Information and Organization 33 (2023) 100478

Contents lists available at ScienceDirect

Information and Organization

journal homepage: www.elsevier.com/locate/infoandorg

Editorial

Risk and the future of AI: Algorithmic bias, data colonialism, and marginalization

1. Introduction

Artificial Intelligence (AI), and discussions surrounding its potential uses, have escalated into polarised debates about the future, capturing a wide range of utopian and dystopian imaginations (Bove, 2023; Bucknall & Dori-Hacohen, 2022). In this editorial, we focus on the duality of risk around potential benefit and potential harm (Roff, 2019) associated with the future of AI. This is of critical importance in developing our understanding of AI as an emerging technology; one whose uses and effects are still uncertain and have yet to stabilize around a recognizable set of patterns (Bailey, Faraj, Hinds, Leonardi, & von Krogh, 2022). A particular focus is on algorithmic bias which is a direct function of the quality of the data that the AI algorithms are trained on, and that are effective only for those populations where there is access to training data. As we discuss further below, algorithmic bias has important implications for reshaping diversity, inclusion and marginalization. Additionally, we highlight the potential harm experienced by those largely invisible workers in the Global South who clean up data and refine algorithm development for the benefit of those using algorithms in the Global North. This emerging trend of data colonialism (Couldry & Mejias, 2019) is of critical importance given the fundamental reliance of AI technologies on data, and because organizations gain access to, control data, and develop algorithms in ways which not only emphasize the digital divide but increase data inequality (Zheng & Walsham, 2021). Based on these reflections, we propose a relational risk perspective as a useful lens in studying the dark side of AI around algorithmic bias, data colonialism, and increasing marginalization which is largely under-represented in the literature.

Our goal is to build prescient knowledge (Barrett & Oborn, 2018; Corley & Gioia, 2011) regarding how AI and its emerging use may shape possible futures in order to sharpen our critical thinking around the potential risk of emerging data-driven technologies. This is in line with Bailey et al. (2022), placing an emphasis on the importance of considering the consequences of AI in advance, rather than historically, given the rapid rate of its development. Such an emphasis on responsible innovation forewarns us as researchers to not wait until negative consequences are realized, at which point it may be too late to influence the design, intent and outcomes related to the technology.

We start by envisioning AI and specific newfound benefits associated with generative AI, which has triggered the latest hype and debate as to its imagined future. Subsequently, we unpack algorithmic bias and different forms of marginalization as potential harms related to the ongoing development and use of AI technologies. We propose a relational view of risk in conceptualizing AI and emerging futures, that integrates both benefit and harm as a duality, as a useful conceptual approach and conclude by outlining key policy implications.

2. The rise of AI and new forms of risk

AI is conventionally defined as the ability of a computer system to perform tasks that are usually thought to require human intelligence, including learning, reasoning and self-correction (Arora, 2020). Broadly, there are two types of AI: deductive and generative. Deductive AI describes computer systems that analyse large datasets in order to ascertain patterns and derive conclusions. Generative AI, however, learns from existing data in order to produce new content. In recent years, ChatGPT developed by OpenAI, has garnered an incredulous amount of attention since being released to the public in November 2022; over a million users signed up in just five days. In response, rival firms have built similar and competing tools. For instance, Google raced to open public access to its AI

https://doi.org/10.1016/j.infoandorg.2023.100478

Available online 24 August 2023

chatbot service, "Bard" (Alba & Metz, 2023). Similarly, Meta has invested heavily in the development of Llama 2, an open source large language model, with an emphasis on promoting transparency and access (Meta, 2023). With the rise of generative AI, there has been ongoing speculation on the ability of AI to design itself and iteratively improve its own algorithms in a self-improvement loop, leading to an artificial general intelligence (AGI) surpassing that of humans, with implications for how future work becomes organised. Whilst these ideas have historically been limited to the realms of science fiction, the ability of large language models to write computer code surpassing the capabilities of many data scientists, has intensified the excitement and concern about AGI. Further, as AI becomes increasingly capable, there is a push to progressively integrate AI with different technologies to realize synergies, for example, with robotics and medicine.

However, despite these significant developments, one notable limitation is the availability of suitable training data in AI algorithms. Even large language models, which are trained on large swathes of the internet have been heavily scrutinised for their scraping of copyrighted material (Creamer, 2023). The output of an algorithm is a function of the quality of its input (Sarker, 2021). The quality and availability of data, needed to build and train the algorithms has led to growing concerns relating to, for example, algorithmic bias, data ownership and misuse. Taken together, the implementation of AI has important implications for risk and shaping the future of work and society more broadly.

