



A powerful potion for a potent problem: transformative justice for generative AI in healthcare

Nicole Gross¹

Received: 6 March 2024 / Accepted: 29 June 2024
© The Author(s) 2024

Abstract

Generative Artificial Intelligence (AI), as a transformative technology, holds significant promise for applications in healthcare. At the same time, the datafication, AI integration, and commodification of health have opened the floodgates for ethical issues, including those related to fairness, access, beneficence, democracy, solidarity, inclusion, and societal harms. As further the digitalization, innovation, and disruption of healthcare is inevitable, the paper maps out how power, equity, access, identity, participation, and knowledge contribute to creating social injustice issues. It also discusses that current justice approaches—distributive justice, representational justice, restorative justice, and capabilities-centered justice—do not have enough impact to prevent or remedy the many harms and injustices that AI has already created in healthcare or will continue to do so. The paper proposes that a transformative justice approach is needed for generative AI as a transformative technology, focused on (1) peace, emancipation, and eliminating the root causes of injustice, (2) holistic conflict resolution, (3) human rights-based approaches, and (4) the empowerment of agency and actors.

Keywords Generative artificial intelligence · Healthcare AI · Transformative technology · Data justice · Transformative justice

1 The promised land: generative AI in healthcare

How does generative AI impact healthcare? What data justice issues will generative AI, as a transformative technology, bring along? and How can these data justice issues be addressed? These are important questions. The further digitalization, innovation, and disruption of healthcare in Europe is inevitable. National healthcare systems have long been struggling with significant treatment and service gaps, inefficiencies, lack of high-quality care, affordability and availability, and overspending [1]. Policymakers, along with key stakeholders such as healthcare professionals, academics, and patients, believe that digital technologies, including generative AI, will transform healthcare by making it better, cheaper, and more efficient in the future [2, 3]. Generative AI refers to “deep-learning models that can

generate high-quality text, images, and other content based on the data they were trained on” [4, n.p.]. While significant gains in healthcare are expected, many critics believe that the further digitalization, datafication, and AI-ification of healthcare is highly contentious, however [5, 6]. All ‘Big Tech’ companies have moved into healthcare to unlock an \$11 trillion opportunity [7]. Generative AI systems currently found on the market are owned by Big Tech [8] (e.g., Microsoft, Google or Meta), use their foundational models to run (e.g., ChatGPT, Co-Pilot or Gemini), or rely on their data and cloud ecosystem to function [9]. Big Tech companies have now become part and parcel of public and private healthcare [10], even when their technologies, systems, and infrastructures are explicitly connected with data justice issues [11]. Data justice is social justice in an increasingly datafied and AI-ified world, and social justice issues include fairness, access, beneficence, democracy, solidarity, inclusion and harms to society [11]. Healthcare is both a human right and a global public good [12], therefore it is imperative that harms are monitored and prevented, and the implementation and scale of generative AI managed with extreme care.

✉ Nicole Gross
nicole.gross@ncirl.ie

¹ School of Business, National College of Ireland, Mayor Street, IFSC, Dublin 1 D01 K6W2, Ireland

But let us step back and see what the hype is all about. Generative AI technologies hold big promises to transform healthcare in previously unimaginable ways: democratize knowledge; provide accessible, around-the-clock, remote and highly empathetic care; empower people; enable better clinical and personal decisions; and prevent errors [13–17]. The envisioned use-cases of generative AI in healthcare include autonomous primary care (e.g., diagnosis, simulation, documentation, or treatment planning); dealing with patients (e.g., navigation, record searches, messaging, or trial identification); administrative tasks (e.g., scribing, coding, message composition, or translation); and research (e.g., performance analysis, data interpretation, or literacy discovery) [18]. As of 2024, prominent generative AI technologies in healthcare include ChatGPT, Microsoft Co-Pilot, Google Bard (now Gemini), Nuance, Suki Assistant, Corti, and EllenAI for clinical administration support; Glass AI, Regard, RedBrick AI, Paige, and Kahun Medical for clinical decision support; Hippocratic AI and Gridspace patient engagement; and Syntegra Medical Mind and DALL-E2 for synthetic data generation [19].

With 180 million active weekly users as of March 2024 [20], OpenAI's ChatGPT is particularly worth mentioning here. Research suggests that ChatGPT can support clinical practice by automating note-writing or helping with scientific writing and research [13, 16]. Additionally, it could provide other benefits such as cost savings, personalized medicine, and streamlined workflows [16], and it could serve patients living in underserved or rural regions [21]. ChatGPT could also provide patient education, personalized learning, and support patients in the self-management of their conditions [17, 21]. Google's new generative AI tool for the medical domain, Med-PaLM2, makes equally big promises, e.g., to increase efficiency, answer medical questions, and manage complex workflows [22]. Finally, Google AI, ChatGPT, and other generative AI technologies could play a significant role in public health. For instance, these systems could provide essential information about infectious or chronic diseases, or improve community health education and literacy, e.g., by promoting health, answering questions about the prevention of illnesses, discussing the impact of social and environmental factors on community health, or giving information about essential services [23, 24].

However, even with all these promises and evident benefits, this transformative technology comes with complex and layered ethical hurdles that affect health, wealth, and power [25]. Stahl and Eke, [26] broke it down into four categories: environmental impacts (pollution, waste, and sustainability), culture and identity (the good life, discrimination, and social sorting, cultural differences, bias, and forming own opinions), individual needs (human-to-human contact,

safety, autonomy, informed consent, data control, accountability, ownership, and intellectual property, and harm), and social justice and rights— the last category is of particular interest to this paper, even though the paper touches on other ethical issues too. According to Taylor [11, n.p.], data justice is the “fairness in the way people are made visible, represented, and treated as a result of their production of digital data”. Related social justice issues include problems with representation, informational privacy, not sharing the data's benefits, a lack of personal autonomy, inability to challenge bias, and discrimination.

This paper explores if and how data justice issues can become relieved, exacerbated or transformed in and through generative AI technologies in healthcare. It also reviews the justice approaches that are currently in place for dealing with misfires, problems, and harms caused by generative AI in healthcare, before suggesting that a transformative justice approach is necessary to deal with both Big Tech and AI as transformative technology.

2 An inevitable wave of data justice issues for healthcare

AI is poised to create transformative changes in society [25], including in healthcare, thereby making a wave of social justice issues inevitable [11, 27, 28]. Social justice, according to the Global Partnership on Artificial Intelligence (GPAI), is “a commitment to the achievement of a society that is equitable, fair, and capable of confronting the root causes of injustice” [29, p.20]. Data justice seeks to understand the societal implications and misfires created by data-driven technologies - may they be related to democratic procedures, the entrenchment and introduction of inequities, discrimination and exclusion of certain groups, dehumanization of decision-making, or interaction around sensitive issues [28, 30]. While technology providers often argue that data is just data, and data is neither good nor bad, the reality is that data is always situated - it needs to be understood in relation to other social practices [27]. Recently, the GPAI proposed mapping social justice issues against six criteria: power, equity, access, identity, participation, and knowledge [29]. This framework is used next to engage with (real and potential) data justices created in and through generative AI technologies in healthcare.

2.1 The problem of power

To understand where data justice issues come from, it is imperative to understand who holds power, what they control, how it this done, how the community is impacted, and how to empower people [29].

Big Tech has been exerting control over digital ecosystems for more than two decades, establishing market power, making money from data as an asset, and building political influence [10, 31–33]. Big Tech companies have expanded into many parts of public services, including healthcare, and this has significantly increased their market power. Their surveillance capitalist business models and practices have contributed to the commodification of healthcare [5, 6, 34] also increasingly lead to a tragedy of the digital commons in health [35]. Surveillance capitalism is “the unilateral claiming of private human experience as free raw material for translation into behavioural data [32, p.8]. Once data is collected by Big Tech, it is used to predict users’ future behaviour, and sold in markets targeting future actions, such as to advertisers, governments, employers, research institutions, insurers, or any other interested third parties. Data essentially serves as an asset for Big Tech companies like Alphabet, Amazon, Apple, Microsoft, or Meta [33], and generative AI allows these companies to deepen their data shadows and maximize their data-related revenue flows even further [9, 36]. Some additional background here: Microsoft had invested \$13bn into Open AI to compete with Google [37], a company reportedly making \$237.86m [38] in 2023 from the sale of user data to advertisers alone. Open AI has already laid out in its privacy policy that it will collect information from its users and reserves the right to use this data to “carry out business transfers” with third-party vendors, service providers, or affiliates [39, n.p.]. But Google has caught up and now has Gemini (formally Bard) and Med-PaLM2 for healthcare. Not only do these AI platforms compete with OpenAI, but they allow Google to ‘unlock’ even more sensitive data from its users. Google recently amended its T&C to ‘allow’ any data from Google Gemini to become part and parcel of its surveillance business model. And as it turns out, medical information is extremely valuable in the market, with the data reportedly fetching as much as \$1000 per patient per year [40].

