

Towards Socio-Technical Theories for Digital Health

Malcolm Fisk

University of Central Lancashire

Historically, social and technological matters have been regarded as operating in different domains. Consequently, issues about the way that we shape, or are shaped by, technologies have often been overlooked. It follows that, when we consider our increasingly technological, networked and data-rich environment, this separation of domains is no longer tenable. This is especially so for health and care – a domain (or sector) for which there are both technological opportunities and threats.

For our field of digital health, given the historical backdrop (or at least that which prevailed before the turn of the millennium), it may be regarded as unsurprising that very few philosophers or social theorists have, to date, addressed the tricky territory at the interface of the three areas of (a) people and their behaviours; (b) health and care services; and (c) the technologies used in these contexts. This presentation takes, therefore, steps to shape a framework for socio-technical theory that can apply.

After appraising the context and noting key (notably technological) developments that are impacting on health and care, an outline framework is set out. It builds on the limited prior work – with that of undertaken by Deborah Lupton and (though non-technical) Joan Tronto being highlighted. The framework will be set out more fully in the presenter's book (focused on telehealth) to be published by Palgrave Macmillan late in 2023.

Highlights and challenges from public involvement in human-centred medical robotics research

Antonia Tzemanaki

University of Bristol

**Understanding Gaps in Standards, Regulations and Training Needs for
Healthcare Professionals working alongside Physically Assistive Robots**

Praminda Caleb-Solly

University of Nottingham

Autonomies-in-Interaction:

Relational Understandings of Caring Technologies

Darren Reed

University of York

How the Perfect Robot Fails:

Systemic issues in the integration of robots for health-social care

Stevienna de Saille, David Cameron

University of Sheffield

At present, millions of pounds are being poured into research on robots to ameliorate the shortage of workers for social care for health-related issues. As a field of academic enquiry, questions tend to fall into two categories: how to make the robots “trustworthy” (ie. reliable, safe, effective) and how to make them desirable (ie. functions, aesthetics, cost).

What is not generally asked is “How might X robot fit into the health-social care system we presently have (and what changes might be needed)?” This paper addresses that question using data from our recently-concluded project “Imagining Robotic Care: Identifying conflict and confluence in stakeholder imaginaries of autonomous care systems”, funded by the UKRI-Trustworthy

Autonomous Systems hub. The project uses a mix of policy review, targeted expert interviews and focus group data gathered using LEGO Serious Play to investigate the 'socio-technical imaginaries', ie. collectively achieved, systemic visions of social transformation through technology (Jasanoff & Kim 2009), held by diverse stakeholders in the UK care ecosystem with regard to the use of robots to fill social care needs arising from health issues. In this paper we discuss some preliminary findings arising from a cross comparison of themes generated within each of the focus groups, which revealed inter-alia a fundamental conflict not so much between stakeholder imaginaries, but between direct experience of the health-social care ecosystem (whether as professionals, informal carers or care users), and the policy-level sociotechnical imaginary of robotics and AI that can solve critical problems within the social care system as it now stands.

Methods

The focus groups were small (4-7 people), each based on specific expertise. These were: roboticists, HRI academics, care and disability academics, care administrators and council staff, social workers, paid carers, care users and general publics (as future stakeholders) for a total of 97 people in 15 groups. These were carried out online via zoom in three hour sessions, with specialised kits for LEGO® Serious Play® (LSP) being sent to participants in advance. LSP was chosen for its capacity to surface tacit knowledge through rich, situated narratives (Rasmussen 2006) and is increasingly used as a research tool. The LSP methodology is simple: in each iteration a question is posed, a model is built, each participant shares their model's story, then the facilitator guides a discussion to deepen the response.

Preliminary findings

The central task of the focus group asked participants to tell a story about a robot giving care to someone. These narratives could be either positive or negative, set anywhere and involve anyone; deliberately, neither 'care' nor

'robot' were predefined. They were then asked to reverse their scenario, which allowed the key points of the imaginary to be identified and further explored.

Although the scenarios varied widely, certain expected tasks occurred more frequently across the groups, such as transferring people from bed to wheelchair to facilities for personal care. In this paper we discuss findings that reveal major conflicts when these imaginaries are compared to the those of the policy documents, where robotic solutions are habitually imagined in isolation from the effects of other policies, such as continual erosion of council budgets and cuts to disability benefits, and deliberately ignore limitations and dysfunctions in the existing system such as out of date IT, incompatible time and task models, and privatisation and fragmentation of services. In short, our findings suggest that even the most technically perfect, trustworthy and desirable robot cannot be productively deployed within the health-social care system as it now stands. The results will have implications for health and social care policymakers, council commissioning services, roboticists and care academics, as well as for the range of publics we can expect to eventually encounter robots as part of their care package.

