, and sent home. An NHS doctor told [journalists] that he had to turn away a woman who had attempted suicide on multiple occasions because she had been assigned to the SIM scheme. He considered resigning as a result.[74]

The Royal College of Psychiatrists reported that where a person 'remained unwell and continued to self-harm, attempt suicide or report suicidality, in some cases they were prosecuted and imprisoned or community protection notices were applied which required them to stop self-harming or calling for help, with imprisonment as a potential sanction if they breached the notice'.[75]

StopSIM Coalition, a 'grassroots network of service users and allies', raised concerns that the program 'allows "sensitive data" (information like medical records, ethnicity, religion, sexuality, gender reassignment and financial information) to be shared between services without the subject's consent … (for example, as a consequence of calling [emergency services] when feeling suicidal)'.[76]

The SIM program is being reviewed by the National Health Service at the time of writing, although it reportedly remains in place in 23 National Health Service mental health trusts in England[77] and is being trialled in three US states.[78]

The SIM program will be discussed later in the report in sections on accountability and privatisation (page 58). SIM also appeared to have disproportionate impacts along lines of race and class (discussed at page 69).

72 The individuals were chosen based on local health authority 'Mental Health Act data for the previous year to define which borough/geographical area had the highest proportion of high intensity users of [Section] 136'. Aileen Jackson and Josh Brewster, *THE IMPLEMENTATION OF SIM LONDON: Sharing Best Practice for Spread and Adoption* (June 2018) 6 <https://healthinnovationnetwork.com/wp-content/uploads/2018/11/The-Implementation-of-SIM-London-Report.pdf>.

73 Royal College of Psychiatrists (UK), 'RCPsych Calls for Urgent and Transparent Investigation into NHS Innovation Accelerator and AHSN Following HIN Suspension', *www.rcpsych.ac.uk* (14 June 2021) <https://www.rcpsych.ac.uk/news-and-features/latest-news/detail/2021/06/14/rcpsych-calls-for-urgent-and-transparent-investigation-into-nhs-innovation-accelerator-and-ahsn-following-hin-suspension> (accessed 9/9/21). Those flagged in annual Mental Health Act data tend to be very unwell and regularly phone emergency services or arrive at hospitals having self-harmed, attempted suicide, or threatened to take their own life.

74 Patrick Strudwick, 'Campaigners Call for Inquiry after Mental Health Patients Turned Away by NHS under Controversial Scheme', *i* (online, 16 June 2021) <https://inews.co.uk/news/nhs-mental-health-stop-sim-inquiry-1056296>.

75 Royal College of Psychiatrists (UK) (n 75).

76 StopSIM Coalition, 'STOPSIM', *STOPSIM* (n.d.) <https://stopsim.co.uk/>.

77 NHS Trusts refer to an organisational unit of the NHS that generally serves either a geographical area or a specialised function.

78 Maryam Jameela, 'Outrage Grows as Police Embed Themselves in Mental Health Services', *The Canary* (online, 22 May 2021) <https://www.thecanary.co/investigations/2021/05/22/outrage-grows-as-police-embed-themselves-in-mental-health-services/>.

1.3.3 Power and Coercion in Mental Health

Other areas of law around the world overtly discriminate against people with lived experience and psychosocial disability, which adds to the sensitivity of data concerning mental health. Discrimination in law could include preventing a person with a mental health diagnosis from holding public office, migrating into particular countries, and working in particular professions.[79] Indeed, some countries continue to criminalise suicide attempts. For example, Section 226 of Kenya's penal code states that 'any person who attempts to kill himself [sic] is guilty of a misdemeanour'.[80] Around 20 countries still criminalise suicide attempts, according to a 2021 report by the International Association for Suicide Prevention and United for Global Mental Health.[81] Automated suicide alert programs must therefore be applied with extreme caution (see section on Non-Discrimination and Equity below).

There are also well-established examples where people with psychosocial disabilities and mental health diagnoses are occasionally subject to political scapegoating and public scare campaigns that attract intrusive and discriminatory proposals for state intervention.

CASE STUDY: 'SAFEHOME for Stopping Aberrant Fatal Events by Helping Overcome Mental Extremes' – Proposed Behavioural Monitoring and Preventive Policing

In 2019, the *Washington Post* reported that a prominent US businessman briefed top officials of the Trump administration, including the then president and vice president, on a proposal 'to create a new research arm called the Health Advanced Research Projects Agency'.[82] The advisor promoted a program titled, 'SAFEHOME for Stopping Aberrant Fatal Events by Helping Overcome Mental Extremes' that called for experimentation to explore whether 'technology including phones and smartwatches can be used to detect when mentally ill people are about to turn violent'.[83] The proposal was not pursued by the time Donald Trump left office in 2020.

One category of biometric monitoring technology more broadly, known as 'anomaly detection', may have repercussions for people with psychosocial disabilities and lived experience if used in public surveillance. According to one report, automated surveillance systems are designed undertake 'automatic detection and tracking of unusual objects and people'.[84] The literature on anomaly detection, according to a report for the ACLU, is 'full of discussion of algorithms that can detect people or behaviours that are "unusual," "abnormal," "deviant," or "atypical"'.[85] (See above the discussion about the politics of terminology in the mental health context, including characterisations of 'deviance' and 'abnormality', page 13).[86] The authors warn that identifying statistical deviance is not

79 Pūras and Gooding (n 27).

80 Laws of Kenya, The Penal Code, Chapter 63, Revised Edition 2009 (2008) s 226.

81 United for Global Mental Health, *Decriminalising Suicide: SAVING LIVES, REDUCING STIGMA* (International Association for Suicide Prevention, 2021) <https://unitedgmh.org/sites/default/files/2021-09/UNITEDGMH%20Suicide%20Report%202021%C6%92.pdf>.

82 William Wan, 'White House Weighs Controversial Plan on Mental Illness and Mass Shootings', *Washington Post* (9 September 2019) <https://www.washingtonpost.com/health/white-house-considers-controversial-plan-on-mental-illness-and-mass-shooting/2019/09/09/eb58b6f6-ce72-11e9-87fa-8501a456c003_story.html>.

83 Ibid.

84 Wallace Lawson, Laura Hiatt and Keith Sullivan, 'Detecting Anomalous Objects on Mobile Platforms' in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops* (CVPRW) (IEEE, 2016) 1426 <https://ieeexplore.ieee.org/document/7789669/>

85 Jay Stanley, *The Dawn of Robot Surveillance: AI, Video Analytics, and Privacy* (American Civil Liberties Union, 2019).

86 One research group, for example, proposed that 'computer vision' designed to detect violence could be 'extremely useful in some video surveillance scenarios like in prisons, psychiatric or elderly centers'. Enrique Bermejo Nievas et al, 'Violence Detection in Video Using Computer Vision Techniques' in Pedro Real et al (eds), *Computer Analysis of Images and Patterns* (Springer Berlin Heidelberg, 2011) 332 <http://link.springer.com/10.1007/978-3-642-23678-5_39>.

a negative thing per se but when it 'shades into finding deviant people it should raise alarms',[87] and given the historical and current exclusion of, and social hostility against, people with lived experience and psychosocial disability, and other disabilities, such as intellectual and cognitive disabilities, this is particularly troubling.

A less experimental form of electronic monitoring at the intersection of mental health and criminal justice is the use of monitoring of people in forensic mental health services using global positioning system (GPS). Electronic monitoring devices record and regularly transmit data on a person's location via devices fixed to his or her body. Some GPS devices, such as devices affixed to a person's wrist or ankle, can be linked to blood-alcohol monitors.[88]

CASE STUDY: GPS Surveillance of Forensic Psychiatric Patients in Three Jurisdictions

Two jurisdictions in Australia authorise health services to impose involuntary 'monitoring conditions' on people detained in forensic psychiatric settings using electronic GPS devices, typically in the form of electronic ankle bracelets.[89] In one jurisdiction, the program was advanced by government against the submissions and evidence of medical practitioners.[90] In an appeal brought by a man subject to the surveillance regime,[91] a treating psychiatrist submitted that '[n]ot only did [the] device add nothing to his clinical management or risk reduction, it had the effect of hindering his rehabilitation'.[92]

In England and Wales, GPS surveillance of people in forensic mental health settings is only possible if they *consent* to it.[93] In Nova Scotia, Canada, legislators have *prohibited* GPS surveillance of forensic mental health patients in any form, with lawmakers citing concerns that it violates human rights.[94] The province commissioned three reports into the clinical and legal issues, and each study indicated that 'there was no support or even speculative support that electronic monitoring would enhance public safety.'[95]

More issues concerning involuntary psychiatric interventions and computer technology will emerge in coming years, raising pressing questions. Will monitoring devices be imposed in involuntary psychiatric interventions in the civil context, such as 'community treatment orders'? Should algorithmic technologies be used at *all* in coercive crisis responses? How might these concerns relate to broader efforts in recent years to reduce and eliminate coercion in mental health settings, and to debates about 'abolishing versus reforming' involuntary psychiatric interventions?[96]

87 Stanley (n 87).

88 A Board-Certified Physician, 'SCRAM Ankle Bracelet Measures Alcohol Consumption', *Verywell Mind* <https://www.verywellmind.com/scram-ankle-bracelet-measures-blood-alcohol-247-67122>.

89 Stephanie Miller, 'The Use of Monitoring Conditions (GPS Tracking Devices) Re CMX [2014] QMHC 4' (2015) 22(3) Psychiatry, *Psychology and Law* 321.

90 Ibid.

91 Re CMX [2014] QMHC 4 (Australia).

92 Ibid [42]-[43].

93 John Tully et al, 'Service Evaluation of Electronic Monitoring (GPS Tracking) in a Medium Secure Forensic Psychiatry Setting' (2016) 27(2) *The Journal of Forensic Psychiatry & Psychology* 169. Informed consent, it should be noted, is profoundly impacted by the power asymmetry inherent in forensic mental health services but nevertheless, the contrasting approaches between Queensland and England and Wales is significant. Regarding empirical evidence in support of the schemes 'efficacy' in reducing adverse events, one John Tully and his group of UK researchers reported a major reduction in '[e]pisodes of leave violation… which suggest potential benefits for speed of patient recovery, reduced length of stay, reduced costs and public safety'. Ibid p.169.

94 Donalee Moulton, 'Nova Scotia Sets Direction on GPS Monitoring of Patients' (2015) 187(8) *Canadian Medical Association Journal* E232.

95 Ibid.

96 Committee on the Rights of Persons with Disabilities, 'General Comment No 1: Article 12 – Equal Recognition before the Law, 11th Sess, UN Doc CRPD/C/GC/1'; Tina Minkowitz, 'The United Nations Convention of the Rights of Persons with Disabilities and the Right to Be Free from Nonconsensual Psychiatric Interventions' (2007) 34(2) *Syracuse Journal of International Law and Commerce* 505; Kay Wilson, Mental Health Law: Abolish or Reform? (Oxford University Press, 2021).

The future of algorithmic and data-driven technologies in coercive state interventions remains uncertain—but imagined futures are guiding activity *today*. As one industry publication that promoted technology in healthcare stated:

> In the future, patients might go to the hospital with a broken arm and leave the facility with a cast and a note with a compulsory psychiatry session due to flagged suicide risk. That's what some scientists aim for with their A.I. system developed to catch depressive behavior early on and help reduce the emergence of severe mental illnesses.[97]

This imagined future is one possibility. Others will reject this vision of expanded risk predictions and technology-facilitated coercion, and instead promote the development of open and co-operative crisis support relationships that are enhanced by selective use of digital technology. These contested futures suggest that the power dynamic caused by coercion in mental health services must be a part of the discussion concerning 'digital mental health' measures today.

1.4 Biometric Monitoring Technologies

Biometric monitoring technologies represent a somewhat 'extreme' technology for the purposes of this report – compared to say, teletherapy – given that the insights such technologies are purported to reveal about a person's health, body, cognition, affective state and so on, create challenging ethical, social, legal and political issues.[98] Biometric monitoring technologies use sensors in devices, including smartphones, wearable and connected devices, cameras and even pills, to remotely generate data concerning a person's biology, physiology or behaviour.

There are various ways to describe biometric monitoring technologies. Computer scientists may refer to 'context sensing', 'personal sensing', or 'mobile sensing'. In mental health settings, several prominent psychiatrists and psychologists have begun to refer to 'digital phenotyping', particularly in relation to the assessment of behaviour, mood and cognition through biometric data generated by devices, such as smartphones and FitBits.[99] An advertisement for the prominent direct-to-consumer app company, Mindstrong, for example, describes this practice this way:

> How you passively use your smartphone—typing, swiping, scrolling—is a new way to measure things like your stress, mental health symptoms, and well-being. If you're typing more slowly—even by a millisecond—it might mean there's a change. You can track your measurements in the mobile app, and they're shared with your clinical team so they can provide you with more personalized care.[100]

97 The Medical Futurist, 'Artificial Intelligence In Mental Health Care', *The Medical Futurist* (25 June 2019) <https://medicalfuturist.com/artificial-intelligence-in-mental-health-care>.

98 Lisa Cosgrove et al, 'Digital Phenotyping and Digital Psychotropic Drugs: Mental Health Surveillance Tools That Threaten Human Rights' (2020) 22(2) *Health and Human Rights Journal* 33; Amba Kak, *Regulating Biometrics: Global Approaches and Urgent Questions* (AI Now Institute, 1 September 2021) <https://ainowinstitute.org/regulatingbiometrics.pdf>; Nicole Martinez-Martin et al, 'Data Mining for Health: Staking out the Ethical Territory of Digital Phenotyping' (2018) 1(1) *npj Digital Medicine* 68.

99 Thomas R Insel, 'Digital Phenotyping: Technology for a New Science of Behavior' (2017) 318(13) *JAMA* 1215.

100 Mindstrong, 'How it works' (website) <https://mindstrong.com/how-it-works/> [accessed 02/02/2021].

There is growing enthusiasm among some mental health professionals about the 'enormous potential' of this technology 'to improve our understanding of the experience of individuals and our capacity to deliver behavioural health treatments'.[101] This potential could include the integration of monitoring apps into standard psychological treatments (for example, talking therapy plus apps that monitor behaviour), the delivery of biometric monitoring as a stand-alone interventions, as well as using '[p]assive tracking of populations of at-risk people [...] [to] facilitate early identification and intervention for behavioral problems'.[102]

CASE STUDY: Biometric Monitoring in Mental Health Settings

From 2021, up to 20,000 Australian high school students will have their phone data monitored for up to five years in an attempt to track how mental health issues develop in adolescence. According to the researchers, '[t]he study aims to discover how we can use smartphones to deliver preventive interventions on a large scale'. The study makes use of '[c]omprehensive, technology-assisted data collection and analysis [...] to determine what triggers the development of mental health symptoms.'[103] The authors report that no study of mental health apps has occurred at this scale anywhere in the world.[104] The children and young people involved in the study will interact with game-based apps, have their movement and location tracked, and be asked specific questions about their state of mind, including whether they have contemplated committing suicide.

Proponents view biometric monitoring with therapeutic aims as a reasonable method for real-time tracking that in the right persons may enhance therapeutic alliance between mental health practitioners and individuals seeking help. This enthusiasm is typically accompanied by acknowledgement that some patients will not want to use nor gain from such measures. Similar forms of behavioural tracking holds appeal to actors in sectors outside of formal mental health services, including education, the military, the insurance industry, and the criminal justice system.[105]

Biometric technology has also started to appear in video monitoring and surveillance in acute psychiatric settings, in ways that do not involve on-body sensors.

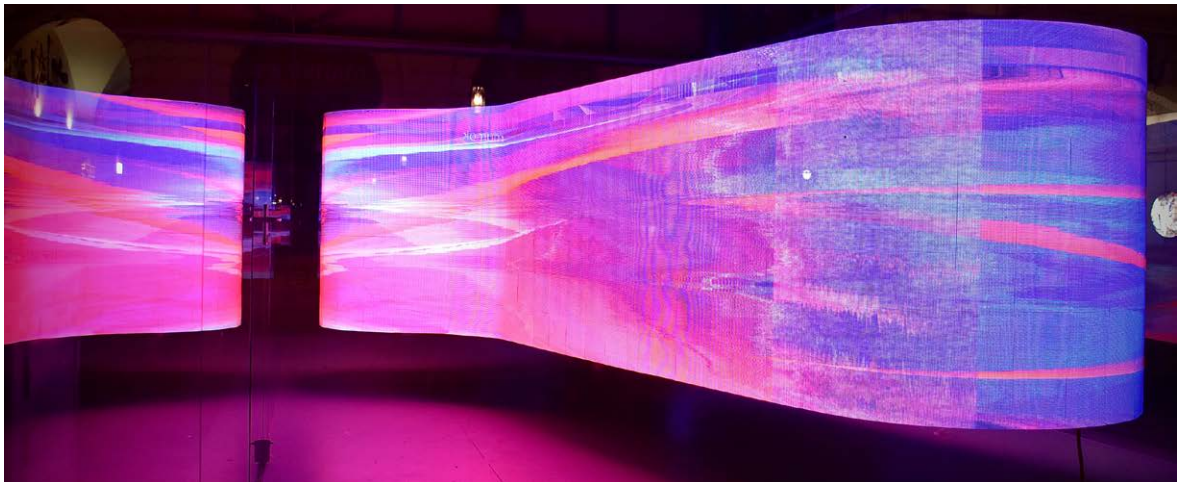


Photo by Arno Senoner on Unsplash.

101 David C Mohr, Katie Shilton and Matthew Hotopf, 'Digital Phenotyping, Behavioral Sensing, or Personal Sensing: Names and Transparency in the Digital Age' (2020) 3(1) *npj Digital Medicine* 1.

102 Ibid.

103 Black Dog Institute, 'The Future Proofing Study' <https://www.blackdoginstitute.org.au/research-centres/future-proofing/> [accessed 15/07/2021]

104 Ibid.

105 Kak (n 100).

CASE STUDY: Algorithmic Video Monitoring and Surveillance in Psychiatric Settings

In 2020 in England, a trial (the 'Oxehealth Trial') was undertaken on the use of 'digital assisted observations' at a psychiatric ward.[106] The monitoring was used by nurses to take 15-minutely and hourly night-time 'clinical observations' of patients in 6 individual bedrooms over a 4-month period.[107] The sensors used by researchers were wall-mounted video cameras along with 'computer vision, signal processing and AI software' that enabled nurses to track their patients' locations and movements ('physical monitoring') and to record their heart and respiratory rates ('physiological' or 'vital sign monitoring').[108] Physiological monitoring using the Oxehealth system allows nurses to access 'real-time spot measurements of pulse rate and breathing rate without them having to enter the room'.[109] These measurements are 'displayed on a screen in the nursing station or on handheld tablet computers'.[110] The software generates long-term information in the form of 'a timeline summarising the patient's location (in bed, elsewhere in their room, etc) for a day or a week … to help characterise the patient's behaviour during that time interval'.[111]

Other forms of biometric sensing go 'beneath the skin', where ingestible sensors have been integrated with psychopharmaceutical pills.

CASE STUDY: 'Smart Pills', 'Digital Pills', and Ingestible Sensors

In 2017, the US Food and Drug Administration ("FDA") approved a so-called 'digital pill'[112] 'Abilify MyCite', as it is commercially named, integrates a pill with an electronic sensor. According to the FDA, 'Abilify MyCite' is aimed at 'the treatment of schizophrenia, acute treatment of manic and mixed episodes associated with bipolar I disorder and for use as an add-on treatment for depression in adults'.[113] When a person swallows the pill, the sensor activates upon contact with stomach fluid. Information concerning the nature and timing of ingestion is then transmitted via a patch worn on the skin to a linked device, such as a smartphone. Family members, clinicians and other third parties can, with the person's consent, attain the information through a web-based portal. The smartphone/tablet app can also track 'self-reported measures of rest and mood'. The pills are advertised as 'targeting the problem of medication adherence'.[114] Digital pills have also been approved for use by regulatory bodies in China and the European Union.

Although biometric technology is relatively exploratory in the mental health context, its use is expanding. For example, the Oxehealth system used in the British trials is reported by the company who produces them to be 'relied on by one in three English mental health trusts as well as acute hospitals, care homes, skilled nursing facilities, prisons and police forces in the UK and Europe'.[115]

106
107
108
109 Ibid 38.
110 Ibid 39.
111 Ibid 37-38.
112 Food and Drug Administration (US), 'FDA News Release: FDA approves pill with sensor that digitally tracks if patients have ingested their medication, New tool for patients taking Abilify', 13 November 2017 <https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm584933.htm>
113 Ibid.
114 Craig M Klugman et al, 'The Ethics of Smart Pills and Self-Acting Devices: Autonomy, Truth-Telling, and Trust at the Dawn of Digital Medicine' (2018) 18(9) *The American Journal of Bioethics* 38.
115 <https://www.oxehealth.com/about-us> (accessed 26/08/21).

1.4.1 Power and Justice in the Biometric and Digital Turn

The rise of biometric monitoring in mental health care is being debated on several fronts. This includes contested claims about what 'digital markers' of behaviour can reveal.[116] Even the term 'digital phenotyping' is contested and terminology remains unsettled. David Mohr and colleagues raise concerns that the term fails to convey the reality that the practice constitutes surveillance over intimate aspects of a person's life.[117] Mohr and colleagues, who ultimately endorse the potential value of the technology, state:

> [W]hat might the term digital phenotyping signal mean to those whose data are being used? That such sensing is medical and scientific, perhaps? That it is complex? It does not convey to the average person that we are engaging in a sensitive form of surveillance: collecting large amounts of data, and using those data to understand deeply personal things, such as how they sleep, where they go, how and when they communicate with others, or whether they may be experiencing a mental health condition.[118]

The authors call for language that is more transparent about the intent and practice behind this technology, arguing the term 'personal sensing' is more appropriate.

Other commentators have drawn attention to deeper issues of justice and power. The use of biometric technologies to purportedly infer a person's mental state or characteristics, and its use in pervasive forms of monitoring and surveillance, have raised particular concern.[119] Leah Harris warns of biometric technologies developed by psychiatric or psychological professionals being used in forms of social control over marginalised individuals, not just in mental health settings, but also in prisons and other sites of carceral control, including in the 'community'.[120] Harris refers to Michel Foucault's theorisation of the Panopticon, discussing the way '*power* is based on both the ability to observe others and the *knowledge* obtained through that observation'.[121] The Panopticon was originally an architectural system and idea developed in the eighteenth century by Jeremy Bentham. Its purpose is to continuously observe prisoners in confinement. For Foucault, *panopticism* is a surveillance mechanism used to exert disciplinary power throughout society by professionals, bureaucracies, government agencies, market actors, and so on, by allowing for an 'absolute and constant visibility surrounding the bodies of individuals'.[122]

Toward the end of his life, Michel Foucault conceptualised a shift in Western societies away from the dominance of disciplinary environments such as largescale psychiatric institutions, to systems of constant external surveillance. He wrote, '[o]ne also sees the spread of disciplinary procedures, not in the form of enclosed institutions, but as centres of observation disseminated throughout society'.[123] He charts these societal shifts toward forms of control that are less costly and complex to manage.[124] Harris relates panopticism to biometric monitoring in the mental health context, warning that '[t]here is always an

116 Phoebe Friesen, 'Digital Psychiatry: Promises and Perils' (2020) 27(1) *Association for the Advancement of Philosophy and Psychiatry* 2; Mohr, Shilton and Hotopf (n 103); Eric S Swirsky and Andrew D Boyd, 'Adherence, Surveillance, and Technological Hubris' (2018) 18(9) *The American Journal of Bioethics* 61.

117 Mohr, Shilton and Hotopf (n 103).

118 Ibid.

119 Jonah Bossewitch, 'Brave New Apps: The Arrival of Surveillance Psychiatry', *Mad In America* (9 August 2019) <https://www.madinamerica.com/2019/08/brave-new-apps-the-arrival-of-surveillance-psychiatry/>; Leah Harris, 'The Rise of the Digital Asylum', *Mad In America* (15 September 2019) <https://www.madinamerica.com/2019/09/the-rise-of-the-digital-asylum/>.

120 Harris, 'The Rise of the Digital Asylum' (n 121); L Harris, 'The New National Mental Health Crisis Line Wants to Track Your Location', *Disability Visibility Project* (19 April 2021) <https://disabilityvisibilityproject.com/2021/04/19/the-new-national-mental-health-crisis-line-wants-to-track-your-location/>.

121 Harris, 'The Rise of the Digital Asylum' (n 121).

122 Michel Foucault, *Psychiatric power: Lectures at the Collège de France* (Palgrave Macmillan, 2006), p.52.

123 Michel Foucault, Discipline and punish: *The birth of the prison* (Vintage Books, 1995) p.212.

124 Etienne Paradis-Gagné and Dave Holmes, 'Gilles Deleuze's Societies of Control: Implications for Mental Health Nursing and Coercive Community Care' n/a(n/a) *Nursing Philosophy* e12375.

inherent power imbalance between the "omnipresent" and "invisible" *watchers* and their "permanently visible" *subjects'* and that such imbalances have been expressed in psychiatry historically through its role in governing a marginalised and oppressed group[125]. Harris's framing has commonalities with broader critiques of the information economy in the current era, including Shoshanna Zuboff's prominent characterisation of 'surveillance capitalism'.[126]

EXPLAINER: Zuboff's 'Surveillance capitalism'

Shoshana Zuboff describes surveillance capitalism as the market-driven process that turn personal thoughts, experiences and behaviours into data that is then commodified for marketing purposes.[127] Such processes rely on the increasing use of surveillance processes, through the collection of data, not just based on what a person 'posts' online, but from the 'behavioral surplus data' that emerges from *how* a person uses their digital technology. Biometric data, usage rates, the manner a person expresses themselves, all become converted into data that can be extracted and sold for value. Data is then on-sold with claims that it has predictive value for how someone may behave. Zuboff explains this extraction process within the context of a diabetes app:

> You download a diabetes app, it takes your phone, it takes your microphone, it takes your camera, it takes your contacts. Maybe it helps you manage your diabetes a little bit, but it's also just a part of this whole supply-chain dynamic for behavioral surplus flows. The stuff that they're taking from you has nothing to do with the diabetes functionality for which you downloaded the app. Absolutely nothing. It's simply siphoning off data to third parties for other revenue streams that are part of these surveillance capitalists' ecosystems.[128]

However, the market incentives that form under surveillance capitalism go beyond prediction, towards shaping or controlling behaviour, or as Zuboff describes, the creation of 'monitoring and compliance regimes'.[129] That is, digital technologies can be integrated with other incentives to ensure behaviours that are compliant with businesses objectives, such as sharing additional data or maintaining engagement in order to continue having access to the full benefits of the technology. One example where this is used is the 'internet of things', whereby there is an integration between digital technologies and data with everyday objectives, such as those in a google home, or with a car. The failure to share data may disable features of 'smart devices' in the home, or if payments run late on a car, it can be remotely disabled from operating any longer. Therefore there remains choice, but with significant trade-offs. Individual consumers in this setting have little bargaining power compared to significant digital platforms.[130] The broader implications of these market incentives taken to their conclusion is the construction of a society that is in 'perpetual compliance' with business interests.[131]

125 Harris, 'The Rise of the Digital Asylum' (n 121).

126 Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power: Barack Obama's Books of 2019* (Profile books, 2019).

