The influence of artificial intelligence within health-related risk work

a critical framework and lines of empirical inquiry

Brown, P. ; van Voorst, R.

DOI
10.1080/13698575.2024.2412374

Publication date
2024

Document Version
Final published version

Published in
Health, Risk & Society

Health, Risk & Society, 2024
Vol. 26, Nos. 7–8, 301–316, https://doi.org/10.1080/13698575.2024.2412374

RESEARCH ARTICLE

The influence of artificial intelligence within health-related risk work: a critical framework and lines of empirical inquiry

Patrick Brown and Roanne van Voorst*

AISSR, University of Amsterdam, The Netherlands

Abstract

In this editorial we highlight the need for empirical studies into the growing use of artificial intelligence (AI) technology in healthcare and social work settings, especially studies which are theoretically informed by critical social science studies of risk and uncertainty. In setting out the importance of interpretative and critical traditions for research into such AI-oriented forms of risk work, we propose three important conceptual lines of inquiry which empirical studies might follow. First, we sketch ways in which the enactment of AI in healthcare work may be changing how risk is handled amid professional decision-making, and creating new categories of patient/service-user. Patients may be evaluated as being at lower or higher risk depending, respectively, upon their engagement or non-engagement with AI-technologies. These questions of (non-)engagement lead us to consider, second, the trust and distrust dynamics around AI-technologies, exploring the potential inequalities that can emerge as a result of (non) engagement. We then consider drivers of this technological embrace in terms of hope and magical thinking in technological-imaginaries, connecting these cultural tendencies to broader structures of ideology and political-economic interests. We conclude this editorial with a plea to social scientists to be cautious to avoid both techno-optimistic narratives and alarmist warnings regarding the implications of artificial intelligence (AI). Instead, we argue that our focus should be a theoretically informed and detailed examining of how expectations (pertaining to risk, trust, and hope) materialise in practice, particularly in the daily experiences of those who develop and enact AI technologies in care settings.

Keywords: Artificial intelligence; healthcare; hope; imaginaries; risk; trust

Introduction

During the process of writing this editorial, the Secretary General of the United Nations António Guterres made the following evaluative statement and plea:

> Artificial Intelligence is being deployed with few guardrails & little caution. Governments, industry, academia & civil society must develop rules & guidelines for AI safety – together & before it is too late. We have no time to waste. (Guterres, Twitter/X post, June 12, 2024)

*Corresponding author. Email: r.s.vanvoorst@uva.nl

Guterres's approach through these words reflects our two central aims in developing this editorial, which similarly involve a plea and various evaluations. First, we seek to highlight a pressing need for more critical social science research into the multiple uses of Artificial Intelligence (AI) as these are applied to contexts of health more broadly, and health care and social work in particular. In doing so we seek to draw attention to a common set of contextual features – not least current resource constraints and anticipated health crises (Gardner, 2023) – which encourage an embracing of AI, and related algorithm and machine learning technologies to assist decision-making amid uncertainty, without a sufficient exploration into the consequences of this 'biotechnical embrace' (to use the concept of Delvecchio Good, 2001). Our second aim is to set out some specific lines of empirical inquiry, together with some conceptual and theoretical underpinnings, along which we hope that critical approaches can be developed.

With the first aim in mind (the call for more critical research into AI in care contexts), *Health, Risk & Society* reflects two important traditions in social science research into risk and uncertainty. Alongside more applied and *analytic* work, as reflected more in other journals in risk studies, this journal has emphasised more *interpretative* and *critical* traditions (as these three traditions – analytic, interpretative and critical – are laid out by Habermas, 1972). Interpretative studies of risk, uncertainty and technology consider the everyday meaning-making and imaginaries around various risk-related technologies, as these shape the (dis)engagement of professionals and publics with these technologies and their wider (often unforeseen) effects pertaining to risk and uncertainty (see for example, Hallowell, 2006; Hallowell et al., 2022). This interpretative tradition is often connected to a more critical tradition which involves illuminating the underlying assumptions and 'illusions' (Habermas, 1972, p. 236) that scientists and social scientists may hold in discussing and engaging with risk-related technology (e.g. Durocher, 2024; Mythen & Weston, 2023). In turn these taken-for-granted notions may need to be 'unlearned' (Habermas, 1987, p. 400; Hallowell et al., 2005).

In pursuing our second aim (to sketch out ideas for future research), we draw on these interpretative and critical traditions in risk and uncertainty research in setting out three potential lines of inquiry, along which these traditions can be further developed around and applied to the study of AI. More specifically, in this editorial we will explore the influence of AI upon everyday professional work within healthcare and social work contexts, paying particular attention to how this influence may be reconfiguring what and who are deemed or categorised as normal (and healthy) or risky (and potentially unhealthy). Building upon this first line, we then move to consider the forms of trust around AI-related technologies which are often assumed and expected within these reconfigured care contexts. Important questions emerge here about who decides to introduce a technology and which other groups (professionals, clients, patients) are expected or obliged to also engage with the technology (Hallowell et al., 2022). Additionally, we need to critically consider who should assume the role of evaluator and human overseer of AI, and whether this role is truly effective in ensuring an ethical application of the technology. In other words: can and should we trust, that human-nonhuman collaboration leads to ethical AI enactments, as is now often assumed in legal frameworks(van Voorst, 2024a)? Recent studies, both within healthcare settings (c.f Carboni et al., 2024; Ghassemi et al., 2021; Tschandl et al., 2020) and in other fields (Peeters & Widlak, 2023; Siles et al., 2019) suggest this is currently not always the case, due to challenges regarding workload (Pols, 2012) and problems concerning the reliability, clarity and transparency of AI ouputs (Ghassemi et al., 2021; Pasquale, 2015).

Assumptions and particular configurations of organisational trust around AI-technology, alongside distrust, may create socio-technical and moral environments where distrustful orientations towards AI technology and/or its applications held by some professionals and patients come to form new and important risk-categories of people within healthcare organisations. In turn, this second line of inquiry into (dis) trust can be located in relation to a third set of considerations regarding socio-technical imaginaries (Jasanoff & Kim, 2015) and hopes, alongside the frustrated and ambivalent experiences of medical staff working with AI (Hoeyer, 2023), or the suffering which can often be obscured amid these hopes (Hallowell, 2006, pp. 11–12). Following Delvecchio Good (2001), we will seek to draw attention to various underlying political-economic structures and interests which, in turn, configure, warp or impede the framings, assumptions and forms of affect which underpin these imaginaries and hopes (Lupton, 2014).