3. AI risk of algorithmic bias on diversity and marginalization

Algorithmic bias is the phenomenon by which an algorithm may perform particularly poorly on a population subgroup if it was not exposed to that subgroup's data during algorithm development and training (Lovejoy, Arora, Buch, & Dayan, 2022). One area where this emerging issue has been highlighted is through health data poverty, whereby marginalised patient groups tend to be underrepresented in data used for health research (Ibrahim, Liu, Zariffa, Morris, & Denniston, 2021). As such, concerns have been raised about how AI and machine learning may be exacerbating healthcare inequalities by underperforming in marginalised patient groups (Panch, Mattie, & Atun, 2019) posing new risks for medical science being developed as well as future treatment effectiveness. For example, algorithmic bias has been well-described in the context of ophthalmological care, which already exhibits social injustice, based on geography within and between countries as well as by socioeconomic status and ethnicity (Campbell et al., 2021).

One of the first real-world implementations of AI in healthcare is the use of machine learning algorithms to diagnose diabetic retinopathy from retinal fundus photos (Ting et al., 2017). While some literature has suggested that the capability of AI to interpret these images may exceed that of human ophthalmologists, recent studies note AI risks systematically underperforming for marginalised patient subgroups, if they are not adequately represented in the data used to train the algorithms (Chu, Squirrell, Phillips, & Vaghefi, 2020). Recently, Burlina, Joshi, Paul, Pacheco, and Bressler (2021) constructed a dataset of retinal fundus photos that specifically excluded dark-skinned patients, and trained a machine learning algorithm on this biased dataset to detect diabetic retinopathy (Burlina et al., 2021). They found that the machine learning algorithm was only 60.5% accurate at detecting retinopathy in dark-skinned patients, compared to 73.0% accurate in light-skinned patients. This is because there are physiological differences between dark-skinned and light-skinned fundi that a machine learning algorithm would not understand if it had not been exposed to the data.

4. AI risk and promoting inclusion through synthetic data

One potential technical solution to the above issues of health data poverty and algorithmic bias has emerged – synthetic data. While traditional deductive machine learning systems learn from data in order to identify patterns and elicit conclusions when presented with similar data in the future, generative systems can learn from data in order to create novel synthetic data that resembles the original dataset. While this method of computer-generated data has been used to create 'deepfakes', with significant potential for harm, in the field of ophthalmology it has been used to mitigate algorithmic bias by creating synthetic fundus photos. In this way, Burlina et al. (2021) attempted to overcome the aforementioned example of algorithmic bias in ophthalmology using synthetic data. This was based on the use of generative adversarial networks (GANs), which function by learning patterns within real datasets in order to produce highly realistic synthetic data not relating to real individuals (Arora & Arora, 2022). In this case, the GAN was used to produce synthetic fundus photos corresponding to dark-skinned patients in order to balance a dataset. These synthetic images could then be used to rebalance the dataset and allow the algorithm to train on dark-skinned patient images. Rebalancing the dataset in this manner increased the accuracy of diagnosing diabetic retinopathy from 60.5% to 71.0% in dark-skinned patients. As such, there is potential for GANs to be used to create data corresponding to marginalised populations for training algorithms, in order to minimise the risks of algorithmic bias towards marginalised populations. Other uses for GANs include privacy preservation, dataset augmentation and debiasing datasets (Arora & Arora, 2022).

In searching for solutions to the problem of algorithmic bias, the focus has been on examining the fairness of decisions made by algorithms, including proactive and reactive oversight (Teodorescu, Morse, Awwad, & Kane, 2021). Whilst such methods of generating synthetic data to balance datasets is promising, it still requires that there is an existing set of data available for the marginalised groups from which to resynthesize more images. Furthermore, if images acquired by low-cost devices in low-middle income countries are not compatible with higher quality fundus photos obtained from high income countries, separate algorithms may be required. This highlights the ongoing lack of solutions that might prevent bias from developing in the first instance. An important influence on risks associated with future work, such as medicine, is to address the underlying causes of bias, which are often rooted in structural inequalities within society (Zheng & Walsham, 2021). For example, the COVID-19 pandemic drew light to the issue of maldistribution of digitalization across populations at a time when it was needed to access basic societal services, with a lack of robust digital infrastructure or structured data emerging as an ongoing risk in AI development (Faraj, Renno, & Bhardwaj, 2021; Lee et al., 2022).

5. AI risk and increased marginalization of the global South

In the above sections, we have stressed the point that the risks around what an AI-based system can do, or represents, is ultimately dependent on the data used to train it. Given this, the issue of data colonialism and marginalization becomes a pertinent point to consider.