References

Jasanoff, S., and S.-H. Kim. (2009). Containing the Atom: Sociotechnical Imaginaries and Nuclear Power in the United States and South Korea. *Minerva* 47 (2): 119-46.

Rasmussen, R. (2006). When You Build in the World, You Build in Your Mind. *Design Management Review*, 17(3), 56–63. <https://doi.org/10.1111/j.1948-7169.2006.tb00053.x>

AI and Health Equity in and Through Primary Care:

A sociotechnical analysis and stakeholder exploration of how to make artificial intelligence a force for health equity in English primary care

Background

Artificial Intelligence (AI)-augmented interventions are currently being rolled out across primary care, but the sociotechnical theory for deploying AI remain relatively neglected [1], and the applied literature focuses on e.g. performance metrics and more recently, formal methods for fairness and obliquely, health equity (HE) by minimising algorithmic bias [2]. However, applying AI in healthcare will affect HE beyond algorithmic bias, through interactions with existing societal health inequities [3]. There is need to understand how the ecosystem in which AI is being implemented can be made to benefit HE through AI.

Aims

To map the ecosystem involved in the implementation of AI in English primary care, and from a sociotechnical perspective assess how this network of actors can be make conducive for improving HE.

Methods

A systematic scoping review followed by a ethnographical inquiry based on 32 interviews with stakeholders including commissioners, decision makers, AI developers, researchers, GPs and patient groups, and complemented by an analysis of UK primary care data to assess the risk of algorithmic bias in big-data applications such as AI systems.

Results

1. AI is likely to impact HE in primary care through a multitude of mechanisms, including both those intrinsic to the AI systems (e.g. algorithmic bias) and wider system- and societal impact (e.g. digital exclusion and enabling privatisation and commercialisation of care provision).

2. Regulation and policy cannot do guarantee equitable implementation of AI, but needs to provide a baseline framework to enable other stakeholders to work equity-promoting: a shared understanding of the causal mechanisms of AI and HE, how to measure HE, and how to share necessary data.
3. All stakeholders needs to be on board for implementation success with regards to above. Currently, innovation typically leaves clinicians and patients behind.
4. Capacity building is needed to enable addressing the above, in particular on commissioning and clinician level.
5. Whilst true for most innovations, the difference with AI the pace of innovation. Previous waves of innovations have happened at a more gradual pace, allowing for a more controlled implementation.

Conclusions

AI in primary care holds great potential. However, if the current implementation is to benefit the health of everyone, careful consideration is needed on the sociotechnical context in which the process is taking place.

Whilst this project is not the first to cover the effects of AI on HE, and the rapid development of AI and related research meant that considerable scholarship has been produced during the during of this project, this thesis with the included studies carves out a niche against the preceding research. Namely, in contrast to preceding works, this thesis takes a systematic, empirical approach specifically focused on the implementation setting that is English NHS primary care, and as such produces an empirically grounded set of recommendations for how AI can be implemented in an equitable manner, as well as sets a methodological precedent on how complex interventions can be assessed “prospectively” from a sociotechnical perspective.

References

1. Mhasawade, V., Zhao, Y. & Chunara, R. Machine learning and algorithmic fairness in public and population health. *Nat Mach Intell* 3, 659–666 (2021). <https://doi.org/10.1038/s42256-021-00373-4>
2. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring Fairness in Machine Learning to Advance Health Equity. *Ann Intern Med*. 2018 Dec 18;169(12):866-872. doi: 10.7326/M18-1990. Epub 2018 Dec 4. PMID: 30508424; PMCID: PMC6594166.
3. d'Elia A, Gabbay M, Rodgers S, et al. Artificial intelligence and health inequities in primary care: a systematic scoping review and framework. *Family Medicine and Community Health* 2022;10
4. Celi LA, Cellini J, Charpignon ML, Dee EC, Dernoncourt F, et al. (2022) Sources of bias in artificial intelligence that perpetuate healthcare disparities—A global review. *PLOS Digital Health* 1(3): e0000022. <https://doi.org/10.1371/journal.pdig.0000022>

Artificial intelligence (AI) and public (dis-)trust?