When it comes to data justice issues, key stakeholders in healthcare—civil society, healthcare providers, and the state—are becoming increasingly disempowered in and through Big Tech’s surveillance, integrated technologies, and data ecosystems. The advent and scale of generative AI technologies now enables these companies to predict and control behaviours to an “unprecedented degree”, all in the interest of maximizing profits [36, n.p.]. As Big Tech is shifting from “surveillance capitalism which curates” to “surveillance capitalism which creates” [36, n.p.], its creations will undermine autonomy, choice, and freedom of thought at an unprecedented scale [32]. In any case, more critical studies of generative AI technologies are needed to consider how the power of individuals, as well as entire states, is being ‘dwarfed’ by the positions that Big Tech companies hold in

business and society, and to explore how technologies can be redirected to serve the public [8, 41].

2.2 The problem of equity

The extraction, processing, and automation of data have significant implications for social justice as they directly impact people’s lives and communities. The transformative power of data but also the measures of justice and actions against biases and discrimination are relevant here [29].

The processes and practices of data extraction, data processing, and data-driven automation [29] tend to be done in pervasive, ubiquitous and so utterly mundane ways, making them for people to notice them [42]. The exact methods are deeply buried in Big Tech’s language, documents, and patent filings too [32, 43]. Yet, these systems have significant impacts on people’s lives and communities, thus affecting equity. According to Health Action International [57, n.p.], “AI has particular risks for key populations who face layering patterns of inequities related to age, gender, sexual orientation, cultural identity, racialized characteristics, literacy, disability, and (mental) health status”. For instance, Open AI’s large language model ChatGPT stereotypes people (e.g., race, gender, and ability) and tends to give heavily biased answers [44–46]. Humans are already biased, but those biases can become amplified when people interact with biased AI tools. Leffner [47] suggests that new or unforeseen biases may even be introduced when people use generative AI technologies, and the effects will last well beyond the user’s initial engagement with the AI. It is worth mentioning here that while some biases may be unproblematic (and could arguably even work to create some equity), many biases are inequitable, which means that “they are based on or lead to the unjust distribution of goods or because they are based on or lead to the undue discrimination of certain people and social groups.” [48, n.p.].

When it comes to healthcare, biased generative AI systems have proven to be very harmful. Some AI systems have become known for their racial biases [49–51]; for instance, an AI system implemented in US hospitals was designed to identify patients who would benefit from additional medical care. However, because of underlying biases, the AI assigned very low-risk scores to black patients, which meant that they missed out on high-risk care management by as much as 46% [52]. Another study found that the use of AI resulted in lower rates of skin cancer detection for people of colour [53]. Many AI technologies in health and medicine ignore the sex and gender dimension too and its impact on health and disease, and these technologies also do not account for bias detection [54]. Sub-optimal results, mistakes, discriminatory outcomes, and decreased equality are inevitable outcomes [54]. Additionally, many generative

AI systems have well-documented age-related biases, and these lead to health inequity for older people [55].

Problematic is that important healthcare decisions are increasingly based on the responses and outputs of biased generative AI technologies. Yet, Big Tech is frequently not held accountable for the harms and inequities that its technologies create ‘in vitro’. Moreover, individuals harmed by these technologies often lack effective measures of justice and recourse against biases and discrimination. Without coherent global AI regulation, generative AI technologies are released to the market based solely on the developers’ subjective judgment of their readiness, even if some of the outputs are nothing short of being ‘legitimately dangerous’ [56].

2.3 The problem of access

Social justice issues in an increasingly datafied and AI-fied world are related to material deprivation, inequality, institutional and structural discrimination, and the maldistribution of resources and social goods [29].

The World Health Organization (WHO) [17] admits that equitable access is not a given, as the use of digital tools is often limited to certain countries, regions, or segments. In many cases, only affluent people, healthcare providers, or even National Healthcare Systems can afford to use them, and the effect is that already disadvantaged or marginalized persons, groups, and even entire nations become disadvantaged even further [57]. To illustrate: data suggests that generative AI technologies are used by people aged 25–44 (around 53%) and predominantly by males (around 66%) [58]. These technologies are mainly used in the US and Europe but also countries like India, Japan, Brazil, and Canada [59]. Looking at generative AI in healthcare specifically, the US and Europe appear as front runners when it comes to trials (e.g., the Brigham and Women’s Hospital and Massachusetts General Hospital in Boston, Vanderbilt University Medical Center in Nashville, K Health in conjunction with Cedars-Sinai in Los Angeles, University of Rochester Medical Center in New York or the Mayo Clinic but also Ada Health in Germany, HeartFlow’s AI in the UK or Icon in Ireland). These fact and figures showcase that many people, communities, and even entire counties (e.g., low- and middle-income countries) are excluded from accessing or using generative AI. It is important here to recognize that history and path dependencies matter, as social justice issues often become entrenched. The increasing digitalization, datafication, AI-fication, and commodification of health, compounded by pervasive access issues, are likely to exacerbate these social justice issues further.

However, even if and when people, communities, or countries have access to generative AI systems, these

systems are not necessarily harmless: in early 2023, a Belgium man died by suicide after AI chatbot ELIZA (ChatGPT-based) encouraged him to ‘sacrifice himself’ [60]. Persistent biases [44, 46, 55], errors and hallucinations up to 27% [61], and misaligned behaviours of the AI to strategically deceive their users [62] are just some of the problems that can eradicate the advantages of privileged access to generative AI. When it comes to healthcare, access (but also the lack of access) can thus become a matter of life and death. It is important to increase transparency, bring injustices into the open, advance access to robust and well-tested technologies, and promote open access and responsible data sharing, for instance.

2.4 The problem of identity

Data justice issues related to identity involve coming to terms with harmful categorizations of self and community, as well as confronting the erasure of identity perpetuated by AI technologies. [29].

Bender et al. [44] talk about the lack of a diverse dataset (e.g., immigrants, domestic abuse victims, sex workers, trans or queer people, neurodivergent people, and other people living with disabilities), a view that is echoed by the WHO [17, p.21], stating that many AI datasets “exclude girls and women, ethnic minorities, elderly people, rural communities and disadvantaged groups”. Much research has recently emerged in this space: Cave and Dihal’s work [63], for instance, has detailed the ‘whiteness of AI’ and Noble’s [48] research has extensively focused on algorithms of oppression. Gross [45] wrote about gender biases in ChatGPT as well as the AI system’s ignorance of transgender people. Satori and Theodorou’s [65] work also provides a socio-technical perspective on AI that details significant inequities related to gender and race. In her ‘Plug and Pray’ Report, Marzin [46] documented how persons living with disabilities are experiencing emerging AI technologies, but also how they can be excluded from their use through a lack of representation.

A consistent message across these reports and research pieces is that generative AI technologies are often not designed with inclusivity in mind. They frequently overlook the needs of vulnerable populations, and this inevitably leads to social harms and injustices. Using inclusive datasets are thus critically important, especially in fields such as healthcare.

2.5 The problem of participation

Participation equates to more democratic data work and sound governance of generative AI. This means challenging

the status quo of power, fair participation, and transforming the system to reflect inclusiveness [29].

Floridi et al. [66] detail why designing AI for the social good matters and how it can be done. For healthcare AI, this entails designing and deploying trustworthy systems that have a meaningful positive impact on human life, while also incorporating safeguards against manipulation right from the outset. It is important to build systems that are contextualized and fully transparent, prioritize data subject consent and privacy protection, display situational fairness, and maintain and foster human autonomy [66]. Yet, many generative AI models such as ChatGPT or DALL-E have harmful and dehumanizing features, many of which are not just “unintentional glitches” but “endemic issues of oppression” [67, n.p.]. In and through mechanisms such as automation, data extraction, analytics, surveillance, and containment, tech companies have managed to build ‘AI empires’ that act to uphold or even reinforce deeply rooted features such as white supremacy, heteropatriarchy, coloniality, and racial capitalism [67].