We will then conclude this editorial with a plea to social scientists to be cautious not to get *swept up* in either techno-optimistic narratives or in alarmist warnings regarding the implications of artificial intelligence (AI). Instead, we argue that our focus should be on examining how expectations (pertaining to risk, trust, and hope) materialise in practice, particularly in the daily experiences of those who develop and enact AI technologies in care settings. Such empirical investigations may often reveal ethical dilemmas and problems that we may currently fear, but it can also uncover surprisingly reassuring outcomes. For instance, Plájas (2024) conducted anthropological research within an AI lab and observed that the employees there approached AI development with considerably more caution and deliberation than prevailing narratives would suggest. This more nuanced understanding highlights the importance of grounding our critical discussions in empirical data reflecting real-world practices and experiences of AI technology, uncertainty and risk, rather than solely relying on overarching theories or media portrayals and narratives.

The influence of AI upon everyday professional work within healthcare contexts: reconfiguring what/who are normal/risky

The rapid advancement of AI in various care contexts has sparked an intense focus on its potential benefits and risks. Scholars have extensively examined the direct consequences of AI on health care, such as issues related to race and inequality. These authors have shown that, while governments worldwide present big data and AI systems as inevitable for the future of health, this narrative typically overlooks the social biases and limitations inherent in AI systems. Research has shown that datasets in AI reproduce vulnerabilities, discriminate against certain groups, and limit personal autonomy (Eubanks, 2018; O'Neil, 2016; Osoba et al., 2017; Passchier, 2021). In response, public and scholarly concerns about algorithmic ethics have grown, and governments worldwide have developed numerous ethical frameworks and guidelines for national health institutions. These frameworks assume that AI can be designed in an ethical way, namely in such a way that human agents can override algorithmic decisions that lead to undesirable outcomes. Scholars also consider 'keeping humans in the loop' as crucial to ensure individual justice (Sleeman & Gilhooly, 2023; Zarsky, 2012).

However, these policy frameworks are problematically insufficient to guarantee an ethical outcome of AI utilisation in daily healthcare practice. Firstly because the evidence that humans indeed do intervene or resist AI, is very thin (Hannah-Moffat, 2013; Monahan & Skeem, 2016; Peeters, 2020). Healthcare and social work professionals

may not understand the workings of AI, may be subconsciously influenced by its perceived objectivity, may not feel they have the authority to challenge recommendations, may not know how to raise concerns, or may simply be under too much time (or other) pressure to constantly double-check the outcomes of AI (van Voorst, 2024a). As we explore later, because tendencies to pursue solutions through AI are more likely in contexts where managers and professionals experience heightened vulnerability, then these workers' capacity for reflexivity and resistance is diminished as a result, while their willingness or need to trust is heightened (Brown, 2021), especially amid discourses where 'digital artefacts and infrastructures have been presented as urgent and necessary' (Pickersgill, 2019, p. 16) and where risk-related decision-making is very challenging.

Secondly and most relevant for this editorial, there are far-reaching societal impacts of AI that are less immediately evident and – presumably for that reason – have received less academic attention. For example, a recent study showed that the introduction of new AI-systems used in radiology changed clinicians' perception of what was a 'tolerable' waiting time for an accurate diagnosis. This transformed professionals' sense of moral responsibility towards patients, leading to a sense that they should work faster – even if that entailed tolerating more uncertainty, and taking more risks, in diagnosing and related treatment (Van Voorst, Forthcoming in 2025).

A third important concern whereby AI reconfigures risk and uncertainty in clinical contexts, is the emergence of new perceived risks and related risk categories. International research on the visions of stakeholders regarding the future of AI in health care has revealed that many of the healthcare professionals who currently already work with AI believe that, in the near future, patients who do not utilise digital tools will themselves come to be deemed a health risk (Ashuri & Van Voorst forthcoming). This perception is linked to the increasing use of preventive and personalised medicine, which involves patients receiving regular or constant guidance on how to maintain optimal health through physical activity, medication, and diet (Boyd & Crawford, 2012; Ceyhan, 2012; Dujardin, 2020). In a study conducted by van Voorst, when asked about their 'future patients', clinicians again and again referred to people who would be unwilling to follow the trend of increased technology. In the words of one cardiologist: 'If we can't monitor them, we also cannot advise them on how to do better. So, naturally their health will be worse off than that of others who do use technology- not providing us with their health data is their free choice, of course, but do be aware that it is going to be these digital refusers that will occupy hospital beds, and cost the state [more] in the future.' That many of these 'digital refusers' might not be able to utilise tools, because they are not digitally literate or, for example, due to older age, are less familiar with it and therefore less trusting of it, remained unacknowledged by the professional in this interview. Indeed most interviewees in the study showed little empathy for other reasons patients might have for not wanting to be traced or diagnosed by AI (Ashuri & van Voorst, Forthcoming in 2025).

A similar shift in perspectives on what and who is risky was shared by stakeholders discussing health incentive applications in another research project. These are self-tracking devices that aim to keep people healthy and productive, for example through digital, quantifiable measurements of movements, like biking or walking. As the word 'incentive' indicates, users of health incentive apps are invited to utilise them in return for presents, gifts, vouchers, discounts or direct financial rewards, and increasingly these are paid for by healthcare insurance companies, employers, and governments. Both scholars and media have expressed fears around privacy, surveillance, inequality and

social exclusion: who has access to users' health data, and with what consequences (MacEachen 2000; Zuboff 2019; Couldry & Meijas 2019; Meskó et al. 2017; Boulos et al. 2014; Lupton 2014)? These are important concerns, but here we want to point to the shift in risk perceptions that can occur after the introduction of such a technology. Research on health incentive apps on workfloors in the UK and the Netherlands found that this technology led to loneliness among, and marginalisation of, employees who, for various reasons, did not feel comfortable using the applications. They were considered as either 'unhealthy' by their colleagues and managers, or as uninterested in improving healthcare (Van Voorst, forthcoming). Such negative effects, or stigmatising interactions, where AI technologies are not embraced, raises important micro-level interactionist questions around the pressures that some people may experience towards using technologies despite misgivings – as 'forced options' (Barbalet, 2009; Delvecchio Good, 2001). Wider questions are apparent here regarding the unequal patterning of such negative effects, and how existing inequalities may deepen amid processes by which marginalised groups may be more likely to refuse particular technologies; through logics informed by the negative histories of these groups with technologies, healthcare institutions and the state (Benjamin, 2016). It is to these questions concerning (dis)engagement, (dis)trust and forced options that we now turn.