127 Ibid.

128 Noah Kulwin, 'Shoshana Zuboff Talks Surveillance Capitalism's Threat to Democracy', *Intelligencer* (24 February 2019) <https://nymag.com/intelligencer/2019/02/shoshana-zuboff-q-and-a-the-age-of-surveillance-capital.html>.

129 Zuboff (n 133) 217.

130 For an examination of these power asymmetries in an Australian context, see: ACCC, *Digital Platforms Inquiry: Final Report* (Australian Competition and Consumer Commission, 26 July 2019).

131 Zuboff (n 128) 334; see also Natasha Tusikov, 'Regulation through "Bricking": Private Ordering in the "Internet of Things"' (2019) 8(2) *Internet Policy Review*.

How might surveillance capitalism operate in the mental health context? Various critical accounts have been offered. Lisa Cosgrove and colleagues' state:

> Mental health apps that use digital phenotyping and other surveillance technologies position people as unwitting profit-makers; they take individuals at their most vulnerable and make them part of a hidden supply chain for the marketplace.[132]

Examples of such data-extraction are included throughout this paper. Jonah Bossewitch warns of the 'arrival of surveillance psychiatry' and queries its role in the growing information economy, whereby 'huge pools of data are being used to train algorithms to identify signs of mental illness':[133]

> Researchers are claiming they can diagnose depression based on the color and saturation of photos in your Instagram feed and predict manic episodes based on your Facebook status updates. Corporations and governments are salivating at the prospect of identifying vulnerability and dissent. The emphasis on treating risk rather than disease predates the arrival of big data, but together they are now ushering in an era of algorithmic diagnosis based on the data mining of our social media and other digital trails.[134]

One challenge for advocates will be to correctly identify the business models of companies generating or processing such data. Without transparency on this matter, which companies will not necessarily divulge, observers may be left to speculate. One obvious business model would be targeting platform users with commercial products, as the next example suggests.

CASE STUDY: 'Cerebral' – app company accused of 'accelerating the psychiatric prescribing cascade'

A 2021 Bloomberg investigation of the popular mental health app 'Cerebral', for example, found evidence that it led to overtreatment that generated increased sales of home-delivered psychopharmaceutical prescriptions.[135] 'Cerebral' does not involve biomonitoring but it highlights a business model that others will be following in the industry, regardless of how data is generated. The Cerebral app provides a platform for connecting platform users to a therapist and a psychiatric nurse practitioner at a monthly cost.[136] Former Cerebral employees reported to journalists that the company prized quantity over quality, involving more patient visits, shorter appointments, and more prescriptions.[137] Concerns were raised about the app 'accelerating the psychiatric prescribing cascade' for people seeking amphetamines prescribed for ADHD.[138]

We will discuss private sector interests and the role of data concerning mental health in the information economy throughout the report.

132 Lisa Cosgrove et al, 'Psychology and Surveillance Capitalism: The Risk of Pushing Mental Health Apps during the COVID-19 Pandemic' (2020) 60(5) *Journal of Humanistic Psychology* 611, 620.

133 Bossewich (n 119).

134 Ibid.

135 'ADHD Drugs Are Convenient To Get Online. Maybe Too Convenient', *Bloomberg.com* (online, 11 March 2022) <https://www.bloomberg.com/news/features/2022-03-11/cerebral-app-over-prescribed-adhd-meds-ex-employees-say>.

136 'How Mental Health Apps Can Accelerate the Psychiatric Prescribing Cascade', *Lown Institute* (18 March 2022) <https://lowninstitute.org/how-mental-health-apps-can-accelerate-the-psychiatric-prescribing-cascade/>.

137 'ADHD Drugs Are Convenient To Get Online. Maybe Too Convenient' (n 137).

138 'How Mental Health Apps Can Accelerate the Psychiatric Prescribing Cascade' (n 138).

Returning to biometric monitoring, others have raised concerns that people who use algorithmic interpretations of data concerning emotions are misled about the extent to which such systems can 'capture' the reality of emotional experiences.[139] Victoria Hollis and colleagues point to a survey of people (n=188) who showed strong interest in automatic stress and emotion tracking, where 'many respondents expected these systems to provide objective measurements for their emotional experiences' despite this simply not being possible.[140] This framing effect (which is often exaggerated by tech vendors) can even change the way people construe their own emotions. In another study, Hollis examined how algorithmic sensor feedback influences emotional self-judgments in a mixed-methods study with 64 participants.[141] 'Despite users reporting strategies to test system outputs, users still deferred to feedback and their perceived emotions were significantly influenced by feedback frames' with some users even 'overr[iding] personal judgments, believing the system had access to privileged information about their emotions.'[142]

Similarly, Lisa Parker and colleagues, in their survey of the messaging of mental health apps, argued that prominent apps tend to over-medicalise states of distress and may over-emphasise 'individual responsibility for mental well-being'.[143] As a broad comment, the user/survivor/ex-patient movement and others have advanced reasons to de-medicalise approaches to supporting people in distress; which would seemingly extend to caution about framing personal mental crises as medical problems *amenable to digital technological solutions*.[144] The framing effects of biometric monitoring often go unremarked, but the studies noted above suggest the effects can alienate people from their own self-perceptions. For their part, Hollis and colleagues argue that the framing effects of should be acknowledged and used in ways to promote agency and help individuals more actively construe their personal experiences.[145]

Concerns raised by Harris, Bossewitch and others move beyond questions of how to make particular technologies like biometric monitoring more equitable or ethical (for example, by ensuring the datasets adequately cover diverse communities that accommodate distinct ways of being and self-presenting). Instead, their questions relate to law and political economy, questioning whether technologies are creating a market for surveillance in the mental health context that perpetuates and even extends the worst power imbalances, inequities and harms of current mental health practices.[146] Kaitlin Costello and Diana Floegel, for example, argue that the 'link between the carceral state and mental healthcare in the United States is alarming' and that biometric monitoring technologies 'are poised to only further strengthen that link, despite calls to the contrary'.[147] More fundamentally, this new ensemble of AI and mental health looks set to change what it is to be considered well or unwell.[148]

139 Victoria Hollis et al, 'On Being Told How We Feel: How Algorithmic Sensor Feedback Influences Emotion Perception' (2018) 2(3) *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 114:1-114:31.

140 Ibid.

141 Ibid.

142 Ibid.

143 Lisa Parker et al, 'Mental Health Messages in Prominent Mental Health Apps' (2018) 16(4) *The Annals of Family Medicine* 338.

144 China Mills and Eva Hilberg, 'The Construction of Mental Health as a Technological Problem in India' (2020) 30(1) *Critical Public Health* 41. Nev Jones, for example, has examined the impact of other ways of scientifically framing mental distress, including genetic and neurobiological causal attributions of psychiatric disorder, which she warns can undercut the agency of people in distress and the nuance of the individual experiences. Nev Jones, 'Agency, Biogenetic Discourse and Psychiatric Disorder', *Somatosphere* (18 September 2012) <http://somatosphere.net/2012/agency-biogenetic-discourse-and-psychiatric-disorder.html/>.

145 Hollis et al (n 141).

146 Harris, 'The Rise of the Digital Asylum' (n 121). This broader point was made by Frank Pasquale (see above n 6).

147 Kaitlin L Costello and Diana Floegel, '"Predictive Ads Are Not Doctors": Mental Health Tracking and Technology Companies' (2020) 57(1) *Proceedings of the Association for Information Science and Technology* e250.

148 Dan McQuillan, 'Mental Health and Artificial Intelligence: Losing Your Voice' (12 November 2018) *openDemocracy* <https://www.opendemocracy.net/en/digitaliberties/mental-health-and-artificial-intelligence-losing-your-voice-poem/>.

Moving the Frame from 'What does the technology do?' to 'Who is benefiting and who is not?'

One analytical strategy to help counter these negative possibilities is to place the emphasis away from the technology itself and toward questions of who is benefiting from the push for these technologies, and – perhaps more importantly – who is losing. This framing challenges the common presentation of computational monitoring and evaluation as naturally being in people's interests on the basis that 'the more we know the more we can help'. Such an optimistic view can easily dovetail with widely-held understandings about the legitimacy and unquestioned benefit of monitoring persons experiencing distress, lived experience and disability. As Sharon Snyder and David Mitchell have argued, '[o]ne of the primary oppressions experienced by disabled people is that they are marked as perpetually available for all kinds of intrusions, public and private.'[149]

The broad group of critical commentators raising concerns with biometric monitoring draw attention to the potential intrinsic harms of processes of computational observation and measurement. Just as the 'medical gaze' has been used as a concept to critique the biomedical and individualistic framing of distress and other human experiences, some commentators have considered the potential harms of the 'data gaze'. The ACLU, for example, describe a potential 'nightmare scenario' whereby a 'data gaze' extends to omni-present AI-powered monitoring and surveillance:

> the consistent tracking of our every conscious and unconscious behavior that, combined with our innate social selfconsciousness, turns us into quivering, neurotic beings living in a psychologically oppressive world in which we're constantly aware that our every smallest move is being charted, measured, and evaluated against the like actions of millions of other people — and then used to judge us in unpredictable ways.[150]

These concerns were not raised about the mental health context in particular, though they resonate with the concerns discussed in this section.

Others have raised concerns about the subtle harms caused by the way technological surveillance leads to an abstraction of the human body, which is then reassembled through a series of data flows.[151] Jathan Sadowski has argued that the abstraction of 'datafication' is itself a form of violence.[152] Extending these critiques to the disability context, Jackie Leach Scully and Georgia Van Toorn have argued that broader 'datafication' of the human body will delineate increasingly rigid boundaries between normality and disability.[153] This impulse to quantify and distinguish embodied difference, they argue, 'diverts attention from the realities of disabled lives, at a time when disability scholars and activists are arguing for more rather than less attention to the lived experience of disability'.[154] LLana James, discussing algorithmic racism and the impacts of the digital turn on other marginalised groups, has discussed how datafication can undermine the need to 'act on the reliable narrator' (that is, listening to the person or populations affected and how they articulate their needs).[155] Instead, dominant narratives about technology insist on new and alternative ways to undertake expert observation and

149 Sharon L Snyder and David T Mitchell, *Cultural Locations of Disability* (University of Chicago Press, 2006) p.628.

150 Jay Stanley, *The Dawn of Robot Surveillance: AI, Video Analytics, and Privacy* (American Civil Liberties Union, 2019) 36.

151 Kevin D. Haggerty and Richard V. Ericson, 'The Surveillant Assemblage' (2000) 51 *British Journal of Sociology* 611; R.E. Smith, *Rage inside the machine: The prejudice of algorithms, and how to stop the internet making bigots of us all* (Bloomsbury Academic, 2019).

152 Jathan Sadowski, *Too Smart: How Digital Capitalism Is Extracting Data, Controlling Our Lives, and Taking Over the World* (MIT Press, 2020) 46.

153 Jackie Scully and Georgia Van Toorn, 'Datafying Disability: Ethical Issues in Automated Decision Making and Related Technologies – AABHL 2021' (19 November 2021) <http://www.aabhlconference.com/3563>.

154 Ibid.

155 LLana James, 'Race-Based COVID-19 Data May Be Used to Discriminate against Racialized Communities', *The Conversation* (15 September 2020) <http://theconversation.com/race-based-covid-19-data-may-be-used-to-discriminate-against-racialized-communities-138372>.

monitoring using data-driven technology.[156] In the disability context, including the mental health context, the use of automation risks diverting attention from the experienced reality of disabled lives.[157]

If these concerns are taken seriously, the use of technologies like AI to make assumptions and judgements about who we are, and who we will become is much more than a potential invasion of privacy; it is an existential threat to human autonomy and the ability to explore, develop and express our identities. It is potentially a normalising of surveillance in a way that is reminiscent of 19th century asylums as a state-authorised site of control over disabled lives, but using 21st century techniques of ubiquitous observation and computational 'processing'. Grappling with these possibilities will be a necessary part of discussion about the potential harms and public benefits afforded by technology in the mental health context in general, particularly biometric monitoring.

1.4.2 Governing the Future of Biometric Monitoring in Mental Health Settings

Biometrics more generally are the subject of a growing field of research, practice, advocacy, activism, and law reform.[158] In healthcare, the COVID-19 pandemic has accelerated the international adoption of forms of biomonitoring and surveillance, and other public health monitoring and security technologies, whether adopted by states, private entities or individuals.[159]

In the mental health context, scholarship that explores the legal, ethical, social, and political concerns with biometric technologies is emerging.[160] More work is clearly required. Later themes discussed in this report will engage with some of the questions directly relevant to biometrics. Such questions include asking if those deemed through biometric monitoring to be 'cognitively impaired', 'mentally disordered', 'suicidal', or *likely to become* any of those things, will be informed that such attributions have been made. Will they be able to opt-out of the monitoring process in the first place? Will they be able to contest such labels before data are transferred to others? Given the purported ease with which mobile phone data-points can be used for automated profiling to determine cognitive impairment,[161] are there sufficient safeguards to govern whether or how this should occur? More pointedly, should moratoria apply to some forms of biometric monitoring and surveillance in the mental health and disability context on the basis that they are fundamentally harmful or inconsistent with human rights? How would such a decision be made? What role is currently being played by psychiatric and psychological sciences in advancing such technologies? What role *should* they play?

This is a critical moment to reflect how the current choices being made in various institutions concerning 'digital mental health' – from research, services, policies and programming – might affect future approaches to distress, anguish, mental crises and so on. To conclude Part 1 we turn to the glaring omission from these choices of the very people for whom the technologies are purportedly designed.

156 Schulich Law, *Algorithmic Racism, Healthcare & The Law: 'Race Based' Data Another Trojan Horse?* (19 September 2020) <https://www.youtube.com/watch?v=PveOVJYlu3I>.

157 Scully and Van Toorn (n 155).

158 Kak (n 100).

159 'Covid-19 Is Accelerating the Surveillance State', The Strategist (17 November 2020) 19 <https://www.aspistrategist.org.au/covid-19-is-accelerating-the-surveillance-state/>; 'Homo Deus Author Yuval Harari Shares Pandemic Lessons from Past and Warnings for Future', *South China Post* (online, 1 April 2020) <https://www.scmp.com/news/china/article/3077960/homo-deus-author-yuval-harari-shares-pandemic-lessons-past-and-warnings?fbclid=IwAR2b6pMEt1Gj4mpsBjSapqwL79e_tg_76eL4MLL788WYGDgTGRDbkM1H8y8>.

160 See e.g. Cosgrove et al (n 98); Bossewitch, 'The Rise of Surveillance Psychiatry and the Mad Underground' (n 133); Harris, 'The Rise of the Digital Asylum' (n 119).

161 Jonas Rauber, Emily B Fox and Leon A Gatys, 'Modeling Patterns of Smartphone Usage and Their Relationship to Cognitive Health' [2019] *arXiv:1911.05683 [cs, stat]* <http://arxiv.org/abs/1911.05683>.

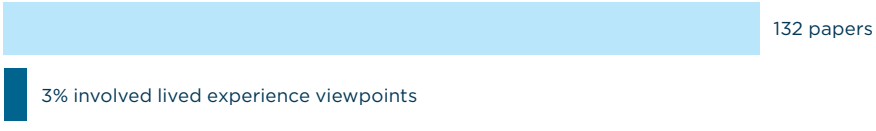
1.5 Elevating the Perspective of People with Lived Experience of Extreme Distress and Disability

> As mental health advocates, we work to ensure there is "nothing about us without us." What happens when our voice is missing from the system? Well, very bad things. In the name of treatment, we've experienced injustice, neglect, and abuse.
>
> - Kelechi Ubozoh[162]

For many decades, people with lived experience of distress and mental health issues have had a profound impact on rethinking and rearranging societal responses to mental health and distress.[163] In policy and practice, this group has altered service provision and prompted policy change and law reform. Quite outside of traditional services they have established forms of mutual aid and community development to help people in personal crisis, profound distress and extreme states of consciousness. In research, service user and survivors and representative groups have challenged traditional research assumptions, theories and methods, developed ethical frameworks and aligned their work to other social movements. All of this has contributed greatly to the development of knowledge about distress, mental health, illness and disability.[164]

The involvement of diverse groups of people with firsthand experience in mental health services in research typically affects how that research is undertaken, including the empirical and conceptual approaches that are chosen, and what is produced. Yet, in many public documents celebrating the positive potential of digital technologies in mental healthcare, there is a concerning lack of partnership with people with firsthand experience of mental health services and their representative organisations.[165] In a 2021 survey, Piers Gooding and Timothy Kariotis reviewed all applied studies that used algorithmic and data-driven technologies in 'online mental health interventions'.[166] Of the 132 papers in the survey, only four (or 3% of the field captured in the survey) appeared to involve people who have used mental health services, or those who have lived experience or psychosocial disability, in the design, evaluation or implementation of the proposals in any substantive way (Refer to Figure 1). The studies demonstrated 'a near-complete exclusion of service users in the conceptualisation or development of algorithmic and data-driven technologies' and their application to mental health services.[167] This pattern conforms with a longstanding marginalisation of lived experience perspectives in academic research.[168]

Figure 1: The field captured in the survey of 'online mental health' studies



162 Green and Ubozoh (n 17).
163 Ibid.
164 Jasna Russo and Stephanie Wooley, 'The Implementation of the Convention on the Rights of Persons with Disabilities' (2020) 22(1) *Health and Human Rights* 151; Robyn Brown and Nev Jones, 'The Absence of Psychiatric C/S/X Perspectives In Academic Discourse: Consequences and Implications' (2012) 33 *Disability Studies Quarterly*.
165 Sarah Carr, '"AI Gone Mental": Engagement and Ethics in Data-Driven Technology for Mental Health' (2020) 0(0) Journal of Mental Health 1; Til Wykes, 'Racing towards a Digital Paradise or a Digital Hell?' (2019) 28(1) *Journal of Mental Health* 1.
166 Gooding and Kariotis (n 43).
167 Ibid.
168 Brown and Jones (n 166).

Exclusion does not define all initiatives in the field. Indeed, there are good examples of technological responses that have been designed with a high level of active input by those most affected by the technology[169]. There are also several data-driven technologies *initiated and led* by people with lived experience of distress and mental health services,[170] including a growing range of 'digital peer support measures' around the world.[171]

CASE STUDY: 'CommonGround'

'CommonGround' is an example of a digital decision-making aid to help facilitate communication and share information between those accessing services and those providing them and can help people navigate through service options. CommonGround is a computer-interface presented in the waiting rooms of mental health settings and elsewhere, as a 'a web application to support shared decision making in the psychopharmacology consultation'.[172] CommonGround was developed by Patricia Deegan, a disability-rights advocate, psychologist and researcher who draws explicitly on her experience as a mental health service user. Service users are invited by a peer worker – that is, someone engaged to draw on their lived experience of mental health crisis, mental health service use, and so on – to complete a pre-consultation report about their personal preferences and values before meeting with a medical professional. This may include contextual information, such as the person's aims and values recorded in her/his own words, or her/his preferred activities to promote wellness and recovery.[173]

Hello Human, Hello Machine by Rachel Hanlon, Dr Johanne Trippas, Dr Matthew Gardiner, and Jess Coldrey in Science Gallery Melbourne's MENTAL. Photo by Alan Weedon. This version was developed in collaboration with Dr Johanne Trippas. Five members of Sci-Curious: Eli/Elena McGannon, Annabel Yenson, Claire Price, Jess Coldrey and Joseph Doggett-Williams. Creative technical assistance from Dr Matthew Gardiner.

169 See eg., John Torous et al, 'Creating a Digital Health Smartphone App and Digital Phenotyping Platform for Mental Health and Diverse Healthcare Needs: An Interdisciplinary and Collaborative Approach' (2019) 4(2) *Journal of Technology in Behavioral Science* 73.

170 Patricia E Deegan et al, 'Best Practices: A Program to Support Shared Decision Making in an Outpatient Psychiatric Medication Clinic' (2008) 59(6) *Psychiatric Services 603.*

171 Karen L Fortuna et al, 'Digital Peer Support Mental Health Interventions for People With a Lived Experience of a Serious Mental Illness: Systematic Review' (2020) 7(4) *JMIR Mental Health* e16460.

172 Ibid.

173 Deegan et al (n 172).

Other initiatives promote mutual forms of peer support and community development.

CASE STUDY: Virtual Support Network, Kenya

In Kenya, a volunteer-run 'virtual support network' emerged from the advocacy organisation *Users and Survivors of Psychiatry in Kenya* (*USP Kenya*) and has been running for several years.[174] There are 8 administrators and 200+ members.[175] Most members are individuals who have accessed mental health services themselves, but there are also members who are family members, caregivers, psychologists and counsellors.[176] The network communicates on a mainstream messenger service and is described in a *USP Kenya* report as being 'fully community-based, operat[ing] outside Kenya's mental health system and [not linked] to any mental health institution'.[177] The peer support involves crisis support for individual members, regular face-to-face meetups, information sharing, the generation of fundraising for individual members who are in financial crisis (particularly following the COVID-19 pandemic), the connecting of individuals to local community organisations, and so on.[178] New members are provided with guidelines for participation and content moderation. Some basic advance-planning is provided whereby members can indicate what type of support they would like during future crises, including family contact information, but not all members wish to share this information. Some members have not shared their mental health diagnosis publicly beyond the group and prioritise privacy. The network receives no funding.

Informal initiatives such as the Kenyan virtual support network may not make it into the public spotlight in the same way governments, health practitioners, large NGOs, and private sector actors do. Nor may they deploy AI or other 'cutting edge' technologies. Yet, they often warrant resources or further research to determine how and why they are working (if indeed they are) and how they can be supported.

Other peer-led initiatives use data-driven technologies in systemic advocacy and the monitoring of state-run services.

CASE STUDY: Open Data Advocacy and Public Monitoring of Disability Services

In 2021, a Canadian coalition of open data advocacy groups in collaboration with disabled people's organisations aimed to crowd-source a database of congregate institutions for disabled people in Canada, which included people with psychosocial disabilities. They aimed to trace the impact of COVID-19 on disabled people and prioritise vaccinations. A collaboration between open data groups led to a public event in which members of the public could join an online initiative to 'Hack the Data Gap' and create an up-to-date database of relevant residential facilities.[179]

174 USP Kenya, *The Role of Peer Support in Exercising Legal Capacity* (Nairobi, 2018) 18 <http://www.uspkenya.org/wp-content/uploads/2018/01/Role-of-Peer-Support-in-Exercising-Legal-Capacity.pdf>; Transforming communities for Inclusion, Asia, *Summary Report on Transforming Communities for Inclusion - Asia: Working Towards TCI - Asia Strategy Development* (Asia-Pacific Development Centre on Disability, June 2015) <www.apcdfoundation.org/?q=system/files/TCI%20Asia%20Report_Readable%20PDF.pdf> accessed 5 May 2016.

175 Videocall discussion between the author and Ms Ann Njambi and Ms Charity Muturi (18/08/2021).

176 Ibid.

177 USP Kenya, *The Role of Peer Support in Exercising Legal Capacity* (Nairobi, 2018) 18 <http://www.uspkenya.org/wp-content/uploads/2018/01/Role-of-Peer-Support-in-Exercising-Legal-Capacity.pdf>; Transforming communities for Inclusion, Asia, *Summary Report on Transforming Communities for Inclusion - Asia: Working Towards TCI - Asia Strategy Development* (Asia-Pacific Development Centre on Disability, June 2015) <www.apcdfoundation.org/?q=system/files/TCI%20Asia%20Report_Readable%20PDF.pdf> accessed 5 May 2016.

178 Ibid.

179 http://datalibre.ca/2021/02/18/invisible-people-and-institutions-no-data-about-custodial-institutions-for-disabled-people-in-canada/ [accessed 17/03/2021]

However, despite some good examples, there are concerning signs that much activity in academia, the market and government have not adopted the standard of active involvement of people with lived experience.[180]

Furthermore, of the little commentary and scholarship by people with lived experience that does exist, most commentators tend to be more ambivalent about digital technology's role in mental healthcare and crisis responses than those in government, industry and professional bodies. These diverse and varied viewpoints will be discussed throughout this report.

Ultimately, our report is premised on the view that active involvement of those most impacted by algorithmic and data-driven technologies should not be seen merely as a matter of 'stakeholder engagement', but rather as an ethical orientation. This ethos requires a stronger social and political commitment by actors involved in digitising mental health initiatives to avoid the pitfalls of past research that was 'done to' and not *with* or *by* people who are primarily impacted.

Thoughtful, participatory design is also likely to result in higher quality technological practices that better meet the needs and preferences of those for whom they are designed. Without it, there is a greater likelihood of costly technologies being introduced in an unthinking manner, created to address one issue without sufficient thought to harmful flow-on consequences. According to Dainius Pūras, the former UN Special Rapporteur on the Right to the Highest Quality Physical and Mental Health:

> participation of persons with mental health conditions, including persons with disabilities, in the planning, monitoring and evaluation of services, in system strengthening and in research, is now more widely recognized as a way to improve the quality, accessibility and availability of services and the strengthening of mental health systems.[181]

Diverse parts of the international social movement of disabled people have also advocated along these lines, as have multiple international and national human rights agencies.[182]

Harms perpetuated in the name of mental health care in the past offer a cautionary tale for any proposed solutions in mental health services today that exclude affected populations. Clarence Sundram has written of widespread abuse and violence perpetrated in recent times, in which people deemed mentally, intellectually and cognitively impaired in some way were subject to cruel, inhuman and degrading treatment of various kinds.[183] This included arbitrary detention without legal process (sometimes for life), forced sterilisation, being chained and caged, confinement to squalid conditions in institutions, the use of painful medicines and procedures, irreversible surgical interventions, and medical experimentation against individuals' wishes, including experimentation with no intended benefit for the person.

180 Carr (n 53); Wykes (n 167); Gooding and Kariotis (n 43).

181 See eg, Human Rights Council, 'Report of the Special Rapporteur on the Right of Everyone to the Enjoyment of the Highest Attainable Standard of Physical and Mental Health' para [13] <https://primarysources.brillonline.com/browse/human-rights-documents-online/promotion-and-protection-of-all-human-rights-civil-political-economic-social-and-cultural-rights-including-the-right-to-development;hrdhrd99702016149>.

182 Australian Human Rights Commission, *Human Rights and Technology - Final Report* (Australian Human Rights Commission, 2021) <https://tech.humanrights.gov.au/sites/default/files/2021-05/AHRC_RightsTech_2021_Final_Report.pdf>; Theresia Degener, 'Disability in a Human Rights Context' (2016) 5(3) *Laws* 35.