AI empires[67], coupled with stellar projections such as a valuation of 17.2 billion for generative AI in the global healthcare market by 2032 - a CAGR of 37.0% [64] [Global Newswire, 2024] or “1 trillion improvement potential” in the US healthcare industry alone [[68], n.p.], showcase just how much the tech industry prioritizes innovation and market growth over participation and designing AI for the social good. To facilitate social justice, a challenge is to take a further look at the government of data, ownership and intellectual property issues, and the governance of the private AI sector more generally [69].

2.6 The problem of knowledge

Lastly, knowledge means pluralistic knowledge, challenging authority across scientific and political structures, interdisciplinary and reflexive thinking, and intercultural learning and sharing [29].

While generative AI is a transformative technology for healthcare, this transformation is driven by a technology-driven agenda as well as Big Tech’s data-hungry business models. Knowledge is frequently concentrated within the tech industry, whereby a mantra ‘move fast and break things’, black boxes, and algorithms as trade secrets have set the tone [32, 36]. Illustrating further: machine learning underlies most generative AI technologies like ChatGPT, Dall-E2, or Google Gemini. Machine learning is made up of three parts: an algorithm or a set of algorithms, training data, and a model. To safeguard their intellectual property, Big Tech frequently withhold the procedures, statistics, and transformer architectures that constitute their algorithms. Equally, the data used to train the algorithms, identify

patterns, and build the neural network often remains hidden or in a ‘black box’. The model is the actual interface that people interact with; however, how it functions or how the AI technology responds and adapts when humans input data or provide feedback remains equally opaque [70]. While black boxes (as opposed to much-called-for ‘glass boxes’) may protect Big Tech’s intellectual property and trade secrets, they also work to reinforce their powerful positions and undermine transparency [70].

Global organizations like the OECD stress that fostering an inclusive AI-enabling ecosystem is imperative to support the development of AI that is both innovative and trustworthy, while also upholding human rights and democratic values [71]. When it comes to healthcare, the WHO [17, p.60] outlines that greater cooperation and collaboration within the United Nations (UN) system is needed to “respond to the opportunities and challenges of deploying AI in health care and of its wider application in society and the economy”. However, Big Tech often falls short in sharing knowledge, engaging in interdisciplinary work, or fostering intercultural learning. These companies pursue their tech-driven agendas and establish contracts or collaborations based primarily on their revenue interests, frequently treading ethical red lines, and violating human rights in the process [72]. When it comes healthcare, policymakers have also been found to prioritize capitalism, entrepreneurship, and market growth over public health and public value [10]. Contestations, social movements, and pushbacks against Big Tech’s approach to thinking, learning, sharing, and politics have been documented in journalistic accounts, academic publications, and releases from civic society organizations. However, these efforts have not been sufficient to dislodge Big Tech from its powerful position. As of mid-2024, the tech field remains dominated by innovation-led strategic agendas, complex yet opaque data extraction processes and practices, and black-box algorithms, leading to significant data justice concerns for citizens [32, 35].

3 Current approaches to data (in)justices

Data justice means coming to terms with how injustices might be remedied, an aspect of interest to this paper, and four justice typologies have been suggested when it comes to advancing research and practice: distributive, representational, restorative, and capabilities-based approaches [73]. This section briefly discusses the developments, practices, and challenges inherent in each justice type before asking the question if the current approaches, practices, and realities are ‘transformative enough’ to remedy the misfires, injustices, and harms created in and through the use of generative AI in healthcare.

3.1 Flaws in distributive justice

Distributive justice concentrates on the fair distribution of social goods, including the risks and benefits, opportunities and burdens, or rights and rewards [29]. As mentioned afore, Big Tech has chosen to ‘gloss over’ the risks, burdens, and social justice issues that are created by their generative AI technologies [44]. Big Tech has raced to launch premature AI technologies and stake out large parts of the market [74] - fully aware that once Pandora’s box is opened, AI cannot be ‘uninvented.’ With generative AI, Big Tech companies are likely to follow the same trajectory established with their search engines, social media, and apps, which involves absorbing user data and selling it on the market for future behaviour (i.e., marketing and advertising). Geiger [75] has called this the ‘ideology of inevitability’. Not only are we talking about the user’s behavioural surplus data here [32]—that of course too— but also all the other materials used to train their AI engines, including some personal, organizational, and even copyrighted materials (32, 76). While the user may get informative, persuasive, and authoritative-sounding answers to their inputs and queries [44], neither the user nor the original creator/owner of the knowledge gets fair value in return.

The risks, harms, and burdens created by generative AI are still being discovered, and the lack of transparency and accountability is often lamented [77]. Yet, ChatGPT has reaped the rewards, achieving a staggering valuation of \$80-90bn in 2023 alone [78]. This valuation is poised to move upwards as AI propels toward a “surveillance capitalism which creates” paradigm and business model [36, n.p.]. What is more, Big Tech—a powerful elite operating in a neoliberal climate—has a good track record of politically influencing regulations and legislation in a way that their business practices remain commercially successful [40, 66], and this has enabled them to continue building their ‘AI empire’ [67].

3.2 Flaws in representational justice

This approach to justice relates to obtaining fairness, dignity, and autonomy in moral, political, cultural, and legal issues [29]. When it comes to representation, known issues relate to the potential for misuse, data biases, data security, and data privacy [77, 79, 80]. In the healthcare space, the issues are mostly centred around human autonomy, explicability, patient privacy, fairness, and the prevention of harm [81] but also around data ownership, informed consent, trustworthiness, and equitable access [82]. Some issues are already well-known and recognized (e.g., bias, exclusion, discrimination, or intellectual property issues) while for other issues significant ‘blind spots’ remain to exist [83].

Entrepreneurial capitalism has allowed generative AI development to get this far and regulations in this space have been noticeably lagging over the last five years. Global governing bodies such as the Global Partnership on Artificial Intelligence, the WHO, the OECD, and the UN made plenty of good recommendations to preserve fairness, dignity, and autonomy but these guidelines are not legally binding. Current data protection regulations, such as the EU’s GDPR for instance do not have enough impact [73]. While some AI-related local laws have been passed right now, these approaches remain both reactive and fragmented according to the AI Index Report [84]. The EU’s AI Act, lauded as “the world’s first comprehensive AI law” [85, n.p.], also does not reach far or deep enough to prevent injustices from happening [86–88]. The Act’s biggest flaw is that it is inherently focused on harmonizing Member State’s national legislation to create a single internal AI market and eliminate obstacles to trade [89]. While the EU AI Act puts forward binding rules on transparency and ethics, it leaves a lot of ‘wiggle room’ for AI companies [90] and fails to deliver sufficient punch to protect people [91]. The Act’s focus on prohibited and high-risk systems leaves many embedded systems utilizing generative AI models unregulated, including those in healthcare (e.g., lifestyle and wellness applications). Many companies using generative AI are still allowed to (invasively) identify and analyse consumers’ feelings via emotion recognition systems. Commentators expect that various forms of surveillance will persist and privacy not to be fully protected across all circumstances and contexts [92, 93]. Likewise, through the enactment of the Digital Services Act 2024, Europe has assumed a pioneering role on the global stage in regulating online platforms and marketplaces, attempting to reclaim societal value and democracy [94]. Yet, the true impact of the Digital Services Act remains unknown, and a great deal of work is still to be done as many surveillance capitalism practices may be illegitimate but not illegal [94]. What is more, other European initiatives such as Europe’s Digital Agenda 2020–2030 or the European Health Data Space may counteract these regulations by lending clear support to data-related commercial and market values rather than acting to uphold public (health) values [10, 95]. It seems that generative AI technologies, with their many problems, misfires, and injustices, remain to stay abreast of ethical standards as well as regulation at this critical time [77].