A case study on using AI in mental health services

Caroline Jones, James Thornton, Age Chapman, Rose Worley, Chloe Harrison, J.C.Wyatt, Donna Chaves

Swansea University

Background

There is huge scope for AI to support overburdened mental health services through effective triage, assessment, engagement and maintenance support. However, there are equally substantial concerns about trust and trustworthiness in the uptake and use of such tools – by clinicians and service-users – which mean potential improvements in care risk being lost.

Indeed, the very concept of ‘trust’ itself is unclear. Relevant policy documents refer to the importance of ‘public trust’ and ‘trustworthiness’ as issues (eg House of Lords Select Committee on AI Report (2018)); but, there is less consideration

of what these terms mean, or little clarity over whether they are being used consistently to mean the same thing. Furthermore, with some notable exceptions, patients' perspectives are often assumed, or ignored altogether. In partnership with the Welsh mental health charity, Adferiad Recovery, our project seeks to address these knowledge gaps.

Methodology

This is a mixed-methods participatory research project. We have engaged with Adferiad Recovery's service-users and therapists to identify key themes for exploration – in the next phase of the project – via surveys and semi-structured interviews. Our aims are to:

- Investigate service-users and therapists' concerns over the use of AI systems in the context of mental health services and support, particularly in relation to trust and trustworthiness.
- Identify the foundational characteristics of trustworthy AI systems, and a hierarchy of importance of relevant factors (for these participants).
- Utilise these findings to draft and disseminate a set of co-produced (with research participants) best practice guidelines for the design, regulation and operation of trustworthy and trusted AI systems in mental health support.

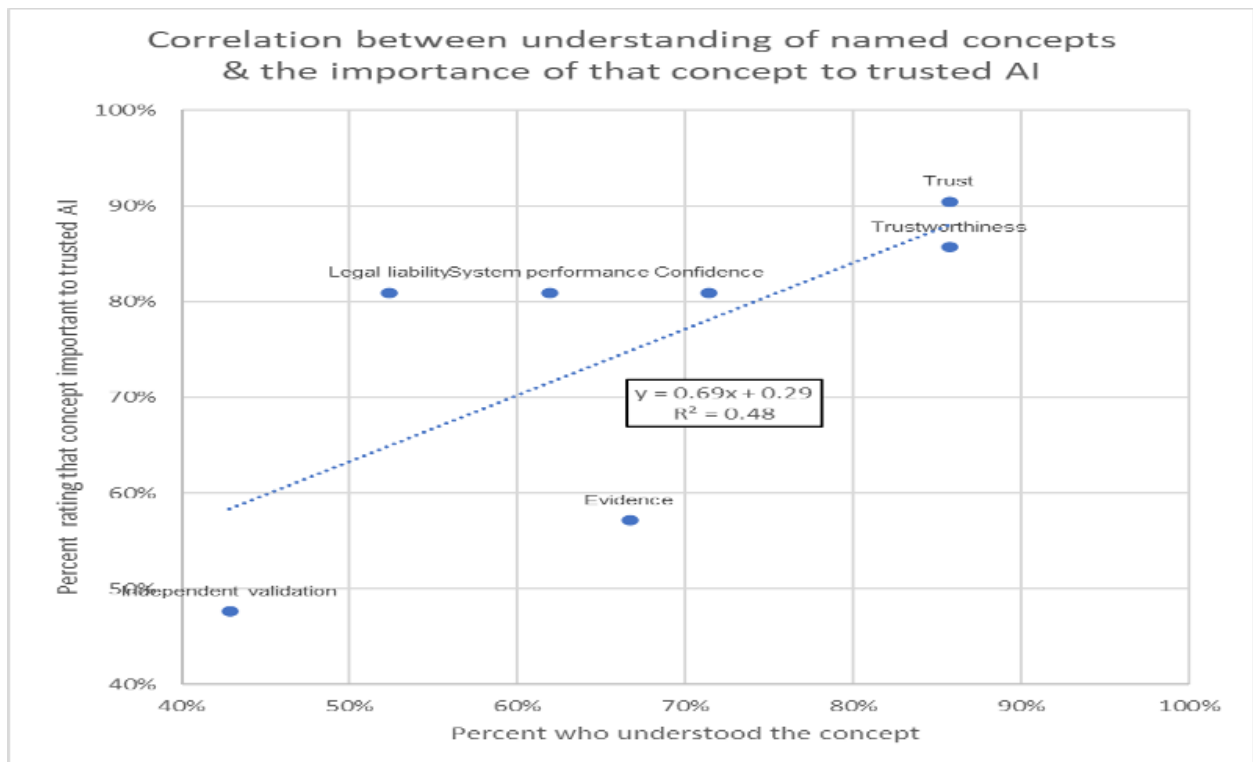
Interviews will be recorded, transcribed and thematically analysed using NVIVO software, as will the open/free text survey responses. Closed question survey responses will be analysed quantitatively, using MS Excel.