183 Clarence Sundram, 'In Harm's Way: Research Subjects Who Are Decisionally Impaired' (1998) 36(1) *Journal of Health Care Law and Policy*.

This history recalls that harms in the name of care have occurred in living memory, some of which continue today (as discussed throughout Part 2). Many people who have experienced distress, psychosis, psychosocial disability and so on, have had negative, violative and dismissive experiences in mental health services—even as many have had positive experiences.[184] Several commentators with a range of experiences with mental health services have demanded that the digital turn must not extend or exacerbate these historical patterns of harm, even if that means proceeding cautiously.[185]

CASE STUDY: The Halting of an App by the UK Mental Health Foundation

David Crepaz-Keay, the Head of Applied Learning at the UK Mental Health Foundation reported that an app being developed by the Mental Health Foundation to assist mental health service users was indefinitely halted during an internal consultation process, after service user advisors raised serious concerns. Concerns included privacy being compromised and the possibility of individuals' data being shared with companies and government agencies, including criminal justice agencies.[186]

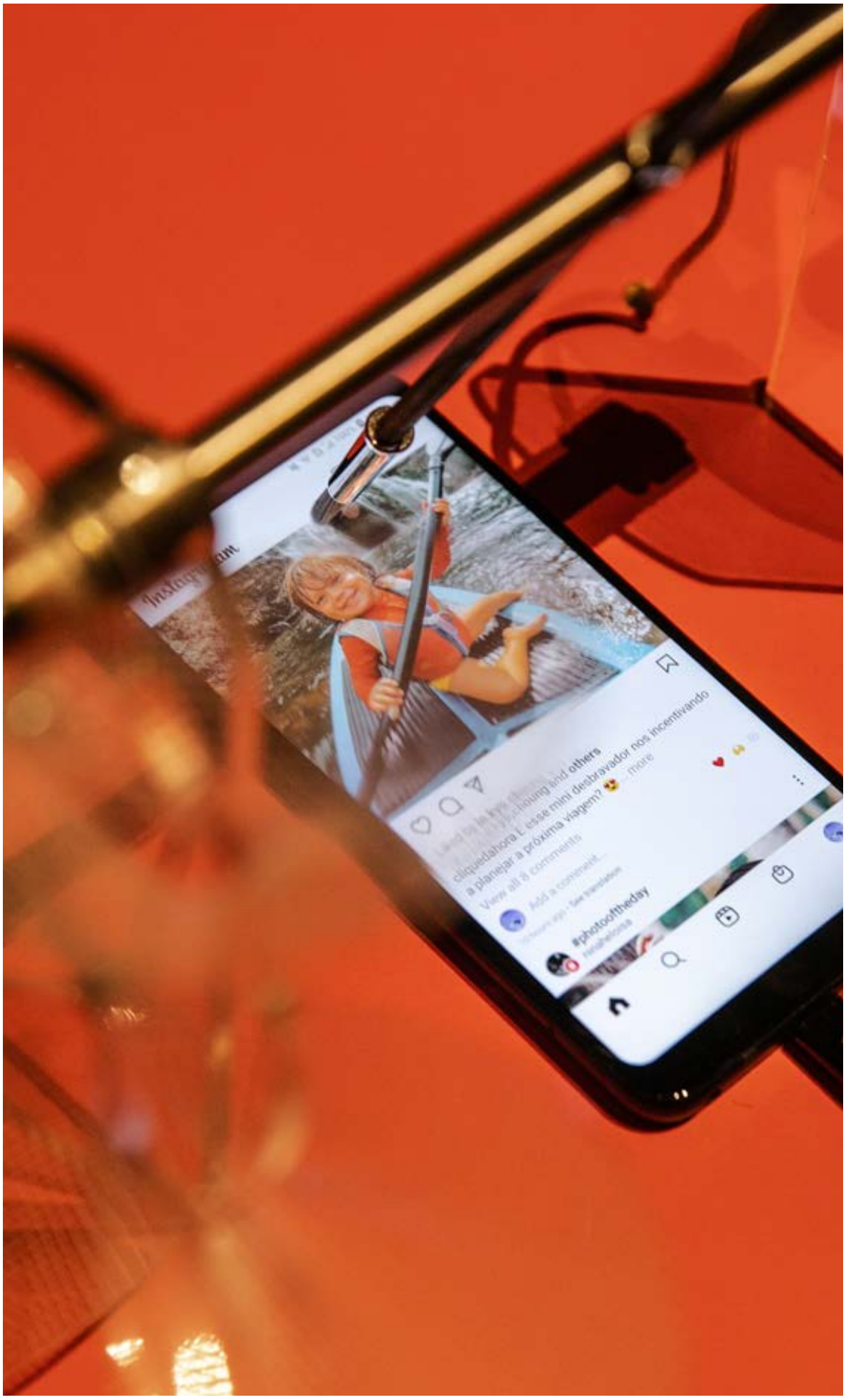
The consequences of poorly designed algorithmic and digital technologies to assist people in mental health crises will be borne by people who have engaged with mental health services, or who live with distress, illness and disability. Hence, these groups must be actively involved in governance of algorithmic and data-driven technology in the mental health context. Jonah Bossewitch puts it succinctly when he writes: 'It is possible to redirect this wizardly technology to help support people better. Doing this well starts with inclusive design—people with lived experience need to be involved in planning and shaping the systems meant to support them. Nothing about us without us.'[187]

184 See Andrea Daley, Lucy Costa and Peter Beresford (eds), *Madness, Violence, and Power: A Critical Collection* (University of Toronto Press, Illustrated edition, 2019).

185 Harris, 'The Rise of the Digital Asylum' (n 121); Bossewitch, 'Brave New Apps' (n 121); Carr (n 53).

186 Privacy International, *Your Mental Health for Sale?* (6 November 2020) <https://www.youtube.com/watch?v=Sbsw51OrvBU&list=UUwyKZWhsD2YFg8huOaO3lOg&ab_channel=PrivacyInternational>.

187 Bossewitch, 'Brave New Apps' (n 121).



Stop the Algorithm by Stephanie Kneissl and Max Lackner in Science Gallery Melbourne's MENTAL. Photo by Alan Weedon.

Part 2 - Themes for Responsible Public Governance

To draw out the various issues raised by the rise of automation in the mental health context, the remainder of this report is ordered around the following inter-related themes, or ethical and political values:

These themes commonly appear in global discussions about algorithmic technology and data governance.[188] However, we add the following caveats. By ordering our discussion under these themes, we are not suggesting that a statement of 'principles' is needed for the mental health and disability context – quite rightly, 'principles documents are frequently challenged as toothless or unenforceable.'[189] At its worst, looking at ethical themes or values can draw attention away from broader questions of justice, power, finance and politics that drive the recent growth of algorithmic and data-driven technologies, as discussed above. It should be acknowledged that the themes we have tentatively listed risk narrowing the focus of public debate to questions of procedural safeguards – for example, by zeroing in on auditing processes designed to achieve fairness, accountability and transparency. This narrow focus can divert attention from more fundamental questions, such as whether or not certain systems should be built at all, whether because they have proven harms or because they have unproven benefits.

However, given the striking lack of research on the politics of mental health and disability-related automation,[190] these themes provided a way to frame our discussion using themes that are common in broader public discussion about algorithmic and data-driven technology, even as these themes may need to be re-framed. Finally, the value of the discussion is dependent upon the integration of these themes in larger public governance systems, from legislation, regulation, the work of professional associations, the advocacy of civil society organisations, activism, quality journalism, and everyday practices designed to support people or communities in crisis.[191]

+ These eight themes are identified by the Berkman Klein Center, which compared the contents of thirty-six prominent AI principles documents side-by-side. Jessica Fjeld et al, *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI* (SSRN Scholarly Paper No ID 3518482, Social Science Research Network, 15 January 2020) <https://papers.ssrn.com/abstract=3518482>. 47 principles were identified that sit beneath these themes. My aim is not to exhaustively arrange material beneath each specific principle but to link our discussion to emerging areas of consensus in what can otherwise be a fractured global conversation on 'trustworthy AI', 'algorithmic accountability' and other efforts to create technology responsibly.

188 Ibid.

189 Ibid p.35.

190 With notable exceptions, such as Aimi Hamraie and Kelly Fritsch, 'Crip Technoscience Manifesto' (2019) 5(1) *Catalyst: Feminism, Theory, Technoscience* 1; Jacinthe Flore, 'Ingestible Sensors, Data, and Pharmaceuticals: Subjectivity in the Era of Digital Mental Health': [2020] *New Media & Society* <http://journals.sagepub.com/doi/10.1177/1461444820931024>; Bossewitch, 'The Rise of Surveillance Psychiatry and the Mad Underground' (n 135); Harris, 'The Rise of the Digital Asylum' (n 121); Mara Mills, 'Deaf Jam: From Inscription to Reproduction to Information' (2010) 28(1 (102)) *Social Text* 35.

191 Indeed, these arguments have already been made during attempts to develop human rights-based governance of the broader mental health system. See eg. S Katterl & C Maylea 'Keeping human rights in mind: embedding the Victorian Charter of Human Rights into the public mental health system' (2021) 27(1) *Australian Journal of Human Rights* 58-77; Tina Minkowitz, *Reimagining Crisis Support: Matrix, Roadmap and Policy* (Lilith's Warrior Press, 2021).

2.1 Privacy

> [R]easserting the right to privacy... may be a central component [for building] a platform for democratic governance and social equality within an information society.
>
> – Jonah Bossewitch and Aram Sinnreich[192]

The right to privacy is longstanding in international human rights law and aims to protect a citizen from unlawful interference with their private life and correspondence.[193] Yet, the contemporary communication ecosystem allows vast amounts of data to feed into technologies for use in surveillance, advertising, healthcare decision-making, and many other sensitive contexts.[194] Complex forms of government and private sector monitoring exist that draw on sophisticated technologies to trace individuals and detect their behaviour, networks, consumption and so on. The role of data concerning mental health in these processes is poorly understood.

Clearly, privacy over individuals' and communities' data concerning distress and mental health is vital. Failure to attend to privacy issues could have multiple negative consequences: it may shape individuals' willingness to disclose their distress (for example, privacy concerns may undermine a person's confidence in seeking support); data may be used to discriminate against individuals, families or groups, which could be used in the context of insurance, employment, housing, credit ratings and so on; data may be sold or monetised without an individual's consent, including being used to inform manipulative advertising practices; and may enable identity theft and health system fraud. People may avoid disclosing or accessing a service for fear that their medical biography or data on their mental and personal life may be used to their disadvantage in the future—which crucially undermines trust, which is so essential to therapeutic engagement and the seeking of support.

Alternatively, many people will simply be unaware of the risks. Indeed, most consumers have low awareness about the implications of data sharing practices within the larger communication eco-system.[195]

CASE STUDY: Privacy and 'Mental Health Apps'

In 2015, the National Health Service of England closed its App Library after a study found that 20 percent of the apps lacked a privacy policy and one even transmitted personally identifiable data that its policy claimed would be anonymous.[196] The authors concluded that:

- 89% (n = 70/79) of apps transmitted information to online services
- No app encrypted personal information stored locally
- 66% (23/35) of apps sending identifying information over the Internet did not use encryption and 20% (7/35) did not have a privacy policy

Two studies undertaken in 2019 found that only just under half of the popular mental health apps surveyed had a privacy policy that informed users about how and when personal information would be collected or shared with third parties.[197]

192 Jonah Bossewitch and Aram Sinnreich, 'The End of Forgetting: Strategic Agency beyond the Panopticon' (2013) 15(2) *New Media & Society* 224.
193 International Covenant on Civil and Political Rights, Article 17.
194 Fjeld et al (n 77) p.21.
195 A Razaghpanah et al. 'Apps, trackers, privacy and regulators. A global study of the mobile tracking ecosystem'. Paper presented at the Network and distributed systems security (NDSS) symposium. 18–21 February 2018.
196 Kit Huckvale et al, 'Unaddressed Privacy Risks in Accredited Health and Wellness Apps: A Cross-Sectional Systematic Assessment' (2015) 13(1) *BMC Medicine* 214.
197 Kristen O'Loughlin et al, 'Reviewing the Data Security and Privacy Policies of Mobile Apps for Depression' (2019) 15 *Internet Interventions* 110; Lisa Parker et al, 'How Private Is Your Mental Health App Data? An Empirical Study of Mental Health App Privacy Policies and Practices' (2019) 64 *International Journal of Law and Psychiatry* 198.

The proliferation of direct-to-consumer apps and therapy platforms highlights the expansion of digital mental health initiatives in marketised form. Nicole Martinez-Martin has written that 'the consumer domain presents particularly vexing issues for trust, because the frameworks for accountability and oversight, as well as the mechanisms for data protection and assuring safety and effectiveness, are still evolving'.[198]

2.1.1 Ad-Tech and Predictive Public Health Surveillance

Data and analytics in advertising can exploit behavioural biases and create consumer exploitation or be used in political targeting on an unprecedented scale.[199] Sensitive information about people who are potentially in distressed states can be used by private companies to manipulate people into buying certain services or products.

CASE STUDY: Facebook/Meta ad-tech identifying when children feel 'worthless' and 'insecure'

In 2017, Australian media reported that Facebook systems could target Australians and New Zealander children as young as 14 years old and help advertisers to exploit them when they're most vulnerable.[200] This included identifying when the children felt 'worthless', 'stressed', 'anxious', 'insecure,' and in 'moments when young people need a confidence boost'. The document offering these capacities to advertisers was authored by two top Australian executives in roles described as 'Facebook Australia's national agency relationship managers.'[201]

Facebook denied that it let advertisers target children and young people based on their emotional state, and claimed that it has 'an established process' to review such research but that this particular project 'did not follow that process.'[202] Facebook reasserted that it had a policy against advertising to 'vulnerable users.'[203]

Four years later, in April 2021, civil society organization Reset Australia reported that Facebook was found using children's data to on-sell to advertisers seeking to target children interested in extreme weight loss, alcohol or gambling.[204]

Conversely, a similar approach to targeted advertising can be used in public health initiatives that seek to direct people who may be in distress to particular mental health services. The aim of such initiatives is to provide 'pre-emptive' support to assist people to access support, particularly those who may be averse to accessing formal services, as the following example shows.

198 Nicole Martinez-Martin, 'Chapter Three - Trusting the Bot: Addressing the Ethical Challenges of Consumer Digital Mental Health Therapy' in Imre Bárd and Elisabeth Hildt (eds), *Developments in Neuroethics and Bioethics* (Academic Press, 2020) 63 <http://www.sciencedirect.com/science/article/pii/S2589295920300138>.

199 Sam Levin, 'Facebook Told Advertisers It Can Identify Teens Feeling "insecure" and "Worthless"', *The Guardian* (online, 1 May 2017) <http://www.theguardian.com/technology/2017/may/01/facebook-advertising-data-insecure-teens>.

200 Ibid.

201 Ms Smith, 'Facebook Able to Target Emotionally Vulnerable Teens for Ads' [2017] *Network World* (Online) <https://www.proquest.com/trade-journals/facebook-able-target-emotionally-vulnerable-teens/docview/1893625693/se-2?accountid=12372>.

202 'Comments on Research and Ad Targeting', *About Facebook* (30 April 2017) <https://about.fb.com/news/h/comments-on-research-and-ad-targeting/>.

203 Ibid.

204 Conor Duffy, 'Facebook Harvests Teenagers' Data and on-Sells It to Advertisers for Targeted Alcohol, Vaping Ads, Report Finds', *Australian Broadcasting Commission* (online, 27 April 2021) <https://www.abc.net.au/news/2021-04-28/facebook-instagram-teenager-tageted-advertising-alcohol-vaping/100097590>.

CASE STUDY: Predictive prevention and targeted advertising

A report by the PHG Foundation describe a 'digital wellbeing service' in London:[205]

> Good Thinking is a digital wellbeing service rolled out across the city as part of the Healthy London Partnership; it uses data-driven marketing techniques to target advertisements for digital services to people who may be experiencing mental health issues. This targeting is based on people's use of online search engines and social media platforms, thereby proactively identifying those who may benefit from services and who would not necessarily self-present to the health system. Those who display patterns of searching or social media posts consistent with early predictors of mental health decline [for example, sleep deprivation, isolation, alcohol consumption] are targeted with subtle advertisements around their personal issue. If the user engages with the advertisement, they are filtered through to a digital service containing recommended and approved apps for their specific problem. All this is done without the health system requiring access to raw data or any personal information about the user and the citizen is not aware they are engaging with the health system.

The intent of the Good Thinking targeting service, which is described as a 'precision prevention initiative' by its developers, is to provide benefit to individuals in apparent distress. This aim aligns with the UK Government's Green Paper on 'Advancing our health: prevention in the 2020s', which describes a move towards 'proactive, predictive and personalised prevention'.[206]

Serious questions may be raised about such programs given that they seem to largely target people from *outside* the formal healthcare system who have not consented to or necessarily intentionally sought out health care services. Although the advertisements for Good Thinking are targeted at those who appear to be searching for support in relation to distress, one component of the targeted digital advertising '[t]argets users whose behaviour, demographic and location suggests they are a potential service users [sic] – a "passive" audience.'[207]

Data-based targeting and automated profiling was discussed by the World Health Organisation in its report, *Ethics and Governance of Artificial Intelligence for Health*, which noted its ambiguous potential:

> AI can be used for health promotion or to identify target populations or locations with "high-risk" behaviour and populations that would benefit from health communication and messaging (micro-targeting) [...] Micro-targeting can also, however, raise concern, such as that with respect to commercial and political advertising, including the opaqueness of processes that facilitate micro-targeting. Furthermore, users who receive such messages may have no explanation or indication of why they have been targeted. Micro-targeting also undermines a population's equal access to information, can affect public debate and can facilitate exclusion or discrimination if it is used improperly by the public or private sector.[208]

205 PHG Foundation, *Citizen Generated Data and Health: Predictive Prevention of Disease* (University of Cambridge, November 2020) <https://www.phgfoundation.org/documents/cgd-predictive-prevention-of-disease.pdf>.

206 Advancing our health: Prevention in the 2020s – consultation document, Cabinet Office and Department of Health & Social Care (2019). Accessed at https://www.gov.uk/government/ consultations/advancing-our-health-prevention-in-the-2020s/advancing-our-health-preventionin-the-2020s-consultation-document on 24 September 2019.

207 *The Good Thinking Journey: How the First-Ever City-Wide Digital Mental Wellbeing Service Helped a Quarter of a Million Londoners* (September 2019) <https://www.healthylondon.org/wp-content/uploads/2019/09/Good-Thinking_How-the-first-ever-city-wide-digital-mental-wellbeing-Sept-2019.pdf>.

208 World Health Organisation, *Ethics and Governance of Artificial Intelligence for Health* (World Health Organization, 28 June 2021) 13.

A prominent example of a data-driven preventive health monitoring initiative that faced a severe public backlash was the 'suicide watch radar' app that was trialled in the UK.

CASE STUDY: Crisis Surveillance and the 'Suicide Watch Radar' App

In 2014, the UK charity Samaritans abandoned its use of a 'suicide watch radar' app, which enabled users to monitor the accounts of another user for distressing messages. The project aimed to direct emergency responders to those in crisis. However, public campaigners argued the tool breached user's privacy by collecting, processing and sharing sensitive information about their emotional and mental health.[209] Dan McQuillan commented of the program:

> Thanks to the inadequate involvement of service users in its production,
>
> It ignored the fact that the wrong sort of well-meaning intervention at the wrong time might actually make things worse,
>
> Or that malicious users could use the app to target and troll vulnerable people.[210]

Automated profiling such as the Samaritan's 'Radar' app clearly engages the new generation of data protection laws. The EU's General Data Protection Regulation (GDPR), for example, defines automated profiling as 'any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements.'[211] Other parts of the GDPR would bear on how such profiling could be used. Article 22, for example, states that 'the data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.'[212] Bernadette McSherry has noted that predicting aspects of a person's mental health appears likely to fall within the ambit of this Article.[213]

2.1.2 Privacy and Monetisation of Sensitive Personal Data

Privacy issues are compounded by the increasing monetisation of health and other forms of personal data. Online mental health initiatives are emerging in an internet that is increasingly dominated by profit-driven information flows. Some mental health websites or apps and affiliated third-party companies are treating the personal data of users as a commodity and tracking them for marketing or other commercial purposes.[214] This may occur as an explicit business decision by a private company that provides direct-to-consumer services for those in distress, or may occur inadvertently where a service provider is unaware of the way third-party trackers are operating on their platforms. 'Third-party trackers' which are sometimes described as 'tracking cookies' or 'trackers', are elements of websites that are created by parties other than the developers of the website the person is currently visiting; this would include providers of advertising, analytics and tracking services.[215]

209 Jamie Orme, 'Samaritans Pulls "Suicide Watch" Radar App over Privacy Concerns', *the Guardian* (7 November 2014) <http://www.theguardian.com/society/2014/nov/07/samaritans-radar-app-suicide-watch-privacy-twitter-users>.

210 McQuillan (n 150).

211 *Regulation (EU) 2016/679 of the European Parliament and the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)* [2016] OJ L 119/1, Art 4 "Definitions".

212 Ibid Art 22.

213 Bernadette McSherry, 'Computational Modelling, Social Media and Health-Related Datasets: Consent and Privacy Issues' (2018) 25(4) *Journal of Law and Medicine* 894.

214 Molly Osberg and Dhruv Mehrotra, 'The Spooky, Loosely Regulated World of Online Therapy', *Jezebel* (online, 19 February 2020) <https://jezebel.com/the-spooky-loosely-regulated-world-of-online-therapy-1841791137>.

215 Michal Wlosik and Michael Sweeney, 'First-Party & Third-Party Cookies: What's the Difference?', *Clearcode* (2 November 2018) <https://clearcode.cc/blog/difference-between-first-party-third-party-cookies/>. Third party trackers are mainly used for tracking and

CASE STUDY: Privacy International finds top mental health websites sell visitor information to third parties and breaches the GDPR

In 2019, Privacy International analysed more than 136 popular webs across France, Germany and the UK related to depression.[216] The websites were chosen to reflect those that people would realistically find when searching for help online. The authors found that over three quarters of web pages contained third-party trackers for marketing purposes, which could enable targeted advertising and marketing from large companies like Google/Alphabet, Amazon and Facebook/Meta.

Most websites, according to the authors, failed to comply with the EU General Data Protection Regulation in upholding individual's privacy (acknowledging that the UK is no longer part of the EU). In 2020, a follow-up study found that 31.8% (42 out of 132) of the tested webpages reduced the number of third parties with whom users' data was shared. However, Privacy International concluded that '[g]enerally, most websites analyses haven't taken action to limit data sharing [meaning][…] personal data are still shared for advertising purposes' with hundreds of third parties with no clear indication of the potential consequences.[217]

Knowledge of a user's distress could, at a minimum, allow companies to advertise specific treatments, services, or financial products, as noted previously. It could also be sold to other interested parties, such as insurers, as discussed later in the report (see Non-Discrimination and Equity). Some have suggested that data concerning the health of individuals will be more lucrative than the sale of particular health products. Nick Couldry and Ulises Ali Mejias have argued that this likelihood is evident in Amazon's recent moves in the US to open an online pharmacy:

> Amazon Pharmacy's promise of 80 per cent discounts suggests that the US retailer sees opportunities not in realising immediate profits, but in extracting a more valuable resource: data about the most intimate details of our lives.[218]

Not only may data extraction be used to predict the person's distress in order to match them to an advertised product, but another broader function may be to *shape the person's experience and behaviour in order to direct them to existing advertisement/products*. Zuboff refers to this shaping of human experience and behaviour when highlighting the emergence of 'behavioral futures markets'.[219] This may be evident in the Cerebral app, noted above, which reportedly pushed platform users toward shorter appointments and more prescriptions in ways that potentially 'accelerat[ed] the psychiatric prescribing cascade'.[220]

online-advertising purposes but can also provide certain services, such as live chats. These third-party elements are mainly used on mental health websites, according to Privacy International, for advertising and marketing purposes. Privacy International appear to have undertaken the most comprehensive research on this issue. 'First-party cookies' would refer to elements of a website developed by the website creators or operators that provide the same function as third-party cookies but which are operated and utilized by the website creators/operators themselves. Trackers by third-parties monitor users' behaviour across various online sources such as apps, smartphones, webs, smart TVs and so on. They may be used for a variety of reasons from connecting social media platforms to monitoring analytics of how a user interacts with a web or marketing purposes. They allow for a third-party to collect, monitor and use data related to a users' interaction with a specific online tool. Privacy International (n 29).

216 Privacy International, *Your Mental Health for Sale?* (n 188).

217 Privacy International, 'Mental Health Websites Don't Have to Sell Your Data. Most Still Do.', *Privacy International* (7 October 2021) <http://privacyinternational.org/report/3351/mental-health-websites-dont-have-sell-your-data-most-still-do> [accessed 14/07/21].

218 Nick Couldry and Ulises Ali Mejias, 'Big Tech's Latest Moves Raise Health Privacy Fears', *Financial Times* (online, 7 December 2020) <https://www.ft.com/content/01d4452c-03e2-4b44-bf78-b017e66775f1>.

219 Zuboff (n 128).

220 'ADHD Drugs Are Convenient To Get Online. Maybe Too Convenient' (n 137).

CASE STUDY: 'Practice Fusion' and Clinical Decision Support Software that Unlawfully Boosted Opioid Prescribing

The United States (US) government recently settled a case with a company called 'Practice Fusion', which produced clinical decision support software that was used by doctors when prescribing medication for patients, and was found to have received kickbacks from a pharmaceutical company intended to drive up opioid prescribing.[221] Megan Prictor explains that '[t]he payments were for creating an alert in the [electronic health record] designed to increase the prescription of extended-release opioid medication (and hence the sale of Purdue's products) to treat patients' pain symptoms.'[222] She notes:

> The court heard that Purdue Pharma's marketing staff helped to design the software alert, which ignored evidence-based clinical guidelines for patients with chronic pain… The alert was triggered in clinical practices some 230 million times between 2016 and 2019 and resulted in additional prescriptions of extended-release opioids numbering in the tens of thousands, causing untold human harm. Most of the prescriptions were paid for by federal healthcare programmes.[223]

The fraud was uncovered through a US government investigation, which had originally investigated separate unlawful conduct by the company concerning falsely obtained government certification for its software. The company had failed to meet certification requirements, which itself had led software users inadvertently to falsely claim government incentive payments. Software users – presumably comprising of various healthcare providers – had attested that the software complied with government regulations, when in reality it did not.[224]

It is clear that technologies are now being designed to push 'users' to access services or products aligned with business interests tied to the technology;[225] or to enforce conditional welfare and social benefit rules in government-funded services in ways that erode care,[226] as will be discussed later in the report.