3.3 Flaws in restorative justice

Restorative justice stands for processes and mechanisms that allow for reconciliation, repair, and ‘righting the wrongs’ [29]. As mentioned afore, ethical principles and guidelines are often just that, and legal measures are either

not there yet or they are fuzzy [77, 88, 96]. In the (extreme) words of Munn [97, p. 869]: most of the ethical and regulatory tools are meaningless, isolated, toothless, and utterly “useless”. As it stands, reconciliation, repair, and wrongdoings are addressed in the following ways: First, the user can report any bad outputs, biases or ‘wrongdoings’ to the AI directly, which allows the AI to engage in learning in real-time. As generative AI systems are intelligent and some are also (lightly) moderated by humans (e.g., ChatGPT and Google Gemini), any feedback given will be used to inform the future responses of the AI system. That said, even a moderated systems can provide misleading information. For instance, Google’s generative AI search (wrongly) advised users to eat rocks to remain healthy [98]. Second, there is collective action - such as the AI Incident Database, which was founded recently to capture the harms, problems, and justice issues created by AI technologies [99]. Third, naming and shaming is also possible—academics but also journalists have been particularly proactive in this space. Fourth is lawsuits, like the ones captured in the US AI-Litigation Database provided by The George Washington University Institute for Trustworthy AI in Law and Society [100]. These lawsuits deal with intellectual property violations (e.g., copyright) caused by the AI training process [101], but they also cover ‘AI crimes’ related to facial recognition and biometric data, breaches of privacy, employment and hiring, disability and unemployment benefits, unfair marketing, immigration, taxes, fraud, and even terrorism [100].

However, AI crimes are not easy to bring forth. Legally speaking, it is often unclear if it is the tech company, AI-designer, or user who is responsible for committing the said AI crime. Furthermore, it is unclear what constitutes an AI crime, whether new types of crimes are emerging, and how these crimes are performed [102]. It is difficult to hold AI accountable when existing laws, such as tort and liability laws, do not cover such emerging technologies, and when the legal status of AI—autonomous, intelligent, and self-learning entities (though not human)—remains undefined. From the outset, AI was “likely to behave in antisocial and harmful ways unless they are very carefully designed” [103, p.303], and this point, AI technologies are becoming increasingly sophisticated, aware, intelligent, and potentially harmful [104]. Yet, restorative justice approaches have not caught up with the clear dangers posed by AI. As any attempts to (legally) regulate remain either reactive or fuzzy [88], “inequality is the name of the game” [96, p. 1].

3.4 Flaws in capabilities-centered justice

The last approach, capabilities-centered justice takes focus on real life, context, and moral regard, in particular, what needs to change in terms of power, equity, access, identity,

participation, and knowledge to allow for fairness, reciprocity, and human flourishing [29, 105].

Big Tech companies understand that generative AI offers clear benefits, including the potential to enhance people’s dignity, well-being, and personal flourishing, while also contributing to a life that is both personally valuable and socially equitable [27, 105]. In her capabilities-approach paper, Nussbaum [104] outlined that human welfare can be broken down into ten different capabilities that create a good life, and only when most of these capabilities are fulfilled, a political order can be regarded as ‘decent’. Buccella’s [105] work outlines that, in theory at least, AI fulfills many of these capabilities. Applying this notion to generative AI in healthcare: the technology can be used to live and prolong life (e.g., public health initiatives and coordinated healthcare); enable better bodily health (e.g., medical diagnoses and treatment but also resource management in hospitals); enhance bodily integrity (e.g., fertility medicine); facilitate senses, imagination, and thought (e.g., empowerment and education); work with affiliation (e.g., social networks, communication, shared experiences); being better with others (e.g., prediction, preservation and protection); and having more control over one’s environment (e.g., good choices, fair treatment and meaningful relationships). Yet, Big Tech companies also know that their AI technologies create so many harms and social injustices [67, 106] that the market nothing short of an “ethical minefield” [108, n.p.].

“AI for all” [105, p. 1143] is not what Big Tech companies are currently pursuing. As it stands, their focus is on the innovation race, marketing, and business models - capitalistic interests, in short [10]. As market rhythms and experiences tip in favour of commercialization, these transformative AI technologies manage to ‘escape’ the very principles that would ensure social justice—moral regard, ethical standards, capabilities, and robust regulation [77].

4 A transformative justice approach to generative AI in healthcare

Borrowing from the wider justice literature, this paper puts forward that a transformative justice approach is needed in the healthcare space to deal with generative AI as a transformative technology. Transformative justice tries not to dismiss other justice approaches but seeks to “radically reform its politics, locus and priorities” [107, p. 340]. As a platform for transformative change, this justice approach moves away from legal (and often restorative) approaches to devise a “range of policies and approaches that can impact on the social, political and economic status of a large range of stakeholders” [107, p. 340]. While other justice approaches tend to be well-meaning in nature, they often

cannot ensure that fairness, entitlement, and equity will be attained— they are too “bounded by greater socio-historical constraints” [108, p. 241]- - especially when it comes to healthcare. A transformative justice approach is about transforming power, community building, and creating positive conflict resolution [109]. Gready and Robins [107] put forward points of several convergence which help to define transformative justice further and these can also be applied to generative AI in healthcare too.

4.1 Peace, emancipation, and eliminating causes of injustice

Gready and Robins [107] outline that it is important to eliminate the causes of conflict, injustice, and drive the notion of peace from ‘above’. When it comes to healthcare, it is about fair access to health for everyone— health is a global public good rather than a commercial one [12, 110]. When it comes to generative AI in healthcare: firstly, states need to intervene by re-establishing a focus on public value [10, 75]. Secondly, and considering the profound transformative impact of AI on the factors contributing to a fulfilling life, it is imperative to ensure equitable access to generative AI, both practically and intellectually, for all individuals, communities, and nations [103, 105]. Thirdly, generative AI, in its current form, must be adapted to address the needs of diverse populations, particularly those who are most marginalized. It should facilitate health literacy, be aware of context-specific empowerment, and incorporate participatory approaches in the technology development, deployment, and scaling. Big Tech companies aiming to launch generative AI in healthcare will need informal as well as formal guidance as to what the needs of diverse populations are, but also answer to regulation and governance grounded in local cultures and contexts. The EU AI Act, Digital Services Act, and European Health Data Space has set good yet also incomplete precedents here. In similar vein, the upcoming US AI Bill of Rights will also leave many gaps when it comes to peace, emancipation, and eliminating the causes of injustice.

It is important that local and contextual guidance and governance is weaved in with wider global discourses and mechanisms to ensure that ethical standards are being scaled across the globe equally, and no pockets of opportunity for exploitation continue to exist. As it stands, global codes and guidance, such as the UN’s Advisory Body on AI, GPAI, or WHO reports, have not changed the trajectory of the AI-race or the business models of Big Tech yet. While all discourses and regulations are instrumental when it comes to working towards peace, emancipation, and eliminating the cause of injustices, and transformational change can only be achieved by investing in and prioritizing human rights

and public value on a local and global level, global institutions like the UN or the WHO will need to be empowered to develop legally-binding, global agreements and regulations for generative AI in healthcare. Such legally-binding, global agreements and regulations should not be impossible to achieve - a glance at the UN’s International Bill of Human Rights or the World Trade Organization’s Trade-Related Aspects of Intellectual Property Rights Agreement showcase working blueprints which could be followed.

4.2 Holistic conflict resolution

It is well-recognized that AI is at a critical juncture as of 2024: generative AI has already caused plenty of harm and misfires, and these technologies are now rapidly expanding into many aspects of healthcare [18]. While ethical guidances, regulations, and laws are either new or still ‘in the making’, it is conflict that drives resolution forward at a personal, relational, structural, and cultural level [107]. Modern conflicts are characterized by inequities of power and status, and in the case of generative AI, such conflict comes to light in the ‘push and pull’ between Big Tech, the state and its citizens. While this push and pull is starting to become captured in academic, journalistic, and policy papers as well as AI incident databases and lawsuits, some conflicts, injustices, and harms may never be captured - even if they have adverse effects in practice. Citizens need to have the right and the opportunity to be heard and act as a community to stand up against Big Tech; however, the reality is that citizens often find themselves without choices in hyperdigital spaces, where automation is pervasive and continually increasing, and the use of technology is unavoidable [111]. To enforce citizen’s rights, states can intervene and resolve conflicts on a case-by-case basis, and some precedents have recently been set. For instance, the U.S. Federal Trade Commission has recently won a lawsuit against Meta, forcing the company to reduce the amount of money it makes from users under 18 [112]. In Europe, and after years of privacy litigation, enforcement, and court rulings, the Court of Justice of the European Union decided in 2023 that Meta was no longer allowed to track and profile its users without their explicit consent [113]. While rulings such as these could put a break on Big Tech’s data capitalism business models, Meta - in a prototypical push and pull move - has already responded to such rulings by providing users with options to either ‘pay us with money’ or ‘pay us with your privacy’. Yet, fundamental rights such as privacy or health cannot be for sale, which makes Big Tech’s tactics up for further conflict and contestation [113].