Exploring forms of trust in AI that are expected and required

The logics, drivers, and processes behind the application of AI in specific healthcare settings require empirical investigation. These settings may vary within particular local organisational contexts or medical specialities. Key factors influencing AI adoption may include 1) The demand for technology due to existing problems and shortages, 2) The identification of a specific type of care context as a potentially profitable market for existing or emerging technologies, 3) The support of an influential advocate, such as a senior healthcare professional or manager. These three factors can exist independently or in combination with one another.

Engagement with science and technology is a complex process which unfolds over different stages and to varying depths (O'Brien & Toms, 2008). Processes of (dis)trust are fundamental to analysing (dis)engagement over time, whereby deeper forms of engagement will only be sustained amid trust relations (P. Brown & Bahri, 2019). In turn, trust relations involve a trustee who forms the basis upon which positive expectations can be inferred, enabling the truster to act 'as if' the future was known (Lewis and Weigert 1985). In healthcare contexts, such trust processes usually involve the development of understandings about specific individual-trustees (e.g. a healthcare professional), alongside understandings of the abstract systems (Giddens, 1990) in which these trustees are embedded (organisational; epistemic; professional; ethics-related; legal). Trust is facilitated by various taken-for-granted assumptions about these systems as well as more explicit reflections on positive outcomes, competence and care amid past experiences of systems and individuals.

Vulnerability is also fundamental to processes of engagement with and trust amid technologies such as AI, in that it is only when one is vulnerable that one needs to trust in the first place (Möllering, 2006). Amid existing vulnerabilities, trusting can form a means of pursuing a solution, as a means of coping, but the process of trusting itself also makes the truster newly vulnerable; to being let down by the trustee. This creates something of a complex relationship between trust and vulnerability which we explore here in an example of the use of algorithmic technologies in child protection social work in England.

Child protection social workers can be understood as a particularly vulnerable group given the nature of the risk work that they are tasked with. Social workers are required to assess risk in contexts where they have limited access to and time with parents (Veltkamp & Brown, 2017), where the base rate of violence is very low and thus where the likelihood of false negatives and false positives are high (Szmukler, 2003), where the stakes of child wellbeing and mortality have become highly politicised and emotive within the UK context (Warner, 2015), and where the local authorities/municipalities who fund this form of social work have had their budgets cut repeatedly and drastically since 2010. It is in this context of heightened vulnerability – of organisations and individual social workers – to mistakes, blame and insufficient resources, that some municipalities have looked to algorithm-driven 'predictive analytics' developed either by the municipality or through private contracts (McIntyre & Pegg, 2018).

While these tools did not involve artificial intelligence as such, they nevertheless show the ways in which algorithm-based technologies can be embraced by organisations amid heightened vulnerabilities. In this context, political-economic factors of the system (resource constraints, accountability demands, heightened politicisation of and accountability for outcomes amid media scrutiny) configure an organisational context in which vulnerability makes engagement with technology more likely, due to the need for a solution. Here, we follow Barbalet (2009), p. 372) in understanding trust as a 'forced option, a situation in which either A trusts B to achieve C, or A cannot have C', whereby under heightened resource constraints and accountability pressures, technologies are pragmatically engaged with by organisations as a means of coping, yet where individual professionals may hold an array of (dis)trusting orientations towards the technology. Examples of differing attitudes towards machine-learning technology is apparent in Carboni and colleagues' (2024) study of different approaches to risk of violence by psychiatric nurses and machine learning algorithmic scores. The nurses' approach to risk was characterised by doubt and alternative explanations of violent interactions in a way that protected patients, and they resisted and reframed the more pre-emptive and punitive logics of the algorithm-based system.

But while Carboni et al. (2024) identified forms of resistance among professionals working with vulnerable patients, professionals within other organisational contexts may feel obliged to work with particular technologies either because other options are made increasingly unfeasible, where a trusted leader within the organisation embraces the technology and becomes a proxy basis for trust relations around the technology ('if *they* think it works, then it must be ok'), or where (as observed in the preceding section) professionals may face stigma and marginalisation for not engaging with the technology. In such contexts, both the processes by which professionals come to trust, or more reluctantly (dis)engage, with an AI technology (amid distrust), as well as questions regarding what this (dis)engagement amid (dis)trust feels like for professionals (what we might call the 'texture' of engagement and trust - P. Brown & Bahri, 2019), are important lines of investigation; as are how these processes develop over time, for individuals and across professional teams. For example, while the introduction of AI is commonly expected to enhance time efficiency in healthcare tasks, therefore potentially freeing-up time to interact with patients, anthropological studies indicate that caregivers can often face increased workloads, as they must engage in additional responsibilities such as reviewing digital outputs and managing false alarms,alongside their regular duties (Pols, 2012, p. 52), leading to an efficiency paradox.

Everyday risk work (risk assessment, or risk communication) itself involves inherent tensions – for example, pertaining to how abstract probabilistic knowledge based on

population data are applied or communicated in individual cases (see P. Brown & Gale, 2018) – and further tensions are likely to emerge in experiences of working with(out) AI technologies. How these tensions are experienced and (not) resolved would therefore become central to understanding professional work with AI, as well as the implications of this work for interactions and relationships with patients and other clients (P. Brown & Gale 2018).

Focusing on these processes, Hallowell and colleagues' (2022) study of understandings, trust and distrust among a range of stakeholders towards the use of 'computational phenotyping' (CP) – whereby machine learning algorithms are used to assist in the diagnosis of rare dysmorphological disease or syndromes – highlighted the complex and dynamic processes of (dis)trust apparent in these contexts. Participants suggested the possibilities of trust emerged gradually over time, but where the patient-doctor relationship remained central and where machine learning was likened to existing technologies (e.g. a stethoscope) which would be controlled and used critically by the clinician, whose interpretation and judgement would remain fundamental:

> Many observed that scientific accuracy or objectivity of CP technology alone is not enough to foster patients' trust in machine-led diagnosis, for all diagnoses need interpretation, explanation and justification and that, according to our interviewees, will require clinician input to inspire patients' trust in the diagnostic process. (Hallowell et al., 2022, p. 6)

This open-ended, non-determinism of AI use in healthcare settings is fundamental to grasping its implications (Hoeyer, 2023) and obvious to those closest to the technologies; yet often overlooked in wider policy and media discussions.

Hallowell and colleagues' (2022) study also shows the reservations expressed by clinicians regarding their capacity to understand the workings of AI (in contrast to simpler technological artefacts), the lack of transparency ('blackbox') and 'explainability' (Bergquist et al., 2024; Ghassemi et al., 2021) as to what an individual assessment or diagnostic probability was based on – 'why the computer is saying what it's saying' (Hallowell et al., 2022, p. 9) – and the limited number of cases and data that the machine learning was based upon. So although there appears to be a growing demand for explainable AI models, particularly in high-risk areas like healthcare, existing explainability techniques fail to provide reliable insights for individual decisions made by AI systems, in order for these to be related to clear and verifiable outcomes (Ghassemi et al., 2021).