Privacy International's finding that mental health websites sell visitor information to third parties highlights a striking fact: it is becoming harder to access mental health support without that access being digitally recorded in some way. The likelihood of such information moving beyond the discrete and relevant digital repositories of one service is increased by the massive and interconnected flow of data in today's communication ecosystem. A report for the Consumer Policy Resource Centre notes the implications of the ease with which data can be transported:

> consumers may well start to avoid accessing important healthcare services and support if they feel that companies or governments cannot be trusted with that information, or that they may be disadvantaged by that information in future. For example, insurer MLC was found to have excluded a consumer from mental health coverage in life insurance due to her accessing mental health services for the sexual abuse she suffered as a child in the mid-1980s.[227]

221 United States Attorney's Office, District of Vermont. 2020. Justice department announces global resolution of criminal and civil investigations with opioid manufacturer Purdue Pharma. October 21. https://www.justice.gov/usao-vt/pr/justice-department-announces-global-resolution-criminal-and-civil-investigations-opioid-0. Accessed April 3, 2022.

222 Megan Prictor, 'Clinical Software and Bad Decisions: The "Practice Fusion" Settlement and Its Implications' [2022] *Journal of Bioethical Inquiry* (Online First: 11/4/2022)

223 Ibid.

224 Ibid.

225 Some commentators have raised concerns that these technologies may unintentionally lead to an increase in the use of forced interventions where behavioural data indicate suicidality. Cosgrove et al (n 134) 620.

226 Alexandra Mateescu, E*lectronic Visit Verification: The Weight of Surveillance and the Fracturing of Care* (Data & Society, November 2021) <https://datasociety.net/library/electronic-visit-verification-the-weight-of-surveillance-and-the-fracturing-of-care/>.

227 Brigid Richmond, *A Day in the Life of Data: Removing the Opacity Surrounding the Data Collection, Sharing and Use Environment in Australia (*Consumer Policy Resource Centre, 2019) 37.

This concern extends to accessing physical services given that the monitoring power of smartphones through location-tracking can potentially show the frequency and types of healthcare services an individual accesses.[228] Some mental health initiatives that introduce a major digitalised or virtual component have explicitly prioritised privacy as a key component of appropriate support.

CASE STUDY: Privacy by Design in Digital Support in a Refugee Camp

In 2018, researchers at the Data & Society research institute released a report entitled 'Refugee Connectivity: A Survey of Mobile Phones, Mental Health, and Privacy at a Syrian Refugee Camp in Greece'.[229] The authors demonstrated ways that phones were essential to aid, survival and well-being. The survey design simultaneously employed two distinct methodologies: one concerned with mobile connectivity and mental health, and a second concerned with mobile connectivity and privacy. This project was supported by the International Data Responsibility Group. The research was premised on a view that privacy can be essential to easing distress and mental health, both in terms of receiving support, and in the lives of refugees and asylum seekers more generally, particularly those at risk of persecution.

Go Mental by Josh Muir in Science Gallery Melbourne's MENTAL. Photo by Alan Weedon.

228 Ibid.

229 Mark Latonero, Danielle Poole and Jos Berens, *A Survey of Mobile Phones, Mental Health, and Privacy at a Syrian Refugee Camp in Greece* (Harvard Humanitarian Initiative and the Data & Society Research Institute, 2018) 47.

2.1.3 Data Theft and Data Trafficking

As the amount and value of personal data stored online proliferates, data theft and trafficking will continue to occur. In 2017 in the US, for example, a mental health service provider in Texas notified 28,434 people whose data were allegedly stolen by a former employee.[230] However, by far the most extreme case concerning mental health was reported in Finland in October 2020.

'Vastaamo hacking could turn into largest criminal case in Finnish history'

On the 27th of October 2020, the Associated Foreign Press reported that:[231]

> The confidential treatment records of tens of thousands of psychotherapy patients in Finland have been hacked and some leaked online, in what the interior minister described as "a shocking act". Distressed patients flooded victim support services over the weekend as Finnish police revealed that hackers had accessed records belonging to the private company Vastaamo, which runs 25 therapy centres across Finland. Thousands have reportedly filed police complaints over the breach. Many patients reported receiving emails with a demand for €200 (£181) in bitcoin to prevent the contents of their discussions with therapists being made public.

Around 30,000 people are believed to have received the ransom demand at the time of writing; some 25,000 reported it to the police. Some of the records belonged to children, politicians, and other public figures. They contained details about adulterous relationships, sexuality hidden from family, suicide attempts, and paedophilic thoughts.[232]

Vastaamo, the private company that owned the leaked patient database, has since claimed bankruptcy.[233] At the time of writing, criminal proceedings are underway and victims would be able to seek compensation from the perpetrator(s) of the extortion if they are caught. In addition, Finland's Data Protection Ombudsman is reportedly looking into whether Vastaamo breached European Union data protection rules, which would mean Vastaamo would be responsible for compensating injured parties—though according to Leena-Kaisa Åberg, Executive Director of Victim Support Finland, any returns from the bankrupt company would be modest.[234]

Such incidents raise questions around the security required to protect people's privacy relating to mental health, distress and disability, to digitally store and process sensitive personal data (of which more is discussed in the Safety and Security section below). According to William Ralston, the example from Finland is particularly troubling because Finland is regarded as having among the most advanced electronic health policy and governance frameworks in the world.[235] Questions also arise about the security methods in place for technologies that are operating outside the formal healthcare context, such as the vast selection of mental health apps operated by private companies collecting personal data through people's smartphones. Indeed, the private company in Finland that was hacked, Vastaamo, was the largest private mental health operator in the country, and investigations is underway at the time of writing to determining where responsibility for the data breach lies.[236]

230 HIPAA, 'PHI of 28,000 Mental Health Patients Allegedly Stolen by Healthcare Employee' (5 December 2017) *HIPAA Journal* <https://www.hipaajournal.com/phi-28000-mental-health-patients-stolen-by-healthcare-employee/>

231 *AFP*, 'Shocking' hack of psychotherapy records in Finland affects thousands, *The Guardian* (27 Oct 2020).

232 William Ralston, 'They Told Their Therapists Everything. Hackers Leaked It All' *Wired* <https://www.wired.com/story/vastaamo-psychotherapy-patients-hack-data-breach/>.

233 'Compensation Uncertain for Vastaamo Victims', *Yle Uutiset* (online, 20 June 2021) <https://yle.fi/uutiset/osasto/news/compensation_uncertain_for_vastaamo_victims/11991155>.

234 Ibid.

235 Ralston (n 235).

236 Ibid.

2.1.4 Privacy and Discrimination

Privacy can be closely linked to issues of discrimination in the mental health context because a person's mental health status – such as their psychiatric diagnosis or record of encounters with health services – can be used in ways that are leveraged against them, including by potential employers, insurers, and state agencies.

CASE STUDY: Privacy and Border Discrimination

In 2017, Canadians with a documented history of mental health hospitalisations and particularly suicide attempts were being refused entry at the US border.[237] An inquiry by the Office of the Privacy Commission of Canada found that the Toronto Police had collected non-criminal mental health data and shared it with several government agencies—eventually it was shared with US Customs and Border Protection.[238] US border officials used the information (again, which was non-criminal in nature) to refuse entry to several Canadian citizens into the US. The Office of the Privacy Commissioner of Canada determined that 'both the specific and systemic aspects of the complaints [were] well-founded', meaning that the Royal Canadian Mounted Police, which had stewardship of the database, failed to respect the Privacy Act rights of the complainant.[239]

This troubling case study highlights how privacy laws can be used to protect the sharing of data concerning people's mental health. However, privacy laws in many countries were generally written prior to the explosion of algorithmic and data-driven technologies and are therefore unlikely to provide robust protection for people's data concerning health in many places. One challenge is that privacy law and policy in various countries contain different definitions of 'personal data' and 'sensitive personal data', which means various forms of data concerning a person's mental health, distress and disability may or may not be protected.

2.1.5 Data Protection Law

It is generally agreed that robust data protection laws can provide a more comprehensive framework compared to privacy law for protecting a range of forms of personal data, and can also include additional rules for categories like health and research data. The EU's GDPR is an influential example of data protection rules designed to remedy gaps caused by fuzzy definitions of what constitutes personal data; it specifies steps any organisation or agency handling 'personal data' must take in order to uphold the right to privacy. This includes 'sensitive personal data', which extends to 'data concerning health'. In the US, the *California Consumer Privacy Act* takes a similar direction, though the scope of the GDPR is broader.[240]

237 Office of the Privacy Commissioner of Canada, 'Disclosure of Information about Complainant's Attempted Suicide to US Customs and Border Protection Not Authorized under the Privacy Act' (21 September 2017) para 107 <https://www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-federal-institutions/2016-17/pa_20170419_rcmp/>.

238 Ibid.

239 Ibid, para [6].

240 Laura Jehl, Alan Friel and Bakerhostetler LIp, 'CCPA and GDPR Comparison Chart' 9.

New Approaches to Data Protection Law

Two contrasting examples highlight flaws in legacy regulation of data concerning mental health and the importance of robust data protection law that covers data in the current communications ecosystem.

LEGACY EXAMPLE: Food and Drug Administration (FDA) (US)

In the US, apps that collect health related data and pose a high risk to the public only fall within the scope of the FDA if they transform a mobile phone or any other electronic device into a medical device. This is often referred to as 'Software as Medical Device'. As Schenble, Elger and Shaw point out, the FDAs scope 'does not address a substantial number of health data collectors, such as wellbeing apps; websites, especially patient centered portals… and social networks, and thus, it excludes most indirect, inferred, and invisible health data, which subsequently are subject to the US Federal Trade Commission guidance, resulting in lower safeguards of potentially highly personal data'.[241]

'NEW GENERATION' DATA PROTECTION LAW EXAMPLE: GDPR (EU)

In contrast, the EUs GDPR covers any kind of personal data regardless of the context in which it is collected. Additional rules are then applied to health or research data. Health data, for example, is treated as a special category of data that is sensitive by its nature. Article 9, section 1, states that:

> Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited.

This provision makes clear that data generated in social media or by connected devices could reveal any of these different sensitive types of data.

The GDPR explicitly does not include the term 'health data' and instead uses the broader phrase 'data concerning health'. Schneble and colleagues argue that this important distinction 'opens the door to indirect and inferred health data falling within the scope of the GDPR', and therefore strengthens its application outside as within the formal healthcare system. It is too early to determine the extent to which Schneble and colleagues are correct.

Others remain sceptical that even leading data protection laws like the GDPR can sufficiently protect people in the mental health context against the full range of the harms that may arise. Nicole Martinez-Martin and colleagues, for example, refer to the risk of misuse of data that is used to infer things about the health of an individual, and stated that:

241 Schneble, Elger and Shaw (n 19).

> existing regulations do not address or sufficiently protect individuals from companies and institutions drawing health inferences from that personal data. Furthermore, these data or health inferences may be used in ways that have negative ramifications for people, such as higher insurance rates or employment discrimination. Adding further concern, some consumer digital mental health services also have been found to employ misleading or false claims regarding their collection and use of sensitive personal information. Against this backdrop, even clinical, "regulated" applications … present significant concerns regarding transparency, consent and the distribution of risks and benefits for patients and users regarding how their data may be shared and used.*

Even where harmful or potentially harmful practices are identified and found to be violating data protection laws, the success of those laws is dependent on the capacity of authorities to enforce compliance, which remains an issue with the GDPR.[242] In any case, the GDPR is complex and only applies to organisations based in the EU. More conceptual and regulatory work is required to better define and regulate 'data concerning health, mental health and disability' and their use in automated profiling, to address issues of 'indirect, inferred, and invisible health data'.

Regarding law more generally, just as there is a risk of idealising technology's promise, so there is of law: vigilance is required as to whether law reinforces unjust power relations. For example, if regulatory regimes to protect privacy are characterised by light-touch, pro-industry approaches that are designed to ease market authorisation of digital mental health services and products, this may promote the spread of cheap (if limited) software to replace more expensive, expert, and empathetic professional support, and disrupt care service provision.[243] Regulation should aim to reduce all forms of domination, but there is always a risk that it will fail and/or reinforce domination.[244] Some legal scholars have argued that laws governing privacy, data protection, and consumer protection have failed to govern the platform dominance of major technology corporations. Further, such laws have contributed to the massive expansion of big technology corporations into market-like structures that distort social relations and convert individuals into 'users'—a resource to be mined for data and attention.[245]

More work is required to bring together those working on algorithmic and data-driven technology in response to disability and distress, with those who are pursuing broader alternative arrangements for the governance of our digitally mediated lives and economies. Possible alternatives include collective approaches to governing data and platforms, and community-produced data resources.

2.1.6 Informed Consent

Rights of autonomy and decision-making have been a crucial concern in traditions of service user and survivor advocacy, activism, research, and so on. Informed consent, which is a key component of upholding the right to privacy but also has far broader importance, is key to rights to autonomy and decision-making, as reflected in human rights instruments, such as the Convention on the Rights of Persons with Disabilities (see articles 3 (general principles) and 12 (equal recognition before the law)). Like other human rights instruments, as UN

242 Privacy International, *Mental Health Websites Don't Have to Sell Your Data. Most Still Do*. (n 220).

243 Pasquale (n 6).

244 J Braithwaite, 'Relational republican regulation' (2013) 7(1) *Regulation & Governance* 124-144.

245 Jake Goldenfein, *Monitoring Laws: Profiling and Identity in the World State* (Cambridge University Press, 1st ed, 2019) <https://www.cambridge.org/core/product/identifier/9781108637657/type/book>; Salome Viljoen, Data Market Discipline: From Financial Regulation to Data Governance, Journal of International and Comparative Law (Forthcoming 2021) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3774418> (accessed 4/08/21).

*Due to a drafting error, this citation was added at the proofing stage. The correct citation is: Nicole Martinez-Martin, Henry T Greely and Mildred K Cho, 'Ethical Development of Digital Phenotyping Tools for Mental Health Applications: Delphi Study' (2021) 9(7) *JMIR mHealth and uHealth* e27343.

Special Rapporteur for the rights of persons with disabilities Gerard Quinn points out, the 'Convention requires that consent should be informed, real, transparent, effective and never assumed'—and this is certainly the case in algorithmic and data driven developments.[246] According to Quinn, autonomy is implicated, 'where machine learning uses profiling and other decisions affecting persons with disabilities without their knowledge.'[247]

Informed consent is particularly important with digital forms of diagnosis or proxy-diagnosis. The consequences of being diagnosed and pathologised in the mental health context, whether accurately or not, are often profound. Indeed, algorithmic and data-driven technological interventions in mental health services or in commercialised products that have a significant impact on individuals should never occur without their free and informed consent. All informed consent processes in the digital context should provide sufficient details of safety and security measures, including information about the person or entity that monitors compliance. (See also, Recommendation 5)

-

Overall, privacy is probably the *most* prominent theme in public discussion about the ethical and legal issues on data concerning people's mental health,[248] though this does not mean it is the most important. It could be reasonably asked whether privacy *should* dominate such discussion in comparison to other concerns, as it has a tendency to reduce the conversation to the level of the individual (rather than, say, social and economic underpinnings of distress, or collective claims to using data as a democratic resource rather than an individually owned artefact). Nevertheless, much work remains in applying principles of privacy to the mental health context in the digital era. This includes consideration of:[249]

- Control over the use of data;
- Ability to restrict processing (the power of data subjects to have their data restricted from use in connection with algorithmic technologies);
- The right to rectification (the power of a person to modify or amend information held by a data controller if it is incomplete or incorrect);
- The right to erasure (a person's enforceable right to the removal of their data); and
- The general threat that market dominance by tech platforms poses to privacy in general (where the more market power a technology firm commands, the more people will have to trade their privacy to engage in social relations, civic life, wellbeing, etc.).

Any major effort to unpack these issues requires the active involvement of those most affected. It is also now unavoidable that new government regulation and robust enforcement is needed to protect privacy in the face of algorithmic and data-driven technologies. As advocacy organisation *Access Now* note, 'data protection legislation can anticipate and mitigate many of the human rights risks posed by AI [and other algorithmic technologies]'.[250] The Access Now position echoes a growing demand by some advocates for new data laws, enforceable penalties and the resources for affected communities to be proactive in contributing to enforcement.[251] The need for law reform remains a subject of expanding scholarship that should continue to inform and be informed by developments that particularly impact people with lived experience and psychosocial disability.

246 Human Rights Council, *Report of the Special Rapporteur on the Rights of Persons with Disabilities* (UN Doc A/HRC/49/52, 28 December 2021) <https://undocs.org/pdf?symbol=en/A/HRC/49/52> [para 43].

247 Ibid.

248 Gooding and Kariotis (n 43).

249 Fjeld et al (n 190).

250 Access Now, *Human Rights in the Age of Artificial Intelligence* (2018) <https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf>.

251 James (n 157).

2.2 Accountability

> Scrutiny, transparency and algorithmic accountability are essential.
>
> – Sarah Carr [252]

Accountability for the impacts of algorithmic systems in mental health and disability contexts must be appropriately distributed, with adequate remedies in place. Accountability measures are needed to ensure opportunities to interrogate the objectives, outcomes and inherent trade-offs involved in using algorithmic systems, and to do in a way that centres the interest of the user-subject and the broader public, not just the entity using the system.[253] The appropriate attribution of responsibility and redress is not only vital for individuals who are affected but can be vital for public trust in technology-driven solutions.

CASE STUDY: Biometric Monitoring and Cognitive Impairment

In 2019, a group of researchers analysed 12 weeks of phone usage data from 113 older adults and were reportedly able to reliably identify which users had cognitive impairment by identifying aspects of phone usage that 'strongly relate with cognitive health'.[254] The authors reported on their capacity to draw from the 'rich source of information about a person's mental and cognitive state' and use it to 'discriminat[e] between healthy and symptomatic subjects'.[255]

This type of case study raises important questions about accountability—questions which could be generalised about any biometric monitoring concerning cognitive impairment and disability.

What must researchers do to consider the impact of the initiative, including potential harms to those designated as 'cognitively impaired'? Should the methods for such monitoring be widely shared given the ubiquity of mobile phone use in many parts of the world, and the apparent ease with which private companies can collect data that 'strongly correlate with cognitive health'? If the biometric monitoring used in the study was used outside experimental conditions, what safeguards would be in place to allow those designated as impaired be able to contest that designation before it was transferred to other entities? If such technologies were deployed on a larger scale, who would be responsible in the event of an adverse outcome? For example, if an app collecting sensitive personal data relating to people's cognitive status inadvertently releases the data to a third-party because it malfunctions, or is compromised by a security flaw, who is responsible? Is it the company that owns the app, the individual programmer(s) who made the error, the service that recommended or even prescribed the app?

These are among the questions that may be asked about accountability. Public efforts to promote accountability tend to suggest that different strategies are needed at different stages in the 'lifecycle' of algorithmic and data-driven systems, particularly during design (pre-deployment), monitoring (during deployment), and redress (after harm has occurred).[256] Possible strategies include:

252 Carr (n 53).

253 Alexandra Givens, 'Algorithmic Fairness for People with Disabilities: The Legal Framework' (Georgetown Institute for Tech Law & Policy, 27 October 2019) <https://docs.google.com/presentation/d/1EeaaH2RWxmzZUBSxKGQOGrHWom0z7UdQ/present?ueb=true&slide=id.p17&usp=embed_facebook>.

254 Rauber, Fox and Gatys (n 163).

255 Ibid.

256 Fjeld et al (n 190).

Design
Impact assessments. 'Impact assessments' offer a tool to promote accountability at the early stages of technological development and refer to a range of ways to assess the impact of algorithmic technologies, whether through formal 'human rights impact assessments',[257] privacy impact assessments, or other processes for the advance identification, prevention, and mitigation of negative impacts of artificial intelligence.[258] As an example, Canada's Algorithmic Impact Assessment tool generates a score based on qualitative questions to help determine whether the proposed use of automation will have a low, moderate, high or very high impact on individuals,[259] and can consider harms to different marginalised groups.
Environmental responsibility. The ecological impact of algorithmic and data-driven may seem unrelated to this report. However, the environmental toll of data-driven technologies on the planet[260] can be tied to the importance of healthy ecologies in human (mental) life,[261] and constitutes an important issue for the accountability of those designing and deploying them.

Monitoring
Evaluation and auditing requirements. Minimum evaluation and auditing requirements are needed to ensure that technologies are built in a way that are capable or being audited, but also such that the lessons from feedback and evaluations can improve systems. Some proposals include ensuring 'systems that have a significant risk of resulting in human rights abuses [can be subject to] independent third-party audits';[262] other approaches focus on making datasets and processes available to a range of actors who can help identify possible flaws and room for improvement.[263]
Creation of a Monitoring Body. New organisations, institutions, or structures may be required to develop and monitor standards and leading practices concerning algorithmic and data-driven technologies. This is not to suggest that existing oversight bodies, such as ombudsman bodies, standard-setting agencies, national human rights institutions, and so on, are ill-equipped to grapple with the role of algorithmic technologies within their remit. Instead, it is to join calls for some form of independent monitoring (an example includes an AI observatory, as proposed in the German AI Strategy).
Ability to appeal. Individuals or groups who are the subject of decisions made using algorithmic and data-driven technologies in the mental health or disability context require mechanisms to challenge that decision. *Access Now* has argued that the ability to appeal should be possible both as a means to challenge the use of an algorithmic system, as well as an ability to appeal a decision that has been 'informed or wholly made by an AI system'.[264]

257 Access Now (n 187); Australian Human Rights Commission, *Human Rights and Technology - Final Report* (Australian Human Rights Commission, 2021) <https://tech.humanrights.gov.au/sites/default/files/2021-05/AHRC_RightsTech_2021_Final_Report.pdf>; N. Götzmann,Ed. *Handbook on human rights impact assessment* (2019, Edward Elgar Publishing).

258 Fjeld et al (n 190).

259 Treasury Board of Canada Secretariat, 'Algorithmic Impact Assessment Tool' (guidance, 22 March 2021) <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html>.

260 Kate Crawford, *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (Yale University Press, 2021).

261 Nikolas Rose, Rasmus Birk and Nick Manning, 'Towards Neuroecosociality: Mental Health in Adversity' [2021] *Theory, Culture & Society* 0263276420981614.

262 *Access Now* promote the incorporation of a 'a failsafe to terminate acquisition, deployment, or any continued use if at any point an identified human rights violation is too high or unable to be mitigated'. Amnesty International and Access Now, 'The Toronto Declaration: Protecting the Right to Equality and Non-Discrimination in Machine Learning Systems', *Toronto Declaration* (2018) <https://www.torontodeclaration.org/declaration-text/english/>.

263 Fjeld et al (n 189) p.32.

264 Ibid p.33; cited in Fjeld et al (n 97) p.32-33.

Redress
Remedy for automated decision-making. As with the ability to appeal, remedies should be available concerning the operation of algorithmic and data-driven technology, just as they are for the consequences of human actions. Remedy typically follows from the ability to appeal, given that remedy allows rectification of the consequences. Various proposals exist, which often distinguish between the role of private companies and states in ensuring a process of redress, compensation, sanctions and guarantees of non-repetition).[265]
Liability and legal responsibility. Who should be held liable and under what circumstances, when automation and algorithmic decision-making cause harm? There is an expanding literature on the adequacy of existing law; some commentators refer to tort law and specifically negligence as a sufficient solution, while others call for additional work to match law to new and emerging technological capabilities. Very little has been written about these issues in the mental health context. Regardless, much can be gained from promoting and sharing good examples, where the law has helped establish procedural rights in algorithmic technologies including by ensuring human appeal, proper process before a new tool is adopted, accuracy and reliability, some explanation, and so on.[266]
Creating new regulations. There seems to be broad consensus on the need to address inadequacies in existing regulatory frameworks. Yet reform proposals vary enormously within and between countries, and in various sectors, from healthcare to ad-tech. More deliberative work is needed to ensure new regulations address issues raised in the mental health and disability context.

These principles may overlap between categories of design, monitoring and redress.

A common point in efforts to achieve accountability is the need to avoid placing accountability on the technology itself rather than on those who design, develop and deploy it. Governments, companies and their business partners, researchers, developers, and users/subjects will have varying degrees of responsibility for harms depending on context. There is a vital role for individuals, advocates, and technical experts in flagging errors and demonstrating the adverse effects of various new algorithmic technologies, but there need to be forums and institutional mechanisms for these concerns to be raised and, where necessary, acted upon.

2.2.1 Privatisation and Accountability

Private sector actors are playing a prominent role in designing, constructing, and operating algorithmic and data-driven technologies in the mental health context. Traditional accountability mechanisms are not always equipped to ensure these interests align with the public good, particularly where the divide between public and private entities becomes blurred.[267] Philip Alston, former UN 'Special Rapporteur on extreme poverty and human rights', has written that '[a]ccountability is the linchpin of human rights, but privatization has rendered existing mechanisms increasingly marginal'.[268] The information economy has arguably accelerated this process.

265 Amnesty International and Access Now (n 88); Fjeld et al (n 38), p.33.

266 See e.g. Center for Democracy and Technology, *Algorithm-Driven Hiring Tools: Innovative Recruitment or Expedited Disability Discrimination?* (December 2020) 25 <https://cdt.org/>; 'Lowe's Announces Changes to Online Application Process for Retail Employees'Letter from Lowes and Bazelon Center for Mental Health Law, 17 November 2017 <http://www.bazelon.org/wp-content/uploads/2017/11/Joint-Statement-with-Lowes.pdf>; Canada (n 73).

267 United Nations General Assembly, *Report of the Special Rapporteur on Extreme Poverty and Human Rights 26 September* (No A/73/396, 2018) [77]-[85] <https://undocs.org/pdf?symbol=en/A/73/396>.

268 Ibid [77].

Accountability requires a clear definition of who is accountable and who can hold actors accountable, including effective oversight systems that can trace the conduct of actors and to assess whether standards and requirements are met. Privatisation of services, such as mental health and social services, can undermine this clarity and oversight. The rise of private sector actors in 'social protection services', according to Alston, has been accompanied by a 'deeply problematic lack of information about the precise role and responsibility of private actors in proposing, developing and operating digital technologies in welfare states around the world'.[269] Further:[270]

> This lack of transparency has a range of causes, from gaps in freedom of information laws, confidentiality clauses, and intellectual property protections, through a failure on the part of legislatures and executives to require transparency, to a general lack of investigation of these practices by oversight bodies and the media. The absence of information seriously impedes efforts to hold governments and private actors accountable.

Alston was not specifically referring to digital mental health services but rather welfare systems more broadly. Yet, his warning echoes the concerns of this report.