Generative AI companies in healthcare watching the push and pull between Big Tech, the state, and citizens can learn from this dynamic and develop holistic approaches to

navigate and transform both current and future conflicts. Holistic conflict resolution, according to Greedy and Robins [107], means using local resources, addressing the root causes of (healthcare and technology-based) inequality, and adopting holistic responses to transform the conflict. Big Tech companies, but also other generative AI companies which rely on Big Tech's foundational models and data ecosystems, ought to consider the insights that come from conflict and take these tension points as opportunity for change. This, for instance, means commercial companies moving beyond the digital extraction, capture, and transformation of personal digital data into a private asset [114]. For healthcare, it means a value-shift towards data altruism and data solidarity [115]. For policymakers, it means co-creating regulation with citizens but also enforcing regulations with political will. For citizens, it means coming forward when data justice issues arise and uniting their voices through social movements. The generative AI field is fast evolving at this stage, which means the collaboration between tech companies, the state and citizens needs to be ongoing rather than characterized by a linear engagement.

4.3 Human-rights based approaches

The WHO clearly states that health is a human right [116]. Generative AI technologies that operate in the healthcare space must thus follow human rights principles, i.e., the 'PANEL principles' - participation, accountability, non-discrimination, empowerment, and international human rights law [117]. Moreover, there should be a focus on building the capacity to ensure that all key stakeholders fulfil their human rights obligations and responsibilities [107]. Before the EU AI Act was agreed, Casolari, Buttaboni and Floridi [87] forcefully stated that the regulation should not adopt a market-based approach but human rights-based one instead. When it comes to capability-building, tech companies need to move away from their current technical and entrepreneurial approaches. As mentioned afore, alternative approaches could be focused on listening to advocacy, fostering active participation, and working to empower victims of rights abuses. This can be achieved through incorporating user feedback but also insights from reports, the AI Incident database, and legal actions, and embedding these considerations within algorithmic frameworks. It is equally important that Big Tech provides transparency and accountability when it comes to these matters, e.g., by turning 'black boxes' into 'glass boxes'.

While it is important to challenge the power relations, doing so is not an easy feat. Commentators from the Corporate Europe Observatory agree that Big Tech's powerful positioning and lobbying stances such as "there is no AI without Big Tech" have derailed the EU AI Act [41, n.p].

This begs the question if they have now become too influential to even be regulated. To ensure due diligence and protect public health interests, the Observatory had initially suggested to exclude Big Tech from these negotiations—similarly how Big Tobacco was (eventually) excluded after many "years of dirty lobbying tactics" [41, n.p.]. Some commentators even suggested that generative AI in healthcare should be regulated akin to nuclear energy [118], where a self-regulation approach is simply unthinkable. However, as the EU chose to keep Big Tech closely tied into the negotiations and allowed for a self-declaration of risk, the AI Act will predominantly work to create and harmonize trade in the European AI market [89]. This means that deep-rooted inequities, exclusions, and unequal power relations will be inevitably reinforced too. Summing up, the EU's innovation and commerce-friendly stance when it comes to AI needs to be critically reviewed and challenged to make these transformations happen [10].

4.4 The power of agency and actors

Greedy and Robins [107] have put forward that actor-oriented approaches to human rights are anchored in everyday perspectives and local contexts. To that end, human rights are shaped by conflicts and struggles but also by understanding those who claim them. For instance, the European Public Health Alliance ran a workshop in late 2023 to give a voice to medical doctors in Europe involved in AI-based innovation. It also launched another workshop in early 2024 focused on the digital inclusion of all citizens when it comes to AI in healthcare, whereby representatives from the European Network on Independent Living, European Agency for Fundamental Rights, European Federation of National Organisations Working with the Homeless, and European Sex Workers Alliance came forward [119]. The European Disability Forum published a report which detailed the struggles, risks, and harms that people with disabilities are facing with AI technologies [46]. This report later informed the UN's wider stance on AI and the rights of persons with disabilities [120]. The Irish Platform for Patients' Organisations, Science & Industry has organized a 'Citizen's Jury' for 2024 to give Irish people and patients a voice, and help shaping shape policy going forward [121]. These examples show that transformation requires the interplay between high-level debates and actual/specific situations. Non-profit organizations are instrumental in bridging this divide as they put forward how and for whom a given strategy may work. What is more, they also highlight what is required to change policies or introduce new regulation in response to emerging human rights issues rooted in real-life struggles and experiences of deprivation and oppression. In any case, interventions need to be set up with an intent to challenge

power differentials— though in some cases, such interventions have misfired, thus reinforcing power asymmetries and marginalizing the marginalized even further [107]. The persistence of Big Tech’s data capitalist business model - despite increasingly stringent regulation and fines, but also their consistent lobbying efforts to water down AI regulation, speak to this point.

5 Concluding remarks

The paper highlights the substantial impact of generative AI in healthcare, underscoring its transformative potential, while also acknowledging the data justice issues that accompany this technology as well as Big Tech’s involvement in healthcare. As current approaches - distributive justice, representational justice, restorative justice, and capabilities-centered justice - do not address data justice issues sufficiently or comprehensively, a transformative justice approach has been suggested. The paper has provided a conceptual framework encompassing peace and emancipation, holistic conflict resolution, human rights-based approaches, and re-distributing power between agency and actors. This framework can guide the definition of a future that harnesses the potential benefits of generative AI in healthcare while simultaneously addressing data justice concerns. However, the paper also acknowledges that considerable challenges exist in making such significant leaps forward. To move towards emancipation, empowerment, and transformation, it is fundamentally important to engage with AI-induced social justice issues thoroughly. Equally, it is important to pay attention to how market processes and power relations unfold in time, space, and context. After all, this is where harms and justices happen but also where contestations take place and new democratic spaces can be developed. The generative AI landscape in healthcare is still rapidly evolving, and this paper on transformative justice aims to offer guidance toward defining a future where social justice issues can be enhanced through transformative technologies.

Funding This research has been developed with funding from the Irish Research Council (New Foundations) grant number NF/2023/1782.

Declarations

Ethical approval This project has received full ethical approval by the National College of Ireland in November 2023 under approval number 23112308.

Conflict of interest None.