A similar problem exists regarding the 'co-creation' of ethical AI design, where professional caretakers or patients are expected to collaborate with coders and AI developers. Applications of AI/Machine Learning (AI/ML) in healthcare are rapidly evolving, and involving patients and the public in the design process is proposed as a strategy to address ethical challenges. While co-creation is vital for aligning algorithms with medical practices, research indicates that these groups often employ different discourses and hold varying views on concepts like 'error' or what constitutes a well-functioning algorithm (Bienefeld-Seall et al., 2023). This suggests that although co-creation is preferable to isolated development and testing, expectations for fully reliable outcomes should be tempered. Moreover, the dynamic nature of AI/ML introduces new, complex challenges for co-design that are frequently underestimated. Donia and Shaw (2021) argue that co-design both amplifies existing challenges and creates entirely new ones. They emphasise the importance of 'design humility', meaning that designers should recognise the limits of what their work can achieve. For example, we need to

consider who gets left out of the design process and think critically about when it might be better not to design at all. Finally, most discussions around co-design come from Western perspectives, raising questions about how to design for people with different backgrounds and experiences. Designers should start with these questions to ensure that co-design effectively leads to ethical AI and machine learning in healthcare.

These significant challenges in explainability and co-design suggest that problems of AI/ML as a 'blackbox', lacking transparency and broader input, are not easy to solve. Yet some participants in Hallowell et al. (2022) study (noted above) analogised the 'black-box' nature of machine learning algorithms with the way senior clinicians could also be untransparent themselves and how, as with colleagues: 'Ultimately you have to trust in the technology, in the knowledge or service it provides' (Hallowell et al., 2022, p. 9). Implicit in these perspectives were therefore different narratives regarding expertise, continuity and change. So while Beck's (1992) classic work encourages us to consider how new technologies configure new social relations and new identities, amid political (re)configurations of risk and responsibility, important empirical questions remain as to whether and how technologies are framed as new or different.

Apparent thus far in the analysis are thus potential chains of trust, 'whereby the expert professional acts as a trustee (the key actor being depended upon by the patient), who is in turn reliant on accurate information' and safe, effective technologies (P. Brown & Calnan, 2010, p. 66). But while trust inevitably involves an insufficiency of knowledge (otherwise there would be no need to trust), the ways in which ignorance and uncertainty may be transcended (Möllering, 2001, p. 414) by inflated expectations and imaginaries (see next section), amid wider systems of trust and hope (P. Brown et al., 2015), represents a potential dark side of trust amid these settings. While we have noted how those closest to, and with most expertise in, AI technologies can be highly reflexive about the capacities and limits of the technologies (Plájas, 2024; Sanderson et al., 2023), it can also be the case that those more distant from, yet with interests attached to the success of, these technologies may wittingly or unwittingly foster a misguided optimism or trust and a corresponding lack of reflexive judgement in their use (see next section). Where such advocates (individuals or teams) are themselves trusted authority figures within a wider organisation, this can shape a wider and potentially dangerous lack of critical reflexivity in the use of new AI-technology.

As we noted earlier, alongside zooming in to explore how trust is constructed in everyday meaning making and interactions with and around new technologies, it is also important to zoom out to further consider different tendencies towards (dis)trust and (dis) engagement across (and within) different professional groups, age cohorts, class and ethnic backgrounds (among other categories), as a way of considering how these broader locations in social space affect the specific processes of experiences and meaning-making upon which (dis)trust in a technology emerges over time. Following on from the theme in the previous section, we want to question how distrust in AI may be deemed risky/naive and problematised, and how some groups' reticence or refusal of these technologies may 'follow the grooves etched by traditional forms of stratification' and inequality around risk (Mythen, 2005, p. 129; Benjamin, 2019, p. 15). Ruha Benjamin's (2019) work on 'the employment of new technologies that reflect and reproduce existing inequities' (p.5), as institutional structures around these technologies shape professional- and citizen- 'subjects who prioritize efficiency over equity' (p.31), is one important and lauded line of inquiry around the use of AI in public organisations (see Zajko, 2022). However, it is Benjamin's somewhat lesser known (Benjamin, 2016) article on informed

dissent and distrust in science among communities who have, historically and to this day, been harmed by science, that may be even more pertinent here. Following Benjamin (2016) and Douglas (1992), we would expect that the categorising and potential problematisation or 'othering' of AI-'resisters' – and the labelling of this resistance as, in itself, risky – will be more likely when the fault-lines around engagement and refusal follow the patterns of wider historical relations of inequality between centre and periphery.

Hope and magical thinking around AI futures

The widespread belief in the potential of Artificial Intelligence (AI) to revolutionise healthcare processes is not only a reflection of technological advancements, but also a manifestation of magical thinking, or techsolutionism (Gardner, 2023; Morozov, 2013). This phenomenon, with various parallels with use of technology in obstetric care (e.g. Davis-Floyd, 1994), is characterised by the attribution of extraordinary powers or abilities to a (technological) object or system, often accompanied by a sense of inevitability, promise and optimism (see Cook, 2016; Watson and Lupton, Wozniak-O'Connor and Watson, 2024). In the context of AI, this thinking is exemplified by the widespread conviction that the scaling up of datasets will propel the field forward, with AI then supporting humans in making more accurate predictions and policy decisions. This 'AI orthodoxy' (McQuillan, 2022) has been reinforced by the proliferation of claims from software developers and technology companies, who are often portrayed as the new 'priests' (Thompson, 2019) driving the development of AI. This vision of AI's future capabilities has led to significant investments in its development and implementation, with significant financial and planetary resources aiding its creation and development (Baas, 2024).

Decidedly hopeful narratives have, however, been countered by critics who warn about the potential threats of AI, including its ability to make autonomous decisions that could have disastrous consequences. What these tech-optimistic and more critical frames have in common is that they both suggest that the future of humankind will be defined by AI (Hoeyer, 2023).