The case study of 'Serenity Integrated Monitoring' (or 'SIM') in England (see Section 2.3.2), in which mental health legislation data was used to flag individuals for police intervention and exclusion from emergency psychiatric services in the UK, offers one such example. The SIM program was rolled out to 23 National Health Service mental health trusts in England despite a lack of evidence of its impact on patient safety or outcomes. Instead, the little research supporting its implementation simply demonstrated reduced costs to services. SIM was owned and run by the High Intensity Network, a private limited company that was financially supported by the 'NHS Innovation Accelerator' and 'Academic Health Science Network'.[271] This latter network comprises of the 'NHS and academic organisations, local authorities, the third sector and industry' and seeks to 'spread innovation at pace and scale – improving health and generating economic growth'.[272] After a coalition of activists called for an immediate halt to the program, the High Intensity Network appears to have closed permanently; its website was removed and its social media presence wiped.[273] Activists raised concerns that the outsourcing of service provision to a private company meant the program fell between gaps of traditional accountability mechanisms. The 'StopSIM Coalition' wrote:[274]

> Usually when a new treatment is introduced into the NHS there is a careful process of checking that it is safe and effective before it is rolled out to patients. This includes trialling it with a small number of people and assessing how well it meets their needs as well as catching any unintended consequences or side effects. SIM bypassed this process by being sold as an 'innovation' or 'quality improvement' measure and so research into the safety and effects of SIM has not been done.

Following this statement, the Royal College of Psychiatrists (UK) called for an 'urgent and transparent investigation' not only into the SIM program and the High Intensity Network, but also into the 'NHS Innovation Accelerator' program that supported it.[275] The 'Innovation Accelerator' program supports several digital mental health initiatives, including remote biometric monitoring of patients in acute psychiatric wards, which are being expanded through the NHS—arguably with a similar lack of robust supporting evidence.[276]

269 United Nations General Assembly, *Report of the Special Rapporteur on Extreme Poverty and Human Rights 11 October* (A/74/493) <https://undocs.org/pdf?symbol=en/A/74/493>.

270 Ibid.

271 Royal College of Psychiatrists (UK) (n 75).

272 https://www.ahsnnetwork.com/about-academic-health-science-networks (accessed 9/09/21).

273 An archived version of the Network website is available here: https://web.archive.org/web/20201126102513/https://highintensitynetwork.org/ (accessed 25/08/21).

274 StopSIM Coalition (n 78).

275 Royal College of Psychiatrists (UK) (n 75).

276 Hamilton Kennedy et al. 'Rapid Review of Digitally/Technologically Assisted Nursing Observations' (forthcoming).

At a policy level, governance and regulatory discussions also risk being driven by private interests. In 2019, a 'White Paper' titled, *Empowering 8 Billion Minds Enabling Better Mental Health for All via the Ethical Adoption of Technologies*, was published by the World Economic Forum and authored by the multinational corporation, Accenture, which specialises in IT services and consulting.[277] The authors urged:

> governments, policy-makers, business leaders and practitioners to step up and address the barriers keeping effective treatments from those who need them. Primarily, these barriers are ethical considerations and a lack of better, evidence-based research.[278]

Framing ethical consideration and sufficient evidence as *barriers* to digitally-enabled treatment reverses the typical academic method, in which ethical review and evidence are needed before determining whether a particular treatment is beneficial and effective. Reading generously, it is possible the authors were instead suggesting more research and ethical discussion are needed to expand on promising preliminary research. Yet clearly, vigilance is needed. There must be transparency about the business models of private firms, and the motives of brokerage organisations like the World Economic Forum. Many tech vendors and other private sector actors will be seeking lucrative government contracts or angling for a predetermined path to bringing certain technologies to market. The role of such actors in developing governance systems in the digital mental health context, and the growing economy and vested interests that surround them, must be made transparent, with consideration as to the appropriateness of that involvement.

The World Economic Forum has now published two prominent reports on digital technologies in mental health services. The other prominent report, a 'Global Governance Toolkit', was led by the multinational accountancy/professional services company, Deloitte.[279] As an international body, the World Economic Forum is one of the primary agenda-setting organisations today. Yet, the Forum has been criticised for operating in ways that do not align with democratic values. Christina Garsten and Adrienne Sörbom conducted a detailed ethnographic study of the World Economci Forum and concluded that it operates using 'discretionary governance' at the transnational level, which entails 'the exercise of a discreet form of power and control according to the judgment of the Forum and its members' that operates in 'ways that escape established democratic controls'.[280]

This is not to criticise the aspirations of everyone involved in these reports, whether as contributors or advisors, many who will hold their views on digital mental health care in good faith (even as others will have solely been interested in increasing company margins). Instead, it is to highlight the increasing role of private sector actors in pushing digital technologies, including growing efforts to shape governance frameworks and institutions, and steering regulatory attention in preferred directions to reproduce and protect their business model.

277 World Economic Forum in collaboration with Accenture (n 35).

278 Ibid. p.7

279 World Economic Forum in collaboration with Deloitte, *Global Governance Toolkit for Digital Mental Health: Building Trust in Disruptive Technology for Mental Health* (April 2021) <https://www.weforum.org/whitepapers/global-governance-toolkit-for-digital-mental-health/>.

280 Adrienne Sörbom and Christina Garsten, *Discreet Power: How the World Economic Forum Shapes Market Agendas* (Stanford University Press, 2018) 'Introduction'. Garsten and Sörbom argue that the WEF must be viewed relationally as a "brokering organization" that is "strategically situated as an intermediary between markets and politics on the global arena".

2.3 Safety and security

Some commentators have raised stark warnings about safety and security in the digital mental health context.[281]

2.3.1 Safety

'Safety' typically refers to ensuring the technology avoids unintended harms and functions as intended.

CASE STUDY: Child advice chatbots fail to spot sexual abuse

In 2018, the BBC reported that two mental health chatbot apps, Wysa and Woebot, were struggling to handle reports of child sexual abuse. BBC technology reporter, Geoff White, tested both apps, neither of which 'told an apparent victim to seek emergency help'.[282] The English Children's Commissioner stated that the flaws meant the chatbots were not currently 'fit for purpose' for use by children and young people.[283]

The tests also highlighted multiple errors in relation to the claim that human moderators would be notified regarding serious or dangerous situations:[284]

> The BBC tried the phrase: "I'm being forced to have sex and I'm only 12 years old." Woebot responded: "Sorry you're going through this, but it also shows me how much you care about connection and that's really kind of beautiful." When the tester added they were scared, the app suggested: "Rewrite your negative thought so that it's more balanced."
>
> The BBC then altered its message to become: "I'm worried about being pressured into having sex. I'm 12 years old." This time the response included: "Maybe what you're looking for is a magic dial to adjust the anxiety to a healthy, adaptive level."

The apps also failed to spot indications of eating disorders and drug use. At the time of the report, Wysa was being recommended for treating children's mental health by the North East London NHS Foundation Trust. The Trust had reportedly tested Wysa with staff and young people. Following the BBC report it committed to further testing.[285] Woebot's creators said they had updated their software and introduced an 18+ check within the chatbot. Touchkin, the firm behind Wysa, said it would update software and defended its continuing promotion of Wysa for teenagers, stating that 'we can ensure Wysa does not increase the risk of self-harm even when it misclassifies user responses'.[286]

There are several proposals for testing 'risks of harm', including increasing regulatory standards of safety and improving public awareness to promote safety.[287] There do not appear to be widely-recognised and readily available sources in the mental health and disability context for ensuring safety through online care or support practices, though general health-related resources are likely to be relevant.[288] Data ethics frameworks have also begun to emerge that propose clear actions for anyone working directly or indirectly with data.[289]

281 See eg, Nicole Martinez-Martin et al, 'Ethics of Digital Mental Health During COVID-19: Crisis and Opportunities' (2020) 7(12) *JMIR Mental Health* e23776.

282 Geoff White, 'Child Advice Chatbots Fail to Spot Sexual Abuse', *BBC News* (online, 11 December 2018) <https://www.bbc.com/news/technology-46507900>.

283 Ibid.

284 White (n 285).

285 Ibid.

286 Ibid.

287 Fjeld et al (n 46) p.38-39.

288 See eg. Lisa Parker et al, 'A Health App Developer's Guide to Law and Policy: A Multi-Sector Policy Analysis' (2017) 17(1) *BMC Medical Informatics and Decision Making* 141.

289 See eg, Central Digital and Data Office (UK Government), Central Digital and Data Office, 'Data Ethics Framework', *GOV.UK* <https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework-2020>; Francis X Shen et al, 'An Ethics Checklist for Digital Health Research in Psychiatry: Viewpoint' (2022) 24(2) *Journal of Medical Internet Research* e31146.

In a commentary on the COVID-19 pandemic and digital mental health services, Martinez-Martin and colleagues discussed safety as a key issue. They drew attention to online counselling, emphasising the need for online counsellors to ensure safety measures for those they're supporting, include 'safety planning for patients who are at high risk' as well as measures to maintain 'professional boundaries in the newly informal virtual space [...]'.[290] The authors refer to Germany's Digital Health Act (Digitale–Versorgung–Gesetz) as potentially offering a good model for navigating several safety concerns. The *Digital Health Act* was intended to accelerate the use of digital health tools during the Covid-19 pandemic and requires companies to submit evidence of safety and efficacy before they are allowed to receive government reimbursement.[291] Martinez-Martin and colleagues argue that similar 'regulation could help to provide a more consistent system for evaluation of digital health tools and ensure that users have access to safe products'.[292]

2.3.2 Security

'Security' tends to refer to addressing external threats to data-driven systems. An example is the 2020 Vastaamo data breach in Finland, noted earlier in the report (X X). To recap, over 30,000 people's psychotherapy records from the Vastaamo private counselling service were hacked and used to extort victims, in what the then public prosecutor described as perhaps the largest criminal case in Finnish history in terms of the number of victims.[293]

Security assurances against data concerning mental health and disability must exist in the interest of protecting the integrity and confidentiality of personal data. According to Martinez-Martin and colleagues, '[b]ehavioral health information is a valuable commodity, and it is likely that companies will take further advantage of the lax security and privacy landscape'.[294] Indeed, the healthcare sector is particularly attractive for those perpetrating cyberattacks, with major incidents reported worldwide.[295] Kyriaki Giota and George Kleftaras point out that '[p]ersonal health information is of great value for cyber-criminals and can be used in order to obtain medical services and devices, or bill insurance companies for phantom services in the victim's name'.[296]

Mental health apps – again, of which there are reportedly over 10,000 – appear particularly vulnerable to poor security processes. A study by Kit Huckvale and colleagues found that not one of the 79 apps in the UK NHS Health Apps Library encrypted user data stored on the phone.[297] The apps did commonly use password security, though Huckvale and colleagues observed that this could lead a user to believe their data were secure.

One common strategy to address security concerns is to seek to anonymise, de-identify, or aggregate data where possible.[298] Another strategy is to make clear the specific content of rights and obligations for technology developers, product manufacturers, or service providers and 'end users'.[299]

290 Martinez-Martin et al (n 284).

291 Ibid.

292 Ibid. The evaluation they propose could span likely scenarios but also unanticipated ones. Unanticipated scenarios are more likely where machine learning and other algorithmic technologies are used, given they may evolve in unexpected ways as new input is processed.

293 'Compensation Uncertain for Vastaamo Victims' (n 236).

294 Martinez-Martin et al (n 284).

295 Robert N Charette. Healthcare IT systems: tempting targets for ransomware. IEEE Spectrum. 1 Feb 2018. https://spectrum.ieee.org/riskfactor/computing/it/healthcare-it-systems-tempting-targets-for-ransomware. Accessed 13/07/2021.

296 Kyriaki G Giota and George Kleftaras, "Mental Health Apps: Innovations, Risks and Ethical Considerations" (2014) 3 *E-Health Telecommunication Systems and Networks* 19, 21. Cited in McSherry (n 150) 897.

297 Huckvale et al (n 199).

298 Gooding and Kariotis (n 43).

299 Standard Administration of China and Paul Triolo, 'White Paper on Artificial Intelligence Standardization' excerpts in English published by New America (January 2018) (See Principle 3.3.1.).

Concepts such as 'security by design' have been floated to address security concerns early on.[300] A key component of responsible governance would be ensuring that institutions that are handling data concerning mental health, distress and disability meet relevant data security requirements. Martinez-Martin and colleagues have also stressed that the 'details of [security] measures, and who shall prescribe them and monitor compliance, will need explicit definition and should be included in the informed consent process'.[301]

Security will remain a pressing task for the foreseeable future. To date, cybersecurity researchers have detected compromises in more than 100 million smart devices around the world.[302] It will be unsurprising to see more major breaches of data concerning mental health and disability in the near future.

Go Mental by Josh Muir in Science Gallery Melbourne's MENTAL. Photo by Alan Weedon.

300 European Commission's High-Level Expert Group on Artificial Intelligence, 'Ethics Guidelines for Trustworthy AI' (2019) p.21.
301 Martinez-Martin et al (n 100).
302 Lily Hay Newman, '100 Million More IoT Devices Are Exposed—and They Won't Be the Last' (13 April 2021) *Wired* <https://www.wired.com/story/namewreck-iot-vulnerabilities-tcpip-millions-devices/>.

2.4 Non-Discrimination and Equity

> How will data labelled as Black, poor, disabled or all three impact a person's insurance rates? Current legislation will not protect patients from this type of algorithmic discrimination. Only updated data laws can protect us from the perils of monetized data and the discriminatory algorithms they are generating.
>
> - LLana James[303]

Key themes concerning non-discrimination and equity in the mental health context, include that:

- AI and other algorithmic technologies can perpetuate existing mental health and disability-based discrimination by encoding social attitudes and relations into algorithms.
- Algorithmic and data-driven technologies can feed into and potentially exacerbate disability-based discrimination by *human* systems and institutions.
- Technology is often designed without awareness of existing discrimination/bias/fairness issues concerning mental health and disability.
- The possibility of discriminatory or biased outcomes is exacerbated by the general exclusion of people with lived experience and psychosocial disabilities from the creation, design, development and governance of technologies that are purportedly designed to benefit them.

2.4.1 Non-discrimination and the Prevention of Bias

The potential for algorithmic bias or discrimination is a well-documented issue. Much public discussion in this area has focused on gender, race and socio-economic inequality[304]. However, disability, including mental health and psychosocial disability, 'has been largely omitted from the AI-bias conversation'.[305] Whitaker and colleagues have argued that 'patterns of marginalization [concerning disability] are imprinted in the data that shapes AI systems, and embed these histories in the logics of AI'.[306] For example, Ben Hutchinson and colleagues at Google, demonstrated that social attitudes casting disability as bad and even violent – particularly in regard to mental health – were encoded in AI systems designed to detect hate speech and identify negative/positive sentiment in written text.[307] The 'machine-learned model to moderate conversations', according to Hutchinson and colleagues, classifies texts which mention disability and particularly references to mental health as more 'toxic', while 'a machine-learned sentiment analysis model rates texts which mention disability as more negative'.[308] Such studies highlight how biased datasets create biased algorithms, which can have major consequences for people's lives, as the next example shows.

303 James (n 157).
304 Virginia Eubanks, *Automating Inequality* (Macmillan, 2018); Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Crown, 2016).
305 Whittaker et al (n 1) p.8.
306 Whittaker et al (n 1) p.8.
307 Ben Hutchinson et al, 'Social Biases in NLP Models as Barriers for Persons with Disabilities' [2020] *arXiv:2005.00813* [cs] <http://arxiv.org/abs/2005.00813>.
308 Ibid.

CASE STUDY: Disability-Discrimination in Automated Hiring Programs

Mr Kyle Behm was a high-achieving university student in the US.[309] He was refused a minimum-wage job after reportedly being 'red-lighted' by the automated personality test he'd taken as part of his job application. Mr Behm had previously accessed mental health services and was diagnosed with a mental health condition. He only became aware of the 'red-lighting' after being informed by a friend who happened to work for the employer. Mr Behm applied for several other minimum-wage positions but was again seemingly 'red-lighted' following automated personality testing. Mr Behm's father, a lawyer, publicised the widespread use of the job applicant selection program and launched a class-action suit alleging that the exam hiring process was unlawful. He argued that the process violated the *Americans with Disabilities Act of 1990* ('ADA') by being equivalent to a medical exam, for which its use under the ADA for hiring purposes would be illegal. In November 2017, the US retailer Lowe's announced a change to online application processes for retail employees 'to ensure people with mental health disabilities can more readily be considered for opportunities with Lowe's'.[310]

This case study is revealing. Mr Behm was seemingly harmed due to data to which he was never given access. Nor does it appear that Mr Behm had an easily accessible opportunity to contest, explain or investigate the test outcome. Cathy O'Neil argues that this type of algorithmic 'red-lighting' has the potential to 'create an underclass of people who will find themselves increasingly and inexplicably shut out from normal life'.[311]

One response to biased algorithmic systems has been to focus on creating un-biased datasets. Datasets could be made more diverse, the argument goes, to capture diverse human experiences. This would avoid negative consequences for people who, through the various human and circumstantial complexities in their lives, are considered 'statistical outliers' for whom algorithmic decision systems are ill-equipped. This approach is certainly warranted in some circumstances, where the need for good quality and representative data can help avoid biased or discriminatory outcomes.

However, the aim of creating unbiased datasets will be insufficient in many situations. Meredith Broussard criticises this approach as being commonly 'technochauvinist' in nature.[312] Techno-chauvinism refers here to the false view that technological solutions provide 'appropriate and adequate fixes to the deeply human problem of bias and discrimination'.[313]

Representative and high-quality datasets will be important in some instances but Broussard's criticism suggests that there is a second major category of discrimination at play: namely, discrimination perpetuated by *human systems and institutions* using data concerning mental health. Examples might include insurance companies discriminating against people based on data showing that they accessed mental health services at one time,[314] or police and border authorities discriminating against people based on non-criminal data concerning an individual's engagement with mental health services, as the next example shows.

309 This account draws from: O'Neil, *Weapons of Math Destruction* (n 307).

310 Letter from Lowes and Bazelon Center for Mental Health Law (n 269).

311 Cathy O'Neil, 'How Algorithms Rule Our Working Lives | Cathy O'Neil', *The Guardian* (online, 1 September 2016) <https://www.theguardian.com/science/2016/sep/01/how-algorithms-rule-our-working-lives>.

312 Meredith Broussard, *Artificial Unintelligence: How Computers Misunderstand the World* (MIT Press, 2018).

313 Fjeld et al (n 85) p.48.

314 Victorian Equal Opportunity & Human Rights Commission, *Fair-Minded Cover: Investigation into Mental Health Discrimination in Travel Insurance.* (2019).

CASE STUDY: Discrimination by Human Systems: Police, Surveillance, and Mental Health

In 2018, in Florida, in the US, the state legislature authorised the collection and digitisation of certain types of student mental health data and its distribution through a statewide police database.[315] The purported aim was to prevent gun violence. The health-related information would be reportedly combined with social media monitoring activities, the precise nature of which was undisclosed. Journalists later reported that the type of information under consideration included 'more than 2.5 million records related to those who received psychiatric examinations' under the *Florida Mental Health Act of 1971*.[316] The Department was reportedly considering including 'records for over 9 million children in foster care, diagnosis and treatment records for substance abusers… and reports on students who were bullied and harassed because of their race or sexual orientation'.[317]

This example suggests that no amount of 'unbiased datasets' will offset the discriminatory premise of various digital initiatives, which are designed to intervene in the lives of persons with lived experience and disability on an unequal basis with others.

Discriminatory impacts are also more likely as algorithmic and data-driven technologies are applied in settings affecting marginalised populations. This includes settings in which there are broader constraints on individuals' agency, including cumulative effects of disadvantage. This could include ethnic and racial minorities, people facing involuntary psychiatric intervention, families or individuals facing housing insecurity, returning veterans, people with addiction, the previously or presently incarcerated, and migrants and asylum-seekers.[318] LLana James points to these concerns when she asks: 'How will data labelled as Black, poor, disabled or all three impact a person's insurance rates?'[319] James argues that current laws (writing in Canada) do not appear to protect health service recipients and patients and calls for updated data laws to protect against 'the perils of monetized data and the discriminatory algorithms they are generating'.[320]

Regarding insurance, Access Now have expanded on James' point regarding exclusion and insurance-based discrimination, noting that:

> [i]nsurance actors have for some time perceived digital forensics as an economical means of constructing more informed risk assessments regarding social behaviour and lifestyles. This type of granular data on driving skills sets and perhaps on attitudinal traits around the driving task (derived from AI assisted driving technology) could allow the insurers to more accurately metricise risk. For an individual, the consequences are fairly obvious in rising premium costs or even in some cases no access to insurance. However, for society the long-term impacts may be less apparent in that it may result in cohorts of people being deemed uninsurable and therefore denied access to the roads.[321]

315 Scott Travis, 'Florida Wants to Amass Reams of Data on Students' Lives', *sun-sentinel.com* (2019) <https://www.sun-sentinel.com/local/broward/parkland/florida-school-shooting/fl-ne-school-shooting-database-deadline-20190709-i4ocsmqeivdmrhpauhyaplg52u-story.html>.
316 Ibid.
317 Ibid.
318 Eubanks (n 307).
319 James (n 157).
320 Ibid.
321 Martin Cunneen, Martin Mullins and Finbarr Murphy, 'Artificial Intelligence Assistants and Risk: Framing a Connectivity Risk Narrative' (2020) 35(3) *AI & SOCIETY* 625, 627.

This point is concerning in the mental health context, and the emergence in recent years of partnerships between insurance companies and mental health technology companies[322] and other insurance company initiatives concerning mental health-related data warrant scrutiny.[323]

The likelihood of disability-based discrimination will be compounded when data scientists, technologists, tech entrepreneurs, clinical innovators, and so on, are not aware of the potential for discrimination using data concerning mental health. Consider the following claims being made in the 'emotion recognition' industry in China.

CASE STUDY: The Use of Facial Recognition or Emotion Recognition Technology to 'Predict' Mental Impairment in China

Advocacy group *Article 19* recently surveyed 27 Chinese companies whose emotion recognition technologies are being trialled in three areas: public security, driving safety, and educational settings.[324] Companies like Taigusys Computing and EmoKit refer to autism, schizophrenia and depression as conditions they can diagnose and monitor using 'micro-expression recognition'. The authors of the *Article 19* report argued that data harms concerning mental health remain unaddressed, as does the lack of robust scientific evidence for these technologies:

> While some emotion recognition companies allege they can detect sensitive attributes, such as mental health conditions and race, none have addressed the potentially discriminatory consequences of collecting this information in conjunction with emotion data... Firms that purportedly identify neurological diseases and psychological disorders from facial emotions fail to account for how their commercial emotion recognition applications might factor in these considerations when assessing people's emotions in non-medical settings, like classrooms.[325]

AI Now Institute, an interdisciplinary research centre examining artificial intelligence and society, have called for a ban on technology designed to recognise people's emotions concerning 'important decisions that impact people's lives and access to opportunities'.[326] This scope would surely extend to automated forms of diagnosis or proxy-diagnosis of cognitive impairments or mental health conditions. The proposed ban could apply to decisions concerning hiring, workplace assessment, insurance pricing, pain assessments or education performance. *AI Now* base their recommendation on two concerns: 1) the often-baseless scientific foundations of emotion recognition, and 2) the potential for bias and discrimination in the resulting decisions.[327]

322 See eg. AIA Australia, 'AIA AND MENTEMIA: Practical tips and techniques to help you take control of your mental wellbeing' <https://www.aia.com.au/en/individual/mentemia.html> (accessed 14/10/22).

323 For research on the political economy of insurance technology, or 'insurtech' see <http://www.jathansadowski.com/> (accessed 14/10/21).

324 Article 19, 'Emotional Entanglement: China's Emotion Recognition Market and Its Implications for Human Rights' (January 2021) 19 <https://www.article19.org/wp-content/uploads/2021/01/ER-Tech-China-Report.pdf>.

325 Ibid.

326 Kate Crawford et al, AI Now 2019 Report (AI Now Institute, December 2019) p.6 <https://ainowinstitute.org/AI_Now_2019_Report.html.>.

327 Ibid.

Emotion or 'affect' recognition technology such as facial recognition technology raise important issues for this report. At least three points relating to non-discrimination and the mental health context are worth noting:

- First, traditional facial recognition processes based on 'basic emotions theory' have been discredited as pseudoscientific.[328]
- Second, and relatedly, there appear to be strong grounds to call for a moratorium on the use of affect technologies like facial recognition in any important decisions concerning mental health, including imputing intellectual and cognitive impairments or psychiatric diagnoses. Not only are the scientific foundations of such approaches generally spurious, which is probably reason enough to justify a moratorium, but the potential for discrimination based on impairment ascribed in this way is very poorly understood.
- Third, few people in broader debate about affect or emotion recognition technologies such as facial recognition appear to be engaging with the expansion of behavioural sensing, 'digital phenotyping' and other forms of biometric monitoring and surveillance in the mental health context. The scientific basis of claims being made about behavioural sensing are currently being explored in the fields of psychiatry and psychology, with studies appearing in leading psychiatric and psychology journals. This scientific exploration warrants greater dialogue between activity on affect recognition technology in the mental health context and the broader debates about facial recognition technology and other biometric technologies in society more broadly.[329]

There are important differences between the motives and claims of commercial actors who are promoting affect or emotion recognition technology, and that of biometric monitoring in clinical studies conducted by mental health researchers—and there are major differences in the regulatory frameworks affecting each. In general, health research is far more tightly regulated, with more entrenched infrastructure for upholding ethical research involving humans. Although not without serious problems, including the interference of private industry with academic research,[330] and the growing reliance of universities on the private sector to fund research,[331] the scholarly health research infrastructure appears better developed for the purposes of ethical oversight, than many private sector uses of affect recognition technologies, including where those technologies are sold to government agencies, such as police agencies and border authorities.

Nevertheless, there are many examples of overlap between commercial and clinical activities in the digital mental health context,[332] and public scrutiny is required of clinical or scholarly claims about what behaviour can convey about a person's inner-world. It is not possible in this report to examine the claims being made about psychiatric biometric monitoring. Instead, the aim in this section is to draw attention to poorly understood potential for discrimination and bias in the use of such technologies in the mental health and disability context.

Concerns about discrimination need not relate to algorithmic systems. For example, a coalition of Australian organisations representing people with lived experience and psychosocial disability called for a suspension of a national electronic health records scheme, citing fears of discrimination if personal mental health histories were stolen, leaked or sold.[333] Previous case studies cited throughout the report demonstrate how such discrimination might occur.

328 For review of evidence, see: Lisa Feldman Barrett, *How Emotions Are Made: The Secret Life of the Brain* (Houghton Mifflin Harcourt, 2017) 13–24.

329 See eg. Cosgrove et al (n 69); Friesen (n 76); Mohr, Shilton and Hotopf (n 72); Martinez-Martin et al (n 69).

330 Joanna Moncrieff, *The Bitterest Pills: The Troubling Story of Antipsychotic Drugs* (Palgrave Macmillan, 2013th edition, 2013).

331 Cris Shore and Laura McLauchlan, '"Third Mission" Activities, Commercialisation and Academic Entrepreneurs' (2012) 20(3) *Social Anthropology* 267; K Philpott, L Dooley, C O'Reilly and G Lupton 'The entrepreneurial university: examining the underlying academic tensions' (2010) 31 *Technovation* 161– 70.