Data Availability Not applicable since this is a conceptual paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. OECD: Health at a Glance: Europe 2022. (2022). <https://www.oecd.org/health/health-at-a-glance-europe/> Accessed 17 November 2023
2. DigitalHealthEurope: Recommendations on the European Health Data Space. (2023). <https://digitalhealtheurope.eu/> Accessed 17 November 2023
3. EU Commission: eHealth: Digital health and care. (2023). https://health.ec.europa.eu/ehealth-digital-health-and-care_en Accessed 17 November 2023
4. IBM: What is generative AI? (2024). <https://research.ibm.com/blog/what-is-generative-AI> Accessed 23 January 2024
5. Hoeyer, K., Bauer, S., Pickersgill, M.: Datafication and accountability in public health: Introduction to a special issue. *Soc. Stud. Sci.* **49**(4), 459–475 (2019)
6. Ruckenstein, M., Schüll, N.: The datafication of Health. *Annu. Rev. Anthropol.* **46**(1), 1–18 (2017)
7. CB Insights:: The Big Tech in Healthcare Report: How Amazon, Google, Microsoft, Apple, & Oracle are fighting for the \$11T market. Research Report. November 30. (2022). <https://cdt.org/insights/eu-ai-act-brief-pt-2-privacy-surveillance/?s=03> (2022) Accessed 24 June 2024
8. Kak, A., Myers West, S., Whittaker, M.: Make no mistake—AI is owned by Big. *MIT Technology Review*. 5 December. (2023). <https://www.technologyreview.com/2023/12/05/1084393/make-no-mistake-ai-is-owned-by-big-tech/amp/?s=03> Accessed 17 December 2023
9. Birch, K.: *Data Enclaves*. Pavgrave MacMillan, Cham (2023)
10. Gross, N., Geiger, S.: Choreographing for Public Value in Digital Health? *Big Data Soc.* **10**(2) (2023). <https://doi.org/10.1177/20539517231220622>
11. Taylor, L.: What is data justice? The case for connecting digital rights and freedoms globally. *Big Data Soc.* **4**(2) (2017). <https://doi.org/10.1177/2053951717736335>
12. Chen, L.C., Evans, T.G., Cash, R.A.: Health as a Global Public Good, In: Kaul, I., Grunberg, I., Stern, M. (eds), *Global Public Goods: International Cooperation in the 21st Century*, Oxford Academic, New York (1999). <https://doi.org/10.1093/0195130529.003.0014> Accessed 17 Nov. 2023
13. Ayers, J.W., Poliak, A., Dredze, M., et al.: Comparing physician and Artificial Intelligence Chatbot responses to patient questions posted to a Public Social Media Forum. *JAMA Intern. Med.* (2023). <https://doi.org/10.1001/jamainternmed.2023.1838>
14. Cascella, C., Montomoli, J., Bellini, V., Bignami, E.: Evaluating the feasibility of ChatGPT in Healthcare: An analysis of multiple clinical and research scenarios. *J. Med. Syst.* (2023). <https://doi.org/10.1007/s10916-023-01925-4>
15. Pearl, R.: 5 Ways ChatGPT Will Change Healthcare Forever, For Better. *Forbes Magazine* 13 February. (2023). <https://>

- www.forbes.com/sites/robertpearl/2023/02/13/5-ways-chatgpt-will-change-healthcare-forever-for-better/?sh=7651d1cc7bfc Accessed 17 November 2023
16. Sallam, M.: ChatGPT Utility in Healthcare Education, Research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare*. **11**(6), 887 (2023)
 17. WHO: Ethics and governance of artificial intelligence in health: guidance on large language models. (2024). <https://www.who.int/publications-detail-redirect/9789240084759> Accessed 23 February 2024
 18. Gartner: Use-Case Prism: Generative AI for U.S. Healthcare Pay-ers. <https://www.gartner.com/en/documents/4535399> Accessed 15 April 2024
 19. Zhang, P., Kamel Boulos, M.N.: Generative AI in Medicine and Healthcare: Promises, opportunities and challenges. *Future Internet*. **15**, 286 (2023)
 20. Shewale, R.: ChatGPT Statistics — User Demographics (March 2024). 12 January. <https://www.demandsage.com/chatgpt-statistics/> Accessed 4 March 2024
 21. Javaid, M., Haleem, A., Singh, R.P.: ChatGPT for healthcare services: An emerging stage for an innovative perspective. *Benchmark Council Trans. Benchmarks Stand. Evaluations*. **3**(1) (2023). <https://doi.org/10.1016/j.tbench.2023.100105>
 22. Google: Med-PaLM- A large language model from Google Research, designed for the medical domain. <https://sites.research.google/med-palm/> Accessed 27 June 2024
 23. Biswas, S.S.: Role of Chat GPT in Public Health. *Ann. Biomed. Eng.* **51**(5), 868–869 (2023)
 24. Parray, A.A., Inam, Z.M., Ramonfaur, D., Haider, S.S., Mistry, S.K., Pandya, A.K.: ChatGPT and global public health: Applications, challenges, ethical considerations and mitigation strategies. *Global Transitions*. **5**, 50–54 (2023)
 25. Gruetzemacher, R., Whittlestone, J.: The transformative potential of artificial intelligence. *Futures*. **135** (2022). <https://doi.org/10.1016/j.futures.2021.102884>
 26. Stahl, B.C., Eke, D.: The ethics of ChatGPT— exploring the ethical issues of an emerging technology. *Int. J. Inf. Manag.* **74** (2024). <https://doi.org/10.1016/j.ijinfomgt.2023.102700>
 27. Dencik, L., Hintz, A., Redden, J., Trere, E.: Exploring Data Justice: Conceptions, applications and directions. *Inform. Communication Soc.* **22**(7) (2019). <https://doi.org/10.1080/1369118X.2019.1606268>
 28. Gabriel, I.: Toward a theory of Justice for Artificial Intelligence. *Daedalus*. **151**(2), 218–231 (2022)
 29. Global Partnership on Artificial Intelligence (GPAI): Data Justice in Practice: A Guide for Impacted Communities. (2022). <https://gpai.ai/projects/data-governance/data-justice-in-practice-a-guide-for-impacted-communities.pdf> Accessed 17 November 2023
 30. Dencik, L., Hintz, A., Cable, J.: Towards data justice? The ambiguity of anti-surveillance resistance in political activism. *Big Data Soc.* **3**(2) (2016). <https://doi.org/10.1177/2053951716679678>
 31. Birch, K., Bronson, K.: Big tech. *Sci. As Cult.* **31**(1), 1–14 (2021)
 32. Zuboff, S.: *The Age of Surveillance Capitalism*. Profile Books, London (2019)
 33. Birch, K., Cochrane, K., Ward, C.: Data as asset? The measurement, governance, and valuation of digital personal data by Big Tech. *Big Data Soc.* **8**(1) (2021). <https://doi.org/10.1177/205395172111017308>
 34. Timmermans, S., Almeling, R.: Objectification, standardization, and commodification in health care: A conceptual readjustment. *Soc. Sci. Med.* **69**(1), 21–27 (2009)
 35. Greco, G.M., Floridi, L.: The tragedy of the Digital Commons. *Ethics Inf. Technol.* **6**(2), 73–81 (2003)
 36. Jacobs, B.: POV: How generative AI is changing surveillance capitalism. *FastCompany*. 28 March. (2023). <https://www.fastcompany.com/90871955/how-generative-ai-is-changing-surveillance-capitalism> Accessed 27 June 2024
 37. Novet, J.: Microsoft's \$13 billion bet on OpenAI carries huge potential along with plenty of uncertainty. *CNBC*. 8 April. (2023). <https://www.cnbc.com/2023/04/08/microsofts-complex-bet-on-openai-brings-potential-and-uncertainty.html> Accessed 27 June 2024
 38. Statista: Advertising revenue of Google from 2001 to 2023 (2024). <https://www.statista.com/statistics/266249/advertising-revenue-of-google/> Accessed 27 June 2024
 39. Open A.I., About: <https://openai.com/about> Accessed 23 November 2023. (2023)
 40. Cherian, S.A.: Healthcare Data: The Perfect Storm. *Forbes Magazine*. 14 January. (2022). <https://www.forbes.com/sites/forbestechcouncil/2022/01/14/healthcare-data-the-perfect-storm/> Accessed 17 November 2023
 41. Vranken, B.: Big Tech lobbying is derailing the AI Act. *Corporate Europe Observatory*. 24 November. (2023). <https://corporateeurope.org/en/2023/11/big-tech-lobbying-derailing-ai-act?s=03> Accessed 1 June 2024
 42. Beverungen, A., Beyes, T., Conrad, L.: The organizational powers of (digital) media. *Organization*. **26**(5), 621–635 (2019)
 43. Kalluri, P.R., Agnew, W., Cheng, M., Ownens, K., Soldaini, L., Birhane, A.: The Surveillance AI Pipeline. *ARXIV*, Retrieved: April 8, from (2024). <https://arxiv.org/abs/2309.15084> (2023) Accessed 8 June 2024
 44. Bender, E.M., Gebru, T., McMillan-Major, A., Shmitchell, S.: The Dangers of Stochastic Parrots: Can Language Models Be Too Big? *ACM Conference on Fairness, Accountability, and Transparency*. Virtual Event. March 3–10: pp. 610–623 (2021)
 45. Gross, N.: What ChatGPT tells us about gender: A cautionary tale about performativity and gender biases in AI. *Social Sci.* **12**(435) (2023). <https://doi.org/10.3390/socsci12080435>
 46. Marzin, C.: Plug and Pray? *European Disability Forum*. (2020). <https://www.edf-feph.org/publications/plug-and-pray-2018/> Accessed 29 November 2023
 47. Leffner, L.: Humans Absorb Bias from AI—And Keep It after They Stop Using the Algorithm. (2023). <https://www.scientificamerican.com/article/humans-absorb-bias-from-ai-and-keep-it-after-they-stop-using-the-algorithm/> Accessed 29 February 2024
 48. Pot, M., Kieusseyan, N., Prainsack, B.: Not all biases are bad: Equitable and inequitable biases in machine learning and radiology. *Insights Imaging*. **12**(13) (2021). <https://doi.org/10.1186/s13244-020-00955-7>
 49. Benjamin, R.: *Race after Technology: Abolitionist Tools for the New Jim Code*. Polity, Cambridge (2019)
 50. Ledford, H.: Millions of black people affected by racial bias in health-care algorithms. *Nature* 26 October. (2019). <https://www.nature.com/articles/d41586-019-03228-6> Accessed 17 November 2023
 51. Noble, S.: *Algorithms of Oppression*. NYU, New York (2018)
 52. Pulido, F.: Why AI bias can hurt your business. *Wired*. 15 June. (2023). <https://www.wired.co.uk/bc/article/why-ai-bias-can-hurt-your-business-ey> Accessed 22 November 2023
 53. Aggarwal, P., Papay, F.A.: Artificial intelligence image recognition of melanoma and basal cell carcinoma in racially diverse populations. *J. Dermatolog Treat.* **33**(4), 2257–2262 (2022)
 54. Cirillo, D., Catuara-Solarz, S., Morey, C.: Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. *Npj Digit. Med.* **3**(81) (2020). <https://doi.org/10.1038/s41746-020-0288-5>
 55. van Kolfshootten, H.: The AI cycle of health inequity and digital ageism: Mitigating biases through the EU regulatory framework on medical devices. *J. Law Biosci.* **10**(2) (2023). <https://doi.org/10.1093/jlb/lsad031> July–December