A closer examination of the discourses surrounding AI reveals that it is not just a reflection of technological advancements, but also a manifestation of a broader cultural ideology (Steinbrook & Redberg 2015; Sharon, 2018). The popular narrative surrounding AI is characterised by a sense of inevitability (Watson and Vaughan Wozniak-O'Connor, 2024), where technology is seen as the only way forward: it is this belief or hope, that stirs investments, and hence, that eventually turns into reality (Delvecchio Good, 2001). This ideology is, for example, reflected in statements of policy-makers, professional caregivers and other stakeholders in healthcare, who often use a discourse about a collapsing health or social work system, and under-staffing, to justify their decision to increasingly utilise AI in their work, as is often reflected in news media representations (Watson and Vaughan Wozniak-O'Connor, 2024; e.g. McIntyre & Pegg, 2018).

The magical thinking surrounding AI has in some cases led to convergences between professional caregivers, hospital board members, investors and the tech industry, whose members are pitching their products at healthcare conferences and other relevant events (Hoeyer, 2023; Baas and van Voorst, forthcoming in 2025). These studies show that professional caregivers often acknowledge their lack of AI-related technical expertise whilst already working with AI, but dismiss these concerns as unnecessary, claiming that

any problems will be solved in the near future by the technology industry. In this – commonly echoed – narrative, only the innovative and resourceful corporate sector, in collaboration with pioneering medics, can solve the issue of public healthcare. Such tech-optimist narratives portray the merging of the medical and technological fields as unavoidable (Watson and Vaughan Wozniak-O'Connor, 2024), therefore veiling the politics behind every technological innovation (Pfaffenberger, 1992).

However, this discursive practice, as well as the ongoing investment in AI's development and implementation, may have massive negative implications for society. If an algorithmic outcome seems right, but is based on false processes or weak data, the effects on individual patients and public health can be enormous. That healthcare stakeholders acknowledge their lack of technical expertise but continue to work with technology companies highlights the complex interplay between hope, expectation, vulnerability, trust and risk. Earlier work on how processes of hope in healthcare contexts shape engagement with technologies (Delvecchio Good, 2001; van Dantzig & De Swaan, 1978) has emphasised the subtle mutual complicity between different actors involved, who each have something to gain from embracing hope. Hoping can form a powerful medium for coping amid vulnerability: whether this be for patients' vulnerability to uncertain outcomes or death; for professionals struggling to provide quality care amid under-resourced care contexts, and where these same individual professionals (rather than the system) are held accountable for managing risks of poor outcomes; or for academic researchers under pressure to attract funding by pursuing the next breakthrough (Damhof & Gulmans, 2023). Where similar hopes in technology help in resolving these diverse vulnerabilities across an organisation, then insidious 'systems of hope' (van Dantzig & De Swaan, 1978) are likely to emerge. These mutually reinforcing tendencies towards hope involve implicit organisational dynamics (see Delvecchio Good, 2001), but it may often be the senior clinicians and policy-makers who are most influential in fostering positive imaginaries. For example, in addressing the potential for machine learning to better focus (shrinking) financial resources in targeting interventions for vulnerable families, one senior UK-bureaucrat speculated:

> It's not beyond the realms of possibility that one day we'll know exactly what services or interventions work, who needs our help and how to support them earlier. (McIntyre & Pegg, 2018)

Here the focus on a hopeful narrative about possible technological futures elides all sorts of practical and ethical issues; from data protection concerns and laws, to mathematical problems (low incidence, thus low sensitivity and specificity), to the 'appalling' current state of IT infrastructure facing social workers in England (Dixon, 2023, p. 65), not to mention a range of studies which show that the recording of data by healthcare and social work professionals is driven by a pragmatic pursuit of various objectives (Bevan & Hood, 2006; Warner & Gabe, 2004) which do not include compiling an accurate evidence-base for AI prediction.

This perspective would emphasise the 'vulnerability' of AI to poor quality data, though as Gallistl et al. (2024) warn, software companies are finding profitable ways of developing synthetic data about vulnerable populations, in an attempt to overcome the weaknesses and gaps in real-world data. So in their study of AI-based interventions on falls among older adults in long-term care, 'software designers turned to synthetical data creation, where visual data on older adults falling were not observed in real-life settings but synthetically created through automated software and data gathered from software designers using motion capture suits in which they tried to imitate different situations of falling down' (Gallistl et al., 2024, p. 6). Such data is unlikely to be valid for modelling risk of falls, while also excluding the voice and experiences of the vulnerable populations they are supposed to serve (Gallistl et al., 2024, p. 6.).

Yet, despite these many wide caveats, the hopeful claim quoted above could also serve a political-economic function (Delvecchio Good, 2001) as a useful (for the policy-maker) distraction – from the recent history of chronic under-funding of these services and the safety concerns this underfunding generated, and large-scale and costly failures in IT implementation in the English NHS. Supported by the wider quasi-magical cultural imaginaries and promises (noted above) around novel technologies (Cook, 2016; Watson and Vaughan Wozniak-O'Connor, 2024), not least AI, these narratives reproduce 'cultures of hope' which in turn may become embraced by some front-line professionals, whose willingness to hope in these novel techologies, and to trust in the developers behind them, is compelled by a need to seek solutions amid highly limited resources and heightened organisational pressures (van Dantzig & De Swaan, 1978). This hopeful 'embrace' of novel technologies among front-line workers should, therefore, by understood amid the wider political-economy or system of hope which compels it (Delvecchio Good, 2001), and in turn the wider cultural and ideological bases upon which such hopeful narratives around technology and technocracy are configured (Davis-Floyd, 1994; N. Brown, 2005; Pasquale, 2015; Watson and Vaughan Wozniak-O'Connor, 2024).

Concluding points and lines for further investigation

Critically interrogating these wider cultural scripts around AI and related technologies, and then tracing their influences as they permeate through and influence healthcare systems, national organisations and local contexts – from policy-pronouncements to frontline-framings – should thus be of fundamental importance to an effective interpretative social science of AI in healthcare contexts. In turn, interpretivist and interactionist social scientists would furthermore explore sense-making around AI and related technologies in everyday risk work, especially how hopes and magical thinking may form an implicit basis, or a more explicit set of cultural scripts, for professionals', managers' and patients'/clients' handling of AI-driven risk-information and risk-informed decisions. Hopeful and magical imaginaries, or indeed more pessimistic orientations towards technologies, may be powerful in shaping different formats of (dis)engagement with AI-technologies and the more or less critical ways in which they are enacted or resisted.