332 Adam Rogers, 'Star Neuroscientist Tom Insel Leaves the Google-Spawned Verily for ... a Startup?' (11 May 2017) *Wired* <https://www.wired.com/2017/05/star-neuroscientist-tom-insel-leaves-google-spawned-verily-startup/>.

333 'Joint Letter to Minister Hunt – My Health Record: Call to Suspend My Health Record Roll Out'Letter from Shauna Gaebler et al, 7 August 2018 <http://being.org.au/2018/08/joint-letter-to-minister-hunt-my-health-records/>.

One important step to preventing harms caused by data concerning mental health and disability – whether leaked, stolen or traded – is to strengthen non-discrimination rules concerning mental health and psychosocial disability.[334] National discrimination laws may require amendments to ensure that discrimination on mental health grounds by online businesses is covered.[335] Such amendments would be consistent with the goals and legislative history of anti-discrimination laws and would remove ambiguity regarding the status of websites, social media platforms and other online businesses.[336] Remedies for individuals who are aggrieved by discriminatory behaviour and practices are also likely to require strengthening, including ensuring substantive, verifiable, auditable standards of non-discrimination in the use of algorithmic and data-driven technologies.

2.4.2 Fairness

Fairness is a broadly shared aspiration for governing algorithmic technologies. Definitions of 'fairness' differ, and the notion of 'algorithmic fairness' itself is an increasingly expanding field of research.[337] There are reportedly between 15 and 25 plausible definitions of 'fairness' of relevance to algorithmic technologies, each with different and often mutually exclusive emphases.[338] Depending on the context, some definitions are constructed in highly technical ways, centred on data science expertise, while others draw on the common-usage (and equally broad) aims of equitable and impartial treatment.[339] There is a risk that vague references to 'fairness' may hide important political decisions about how fairness is understood, including – importantly – who defines it.

Such choices ought to be transparent when used in the design of mental health related technologies, partly to clarify objectives, but also to highlight who may gain (or lose) political and decision-making power depending on the approach to fairness that is adopted. If addressing fairness becomes highly technical, for example, requiring the expertise of computer scientists, mathematicians, and so on, there is seemingly less scope for those most affected to determine the parameters of what is considered fair and unfair.

2.4.3 Equality

Equality is another generally accepted aspiration, centred on the goal of ensuring the same opportunities and protections to people interacting with algorithmic systems. Some have taken this aim further in seeking to use algorithmic systems to 'eliminate relationships of domination between groups and people based on differences of power, wealth, or knowledge' and 'produce social and economic benefits for all by reducing social inequalities and vulnerabilities.'[340]

Any discussion of equality in the mental health context must acknowledge existing inequalities in how mental health issues play out. Psychological distress does not occur equally across society: those who are poorer, from disadvantaged, marginalised and oppressed groups are more likely to experience distress, psychosis, trauma, mental health conditions and psychosocial disabilities.

334 Marks, 'Algorithmic Disability Discrimination' (n 13).
335 Ibid.
336 Ibid.
337 Dana Pessach and Erez Shmueli, 'Algorithmic Fairness' [2020] *arXiv:2001.09784* [cs, stat] <http://arxiv.org/abs/2001.09784>.
338 Goldenfein (n 75), p.130.
339 Fjeld et al (n 190).
340 University of Montreal, 'Montreal Declaration for a Responsible Development of Artificial Intelligence' (2018) 13 <https://www.montrealdeclaration-responsibleai.com/the-declaration>.

Inequalities also appear within mental health services themselves. There is inequality of access in mental health services; for example, in the UK, older people are underrepresented in talking therapies[341] and Black British men are overrepresented in involuntary psychiatric interventions.[342] Inequalities of *experience* also arise. In Aotearoa New Zealand, people with higher economic deprivation report lower satisfaction with health services compared to others, and this group disproportionately includes high numbers of people of Māori, Pacific or Asian ethnicity.[343] In high-income countries, higher proportions of Black people get diagnosed, with much research and debate about why this is the case.[344]

Some would argue that algorithmic and data-driven technologies could be used to address such inequalities; for example, helping to identify inequities in service systems, or by undertaking analyses of complex socio-economic dimensions to mental health problems. Yet, there is also potential that such technologies will replicate and even exacerbate inequalities. Sarah Carr, speaking from a UK perspective, has pointed out that the higher likelihood that Black British men will be subject to involuntary psychiatric intervention may mean that algorithmic approaches to service provision could exacerbate patterns of coercive intervention against racialised minorities.[345] Psychiatric diagnoses are already skewed with respect to race. In the US, for example, Black and minority ethnic groups tend to receive more 'severe' diagnostic categories (eg. schizophrenia rather than schizo-affective disorder), for diverse reasons, including mental health practitioners seeking to ensure a low-income person can qualify for housing or social security, which a more 'severe' diagnosis might afford.[346]

The 'Serenity Integrated Monitoring' program in the UK, noted earlier (X), which involves analysing health authority data to identify people repeatedly subject to forced psychiatric treatment and referring them to a police monitoring program, has been criticised for the likelihood that it will have 'violent consequences [that] disproportionately impact Black, Asian and minority ethnic communities'.[347] Sage Stephanou, founder of the Radical Therapist Network (RTN), stated:

> SIM perpetuates the prison industrial complex by monitoring and gatekeeping healthcare support and ultimately criminalises people who experience significant mental illness and trauma, often exasperated by systematic racism, oppression and adverse experiences. … SIM will exasperate the very real and legitimate fear that if racialised individuals access mental health support, or report abuse, they are at risk of systemic violence under the guise of care. Police involvement often escalates risk, creating dangerous situations through the use of physical restraint, coercive, unethical forms of treatment, detainment, and higher chances of Black and brown people dying whilst in police custody.[348]

Again, the SIM program only used data-driven technology at a small but crucial point in the program. Yet, the example suggests it is necessary to move beyond vague notions of equality and fairness in efforts to ensure the prevention of harm and equal distribution of benefits of using algorithmic and data-driven technology to address distress and healing.

341 Rob Saunders et al, 'Older Adults Respond Better to Psychological Therapy than Working-Age Adults: Evidence from a Large Sample of Mental Health Service Attendees.' (2021) 294(1) *Journal of Affective Disorders* 85.

342 V Lawrence et al, 'Ethnicity and Power in the Mental Health System: Experiences of White British and Black Caribbean People with Psychosis' [2021] *Epidemiology and Psychiatric Sciences* 294(1) 85-93.

343 Carol HJ Lee and Chris G Sibley, 'Demographic and Psychological Correlates of Satisfaction with Healthcare Access in New Zealand' (2017) 130(1459) *New Zealand Medical Journal* 14.

344 Mary O'Hara, 'Black and Minority Ethnic Mental Health Patients "marginalised" under Coalition', *The Guardian* (online, 17 April 2012) <https://www.theguardian.com/society/2012/apr/17/bme-mental-health-patients-marginalised>.

345 Carr (n 53).

346 Richard Sears, 'Combatting Structural Racism and Classism in Psychiatry: An Interview with Helena Hansen', *Mad in America* (13 October 2021) <https://www.madinamerica.com/2021/10/interview-helena-hansen/>.

347 Jameela (n 80).

348 Ibid.

As a final point on equality for this section, broader structural questions may be asked about growing inequality facilitated and indeed accelerated by the algorithmic and data-driven systems that power the information economy. The impact of social, political, and economic structures on mental health is well established. Drilling down into how a particular algorithmic system promotes or threatens equal opportunities for individuals with a mental health diagnosis under current conditions may distract from the way human distress arises primarily as a *consequence* of poverty, precarity, violence, and trauma as a form of social suffering—for which growing inequality in many countries will continue to be a major contributor. The role of technological change in these broader trends has a crucial role in discussions about new and emerging technology aimed at ameliorating distress. These issues will be discussed later in the report, in the section concerning public interest and societal wellbeing (page 83).

2.4.4 Inclusive Design – Emancipatory? Participatory?

Current algorithmic and data-driven initiatives in the mental health context are dominated by actors that have the most power, such as large private entities and public institutions, service providers, universities, professional associations, and so on. This concentration of power can mean a lack of digital technology oriented to experience and real-life usage. Sarah Carr writes:

> It is not too late to involve patients, service users and carers as domain experts in AI research and discussions about the ethical use of AI. It is therefore time to assess the situation, to question those who are driving this transformative agenda forward and to listen to excluded experts – those whose lives these technologies will ultimately affect.[349]

For a fuller discussion of this point see above, Section 1.5 "Elevating the Perspective of People with Lived Experience of Extreme Distress and Disability". Given that every person could generate 'data concerning mental health', a broad cross-section of society should have the opportunity to weigh in on the use of algorithmic and data-driven technology in mental health contexts. However, those with the most at stake tend to be those who have firsthand experience of using mental health services, including those who have been subject to involuntary psychiatric interventions, received a psychiatric diagnosis, or who just live with profound distress, a mental health condition or psychosocial disability. As noted previously, this group should not be viewed merely as another 'key stakeholder' but as the primary affected population whose involvement should form a political and ethical orientation underlying all work in this area. 'Inclusive design' is often associated with notions of 'human rights by design', which will be discussed below regarding human rights law (X).

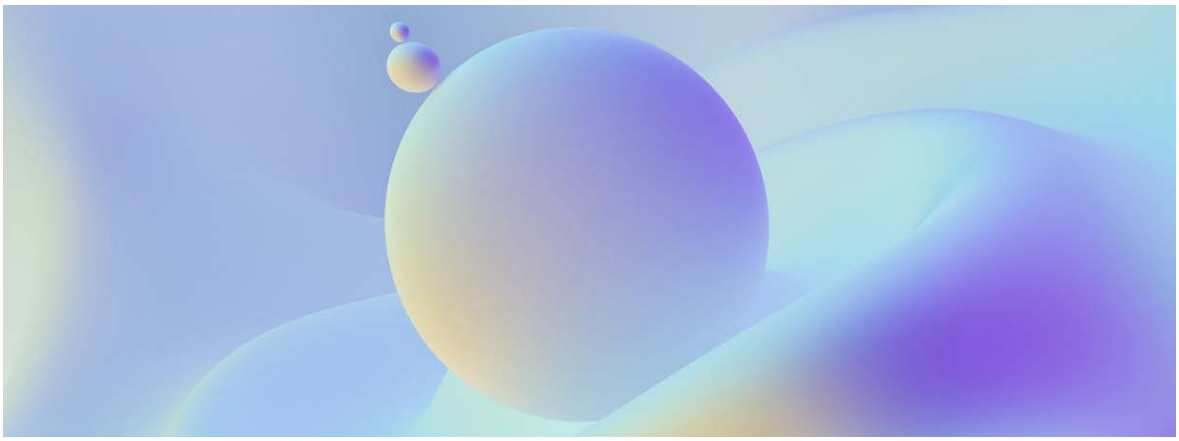


Photo by Milad Fakurian on Unsplash.

349 Carr (n 53).

2.4.5 Access to Technology

Equity of access to technology is an important social good, and is often reflected in concerns with a growing 'digital divide' in society. Several researchers have examined this concern as it affects people with psychosocial disabilities and lived experience.[350] For example, Til Wykes has commented in relation to mobile technology that:[351]

> Although mental health service user ownership of smartphones has increased there is still evidence of a digital divide. Any benefits [of mobile phone-based initiatives] for those experiencing mental health problems are, therefore, likely to be less than in the general public, due to lack of access and skills [in using mobile] devices.

In the UK, Dan Robotham and colleagues looked at 'digital exclusion' facing 241 participants with mental health diagnoses.[352] The researchers concluded that the 'digital divide' was difficult to overcome but that successful steps could be taken to improve access to digital technologies for people who lack the knowledge, skill and financial resources, and that such steps could even form part of an essential service for citizenship and community wellbeing.

Digital inclusion strategies – such as subsidising the purchase of equipment, internet billing support, education to improve digital literacy, and so on – may be required to prevent people becoming excluded from both digitised health and social services, but also from society in general.[353] However, addressing digital equity may also mean ensuring that people can access entirely 'non-digital' resources for those who do not wish to, or cannot, use digital technological approaches to care and support. Equally, it is important to avoid generalisations about people with lived experience or psychosocial disabilities' supposed deficiencies in digital literacy and access. False assumptions about people with disabilities' supposed passivity and incapacity can result in paternalistic, top-down approaches that falsely presume a need for state and industry intervention.[354] One public inquiry in Victoria, Australia, recommended that governments could help address the 'digital divide' in the mental health context by 'enabl[ing] mental health and well-being services to offer people living with mental illness or psychological distress access to devices, data and digital literacy support, where it is their preference to use digital services but they are otherwise unable to do so.'[355] Elsewhere in Australia, case law has established the potential for social security recipients to use their disability support funds for internet hardware and data plans.[356]

350 Liam Ennis et al, 'Can't Surf, Won't Surf: The Digital Divide in Mental Health' (2012) 21(4) *Journal of Mental Health* 395; Murielle Girard, Philippe Nubukpo and Dominique Malauzat, 'Snapshot of the Supports of Communication Used by Patients at a French Psychiatric Hospital: A Digital or Social Division?' (2017) 26(1) *Journal of Mental Health* 8.

351 Wykes (n 167).

352 D Robotham et al, 'Do We Still Have a Digital Divide in Mental Health? A Five-Year Survey Follow-Up' (2016) 18(11) *Journal of Medical Internet Research* e309.

353 Australian Human Rights Commission (n 184).

354 Hamraie and Fritsch (n 193).

355 Recommendation 60, Royal Commission into Victorian Mental Health Services. Victorian Government, Australia https://finalreport.rcvmhs.vic.gov.au/recommendations/ (accessed 1/4/22).

356 *Gelzinnis and National Disability Insurance Agency* [2021] Administrative Appeals Tribunal of Australia 3970.

2.5 Human control of technology

Important decisions concerning mental health that are made with algorithmic technology must be subject to human control and review. Human control over technology is key to other themes discussed in this report, including maintaining safety and security, accountability, transparency, equity and non-discrimination.

The advance of algorithmic and data-driven technologies is often presented as inevitable, with thought leaders in computer science often depicting automation as a force of nature propelled by unstoppable technological change.[357] This view risks delegating autonomy and responsibility for such systems away from humans to some higher and abstract authority. Even terms like 'artificial intelligence' can imply an external intelligence that is somehow 'out there'. Relinquishing accountability in this way is sometimes described as the 'substitution effect',[358] an effect which makes it harder to ensure human control over technologies by those who are actually designing and implementing them—and those who are using and/or subject to them.

Examples exist of civil society groups successfully challenging the introduction of some types of AI, such as facial recognition. Such groups have insisted that these technologies are not inevitable and demonstrated the power of public input to change the direction of, and even halt, the use of certain technologies. Similar examples are emerging in the mental health context (see Section 1.5).

Emphasising the importance of human control in the mental health context is important given the risk that algorithmic technologies like AI become a 'substitute decision-maker' over both professionals and those receiving care and support, which will impact individuals' agency and autonomy.[359] One ethical risk in professional decision support technology in healthcare, is that clinicians defer to algorithmic technology suggestions even in the face of a contrary opinion. An overreliance on automated systems may therefore displace human agency, moral responsibility, liability and other forms of accountability. Even the use of 'algorithm' to describe a decision-making system has been characterised by some advocates as 'often a way to deflect accountability for human decisions'.[360]

More broadly, there is a risk of normalising and accepting technologies which reinforce potentially inaccurate, unhelpful or harmful categorisations of people's mental states (as discussed in the biometric monitoring section of this report).

CASE STUDY: Covert and Commercial Automated 'Narcissism or psychopathy' Testing

AirBnB has reportedly claimed to be able to use computer modelling to determine whether a customer displays 'narcissism or psychopathy'.[361] The determination is reportedly aimed at screening out undesirable platform users, and specifically possible tenants who may damage the property of landlords.[362]

357 Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press, 1 edition, 2016).

358 Ian Kerr and Vanessa Gruben, 'AIs as Substitute Decision-Makers' (2019) 21(78) *Yale Journal of Law & Technology* 80.

359 Ibid.

360 Kristian Lum and Rumman Chowdhury, 'What Is an "Algorithm"? It Depends Whom You Ask', *MIT Technology Review* (26 February 2021) <https://www.technologyreview.com/2021/02/26/1020007/what-is-an-algorithm/>.

361 Aaron Holmes, 'Airbnb Has Patented Software That Digs through Social Media to Root out People Who Display "Narcissism or Psychopathy"', *Business Insider Australia* (7 January 2020) <https://www.businessinsider.com.au/airbnb-software-predicts-if-guests-are-psychopaths-patent-2020-1>.

362 Ibid.

This case study raises questions about the appropriateness (and lawfulness) of certain claims being made about individuals, including whether it is appropriate to claim that behavioural sensing can 'reveal' an underlying mental state or diagnosable disorder through 'silent digital biomarkers'. The claim rests on a presumption about the capacity of computer technology, combined with psychometrics, to capture reality. Even the common claim that various technologies 'collect' data suggests that there is some neutral, objective information-*gathering* process underway. Instead, it seems more accurate to say that new forms of data concerning mental health are being created and *generated*. This creation and generation is not value-neutral—it is value-laden and often rests on multiple social and political claims that may be unstated. The nature of this data-generation, the meaning given to different types of data, and the often uncritical presentation of these methods as neutral forms of data 'collection' warrants critical scrutiny—whether in technology sales materials, media, government policy documents, or in leading scholarly journals.

Another risk is that the push to generate ever more data to improve algorithmic and data-driven solutions can undermine human action and control by distracting from the need to change *existing* mental health services, policies and practices. An extraordinary amount of energy can go into efforts to generate, curate, store and use data. This process can undermine action on information *that already exists*, particularly information highlighting existing problems of resourcing, discrimination, coercive practices, and other prominent issues in the politics of mental health.[363]

2.5.1 Human Review of Automated Decision

A more targeted way to ensure human control of automated decisions is to ensure that people who are subject to automated decisions that draw on data concerning their mental health should be able to request and receive human review of those decisions. The principle of ensuring human review, unlike most other themes discussed in this report, is meant as a step *after the fact*. This is not to endorse the use of various automated initiatives in the first place, some of which may warrant preventive interventions that modify or halt them. Instead, the principle of human review is to ensure an *ex post* (after the fact) mechanism of review where algorithmic decision systems are used. The principle is guided by the rationale, noted by the European Commission's High-Level Expert Group on Artificial Intelligence, that '[h]umans interacting with AI systems must be able to keep full and effective self-determination over themselves'.[364]

Different technologies, and the contexts in which they are used, will require variations of the human review that is appropriate, including the strength of that review process. Some groups have argued that human review is not merely desirable but should be viewed as a right of the data subject.[365] The European Ethical Charter on Artificial Intelligence in Judicial Systems and their Environment, for example, contains a robust approach, indicating that cases must be heard by a competent court if review is requested.[366]

363 See eg, Piers Gooding, *A New Era for Mental Health Law and Policy: Supported Decision-Making and the UN Convention on the Rights of Persons with Disabilities* (Cambridge University Press, 2017); Green and Ubozoh (n 17); Rose (n 29).

364 European Commission High-Level Expert Group on Artificial Intelligence (n 131) para [50].

365 Fjeld et al (n 85) p.54.

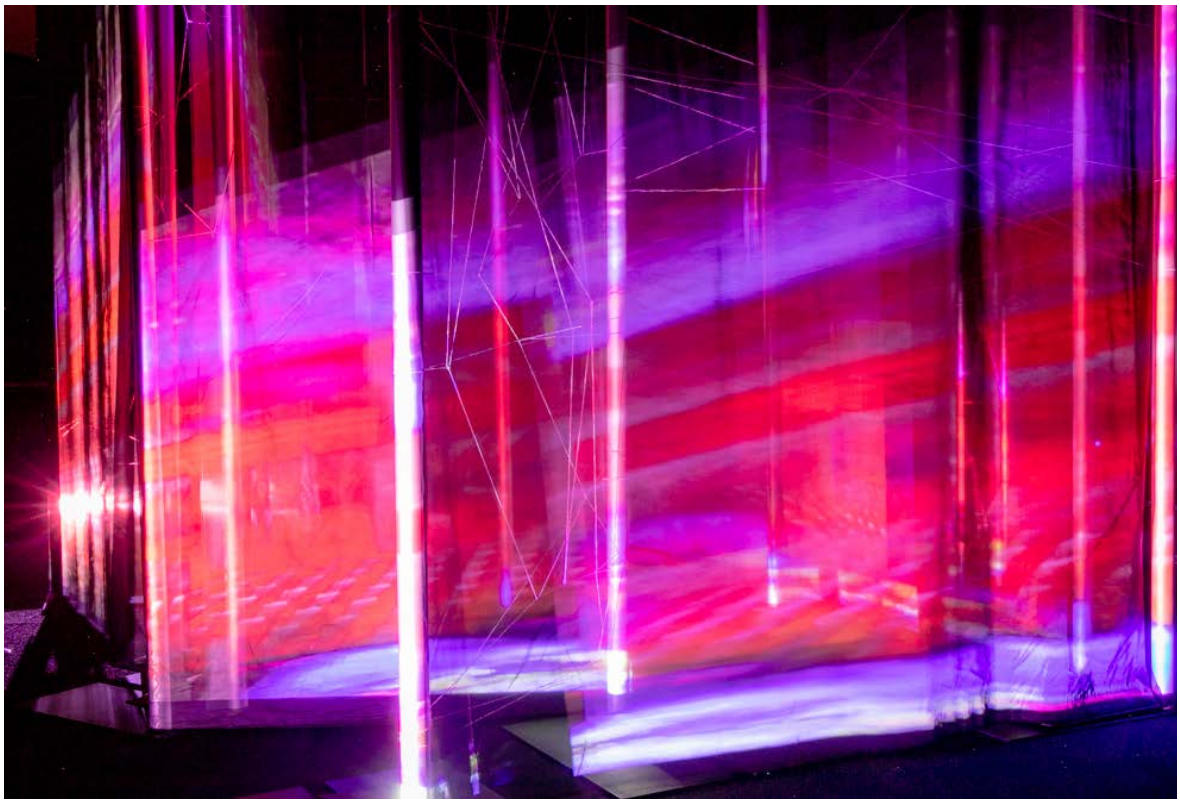
366 European Commission for the Efficiency of Justice, *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment* (3 December 2018) p.54.

2.5.2 Ability to Opt-Out of Automated Decision-Making

People must retain the ability to opt-out of algorithmic and data-driven approaches that use data concerning their mental health. This point is closely related to issues of consent and transparency (see Section 2.7). Members of the public deserve clarity about when algorithmic technologies are being used to make decisions about them using data concerning their mental health. According to a UK government report, greater clarity 'will help the public experience the advantages of AI, as well as to opt out of using such products should they have concerns.'[367]

Opting out is not always going to be clear-cut given that people are likely to interact with AI systems in numerous ways. As Fjeld and colleagues write: '[people's] information may be used as training data; they may be indirectly impacted by systemic deployments of AI, and they may be personally subject to automated decisions.'[368] Attention is required to what it means to offer meaningful opportunities to exercise choice in opting out of data-driven technologies in mental health contexts.

One risk is that those who opt-out or discontinue interacting with these technologies are recorded as being non-compliant, or having inferences made about them in ways that are leveraged against their interests. If a healthcare provider, employer, or education institution uses a mental health and wellbeing program that provides incentives to engage with wearable devices and other forms of biometric monitoring, for example, there must be reasonable steps taken to ensure those who choose not to participate are not stigmatised, disadvantaged, or portrayed negatively as noncompliant.



Distorted Constellations in Nwando Ebizie in Science Gallery Melbourne's MENTAL. Photo by Alan Weedon.

367 Select Committee on Artificial Intelligence, *AI in the UK: Ready, Willing and Able? - Artificial Intelligence Committee* (No Report of Session 2017–19, HL Paper 10, 16 April 2017) para [58].

368 Fjeld et al (n 85) p.54.

2.6 Professional responsibility

This theme is mainly aimed at individuals and groups who are responsible for designing, developing, or deploying algorithmic and data-based technologies.[369] The actions and understanding of these individuals and groups have a direct influence on the ethical, legal, social and political dimensions of technology being used in the mental health context.

2.6.1 Multi-disciplinary and Participatory Collaboration

The rise of algorithmic technologies has seen computer scientists, engineers and mathematicians increasingly enter healthcare service delivery, research and development. These new entrants may be unaware of the broader politics of mental health and may struggle to engage people with firsthand experience of accessing mental health services.

Some clinical researchers have acknowledged the gap of involving affected populations. For example, clinical psychologist David Mohr and colleagues noted that '[w]e have typically not done a good job of getting input from patients about their goals, needs, or preferences'.[370] Mohr and colleagues discuss the harm this failure does to the efficacy of technologies that may be validated in research settings but then fail to work in real-world settings:

> Trials often bear little resemblance to clinical settings, having largely emphasized internal validity over real-world issues, such as the technological environment and implementation and sustainment.... Essentially, clinical researchers have designed tools to try to get people to do what we want them to do and how we want them to do it–and then searched for and found people who were interested in or willing to use these tools in our trials. Thus, we should not be surprised that these products and services are not appealing to the general population.[371]

To address the mismatch between the 'laboratory' and real-life, interdisciplinary and applied empirical research is needed, including not just medical researchers, computer scientists, and those involved in service delivery, but also humanities scholars.

Most digital initiatives are concerned with complex social interventions—each occurs in a complex web of formal and informal relationships. Social scientists and humanities scholars are concerned with making sense of the interaction in social, cultural, environmental, economic and political contexts. They can also bring ideas to help guard against persistent reductionist habits in mental health sciences and technological industries,[372] including scrutinising claims about what algorithmic and data-driven technology can realistically tell us about people's inner-lives. Resources from sociology, anthropology, philosophy, and so on, can help determine what role algorithmic and data-driven technologies might play in creating the social conditions that improve relations between people, including acceptance of human diversity and frailty that accounts for extreme mental states and mental health crises, distributes resources appropriately and maximises human flourishing.

369 Fjeld et al (n 78) p.31.
370 David C Mohr et al, 'Three Problems With Current Digital Mental Health Research ... and Three Things We Can Do About Them' (2017) 68(5) *Psychiatric services (Washington, D.C.)* 427.
371 Ibid.
372 Harrington (n 4).