56. Franzen, C.: 'Legitimately dangerous': Google's erroneous AI Overviews spark mockery, concern. *Venture Beat*. May 24, (2024). <https://venturebeat.com/ai/legitimately-dangerous-googles-erroneous-ai-overviews-spark-mockery-concern/> Accessed 27 June 2024
57. Health Action International. AI &, Healthcare: (2023). <https://healthai.haiweb.org/> Accessed 17 November 2023
58. Statista: Global user demographics of ChatGPT in 2023, by age and gender. (2023). <https://www.statista.com/statistics/1384324/chat-gpt-demographic-usage/#:~:text=ChatGPT%20is%20used%20most%20widely%20among%20those%20between,account%20for%20over%2060%20percent%20of%20ChatGPT%20users.> Accessed 27 June 2024
59. Adebisi, A.: Generative AI Industry Report 2023: Statistics, Trends, and Market Size, *GadgetAdvisor*. July 11. (2023). <https://gadgetadvisor.com/ai/generative-ai-industry-report-2023-statistics-trends-and-market-size/> Accessed 27 June 2024
60. Walker, L.: Belgian man dies by suicide following exchanges with chatbot. *Brussels Times*. 28 March. (2023). <https://www.brusselstimes.com/430098/belgian-man-commits-suicide-following-exchanges-with-chatgpt> Accessed 17 November 2023
61. Cabrera, S.: How AI companies are trying to solve the LLM hallucination problem. *FastCompany*. January 8. (2024). <https://www.fastcompany.com/91006321/how-ai-companies-are-trying-to-solve-the-llm-hallucination-problem> Accessed 27 June 2024
62. Scheurer, J., Balesni, M., Hobbhahn, M.: Technical Report: Large Language Models can Strategically Deceive their Users when Put Under Pressure. *arXiv (preprint)* (2023). <https://doi.org/10.48550/arXiv.2311.07590>
63. Cave, S., Dihal, K.: The whiteness of AI. *Philosophie Technol.* **33**, 685–703 (2020)
64. Global News Wire: Generative AI in Healthcare Market Set to Reach a Valuation of USD 17.2 Bn by 2032. April 3. (2023). <https://rss.globenewswire.com/en/news-release/2023/04/03/2639957/0/en/Generative-AI-in-Healthcare-Market-Set-to-Rreach-a-Valuation-of-USD-17-2-Bn-by-2032-Data-Analysis-by-Experts-at-Market-us.html> Accessed 27 June 2024
65. Satori, L., Theodorou, A.: A sociotechnical perspective for the future of AI: Narratives, inequities, and human control. *Ethics Inf. Technol.* **24**(4) (2022). <https://doi.org/10.1007/s10676-022-09624-3>
66. Floridi, L., Cowls, J., King, T.C., Taddeo, M.: How to design AI for Social Good: Seven essential factors. In: Floridi, L. (ed.) *Ethics, Governance, and Policies in Artificial Intelligence*. *Philosophical Studies Series*, vol. 144. Springer, Cham (2021)
67. Tacheva, J., Ramasubramanian, S.: AI empire: Unraveling the interlocking systems of oppression in generative AI's global order. *Big Data Soc.* **10**(2) (2023). <https://doi.org/10.1177/20539517231219241>
68. Bhasker, S., Bruce, D., Lamb, J., Stein, G.: Tackling healthcare's biggest burdens with generative AI. *McKinsey Report*. July 10. (2023). <https://www.mckinsey.com/industries/healthcare/our-insights/tackling-healthcares-biggest-burdens-with-generative-ai> Accessed 27 June 2024
69. WHO: Benefits and risks of using artificial intelligence for pharmaceutical development and delivery. March 25 (2024). <https://www.who.int/publications/i/item/9789240088108> Accessed 27 June 2024
70. Bagchi, S.: Why We Need to See Inside AI's Black Box. *Scientific American*, May 26 (2023). <https://www.scientificamerican.com/article/why-we-need-to-see-inside-ais-black-box/> Accessed 27 June 2024
71. OECD: Governing with Artificial Intelligence: Are governments ready? In: *OECD Artificial Intelligence Papers*. No 20. OECD Publishing, Paris (2024)
72. Madhumita, M., Siddarth, S.: How governments are beginning to regulate AI. *Financial Times*, May 30. (2019). <https://www.ft.com/content/025315e8-7e4d-11e9-81d2-f785092ab560> Accessed 27 June 2024
73. Leslie, D., Katell, M., Aitken, M., Singh, J., Briggs, M., Powell, R., Rincón, C., Chengeta, T., Birhane, A., Perini, A., Jayadeva, S., Mazumder, A.: Advancing data justice research and practice: an integrated literature review. The Alan Turing Institute in collaboration with The Global Partnership on AI (2022). <https://advancingdatajustice.org/wp-content/uploads/2022/04/advancing-data-justice-research-and-practice-integrated-literature-review.pdf>
74. Waters, R.: Big Tech is racing to claim its share of the generative AI market. *Financial Times*, July 27. (2023). <https://www.ft.com/content/be81fc62-49eb-40c9-a66a-2dc652e9b400> Accessed 27 June 2024
75. Geiger, S.: Silicon Valley, disruption, and the end of uncertainty. *J. Cult. Econ.* **13**(2), 169–184 (2020)
76. Coffman, C.: Does the Use of Copyrighted Works to Train AI Qualify as a Fair Use? Copyright Alliance. 11 April. (2023). <https://copyrightalliance.org/copyrighted-works-training-ai-fair-use/> Accessed 17 November 2023
77. Ayling, J., Chapman, A.: Putting AI ethics to work: Are the tools fit for purpose? *AI Ethics.* **2**, 405–429 (2022)
78. Schuetz, M.: OpenAI Seeks \$90 Billion Valuation in Possible Share Sale, *WSJ Says*. *Bloomberg*. 26 September. (2023). <https://www.bloomberg.com/news/articles/2023-09-26/openai-seeks-90-billion-valuation-in-possible-share-sale-wsj?embedded-checkout=true#xj4y7vzkg> Accessed 17 November 2023
79. Kaszım, E., Koshiyama, A.S.: A high-level overview of AI ethics. *Patterns.* **2**(9), 100314 (2021)
80. Siau, K., Wang, W.: Artificial Intelligence (AI) Ethics: Ethics of AI and ethical AI. *J. Database Manage.* **31**(2), 74–87 (2020)
81. Karimian, G., Petelos, E., Evers, S.M.A.A.: The ethical issues of the application of artificial intelligence in healthcare: A systematic scoping review. *AI Ethics.* **2**, 539–551 (2022)
82. Cohen, G.: What should ChatGPT Mean for Bioethics? *Am. J. Bioeth.* **23**(10), 8–16 (2023)
83. Hagedorff, T.: Blind spots in AI ethics. *AI Ethics.* **2**, 851–867 (2022)
84. AI Index Report: Policy and Governance. (2023). https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report-2023_CHAPTER_6-1.pdf Accessed 17 November 2023
85. Parliament, E.U.: EU AI Act: first regulation on artificial intelligence. (2023). <https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence> Accessed 17 November 2023
86. Petrozzino, C.: Who pays for ethical debt in AI? *AI Ethics.* **1**, 205–208 (2021)
87. Casolari, F., Buttaboni, C., Floridi, F.: The EU Data Act in Context: A Legal Assessment. SSRN (2023). <https://doi.org/10.2139/ssrn.4584781> 27 June 2024
88. Kazım, E., Güçlütürk, O., Almeida, D., et al.: Proposed EU AI Act—Presidency compromise text: Select overview and comment on the changes to the proposed regulation. *AI Ethics.* **3**, 381–387 (2023)
89. Smuha, N.A., Yeung, K.: The European Union's AI Act: beyond motherhood and apple pie? SSRN (2024). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4874852_s=03 Accessed 27 June 2024
90. Heikkilä, M.: Five things you need to know about the EU's new AI Act. *MIT Technology Review*. 11 December. (2023). <https://www.technologyreview.com/2023/12/11/1084942/five-things-you-need-to-know-about-the-eus-new-ai-act/> Accessed 14 December 2023