Within this interaction – between hopeful and/or critical AI-imaginaries and everyday caring practices and risk work – processes of (dis)trust also come to the fore. As we have considered above, processes of trust and distrust are vital to analysing engagement with, and/or resistance to, new technologies. Processes of trust always require some 'leap of faith' in order to transcend the inevitable uncertainty that exists about future outcomes (Möllering, 2001). Narratives of hope may facilitate this 'leap', and we have also explored how this momentum towards trusting in, and thus embracing, technology may be further aided by the heightened contexts of real, and anticipated vulnerability that often exist (for professionals and patients) and the need or 'will to trust' which accompanies this (P. Brown, 2009). It may therefore be useful to compare the enacting of AI-technologies in contexts of different formats of professional power and vulnerability (across specialities and/or geographical contexts, for example), in order to explore how trust and hope may, potentially, reinforce one another (P. Brown et al., 2015) or where, alternatively, distrust and tensions in hoping (between hope-as-desire, and hope-as-expectation) sit awkwardly alongside one other.

We have argued that the introduction of AI in care contexts therefore raises important questions about trust, specifically regarding who decides to implement these technologies and which stakeholders (professionals, clients, patients) are expected to trust them, as different levels of trust and distrust can shape healthcare relationships and outcomes. We have seen that the 'who decides?' question highlights that these processes of risk, trust, hope and vulnerability, are importantly shaped by existing power relations. We have argued that existing inequalities can shape the influence of AI on health and social work practices, particularly in reshaping definitions of 'normal' and 'risky.' They may also influence professionals themselves as they embrace or resist AI. This, in turn, can impact professionals' and patients' self-identity and interactions with others, potentially leading to stigma for those who resist.

These latter questions regarding the potential emergence of inequalities resulting from the application of AI, remind us that these processes of meaning-making (of hope, trust and risk assessment) may result in the kinds of differential outcomes that lend themselves to more analytic-quantitative analyses. There is also the need for social scientists to step back and question their own assumptions (more hopeful or cynical) towards AI-technology – one important way to do this, is to study how AI is implemented, and with which effects, in different contexts: how do real-life practices and experiences, match up with expectations and hopes? This critical approach, resisting simplistic assumptions and characterisations of AI, should form the basis of detailed empirical-ethnographic inquiry, for example combining interviews, to explore the meanings of various stakeholders, alongside observations of everyday practices and interactions involving AI.

Finally, we hope that social scientists continue to investigate the underlying political and economic structures that influence imaginations of and implementations of AI, both within the care sector and within the social sciences themselves. Approaching AI from a more critical perspective will require careful attention to the political and economic interests surrounding AI adoption (c.f. Delvecchio Good, 2001). It will also involve recognising the suffering associated with AI, as researchers examine how AI imaginaries may reflect or obscure deeper systemic issues within the care and academic sectors, which are often easily overlooked in overly optimistic, tech-driven narratives (see Hallowell, 2006).

Acknowledgement

The authors are grateful to Jeremy Dixon (Cardiff) and Veronica Moretti (Bologna) for their excellent and insightful feedback and suggestions. Any mistakes and weaknesses are the authors alone. Patrick would also like to acknowledge the important contributions of Prof. Nina Hallowell within this emerging literature and indeed her contribution to the journal. You can read more about Nina's work in Andy Alaszewski's In Memoriam essay in the first issue of the journal in 2024.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The research and writing of this editorial on the part of Roanne van Voorst has been funded by the European Union (ERC, Health-AI, grant 101077251). Views and opinions expressed are, however, those of the author only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

References

Ashuri, T., & van Voorst, R. (forthcoming). Networked biopower: Personalized healthcare in a datafied world. *Communication & society.*

Ashuri, T., & van Voorst, R. (in press). *Networked biopower: Personalized healthcare in a datafied world*. New Media and Society.

Baas, M. (2024). Artificial intelligence and the question of creativity: Art, data and the socio-cultural archive of ai-imaginations. *European Journal of Cultural Studies*, *27*(4), 788–795.

Baas, M., & van Voorst, R. (forthcoming in 2025). The magic of ai. Unveiling the sociopolitical implications of belief in artificial intelligence. Special Issue in Anthropological Theory.

Barbalet, J. (2009). A characterization of trust, and its consequences. *Theory & Society*, *38*(4), 367–382. https://doi.org/10.1007/s11186-009-9087-3

Beck, U. (1992). *Risk society: Towards a new modernity.* Sage.

Benjamin, R. (2016). Informed refusal: Toward a justice-based bioethics. *Science, Technology, & Human Values*, *41*(6), 967–990. https://doi.org/10.1177/0162243916656059

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. Wiley.

Bergquist, M., Rolandsson, B., Gryska, E., Laesser, M., Hoefling, N., Heckemann, R., Schneiderman, J. F., & Björkman-Burtscher, I. M. (2024). Trust and stakeholder perspectives on the implementation of AI tools in clinical radiology. *European radiology*, *34*(1), 338–347. https://doi.org/10.1007/s00330-023-09967-5

Bevan, G., & Hood, C. (2006). What's measured is what matters: Targets and gaming in the English public healthcare system. *Public Administration*, *84*(3), 517–538. https://doi.org/10.1111/j.1467-9299.2006.00600.x

Bienefeld, N., Boss, J. M., Lüthy, R., Brodbeck, D., Azzati, J., Blaser, M., Willms, J., & Keller, E. (2023). Solving the explainable AI conundrum by bridging clinicians' needs and developers' goals. npj digit. *Med*, *6*(94). https://doi.org/10.1038/s41746-023-00837-4

Boulos, M. N. K., Brewer, A. C., Karimkhani, C., Buller, D. B., & Dellavalle, R. P. (2014). Mobile medical and health apps: State of the art, concerns, regulatory control and certification. *Online Journal of Public Health Informatics*, *5*(3), e229.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information Communication & Society*, *15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Brown, N. (2005). Shifting tenses: Reconnecting regimes of truth and hope. *Configurations*, *13*(3), 331–355. https://doi.org/10.1353/con.2007.0019

Brown, P. (2009). The phenomenology of trust: A Schutzian analysis of the social construction of knowledge by gynae-oncology patients. *Health, Risk & Society*, *11*(5), 391–407. https://doi.org/10.1080/13698570903180455

Brown, P. (2021). *On vulnerability.* Routledge.