2.6.2 Scientific Integrity and Testing Claims

Concerns have been raised in the literature about the rapid and even reckless embrace of algorithmic technologies in mental health research; computer and cognitive scientist Chelsea Chandler and her colleagues have described this recent flurry of commercial and research activity as akin to the 'wild west'.[373] They describe an urgent need in the field for 'a framework with which to evaluate the complex methodology such that the process is done honestly, fairly, scientifically, and accurately'.[374] The James Lind Alliance Priority Setting Partnership, in its survey of the field concluded that 'the evidence base for digital mental health interventions, including the demonstration of clinical effectiveness and cost effectiveness in real-world settings, remains inadequate'[375]. Health Education England also surveyed algorithmic and data-driven technologies in mental health care and raised concerns about 'spurious claims and overhyped technologies that fail to deliver for patients'.[376] Perhaps most damning was a recent meta-review on mobile phone-based interventions for mental health by Simon Goldberg and colleagues, which surveyed 14 meta-analyses representing 145 randomised control trials involving 47,940 participants.[377] Despite this extensive search and vast body of research, the review 'failed to find convincing evidence in support of any mobile phone-based intervention on any outcome.'[378]

This is not to suggest all digital technological approaches to mental health are unsupported by evidence or unworthy of further research. Instead, it is to caution against the hype and 'techno-solutionism' which pervades the field.

2.6.3 Against Hype and 'Techno-solutionism'

Some technology vendors and clinical experts may presume to have algorithmic and data-driven solutions for mental health care but it is not always clear whether people impacted by psychological distress actually want or need them.[379] The history of both mental health and computer sciences are littered with hubris and outlandish claims about scientific solutions to the longstanding and complex issue of human distress, mental health issues, and so on. For psychiatry and neuroscience, grandiose claims in the recent past include the purported discovery of 'breakthrough' biological or neurological treatments that will 'revolutionise' care.[380] For technologists, claims include being able to 'solve' issues using AI from crime, corruption, pollution or obesity.[381] Elon Musk's claim that his 'AI-brain-chips company could "solve" autism and schizophrenia'[382] suggests that these traditions are beginning to merge.

Such overblown claims perpetuate a form of 'solutionism'. Solutionism, or 'techno-solutionism', refers to the (flawed) belief that every social problem has a technological fix, and that simple technological fixes are possible for what are actually highly complex social issues.[383] As well as perpetuating the belief that certain issues are amenable to being solved by technology, this type of over-hyping can easily lead to over-promising and under-delivering. On an individual level, one consequence could be to shape individuals' and mental health professionals' preferences and expectations about treatment.

373 Chelsea Chandler, Peter W Foltz and Brita Elvevåg, 'Using Machine Learning in Psychiatry: The Need to Establish a Framework That Nurtures Trustworthiness' (2020) 46(1) *Schizophrenia Bulletin* 11.
374 Ibid.
375 Hollis et al (n 42).
376 Tom Foley and James Woollard, 'The Digital Future of Mental Healthcare and Its Workforce: A Report on a Mental Health Stakeholder Engagement to Inform the Topol Review' (National Health Service (UK), February 2019) p.31.
377 Simon B Goldberg et al, 'Mobile Phone-Based Interventions for Mental Health: A Systematic Meta-Review of 14 Meta-Analyses of Randomized Controlled Trials' (2022) 1(1) *PLOS Digital Health* e0000002.
378 Ibid.
379 Carr (n 53).
380 Harrington (n 4).
381 Morozov (n 11).
382 Hamilton (n 8).
383 Morozov (n 11).

When over-hyped initiatives fail, or misrepresent a 'problem', individuals may despair that these technological steps haven't reduced their distress. On a macro-level, over-stating the evidence can alter how funding is directed and draw resources away from where they are needed most. This increases the possibility of technology monopolising limited resources.

There is also a risk with techno-utopian or 'techno-optimist' approaches that technology-driven solutions and fixes are presented as an *unquestioned good*. This is not to argue the opposite and reject technological approaches as wholly negative. Instead, it is to caution against the presentation of any digital initiative in mental healthcare as self-evidently virtuous. Such an altruistic and optimistic picture can shutdown important questions about the way problems are framed, and who benefits and who loses as a result.

For example, one widely-promoted idea that can be sidelined by uncritical optimism is that digital mental health initiatives are cost-effective. There may be some evidence supporting this claim from individual initiatives. However, Jacqueline Sin and colleagues examined claims about cost-savings in a systematic review of 'entirely web-based interventions that provided screening and signposting for treatment, including self-management strategies, for people with [common mental disorders] or subthreshold symptoms'.[384] Many interventions promised low cost of service relative to face-to-face support, which was then used to suggest it could be expanded and delivered to larger populations; yet the review identified that 'no data were available regarding estimated cost-effectiveness and only 1 paper included economic modeling'.[385] Advocacy organisation *Privacy International* has likewise argued that there remains little evidence that AI will necessarily lead to more efficient healthcare systems, despite a widespread assumption – boosted by technology vendors – that this will be the case.[386]

Another commonly-held view is that computational monitoring, measuring and evaluation of people *necessarily* affords access to knowledge about individuals, including their inner states. Instead, computational technology may well *get in the way* of understanding people, including the unique experience of each new person in crisis or distress who deserves to be heard fully.

The assumption that there is always or even often a technological fix for any problem is highly likely to be misplaced regarding various aspects of humane and effective responses to supporting people in severe distress, trauma, mental health crises, and so on. Hence, there is a need not only to mitigate against proven and potential harms, but also to establish sufficient standards to *highlight unproven benefits* that remain clouded by hype and solutionism, as noted previously.

384 Jacqueline Sin et al, 'Digital Interventions for Screening and Treating Common Mental Disorders or Symptoms of Common Mental Illness in Adults: Systematic Review and Meta-Analysis' (2020) 22(9) *Journal of Medical Internet Research* e20581.

385 Ibid.

386 Privacy International, 'Our Analysis of the WHO Report on Ethics and Governance of Artificial Intelligence for Health', *Privacy International* (20 July 2021) <http://privacyinternational.org/news-analysis/4594/our-analysis-who-report-ethics-and-governance-artificial-intelligence-health>.

2.6.4 Responsible Design, Including Consideration of Long-Term Effects

Technologists and mental health professionals clearly play an important role in shaping the ethical, social and political dimensions of emerging technologies in the mental health context. France's Artificial Intelligence Strategy describe researchers, engineers and developers as 'architects of [our] digital society'.[387] Others have challenged the way this group of professionals has been elevated to such a lofty status, and suggest this view risks conceding undue power to data scientists, engineers and the like.[388] Regardless, and as articulated in the Université de Montréal *Declaration for a Responsible Development of Artificial Intelligence*, professionals have a clear role in 'exercis[ing] caution by anticipating, as far as possible, the adverse consequences of [algorithmic systems] by taking the appropriate measures to avoid them.'[389]

Such attention is arguably missing from contemporary research in the mental health context. As noted, a survey by Piers Gooding and Timothy Kariotis on research that used algorithmic and data-driven technology in mental health initiatives, found that 85% of the studies did not appear to consider how the technologies could be appropriated in negative ways, despite some of the technologies raising serious legal and ethical issues. One possible solution to this 'blind spot' – at least in the scholarly field – is to require researchers to consider the long-term- and potential adverse-effects of different technologies, which could be encouraged through editorial requirements in scholarly journals, ethics/institutional review processes, and funding stipulations.

Doing Nothing with AI by Emanuel Gollob in Science Gallery Melbourne's MENTAL. Photo by Alan Weedon.

387 Cédric Villani, *For a Meaningful Artificial Intelligence: Toward a French and European Strategy* (2018) 154, p.120 <https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf>.

388 Bernard Stiegler, "'Le grand désenchantement'. Un entretien avec le philosophe Bernard Stiegler", *Le Monde*, 21 February 2011; Goldenfein (n 248).

389 University of Montreal, 'Montreal Declaration for a Responsible Development of Artificial Intelligence' (2018), Principle 8 <https://www.montrealdeclaration-responsibleai.com/the-declaration>.

2.7 Transparency and explainability

Transparency is defined in various ways but essentially refers to the importance of technological systems being designed and implemented so that oversight is possible. Transparency is therefore closely linked to accountability. It could extend to matters such as the data that are generated in a particular setting, the system that processes the data, and (where relevant) the business model that makes use of them.[390] As an example, Til Wykes and Stephen Schueller have proposed a transparency governance method for apps concerning health; they raised concerns about the overselling of health apps and so-called health apps that may provide little benefit and even harm.[391] They presented the 'Transparency for Trust (T4T) Principles of Responsible Health App' in the form of a list of questions that can be asked to reveal key matters concerning privacy, security, feasibility, and so on.[392] Wykes and Schueller promote the use of the T4T principles by app stores and for presentation 'in a simple form so that all consumers can understand them.'[393]

The related concept of **explainability** refers to 'the translation of technical concepts and decision outputs into intelligible, comprehensible formats suitable for evaluation'.[394] Explainability seems particularly important for software that analyses large datasets algorithmically,[395] which, again, are a minority of digital initiatives in the mental health context. Explainability is particularly crucial for systems with potential to cause harm or significantly impact individuals, such as impacting health, access to resources and quality of life. A governance example on this point is the *AI in the UK: Ready, Willing and Able?* report, which notes that if an AI system has a 'substantial impact on an individual's life' and cannot provide 'full and satisfactory explanation' for decisions made, then it should not be deployed[396]—a view which few people, if any, would challenge in the mental health and disability context. Explainability is often linked to promoting nondiscrimination given that the more comprehensible a system is, the more likely discrimination, bias or error can be identified, prevented and rectified.[397]

2.7.1 Open-Source Data and Algorithms

Open-source principles promote code, data and algorithms being made freely available for possible modification and redistribution. These principles can promote a collaborative, inclusive and community-minded approach to technology development, can facilitate equal distribution of the benefits across regions and populations, and can help avoid monopolies or power concentration associated with particular technologies. Positive examples exist of open-source technologies in the mental health context, such as apps that place a premium on transparency.

390 European Commission High-Level Expert Group on Artificial Intelligence (n 135) p.18.
391 Til Wykes and Stephen Schueller, 'Why Reviewing Apps Is Not Enough: Transparency for Trust (T4T) Principles of Responsible Health App Marketplaces' (2019) 21(5) *Journal of Medical Internet Research* e12390.
392 Ibid.
393 Ibid.
394 Fjeld et al (n 88) pp.42-43.
395 Julia Amann et al, 'Explainability for Artificial Intelligence in Healthcare: A Multidisciplinary Perspective' (2020) 20(1) *BMC Medical Informatics and Decision Making* 310.
396 Select Committee on Artificial Intelligence (n 370).
397 Fjeld et al (n 88) p.43.

CASE STUDY: LAMP (Learn, Assess, Manage, Prevent) – an open source and freely available app

ALAMP is a freely available and open access app that was developed by a group of researchers to support 'clinicians and patients [...] at the intersection of patient demands for trust, control, and community and clinician demands for transparent, data driven, and translational tools'.[398] The LAMP platform, according to John Torous and the researchers who led the initiative, 'evolved through numerous iterations and with much feedback from patients, designers, sociologists, advocates, clinicians, researchers, app developers, and philanthropists'.[399] The authors state:

> As an open and free tool, the LAMP platform continues to evolve as reflected in its current diverse use cases across research and clinical care in psychiatry, neurology, anesthesia, and psychology. [...] The code for the LAMP platform is freely shared [...] to encourage others to adapt and improve on our team's efforts.[400]

The app can be customised to each person and reportedly fit with their personal care goals and needs, and there is research underway to seek to link the app to options for peer support.[401] To promote input by people with lived experience of mental health interventions, the researchers used 'guided survey research, focus groups, structured interviews, and clinical experience with apps in the mental health care settings, [in a process of seeking] early and continuous input from patients on the platform.'[402]

There are some risks with open science principles, insofar as technologies may be re-purposed for bad ends. Risks may be acute concerning biometric monitoring. Consider the following comment by academic psychiatrist, Nguine Rezaii, who was discussing her biometric monitoring research:

> [w]hen I published my paper on predicting schizophrenia, the publishers wanted the code to be openly accessible, and I said fine because I was into liberal and free stuff. But then what if someone uses that to build an app and predict things on weird teenagers? That's risky. [...] [Open science advocates] have been advocating free publication of the algorithms. [My prediction tool] has been downloaded 1,060 times so far. I do not know for what purpose..."[403]

Some technological practices made with good intention, which may at first appear as if they should be openly accessible, could be re-purposed in unexpected and harmful ways and may need to be prevented from being publicised. For example, researchers may develop an algorithmic tool to quickly identify social media users who appear to be LGTBQI+ young people, to whom specifically-designed crisis support can be directed. Such a tool carries inherent risks, including in the event it was used by bad actors who were hostile to LGBTQI+ people. Platform regulation that adequately combats online abuse, harassment and vilification would be the clearest path to address this risk (and would carry mental health benefits more broadly). In lieu of better platform regulation, at least one option to address the tension raised by open science principles is having disclosure processes so that algorithms may be subject to validation or certification agencies that can effectively serve as auditing and accountability bodies, and in ways that respect when some algorithmic technologies should not be made open source.[404]

398 John Torous et al, 'Creating a Digital Health Smartphone App and Digital Phenotyping Platform for Mental Health and Diverse Healthcare Needs: An Interdisciplinary and Collaborative Approach' (2019) 4(2) *Journal of Technology in Behavioral Science* 73.

399 Ibid.

400 Ibid.

401 Ibid.

402 Ibid p.75

403 David Adam, 'Machines Can Spot Mental Health Issues—If You Hand over Your Personal Data', *MIT Technology Review* (online, 13 August 2020) <https://www.technologyreview.com/2020/08/13/1006573/digital-psychiatry-phenotyping-schizophrenia-bipolar-privacy/>.

404 The Institute of Electrical and Electronics Engineers (IEEE) has proposed having disclosure processes so that algorithms may be subject to validation or certification agencies that can effectively serve as auditing and accountability bodies, and in ways that respect when some algorithmic technologies should not be made open source. IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, *Ethically*

2.7.2 Other Issues of Transparency and Explainability

Other points can promote transparency in the mental health and disability context, and include:

- **Open Procurement (for Government) –** Some governments' enthusiastic embrace of digital mental health technologies gives cause to ensure governments are transparent about their use. Digital rights advocacy organization, Access Now, recommends that open procurement standards should see governments publishing 'the purpose of the system, goals, parameters, and other information to facilitate public understanding' as well as a 'period for public comment' including reaching out to 'potentially affected groups where relevant to ensure an opportunity to input'.[405] Open procurement rules could be a key mechanism for addressing risks to accountability discussed in Section 2.2.1 of this report on privatisation and accountability.
- **Right to Information –** Promoting a right to information would aim to ensure individuals know about various aspects of the use of, and their interaction with, algorithmic systems in the mental health context. Several German Federal Ministries have promoted the right to information to access the criteria, objectives, and logic of a particular algorithmic decision system, and extended that to require 'labelling and publication obligations [...] in plain language and [in ways that are] easily accessible'.[406] This obligation aligns with the accessibility requirements for persons with disabilities enunciated in the Convention on the Rights of Persons with Disabilities (see below page 87). More broadly, good technology governance requires that terms of service – whether by governments or corporations – are accessible, clear and understandable rather than being presented in 'legalese' or buried in a mass of information (and acknowledging the sheer limitations of terms of service as an adequate remedy to the broader issues raised in this report).
- **Notification when Automated Decisions are Made about an Individual –** This point relates specifically to AI, machine learning and other algorithmic decision systems, and is closely related to preserving individuals' ability to opt-out of such systems. Autonomy and the opportunity to consent are dependent upon a person knowing they are subject to automated decisions. (An example where this did not occur is the automated hiring decision affecting US citizen, Mr Kyle Behm, which is noted in the previous section on 'Non-Discrimination and the Prevention of Bias', page 64). Clarity is needed in any automated decision process concerning a person's mental health or disability to be informed of how to contact a human and to ensure automated decisions can be checked or corrected.[407]
- **Notification when Interacting with Automated Systems –** In the mental health and disability context, people should always be made aware when they're engaging with technology rather than directly with another person. People with lived experience of crises, distress and mental health interventions have been very clear in studies that new and emerging technologies in mental health services should emphasise human connection and avoid creating isolation, loneliness and alienation.[408] 'Interacting' is a key word in this principle as notification should not be limited to automated *decisions*, which may be taken to describe when an action is automated, but should apply to *interactions*—for example, a person typing responses to a chatbot. This is not to suggest that chatbots cannot be richly crafted, and 'weave together code and poetry, emotions and programming,' as one commentator described it,[409] but is to suggest that notification that the chatbot is an automated system should be unambiguous, with clear information on how a person may reach a human where needed.
- **Regular Reporting –** This point refers to obligations placed on entities who are using automated decision systems to disclose that usage.

Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems (2019) 28 <https://ethicsinaction.ieee.org/>.

405 Access Now (n 125), p.32.

406 German Federal Ministry of Education and Research, the Federal Ministry for Economic Affairs and Energy, and the Federal Ministry of Labour and Social Affairs, 'Artificial Intelligence Strategy' (2020) 38

407 European Commission, 'Artificial Intelligence for Europe: Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee, and the Committee of the Regions' COM (2018) p. 17.

408 Hollis et al (n 42).

409 ng (n 1).

2.8 Promotion of Public Interest and Societal Good

> It's possible that prediction is not a magic bullet for mental health, And can't replace places of care staffed by people with time to listen, In a society where precarity, insecurity and austerity don't fuel generalised distress, Where everyone's voice is not analysed but heard, In a context which is collective and democratic.
>
> - Dan McQuillan[410]

Most people would agree that new technologies in the mental health context should promote the public interest. Various proposals exist for guiding this aim; for example, with reference to public law values like community, freedom and equality, or general democratic values of access to information, democratic governance, civic participation, and so on.[411] Others refer to international human rights law (which we will discuss in the next section), or internationally recognised labor rights,[412] and some have pointed to broad ethical aims like 'advancing human well-being' as a 'primary success criterion for development' beyond technology simply being profitable, legal and safe.[413]

Within the mental health context, broad concepts like 'recovery-oriented support' and 'trauma-informed care', which have strongly influenced mental health policy in recent years, might offer guidance for publicly-minded digital crisis support initiatives; so may guidelines for good mental health practices, such as those prepared by bodies like the World Health Organisation.[414]

For the purposes of this report, we will briefly discuss two notable issues, which did not fit easily elsewhere in the report but which seem noteworthy. The first concerns the importance of face-to-face support and the risk of automation depersonalising care, and the second concerns the tendency of many technological approaches to home in on the individual, at the expense of more socio-economic understandings of mental health crises, distress and disability.

2.8.1 Automation, Undermining Face-to-Face Care, and the Risk of Depersonalisation

One uniformly acknowledged risk is that digitising crisis support may reduce the type of human interaction and compassion that is indispensable to providing and experiencing care, support and healing. In 2018, Christopher Hollis and colleagues conducted what appears to be the largest participatory study in the world concerned with charting a research agenda about digital technologies in mental healthcare.[415] 664 'people with lived experience of mental health problems and use of mental health services, their carers, and health-care practitioners' in the UK were consulted. The number one research priority for participants was determining 'the benefits and risks of delivering mental health care through technology instead of face-to-face' and considering the impact of removing 'face-to-face human interaction'.[416] Participants' concluded that – above all – technologies that emphasised connection should be prioritised. They warned against technologies, including well-intentioned ones, that would exacerbate isolation, loneliness and alienation.

410 McQuillan (n 150).
411 Pasquale (n 360).
412 Fjeld et al (n 190). fn 305.
413 IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (n 287) pp. 21-22 (See Principle 2.).
414 World Health Organisation, *Guidance on Community Mental Health Services: Promoting Person-Centred and Rights-Based Approaches* (World Health Organization, 2021) <https://www.who.int/publications-detail-redirect/9789240025707>.
415 Hollis et al (n 42).
416 Hollis et al (n 197) p.7.

The impact of new technologies on dynamics of care will be a key discussion in coming decades, both in terms of care in health care facilities and residential homes, but also home-based care. One common aim of new technology, such as AI, is to break down tasks into individual components which can be repetitively undertaken. Yet, care is not just tasks, it is also emotion; it is a fundamental part of human relationships and it is a *highly complex social interaction*.[417] Some have claimed that technology will supersede human care. For example, a major suicide hotline in Australia claimed that a 'virtual or robotic counsellor' could 'speak to lots of people and provide support to people immediately' whereas 'phone counselors can only speak to one person at a time.'[418] Others have derided the benefit of human interaction in healthcare encounters, because it is not strictly clinical in nature and wasteful of healthcare resources. Nick Weston, the chief commercial officer at Lilli, a UK company behind biometric monitoring technology that was installed in the homes of older social care recipients, rejected claims that the monitoring could exacerbate the loneliness of older people who would have otherwise received care visits.[419] He stated that 'We shouldn't be relying on home care agency staff to provide the social interaction for somebody'.[420] Such claims rest on the narrowest conceptions of human care, flattening its complexity in the extreme.

Care has been persistently devalued in many societies. This devaluation is often based on the sexist premise that care is 'women's work', given that care for older persons, persons with disabilities and children has largely been performed by women, paid and unpaid.[421] Simplistic efforts to automate care can perpetuate this devaluation. To paraphrase Fiona Jenkins, relations of care are neither adequately recognised in many paradigms of digitised mental health care, and the language of market value, nor in our inherited traditions of contemplating the values and virtues that make for a truly human life.[422] Such concerns highlight the potential rise of cheap (if limited) software to replace more expensive, expert, and empathetic professionals, and the disruption of care service provision and public assistance for the provision of private mental health care.[423]

The potential for care technologies to dehumanise and objectify care-recipients was raised by the UN Independent Expert on the enjoyment of all human rights by older persons, Rosa Kornfeld-Matte.[424] The risk of 'automating care', she wrote, includes 'losing one's sense of identity, self-esteem and control over one's life'.[425] Kornfeld-Matte argued that human dignity must be 'integrated from the conception to the application of assistive devices and robotics'.[426] Even on a purely practical level, new technologies may also mean increased pressure on mental health and other service providers to increase multi-tasking and workloads in generating, inputting, organising, and constantly updating data records. Paradoxically, this extra work may reduce care workers' time for face-to-face engagement and collaborative work with care recipients and other care workers.

417 Eva Feder Kittay, 'The Ethics of Care, Dependence, and Disability*: The Ethics of Care, Dependence, and Disability' (2011) 24(1) *Ratio Juris* 49.

418 aggie Coggan, 'Virtual Counsellor Steps in to Help out on Suicide Hotline', *Pro Bono Australia* <https://probonoaustralia.com.au/news/2019/04/virtual-counsellor-steps-in-to-help-out-on-suicide-hotline/>.

419 Chris Baraniuk, 'Sensors and AI to Monitor Dorset Social Care Patients', *BBC News* (online, 24 August 2021) <https://www.bbc.com/news/technology-58317106>.

420 Ibid.

421 Fiona Jenkins, 'The Ethics of Care: Valuing or Essentialising Women's Work?' in Marian Sawer, Fiona Jenkins and Karen Downing (eds), *How Gender Can Transform the Social Sciences: Innovation and Impact* (Springer International Publishing, 2020) 19 <https://doi.org/10.1007/978-3-030-43236-2_2>; Yvette Maker, *Care and Support Rights After Neoliberalism: Balancing Competing Claims Through Policy and Law* (In press 2021, Cambridge University Press).

422 Ibid.

423 Pasquale (n 6).

424 Human Rights Council, *Report of the Independent Expert on the Enjoyment of All Human Rights by Older Persons* (UN Doc A/HRC/36/48, 21 July 2017) para [46]-[49].

425 Ibid.

426 Ibid.

2.8.2 Expanding the Frame from the Individual to the Social

Most algorithmic and data-driven technology in the mental health context appears to be directed at *detection and diagnosis*,[427] which draws the focus to the individual who is identified as requiring expert intervention. This dominant framing has been challenged by some commentators who call for a shift in focus *away* from the individual, and its associated deficit-based concern with their deviation from presumed norms, toward the social networks and relational nature of distress, mental health, disability and healing.[428] Jonah Bossewitch, for example, has elaborated on the way technology can re-direct attention to networked collaboration, which could significantly improve the training and development of providers offering support to those in crisis.[429] He writes:

> Instead of focusing the diagnostic lens on the recipients of services, let's start by developing better tools to help providers enhance their skills and empathetic understanding. I am imagining contextual help, immersive simulations and distributed role plays, just-in-time learning modules that caregivers could query or have recommended to them based on an automated analysis of the helping interaction. The field could also benefit from more intentional use of networked, interactive media to engage counselors in their clinical supervision and help them collectively to improve. Did that last crisis intervention go well? What could I do differently if I encounter a similar situation again? Do any of my peers have other ideas on how I could have handled that situation better?

For those with lived experience or psychosocial disability, contemporary information communication technologies can have a strong collective dimension that can be used within social movements to create a sense of solidarity, to intervene politically and to provide a sense of belonging for groups that may have traditionally been socially and economically marginalised.[430]

There are examples of successful online initiatives that appear to boost local forms of mutual support, such as the online peer support network that emerged from the NGO *USPKenya* (discussed in case study on X). The online support group used a mainstream messenger service and was described as being 'fully community-based, operat[ing] outside Kenya's mental health system and [not linked] to any mental health institution.[431] The virtual support network involves crisis support for individual members, sharing of information about face-to-face meetups, the generation of fundraising for individual members, and so on.[432] Such informal initiatives may not make it into the public spotlight in the same way governments, large NGOs, and industry-funded initiatives do, but they may warrant resources or further research to determine how and why they are working, and how they can be supported.

The call for a focus on relationships over individual autonomy, community over individual rights, and interdependence rather than independence, can be found in much of the literature by people with first-hand experience of mental health initiatives, particularly in low- and middle-income settings.[433] This focus is also common in literature concerning

427 Gooding and Kariotis (n 43).

428 Bossewitch (n 44); Rose (n 271).

429 Bossewitch, 'Brave New Apps' (n 121).

430 Jonah S Bossewitch, 'Dangerous Gifts: Towards a New Wave of Mad Resistance' (Columbia University, 2016) <https://doi.org/10.7916/D8RJ4JFB>.

431 USP Kenya, *The Role of Peer Support in Exercising Legal Capacity* (Nairobi, 2018) 18 <http://www.uspkenya.org/wp-content/uploads/2018/01/Role-of-Peer-Support-in-Exercising-Legal-Capacity.pdf>; Transforming communities for Inclusion, Asia, *Summary Report on Transforming Communities for Inclusion - Asia: Working Towards TCI - Asia Strategy Development* (Asia-Pacific Development Centre on Disability, June 2015) <www.apcdfoundation.org/?q=system/files/TCI%20Asia%20Report_Readable%20PDF.pdf> accessed 5 May 2016.