91. European Consumer Organization: EU rules on AI lack punch to sufficiently protect consumers. (2023). <https://www.beuc.eu/press-releases/eu-rules-ai-lack-punch-sufficiently-protect-consumers?s=03> Accessed 17 December 2023
92. Cabrera, L.L.: EU AI Act Brief– Pt. 2, Privacy & Surveillance, Center for Democracy & Technology. April 30. (2024). <https://www.cnbc.com/2023/01/10/microsoft-to-invest-10-billion-in-chatgpt-creator-openai-report-says.html> Accessed 24 June 2024
93. Hakobyan, M.: EU: Bloc's decision to not ban public mass surveillance in AI Act sets a devastating global precedent. Amnesty International. (2024). <https://www.amnesty.org/en/latest/news/2023/12/eu-blocs-decision-to-not-ban-public-mass-surveillance-in-ai-act-sets-a-devastating-global-precedent/> Accessed 24 June 2024
94. Zuboff, S.: The EU has fired the starting gun in the fightback against Big Tech. Financial Times. 2 May. (2022). <https://www.ft.com/content/31f49915-0f85-48b0-bf81-131960318967> Accessed 28 February 2024
95. Bradley, S.H., Hemphill, S., Markham, S., Sivakumar, S.: Healthcare systems must get fair value for their data. *Bmj* **377**, e070876 (2022)
96. Sloane, M.: Inequality Is the Name of the Game: Thoughts on the Emerging Field of Technology, Ethics and Social Justice. In Proceedings of the Weizenbaum Conference 2019. Challenges of Digital Inequality - Digital Education, Digital Work, Digital Life, pp. 1–9 (2019)
97. Munn, L.: The uselessness of AI ethics. *AI Ethics*. **3**, 869–877 (2023)
98. Morris, S., Murgia, M.: Google's AI search tool tells users to 'eat rocks' for your health. Financial Times. May 24. (2024). <https://www.ft.com/content/13b5b637-f2bb-4208-bed4-2fa760adfb7f> Accessed 27 June 2024
99. : AI Incident Database, About: (2023). <https://incidentdatabase.ai/about/> Accessed 17 November 2023
100. AI-Litigation Database: AI Litigation Database. (2023). <https://blogs.gwu.edu/law-eti/ai-litigation-database/> Accessed 17 November 2023
101. Walsh, D.: The legal issues presented by generative AI. MIT Sloane. (2023). <https://mitsloan.mit.edu/ideas-made-to-matter/legal-issues-presented-generative-ai#:~:text=The%20case%20alleges%20that%20the,train%20models%20infringes%20on%20copyrights> Accessed 17 November 2023
102. King, T.C., Aggerwal, N., Taddeo, M., Floridi, L.: Artificial Intelligence Crime: An interdisciplinary analysis of foreseeable threats and solutions. *Sci Eng. Ethics*. **26**, 89–120 (2019)
103. Omohundro, S.: Autonomous Technology and the Greater Human Good. *J. Exp. Theor. Artif. Intell.* **26**(3), 303–315 (2014)
104. Bostrom, N.: *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, Oxford University Press (2014)
105. Buccella, A.: AI for all is a matter of social justice. *AI Ethics*. **3**, 1143–1152 (2023)
106. McQuillan, D.: Predicted benefits, proven harms. *The Sociological Review*. June. t (2023). <https://thesociologicalreview.org/magazine/june-2023/artificial-intelligence/predicted-benefits-proven-harms/> Accessed 17 November 2023
107. Gready, P., Robins, S.: From transitional to transformative justice: A New Agenda for Practice. *Int. J. Transitional Justice*. **8**(3), 339–361 (2014)
108. Creary, M.S.: Bounded Justice and the limits of Health Equity. *J. Law Med. Ethics*. **49**(2), 241–256 (2021)
109. Nocella, I.I.: An overview of the history and theory of transformative justice. *Peace Confl. Rev.* **6**(1), 1–10 (2011)
110. Christiansen, I.: Commodification of Healthcare and its consequences. *World Rev. Political Econ.* **8**(1), 82–103 (2017)
111. Dholakia, N., Darmody, A., Zwick, D., Dholakia, R.R., Firat, F.: Consumer Choicemaking and Choicelessness in Hyperdigital Marketspaces. *J. Macromarketing*. **41**(1), 65–74 (2021)
112. Bartz, D.: US federal judge rules against Meta in privacy fight with FTC. Reuters. 27 November. (2023). <https://www.reuters.com/legal/us-federal-judge-rules-against-meta-privacy-fight-with-ftc-2023-11-27/> Accessed 28 November 2023
113. Lomas, N.: CJEU ruling on Meta referral could close the chapter on surveillance capitalism. *Tech Crunch*. 4 July. (2023). <https://techcrunch.com/2023/07/04/cjeu-meta-superprofiling-decision/> Accessed 29 November 2023
114. Birch, K., Chiappetta, M., Artyushina, A.: The problem of innovation in technoscientific capitalism: Data rentiership and the policy implications of turning personal digital data into a private asset. *Policy Stud.* **41**(5), 468–487 (2020)
115. Prainsack, B., El-Sayed, S., Forgo, N., Szoszkiewicz, L., Baumer, P.: Data solidarity: A blueprint for governing health futures. *Lancet*. **4**(11), E773–E774 (2022)
116. WHO: Health is a fundamental human right. (2017). <https://www.who.int/news-room/commentaries/detail/health-is-a-fundamental-human-right> Accessed 27 November 2023
117. UN: General's Strategy on New Technologies. (2018). <https://www.un.org/en/newtechnologies/> Accessed 28 June 2024
118. Khlaaf, H.: How AI Can Be Regulated Like Nuclear Energy. *Time Magazine* 23 October. (2023). <https://time.com/6327635/ai-needs-to-be-regulated-like-nuclear-weapons/?s=03> Accessed 17 November 2023
119. European Public Health Alliance: Events. <https://epha.org/events/> Accessed 24 June 2024 (2024)
120. UN: A/HRC/49/52: Artificial intelligence and the rights of persons with disabilities - Report of the Special Rapporteur on the rights of persons with disabilities. United Nations. 28 December (2021). <https://www.ohchr.org/en/documents/thematic-reports/ahrc4952-artificial-intelligence-and-rights-persons-disabilities-report> Accessed 29 November 2023
121. Irish Platform for Patients' Organisations: Science & Industry: 2024 Citizens' Jury on AI in Healthcare. (2024). <https://ipposi.ie/aicitizensjury/> Accessed 27 June 2024

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.