Summary of main findings

The initial consultation phase was developed to agree a list of trust factors that might be relevant to service-users and therapists when deciding whether to trust an AI system intended to support mental health. 25 responses were received: 24 service-users (3 null forms); and 1 therapist. Of the completed forms, the following results were found:

Trust factor	Importance (%)	Importance Score	Comprehension (%)	Comprehension Score
Trust	80.95	93	76.19	95
Trustworthiness	76.19	94	66.67	93
Confidence	42.86	87	47.62	82
System performance	57.14	92	38.10	80
Independent validation	28.57	77	14.29	66
Legal liability	52.38	89	28.57	74
Evidence	28.57	71	42.86	77

*Importance and Comprehension scores were calculated by summing the values (0 - 5) assigned to each trust factor for each participant (n = 21)



We found a large variation in the proportion of people who understood a concept, and in the rated importance of these concepts. The lowest rated was “independent validation” and the highest were “trust” and “trustworthiness”. There was a positive correlation between importance and comprehension; but “legal liability”, “system performance” and “confidence” were all given the same amount of importance, but varied in their understanding from 53% to 72% – so it is not a perfect correlation.

As a result of these initial findings the definitions provided were re-worked to improve comprehensibility (Flesch reading ease; Flesch-Kincaid grade level); the accompanying invitation to participate was made more user-friendly; and further piloting was undertaken.

Discussion

The meanings attached to phrases such as ‘public trust’ matter, yet it is all too often unclear what is meant – precisely – by such terms. Some stakeholders appear to equate ‘trust’ with empirical correctness (we trust something/someone if it/they give the ‘right’ answer), whereas others appear to have a more normative meaning in mind (we trust something/someone if it/they

can be relied upon to stick to commitments or act in good faith). Such understanding is often implied rather than made explicit. When organisations do address these problems of definition, sometimes the word ‘trust’ is abandoned altogether; e.g. the NHS AI lab preferred to speak in terms of ‘confidence’, rather than ‘trust’.

Further, perspectives of what patients/service-users understand ‘trust’ to mean, and what they might consider important for assessing it, are often assumed or ignored altogether. Our study therefore makes a significant contribution to medical sociology by elucidating their perspectives on these issues; including best-practice guides for the design, regulation and operation of trustworthy and trusted AI systems in mental health support.

Remote monitoring for diabetes mellitus disease:

A Case Study in Turkish Healthcare System

Ceren Yavuz, Burçin Taneri, İdil Kula

Middle East Technical University

The advent of information and communication systems unlocked a new era of smart haptic devices in healthcare. Almost half a century, the acceleration of digitization in healthcare systems and infrastructures enabled discussions of remote and smart monitoring systems for diagnosis, treatment and prognosis. Early thinkers and proponents of remote healthcare monitoring systems come up with enthusiastic arguments on the benefits of these systems. However, building and maintaining responsible, fair and fully-fledged systems of remote healthcare monitoring requires a serious analytical examination. Field research and critical analysis is needed to adequately position remote healthcare monitoring systems.

In order to make a preliminary analysis towards the vast possibilities of remote patient monitoring (RPM) systems in healthcare, we narrowed down our choices to diabetes mellitus disease in Turkey. Portraying a regional

perspective on RPM systems, we stepped down to the field and conducted qualitative research followed with an evaluation of data collected. Our study is constructed on three different pillars of data sources. Firstly, we benefited from publicly available data on statistics databases of state institutions, academic databases such as Web of Science and Scopus and disclosures of companies in the field. While our second data source is semi-structured interviews, the final source stands on primary data gathered from questionnaires.

We conducted 10 interviews and an online questionnaire with stakeholders in the field. Our findings are categorized as economic, legal, social, organizational, and technological. With an aim to make a projection on these five dimensions, we analyzed, evaluated and discussed our findings. Casting upon it, we conducted a technology assessment that is specified for the case of diabetes mellitus disease in Turkey. Following our discussion on possible futures into adoption of RPM systems for diabetes mellitus disease in Turkey, we came up with a roadmapping study on top of our technology assessment investigation.