Brown, P., & Bahri, P. (2019). 'Engagement' of patients and healthcare professionals in regulatory pharmacovigilance: Establishing a conceptual and methodological framework. *European Journal of Clinical Pharmacology*, *75*(9), 1181–1192. https://doi.org/10.1007/s00228-019-02705-1

Brown, P., & Calnan, M. (2010). Braving a faceless new world? Conceptualizing trust in the pharmaceutical industry and its products. *Health*, *16*(1), 57–75. https://doi.org/10.1177/1363459309360783

Brown, P., de Graaf, S., Hillen, M., Smets, E., & Laarhoven, H. W. (2015). The interweaving of pharmaceutical and medical expectations as dynamics of micro-pharmaceuticalisation: Advanced-stage cancer patients' hope in medicines alongside trust in professionals. *Social Science & Medicine*, *131*, 313–321. https://doi.org/10.1016/j.socscimed.2014.10.053

Brown, P., & Gale, N. (2018). Theorising risk work: Analysing professionals' lifeworlds and practices. *Professions & Professionalism*, *8*(1), e1988. https://doi.org/10.7577/pp.1988

Carboni, C., Wehrens, R., van der Veen, R., & de Bont, A. (2024). Doubt or punish: On algorithmic pre-emption in acute psychiatry. *AI & Society*, 1–13. https://doi.org/10.1007/s00146-024-01998-w

Ceyhan, A. (2012). Surveillance as biopower. In B. Kirstie, D. Kevin, & H. D. Lyon (Eds.), *Routledge handbook of surveillance studies* (pp. 38–45). Routledge.

Cook, J. (2016). Young adults' hopes for the long-term future: From re-enchantment with technology to faith in humanity. *Journal of Youth Studies*, *19*(4), 517–532. https://doi.org/10.1080/13676261.2015.1083959

Couldry, N., & Mejias, U. (2019). Data colonialism: Rethinking big Data's relation to the contemporary subject, television and new Media. *20*(4), 336–349.
Damhof, L., & Gulmans, J. (2023). Imagining the impossible: An act of radical hope. *Possibility Studies & Society*, *1*(1–2), 51–55. https://doi.org/10.1177/27538699231174821
Davis-Floyd, R. (1994). The technocratic body: American childbirth as cultural expression. *Social Science & Medicine*, *38*(8), 1125–1140. https://doi.org/10.1016/0277-9536(94)90228-3
Delvecchio Good, M. J. (2001). The biotechnical embrace. *Medicine, Culture and Psychiatry*, *25* (4), 395–410. https://doi.org/10.1023/A:1013097002487
Dixon, J. (2023). *Adult safeguarding observed: How social workers assess and manage risk and uncertainty*. Policy Press.
Donia, J., & Shaw, J. A. (2021). Co-design and ethical artificial intelligence for health: An agenda for critical research and practice. *Big Data & Society*, *8*(2), 205395172110652. https://doi.org/10.1177/20539517211065248
Douglas, M. (1992). *Risk and blame: Essays in cultural theory*. Routledge.
Dujardin, C. (2020). Precision medicine': Critical reflections on Europe's latest healthcare paradigm. In J. Mansnerus, R. Lahti, & A. Blick (Eds.), *Personalized medicine: Legal and ethical challenges* (pp. 60–80). University of Helsinki, Faculty of Law.
Durocher, M. (2024). Food, bodies, health (risks): The biopolitics of organic materiality testing in the context of diet-associated health risk management practices. *Health, Risk & Society*, *26* (5–6), 260–281. https://doi.org/10.1080/13698575.2024.2365636
Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor. St*. Martin's Press.
Gallistl, V., von Laufenberg, R., & Lehner, K. (2024). Vulnerability assemblages: Situating vulnerability in the political economy of artificial intelligence. *Socius: Sociological Research for a Dynamic World*, *10*, 1–10. https://doi.org/10.1177/23780231241266514
Gardner, J. (2023). Imaginaries of the data-driven hospital in a time of crisis. *Sociology of Health & Illness*, *45*(4), 754–771. https://doi.org/10.1111/1467-9566.13592
Ghassemi, M., Oakden Rayner, L., & Beam, A. L. (2021). The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digital Health*, *3*(11), e745–e750. https://doi.org/10.1016/S2589-7500(21)00208-9
Giddens, A. (1990). *The consequences of modernity*. Stanford University Press.
Habermas, J. (1972). *Knowledge and human interests*. Heinemann.
Habermas, J. (1987). *Theory of communicative action (vol 2), lifeworld and system: A critique of functional reason*. Polity.
Hallowell, N. (2006). Varieties of suffering: Living with the risk of ovarian cancer. *Health, Risk & Society*, *8*(1), 9–26. https://doi.org/10.1080/13698570500532322
Hallowell, N., Badger, S., Sauerbrei, A., Nellåker, C., & Kerasidou, A. (2022). 'I don't think people are ready to trust these algorithms at face value': Trust and the use of machine learning algorithms in the diagnosis of rare disease. *BMC Medical Ethics*, *23*(1), 1–14. https://doi.org/10.1186/s12910-022-00842-4
Hallowell, N., Lawton, J., & Gregory, S. (2005). *Reflections on research: The realities of doing research in the social sciences*. Open University Press.
Hannah-Moffat, K. (2013). Actuarial sentencing: An "unsettled. *Proposition. Justice Quarterly*, *30* (2), 270–296.
Hoeyer, K. (2023). *Data paradoxes: The politics of intensified data sourcing in contemporary healthcare*. MIT Press.
Jasanoff, S., & Kim, S. H. (2015). *Dreamscapes of modernity: Sociotechnical imaginaries and the fabrication of power*. University of Chicago Press.
Lewis, J., & Weigert, A. (1985). Trust as a social reality. *Social Forces*, *63*(4), 967–985.
Lupton, D. (2014). Critical perspectives on digital health technologies. *Sociology Compass*, *8*(12), 1344–1359. https://doi.org/10.1111/soc4.12226
Lupton, D., Wozniak O'Connor, V., & Watson, A. (2024). Arts-Based and Sensory Methods to Imagine. In V. Fors, M. Berg, & M. Brodersen (Eds.), *The De Gruyter Handbook of Automated Futures: Imaginaries* (Vol. 2, pp. 395–412).
MacEachen, E. (2000). The mundane administration of worker bodies: From welfarism to neoliberalism. *Health, Risk & Society*, *2*(3), 315–327.

McIntyre, N., & Pegg, D. (2018, September 16th). *Councils use 377,000 people's data in efforts to predict child abuse*. The Guardian. https://www.theguardian.com/society/2018/sep/16/councils-use-377000-peoples-data-in-efforts-to-predict-child-abuse

McQuillan, D. (2022). *Resisting AI: an anti-fascist approach to artificial intelligence*. Policy Press.

Meskó, B., Drobni, Z., Bényei, É., Gergely, B., & Győrffy, Z. (2017). Digital health is a cultural transformation of traditional healthcare. *Mhealth*, *3*, 38. https://doi.org/10.21037/mhealth.2017.08.07

Möllering, G. (2001). The nature of trust: From Georg Simmel to a theory of expectation, interpretation and suspension. *Sociology*, *35*(2), 403–420. https://doi.org/10.1177/S0038038501000190

Möllering, G. (2006). *Trust: Reason, routine, reflexivity*. Elsevier.