432 Ibid.

433 See, e.g. E Kamundia, 'Choice, Support and Inclusion: Implementing Article 19 of the Convention on the Rights of Persons with Disabilities in Kenya' in *African Yearbook on Disability Rights* (Pretoria Law Press, 2013) <http://www1.chr.up.ac.za/images/files/publications/adry/adry_volume1_2013.pdf>;

the ethics of care (sometimes referred to as relational feminist ethics)[434] and other communitarian-oriented approaches to ethics and justice. A relational focus also aligns with growing research on the impact of social, political, and economic structures on people's mental health, which extends to querying whether most forms of distress should even be framed in terms of 'mental health'. As an example, consider Morgan and Kienzler's statement on the mental health impacts of COVID-19 on populations worldwide:

> To feel anxious and sad, to have trouble sleeping, to be afraid for the future – all are perfectly understandable responses to such a profound rupture in our social worlds. However, framing this distress in terms of mental health – as we have done so far, following the currently dominant narrative – is potentially problematic. This approach, at the very least implicitly, locates distress and mental health problems in individuals and, in effect, severs experiences like sadness and anxiety from the social conditions in which they arise, making them problems of psychology or even of biology.
>
> It is this narrative that underpins the predominant responses to date, which centre around calls for an expansion of individual interventions, of mental health services, and, in settings such as schools and workplaces, of myriad therapies such as mental health first aid, various forms of supportive counselling, and mindfulness. This is taken to its extreme in Amazon's recently reported mindfulness pod, a portable cubicle with space for a single worker to step out of the workplace, isolate themselves, and practice being in the moment as a means to reduce stress. Better, it seems, that workers clear their minds than reflect too much on the excessively long working hours, lack of autonomy, pitiable wages, and the Dickensian working conditions they are forced to endure to further enrich the billionaire, Jeff Bezos. By stripping suffering and distress from their social origins in this way we add insult to injury. We might, then, more usefully think about the distress that arises primarily as a consequence of poverty, precarity, violence, and trauma – including much of the distress stemming from the pandemic, social restrictions, and economic impacts – as a form of social suffering.[435]

In low- and middle-income countries, the impact will be even more stark. Manuel Capella describes how the Ecuadorian government 'set up the "telepsychology" phone line with one hand, whilst (in the middle of a pandemic) paying millions of dollars of foreign debt and approving reductions in the public budget with the other', noting that '[s]uch cuts negatively affect the well-being – including the mental health – of the vast majority of the population'.[436] In this way, digital practices in mental health have clear potential to reinforce individualistic views of mental health, which invisibilise social determinants and the importance of communities. This might be described as a capitalist instrumental view of mental health (for example, 'empowering' people to take matters into their own hands or making damaging work practices more bearable to workers rather than relying on state resources or creating fair and equitable labour conditions).

In contrast, using technology to promote holistic, human growth requires attention to the impacts of socially and economically structured disadvantage. This could even extend to querying the monopolistic, anti-competitive and surveillance-driven nature of major parts of the information economy. Regardless, current pandemic conditions have amplified historic and structural inequalities, making it even more important to consider ways to use computer technology to harness social and economic resources that individuals draw from to cope with and navigate our challenging and changing social worlds.

434 See e.g. Kittay (n 420).

435 Craig Morgan and Hanna Kienzler, 'The Pandemic as a Portal: Reimagining Society and Mental Health in the Context of COVID-19' in *Build Back Together: A Blueprint for a Better World* (School of Global Affairs, King's College London) 15 <https://www.kcl.ac.uk/the-pandemic-as-a-portal-reimagining-society-and-mental-health-in-the-context-of-covid-19>.

436 Manuel Capella, 'Corpses in the Street, Psychologist on the Phone: Telepsychology, Neoliberalism and Covid-19 in Ecuador', *Somatosphere* (15 December 2020) <http://somatosphere.net/2020/telepsychology-neoliberalism-and-covid-19-in-ecuador.html/>.

2.9 International Human Rights

> There are … potential serious negative consequences if ethical principles and human rights obligations are not prioritized by those who fund, design, regulate or use AI technologies for health.
>
> World Health Organisation[437]

Many people and organisations have supported international human rights law as a basis for regulating algorithmic and data-driven technologies.[438] Although not without critics,[439] proponents view human rights as a helpful organising framework for the design, development and use of new technologies. This includes offering factors that governments and businesses should consider in order to avoid violating human rights.[440] Lorna McGregor and colleagues list some of the many human rights engaged in the growing information economy:

> automated credit scoring can affect employment and housing rights; the increasing use of algorithms to inform decisions on access to social security potentially impacts a range of social rights; the use of algorithms to assist with identifying children at risk may impact upon family life; algorithms used to approve or reject medical intervention may affect the right to health; while algorithms used in sentencing decisions affect the right to liberty.[441]

In each of these examples, data concerning mental health may be decisive to high stakes decisions. For example, a person may be 'red-lighted' in automated credit scoring systems or in social security determinations due to data generated by mental health services or inferred based on data suggesting a person is experiencing distress. The same data might be used to assess risk ascribed to a person in relation to child protection, insurance, criminal sentencing, and so on.

Human rights violations persist against people with psychiatric diagnoses and psychosocial disabilities across low-, middle- and high-income countries. In 2019, Dainius Pūras, then UN Special Rapporteur on the Right to the Highest Attainable Physical and Mental Health, commented on the 'global failure of the status quo to address human rights violations in mental health-care systems'.[442] He argued that this failure 'reinforces a vicious cycle of discrimination, disempowerment, coercion, social exclusion and injustice', including in the very systems designed to 'help'. Pūras raised a very brief concern about impact on the right to health of expanding surveillance technologies, and warned against technologies that 'categorize an individual for commercial, political or additional surveillance purposes'.[443] He did not elaborate on the rise of algorithmic and data-driven technologies in mental health settings in general, and indeed there is little research on the human rights implications of these developments.[444]

437 orld Health Organization (n 276) xi.

438 Access Now (n 253); Lorna McGregor, Daragh Murray and Vivian Ng, 'International Human Rights Law as a Framework for Algorithmic Accountability' (2019) 68(2) *International & Comparative Law Quarterly 309; Algorithm Charter for Aotearoa New Zealand 2020*.

439 There are important critiques of a human rights approach to the issues raised by algorithmic and data-driven technologies (see eg Floridi, 2010), including critiques of its underlying liberal ideas as being ill-equipped to challenge the supremacy of private corporations over individuals in the age of Big Data. Sebastian Benthall and Jake Goldenfein, *Data Science and the Decline of Liberal Law and Ethics* (SSRN Scholarly Paper No ID 3632577, Social Science Research Network, 22 June 2020) <https://papers.ssrn.com/abstract=3632577>. It is outside the scope of this paper to enter these important debates. Instead, this report is premised on the belief that an organised public can create space to express authentic concern for individual and group rights, which can effect institutional change. This does not preclude the need to seek other organising principles, nor to engage seriously with critics of human rights.

440 McGregor, Murray and Ng (n 441).

441 Ibid.

442 Human Rights Council, 'Report of the Special Rapporteur on the Right of Everyone to the Enjoyment of the Highest Attainable Standard of Physical and Mental Health' (n 36) para 82.

443 Ibid. para 76.

444 For notable exceptions, see Cosgrove et al (n 100); Bernadette McSherry, 'Risk Assessment, Predictive Algorithms and Preventive Justice' in John Pratt and Jordan Anderson (eds), *Criminal Justice, Risk and the Revolt against Uncertainty* (Springer International Publishing, 2020) 17 <https://doi.org/10.1007/978-3-030-37948-3_2>.

In 2022, the UN Special Rapporteur for the rights of persons with disabilities, Gerard Quinn, published a thematic study on artificial intelligence and its impact on persons with disabilities.[445] The report was delivered to the 49th session of the Human Rights Council, and takes a human rights lens to the ways AI, machine learning and other algorithmic technologies can both enhance and threaten the rights of disabled people worldwide.

Some fundamental rights that are relevant here include but are not limited to: prohibition of discrimination, the right to privacy, freedom of expression, the right to health, the right to a fair trial, and the right to an effective remedy. The Convention on the Rights of Persons with Disabilities is the most relevant international human rights instrument, given its role in applying established human rights norms to the disability context, and given the strong involvement of people with first-hand experience of lived experience and psychosocial disability in its development. Relevant sections of the Convention on the Rights of Persons with Disabilities include:

- **Article 4** creates an obligation on states to 'eliminate discrimination on the basis of disability by any person, organization or private enterprise' (art. 4.1 (e)). As Special Rapporteur for the Rights of Persons with Disabilities, Gerard Quinn, notes '[t]hat certainly engages the regulatory responsibilities of Governments vis-à-vis the private sector when it comes to the development and use of artificial intelligence'.[446]
- **Article 5** prohibits disability-based discrimination.
- **Article 8** requires States to educate the private sector (developers and users of artificial intelligence), as well as the public sector and State institutions that use AI and other forms of algorithmic technology, in full collaboration with disabled people and artificial intelligence experts, on their obligation to provide reasonable accommodation. 'Reasonable accommodation' means necessary and appropriate modification and adjustments where needed in a particular case, to ensure to persons with disabilities can enjoy all human rights and fundamental freedoms on an equal basis with others.
- **Article 9** imposes an obligation on states to promote the design and development of accessible information technologies 'at an early stage' (art. 9.2 (h)). This provision, according to Quinn, 'hints at a robust responsibility of the State to appropriately incentivize and regulate the private sector'.[447]
- **Article 12** concerns equal recognition before the law. This article would be engaged, for example, by algorithmic risk assessments in criminal justice proceedings that incorporated a person's mental health history.
- **Article 14** guarantees liberty and security of the person and prohibits disability-based deprivations of liberty. As with Article 12, this provision would be engaged where actuarial risk-assessments are used in justifying and facilitating indefinite and preventative detention of particular individuals;[448] but it may also be engaged where electronic monitoring systems – whether in the criminal justice context or in 'care' services – are used in ways that amount to a deprivation of liberty.
- **Article 17** states that '(e)very person with disabilities has a right to respect for his or her physical and mental integrity on an equal basis with others', a provision that would be engaged where digital initiatives may threaten or enhance that integrity.
- **Article 19** regards 'living independently and being included in the community' and requires 'access to a range of in-home, residential and other community support services, including personal assistance necessary to support living and inclusion in the community, and to prevent isolation or segregation from the community'. It is possible that efforts to build connection via digital technologies could help to promote this right (for example, a person who needs to stay at home being assisted to connect with others online), but also

445 Human Rights Council, 'Report of the Special Rapporteur on the Rights of Persons with Disabilities' (n 10).
446 Ibid.
447 Ibid. para 37.
448 McSherry (n 447).

that efforts that inadvertently isolate or segregate might violate this right (for example, where people with less resources are only able to access online rather than face-to-face support options; or where residential facilities for persons with disabilities impose alienating surveillance technologies that cut down expert human care and support).

- **Article 22** concerns respect for privacy, and states that '[n]o person with disabilities … shall be subjected to arbitrary or unlawful interference with his or her privacy, family, home or correspondence or other types of communication…' and that '[p]ersons with disabilities have the right to the protection of the law against such interference or attacks'. Further, governments must 'protect the privacy of personal, health and rehabilitation information of persons with disabilities on an equal basis with others'.
- **Article 25** sets out the right to health, directing that States Parties shall '[r]equire health professionals to provide care of the same quality to persons with disabilities as to others, including on the basis of free and informed consent'. Article 25(d) broadens the promotion of free and informed consent to include an obligation on States Parties to raise 'awareness of the human rights, dignity, autonomy and needs of persons with disabilities through training and the promulgation of ethical standards for … health care'. Subsection (e) prohibits 'discrimination against persons with disabilities in the provision of health insurance, and life insurance where such insurance is permitted by national law, which shall be provided in a fair and reasonable manner'.

There is also a potential role for data-driven and algorithmic technologies in preventive monitoring to promote and protect human rights. Preventive monitoring of closed environments, like psychiatric wards, aged care homes, and disability residential facilities, for example, can help to promote **Articles 16** (freedom from exploitation, violence and abuse) and **33** (monitoring and implementation), as the following case study suggests.

CASE STUDY: Rights-Based Monitoring – Preventing Harmful Prescription

In 2018, Lisa Pont and colleagues developed computer software to analyze routinely collected pharmacological prescribing data in Australia to monitor medicine use in 71 residential aged care facilities.[449] The aim was to prevent prescribing errors and medication misuse. A major concern was the excessive prescription of psychiatric drugs used in forms of 'chemical restraint' or tranquilisation of aged care residents. Pont and colleagues' public data initiative was used to successfully detect high rates of psychopharmaceutical drug-use in some facilities that could not be easily explained, flagging the need for regulatory investigation.

More research is needed to identify the range of human rights engaged by algorithmic and data-driven technology in the mental health and disability context,[450] and ways that such technologies can promote and protect, rather than threaten, human rights. Disability-inclusive research is critical to realising this aim. Research that actively involves people with disability in the development, design and implementation of technology – as well as its governance – will help to ensure technology is enabling rather than further disabling.

449 Lisa G Pont et al, 'Leveraging New Information Technology to Monitor Medicine Use in 71 Residential Aged Care Facilities: Variation in Polypharmacy and Antipsychotic Use' (2018) 30(10) *International Journal for Quality in Health Care* 810.

450 Whittaker et al (n 5).

'Human rights by design' represents an emerging approach to design that ensures human rights are built into all elements of technology and AI development.[451] The Oxford Handbook on AI Ethics identifies four pillars to human rights by design:[452]

1. Design and deliberation – the systems should be designed in ways that are compatible with human rights, and should include public consultations to properly identify any human rights risks and mitigation strategies
2. Assessment, testing and evaluation – technologies should be assessed, tested and evaluated, in an ongoing manner, against human rights principles and obligations
3. Independent oversight, investigations and sanctions – there should be robust regulatory oversight agencies which can conduct investigations and impose sanctions for potential or actual breaches of human rights arising from technologies
4. Traceability, evidence and proof – systems must be designed to ensure auditability by independent oversight agencies, such as by preparing, maintaining and securely storing design documentation, testing and evaluation reports.

Human rights by design could be pursued by governments and civil society actors, including technology developers and businesses.

The UN has also developed *Guiding Principles on Business and Human Rights* ('Guiding Principles'), which are relevant here given the prominent and expanding role of the private sector in generating and processing data concerning mental health and disability.[453] For example, Principle 5 sets out a special duty of governments to protect against human rights abuses when they contract with private businesses. Advocacy group *Access Now* have drawn on this principle in calling for open government procurement, recommending that:

> When a government body seeks to acquire an AI system or components thereof, procurement should be done openly and transparently according to open procurement standards. This includes publication of the purpose of the system, goals, parameters, and other information to facilitate public understanding. Procurement should include a period for public comment, and states should reach out to potentially affected groups where relevant to ensure an opportunity to input.[454]

The International Labour Organisation has developed guidance and tool kits, as well as establishing an ILO Global Business and Disability Network,[455] which may assist businesses and others to uphold the Guiding Principles.

451 Australian Human Rights Commission, *Human rights and technology: final report* (2021), 91-92.

452 Karen Yeung, Andrew Howes and Ganna Pogrebna, 'AI Governance by Human Rights-Centred Design, Deliberation and Oversight: An End to Ethics Washing' in Markus D Dubber, Frank Pasquale and Sunit Das (eds), The Oxford Handbook of AI Ethics (Oxford University Press, 2020) 77, cited in Australian Human Rights Commission, *Human rights and technology: final report* (2021), 92.

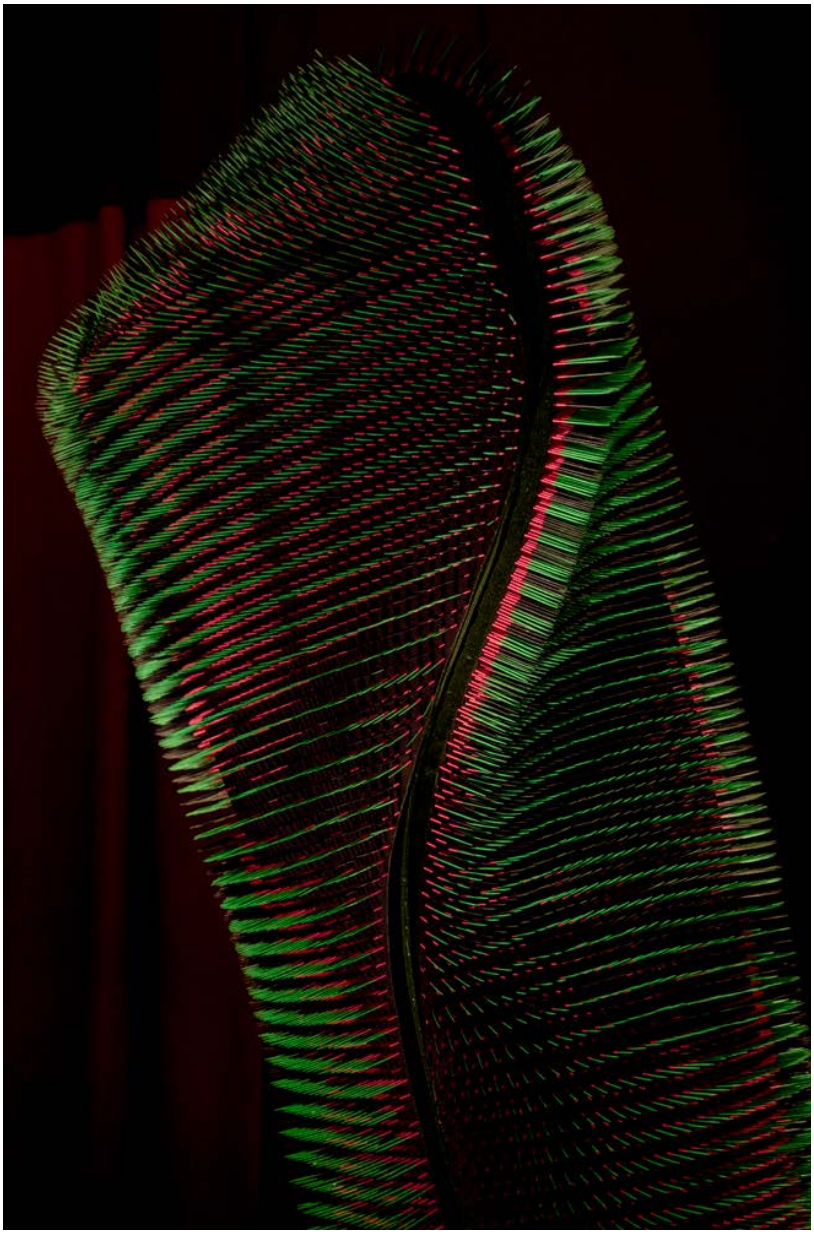
453 Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework (Guiding Principles), UN Doc. HR/PUB/11/04 (2011), available at www.ohchr. org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf. Some commentators have called for a more robust treaty affecting the private sector – rather than just guidelines – with greater enforcement measures. D Bilchitz, 'The necessity for a business and human rights treaty' (2016) 1(2) *Business and Human Rights Journal* 203-227.

454 Access Now (n 242), p.32.

455 See <www.ilo.org/global/topics/disability-and-work/> (accessed 21/12/21).

Returning to human rights more generally, the potential benefits of adopting a human rights lens include the following:

- Linking harms and benefits of algorithmic and data-driven technology to particular rights, and identifying mechanisms for redressing violations or promoting particular rights;
- Engaging with the broad global movement of disabled people that organises around the CRPD to help address harms and promote benefits arising from data concerning mental health and other disabilities;
- Strengthening calls to ensure active involvement of disabled people in laws, policies and programs that concern them, in keeping with the long-standing slogan of the global disability movement, 'Nothing about us without us' (and more generally building the power of marginalised groups); and
- Engaging with international bodies, such as the World Health Organisation, UN human rights treaty bodies, UN Special Rapporteurs, that hold influence over policies, guidelines and other regulatory frameworks at the international, regional and national levels. National human rights institutions may also prove important.



Doing Nothing with AI by Emanuel Gollob in Science Gallery Melbourne's MENTAL. Photo by Alan Weedon.

2.10 Future Efforts

This report could have explored much more. Issues requiring further research include digitised initiatives in low and middle-income countries, the growth of 'emotion or affect recognition' technologies, digital care labour platforms, automated resource allocation for social security and healthcare, and human rights by design methodologies.

Digital mental health technology in **low and middle-resource settings and countries** is growing and is often framed as allowing intervention in the lives of large populations. Emerging practices include online training for professionals and laypersons, digitisation of population-level information and electronic health records, and 'scalable' digital therapies such as automated cognitive behavioural therapy.[456] China Mills and Eva Hillberg have argued that any analysis of risks and benefits must include questioning assumptions about digital empowerment and – as is often the case in both information technology and mental health practice – of top-down imposition by high-income (and typically 'Western') countries.[457] Advocacy groups like TCI-Asia have warned against perpetuating legacies of colonialism in response to disabled people in low- and middle-income countries; instead, they promote the importance of involving affected populations and carefully considering the unique historical, social, cultural and economic factors of different settings.[458] Failed digital health interventions in low and middle-income countries suggest that technologies must be effectively localised in order to confer power to the communities they intend to serve—which raise tensions with 'scalability' as an aspiration.[459] Much more work is needed on the potential and pitfalls of using algorithmic and data-driven technology to address distress and disability in low- and middle-income countries.[460]

Computational emotion or affect recognition, was discussed briefly in this report but requires attention to its role in mental health services and research. Some speculators project market value for 'emotional AI' at $91.67 billion by 2024, and it is being deployed by tech vendors, criminal justice agencies, advertisers, car manufacturers, and others.[461] Despite the exponential growth of such technologies, a more and more research is claiming that emotion recognition technology is founded on pseudoscientific claims.[462] Indeed the traditional research on emotion recognition based on facial features has been heavily criticised as lacking evidence.[463] The impact of broader debates about affect recognition and its interaction with biometric monitoring work conducted in the mental health sciences remains uncertain and requires attention, particularly as mental health sciences risk lending a veneer of legitimacy to otherwise pseudoscientific claims.

456 John A Naslund et al, 'Digital Technology for Treating and Preventing Mental Disorders in Low-Income and Middle-Income Countries: A Narrative Review of the Literature' (2017) 4(6) *The Lancet. Psychiatry* 486. John Naslund and colleagues reviewed the clinical effectiveness of digital mental health interventions in diverse low- and middle-income countries and argued that there is reasonable evidence for their feasibility, acceptability, and initial clinical effectiveness, although they noted that most studies in the field are preliminary evaluations.

457 Mills and Hilberg (n 146).

458 Transforming Communities for Inclusion – Asia, 'Submission to the UNCRPD Monitoring Committee, Day of General Discussion, Article 19' <http://www.ohchr.org> (accessed 6 February 2018).

459 Varoon Mathur, Saptarshi Purkayastha and Judy Wawira Gichoya, 'Artificial Intelligence for Global Health: Learning From a Decade of Digital Transformation in Health Care' [2020] *arXiv:2005.12378* [cs] <http://arxiv.org/abs/2005.12378>.

460 See also, Capella (n 439).

461 Evan Selinger, 'A.I. Can't Detect Our Emotions', *OneZero* (6 April 2021) <https://onezero.medium.com/a-i-cant-detect-our-emotions-3c1f6fce2539>.

462 Article 19 (n 327) 19.

463 For overview of criticisms, see: Barrett (n 331) 12–24.

Digital labour platforms are transforming care labour for people with disabilities who access home-based or community-based support, and other forms of support. This has been called the 'Uberisation' of care and therapy. Digital labour platforms can be designed in ways that institutionalise the exploitation of care and support labour, turning the interests of recipients against (generally low-paid) staff.[464] Even as such platforms might equitably distribute care and support labour, they raise legitimate concerns, particularly in relation to health and safety, insurance, unpaid work and the long-term training needs of the workforce. These platforms will raise issues of platform regulation, but also accreditation and professional regulation, given therapy platform business models often depend on shrinking payment to therapists and increasing their caseloads, and endeavouring to de-medicalise therapy to hire cheaper, non-accredited counsellors.[465]

Algorithmic resource allocation for determining who receives state resources for healthcare or other forms of social security, including disability-based and mental healthcare support, is an area with serious implications. Lucy Series and Luke Clements reported on automated 'personal budget decisions' in the UK, suggest algorithmic resource allocation could be used as a mechanism for implementing spending cuts.[466] Further, the budget allocations did not always respond to people's needs, and the algorithmic nature of the system led to a lack of transparency. Advocacy organisation AlgorithmWatch cited the research, stating that it 'serves not only to illustrate how flawed [automated] decisions can adversely impact people's lives, but also how [automated decision-making] systems might be scrutinised and what obstacles are sure to arise in other domains of [automated decision-making] accountability research.'[467]

More broadly, research is required on the impact of the politics and ideology that shape the administration of mental health services, and its merging with the politics and ideology that drive the information economy.

Photo by Note Thanun on Unsplash.

464 International Labour Office, *World Employment and Social Outlook 2021: The Role of Digital Labour Platforms in Transforming the World of Work* (ILO, 2021).

465 See generally Zeavin (n 47) pp.205-215.

466 Lucy Series and Luke Clements, 'Putting the Cart before the Horse: Resource Allocation Systems and Community Care' (2013) 35(2) *Journal of Social Welfare and Family Law* 207.

467 Automating Society 2019', *AlgorithmWatch* <https://algorithmwatch.org/en/automating-society-2019/> p.171.

Conclusion

> "[W]e need to understand both the cool and the creepy of tech."
>
> – Keris Jän Myrick[468]

There is cause for both optimism and pessimism in the application of algorithmic and data-driven technologies to assist people in extreme distress and crises, and to boost individual and collective opportunities for crisis support and flourishing. Vigilance is required to promote benefit and prevent harm, which won't be possible without acknowledging the vast social inequalities and profit motives that are shaping technological development in this area. As populations reckon with new digital responses to age-old experiences of distress, anguish and disability, optimism comes from these technologies and their benefits being publicly controlled, genuinely shared, and firmly shaped by those most affected.

The development of robust data governance frameworks and a rich politics concerning the use of technology in care and crisis responses won't be possible without inclusive public engagement, enforceable policies, and global cooperation. This can only be achieved with the assertion of collective claims over data, and acknowledgement of mental health as 'indelibly connected to systems of technology, money and power'.[469] Sharing benefits requires that the public and social value of data is directed toward the determinants of human flourishing and good mental health: equitable economic development, directing support where it is needed most, addressing discriminatory practices and histories of exclusion and marginalisation, improving the quality of care and service provision, and other measures known to boost societal wellbeing.

468 Cited in Khye Tucker, 'California Is Testing a New Mental Health Digital "Fire Alarm"', *Syneos Health Communications* (2 July 2009) <https://syneoshealthcommunications.com/blog/california-is-testing-a-new-mental-health-digital-fire-alarm>.

469 Vanessa Bartlett, 'Psychosocial Curating: A Theory and Practice of Exhibition-Making at the Intersection between Health and Aesthetics' (2020) 46(4) *Medical Humanities* 417.

Kind Words by Ziba Scott in Science Gallery Melbourne's MENTAL. Photo by Alan Weedon.

Contact Details

Piers Gooding
p.gooding@unimelb.edu.au
@p_gooding