Monahan, J., & Skeem, J. L. (2016). Risk assessment in criminal sentencing. *Annual Review of Clinical Psychology*, *12*(1), 489–513.

Morozov, E. (2013). *To save everything, click here: The folly of technological solutionism*. PublicAffairs.

Mythen, G. (2005). Employment, individualization and insecurity: Rethinking the risk society perspective. *The Sociological Review*, *53*(1), 129–1 49. https://doi.org/10.1111/j.1467-954X.2005.00506.x

Mythen, G., & Weston, S. (2023). Interrogating the deployment of 'risk' and 'vulnerability' in the context of early intervention initiatives to prevent child sexual exploitation. *Health, Risk & Society*, *25*(1–2), 9–27. https://doi.org/10.1080/13698575.2022.2150750

O'Brien, H., & Toms, E. (2008). What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, *59*(6), 938–955. https://doi.org/10.1002/asi.20801

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Osoba, O. A., Iv, W., & Welser, W. (2017). *An intelligence in our image: The risks of bias and errors in artificial intelligence*. Rand Corporation.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Passchier, R. (2021). *Artificiële intelligentie en de rechtsstaat: over verschuivende overheidsmacht, Big Tech en de noodzaak van constitutioneel onderhoud*. Boom Publisher.

Peeters, R. (2020). The agency of algorithms: Understanding human-algorithm interaction in administrative decision-making. *Information Polity*, *25*(4), 507–522.

Peeters, R., & Widlak, A. C. (2023). Administrative exclusion in the infrastructure-level bureaucracy: The case of the Dutch daycare benefit scandal. *Public Administration Review*, *83*(4), 863–877. https://doi.org/10.1111/puar.13615

Pfaffenberger, B. (1992). Technological dramas. *Science, Technology, & Human Values*, *17*(3), 282–312. https://doi.org/10.1177/016224399201700302

Pickersgill, M. (2019). Digitising psychiatry? Sociotechnical expectations, performative nominalism and biomedical virtue in (digital) psychiatric praxis. *Sociology of Health & Illness*, *41*(s1), 16–30. https://doi.org/10.1111/1467-9566.12811

Plájas, I. (2024, July 22). Slowing down AI futures – countering techno-solutionist/dystopian futures through multimodal storytelling. Paper presented at EASA conference, Barcelona. Abstract available through. https://nomadit.co.uk/conference/easa2024/paper/79909

Pols, J. (2012). *Care at a distance: On the closeness of technology*. Amsterdam University Press.

Sanderson, C., Douglas, D., Lu, Q., Schleiger, E., Whittle, J., Lacey, J. Newnham, G., Hajkowicz, S., Robinson, C., & Hansen, D. (2023). AI ethics principles in practice: Perspectives of designers and developers. *IEEE Transactions on Technology and Society*, *4*(2), 171–187. https://doi.org/10.1109/TTS.2023.3257303

Sharon, T. (2018). When digital health meets digital capitalism, how many common goods are at stake?. *Big Data & Society*, *5*(2). https://doi.org/10.1177/2053951718819032

Siles, I., Espinoza-Rojas, J., Naranjo, A., & Tristán, M. F. (2019). The mutual domestication of users and algorithmic recommendations on netflix, communication. *Culture and Critique*, *12* (4), 499–518. https://doi.org/10.1093/ccc/tcz025 *Sociology Compass*, *8(12)*, 1344–1359.

Sleeman, D., & Gilhooly, K. (2023). Groups of experts often differ in their decisions: What are the implications for AI and machine learning? A commentary on 'noise: A flaw in human judgment' by Kahneman, Sibony, and Sunstein (2021). *AI Magazine*, *4*, 555–567.

Steinbrook, R., & Redberg, R. F. (2015). Reporting research misconduct in the medical literature. *JAMA Internal Medicine*, *175*(4), 492–493.

Szmukler, G. (2003). Risk assessment: 'numbers' and 'values'. *Psychiatric Bulletin*, *27*(6), 205–207. https://doi.org/10.1192/pb.27.6.205

Thompson, C. (2019). *Coders: The Making of a New Tribe and the Remaking of the World*. Penguin.

Tschandl, P., Rinner, C., Apalla, Z., Argenziano, G., Codella, N., Halpern, A., Janda, M., Lallas, A., Longo, C., Malvehy, J., Paoli, J., Puig, S., Rosendahl, C., Soyer, H. P., Zalaudek, I., & Kittler, H. (2020, August). Human-computer collaboration for skin cancer recognition. *Nature Medicine*, *26*(8), 1229–1234. https://doi.org/10.1038/s41591-020-0942-0. Epub Jun 22. PMID: 32572267.

van Dantzig, A., & De Swaan, A. (1978). *Omgaan met angst in een kankerziekenhuis. [Coping with fear in a cancer hospital]*. Spectrum.

van Voorst, R. (2024a). *Challenges and limitations of human oversight in ethical AI implementation in healthcare: Balancing digital literacy and professional strain*. Digital Health.

Van Voorst, R. (2024b, September). The medical tech-facilitator. An emerging position in dutch public healthcare and their tinkering practices. *Medicine Anthropology Theory*, *11*(2), 1–23. https://doi.org/10.17157/mat.11.2.7794

Van Voorst, R. (Forthcoming in 2025). Health incentive apps as technological drama. In P. Hackett, S. Na, & A. Godley-Smith (Eds.), *Edited volume 'handbook of AI and Robotics'*. Routledge.

Veltkamp, G., & Brown, P. (2017). The everyday risk work of Dutch child- healthcare professionals: Inferring 'safe'and 'good' parenting through trust, as mediated by a lens of gender and class. *Sociology of Health & Illness*, *39*(8), 1297–1313. https://doi.org/10.1111/1467-9566.12582

Warner, J. (2015). *The emotional politics of social work and child protection*. Policy Press.

Warner, J., & Gabe, J. (2004). Risk and liminality in mental health social work. *Health, Risk & Society*, *6*(4), 387–399. https://doi.org/10.1080/13698570412331323261

Zajko, M. (2022). Artificial intelligence, algorithms, and social inequality: Sociological contributions to contemporary debates. *Sociology Compass*, *16*(3), e12962. https://doi.org/10.1111/soc4.12962

Zarsky, T. (2012). Governmental data mining and its alternatives. *Pennsylvania State Law Review*, *116*(2), 285–330.

Zuboff, S. (2019). Surveillance capitalism and the challenge of collective action. *New Labor Forum*, *28*(1), 10–29.