6. The rise of data colonialism as a skewed value exchange across North-South

The term "data colonialism" was coined by Couldry and Mejias (2019), as a distinctive twenty-first century parallel to historical colonialism albeit achieved through the abstract quantification methods of computing. The concept refers to the domination and control of data and data flows by powerful countries, corporations or entities in "the Global North" over those in "the Global South". Data, unlike other historical natural products (in relation to colonialism), is developed through a highly socialized process as it becomes captured, selected, and appropriated (Couldry & Mejias, 2019). The process of the Global North appropriating and enabling data extraction for commodification from the Global South creates new human "data relations" and a new social order; even social life – captured through data – can become a resource that is mined and extracted by those with power and access for capital across the globe. In other words, the exploitation of human beings through data is becoming naturalized in data colonialism, creating new forms of risk in relation to the appropriation of data for profit.

This concept is significant in the discussion of AI futures, as it highlights the potential risks around increased social discrimination, behavioural influence, and marginalization of economies. Data colonialism brings a new power imbalance in the global data landscape, where the Global North can disproportionately control global data flows, and thus influence the global digital economy and data landscape (Kwet, 2023). An important dynamic shaping the inherent risks is the exchange of value between the colonizer and the colonized; while the Global North is able to exploit data from less developed countries, they have often been accused of providing inadequate compensation in return.

This skewed value exchange can occur at several levels, and the Sama controversy brought to light by a TIME investigation is one case that illustrates this well. Sama is a self-proclaimed "ethical AI" company headquartered in California. They provide image, video and sensor data content moderation, data-labelling and annotation services to train machine learning algorithms for large companies such as Meta, Google, and Microsoft (Perrigo, 2022). Recent controversy surrounds OpenAI's outsourcing of Kenyan laborers via Sama. ChatGPT's predecessor, GPT-3, had issues with generating violent, sexist and racist remarks, due to the AI having been trained on hundreds of billions of words scraped from the Internet (Perrigo, 2023). The vast repository of data would have taken humans far too long to purge manually, and as such OpenAI opted to build an additional AI-powered safety mechanism to detect and eliminate toxicity and bias from the platform. As part of this process, Kenyan Sama workers would feed the AI with labelled examples of inappropriate materials such as child sexual abuse, hate speech, suicide, torture and bestiality (Perrigo, 2023). However, several issues arose from these contracts.

First is the issue with exploitation linked to precarious working circumstances. Sama employees have emphasized the mental trauma arising from their work as well as the alleged suppression of their unionization rights. They have shared how part of their work involves watching distressing videos of murders, suicides, rapes, and child sexual abuse, which has led to mental illnesses (PTSD, anxiety, depression) among workers; at the same time workers are unable to afford quality mental healthcare. Sama workers have expressed how "the work that [they] do is a kind of mental torture…feel like it's modern slavery, like neo-colonialism" (Perrigo, 2022).

Second, there is the exploitation of marginalised workers via unfair monetary compensation. Although Sama claims to be bringing progress and to have lifted more than 50,000 out of poverty in the Global South, workers in the Nairobi office are among the lowest-paid Sama workers globally, and data labellers employed by Sama on behalf of OpenAI were paid $1.32 - $2 per hour (Perrigo, 2023). Put together, Perrigo (2023) highlights the darker picture that this controversy revealed: "that for all its glamor, AI often relies on hidden human labor in the Global South that can often be damaging and exploitative. These invisible workers remain on the margins even as their work contributes to billion-dollar industries".

This leads to a third point, that is how data colonialism widens gaps in digital maturity and technological progress (Lee, Barrett, Prince, Oborn, Li, & Thomas, 2022; Lee & Lee, 2022) via disproportionate technological returns. For instance, according to 2021 UN data, in sub-Saharan Africa, 89% of learners did not have access to household computers and 82% lacked internet access (Mo Ibrahim Foundation, 2021). Putting this into perspective in relation to the Sama example, this illustrates the reinforcement of marginalization via AI developments, particularly via social exclusivity and inequality of access (Lee, Barrett, Prince, & Oborn, 2022). Marginalised workers, especially in developing countries, are working on these cutting-edge AI systems for the Global North to enable further AI and technological development. However, they are not only getting paid low wages and working under precarious circumstances; the benefits from these systems are often not able to be reaped by the marginalised populations themselves. In this way, while such progression leads developed economies onto the next wave of development, the Global South is left further behind without being able to benefit from these developments, and the technological and productivity gap between the two widens.

As such, we suggest that an imbalance of risk and harm in AI development is reinforcing marginalization. As AI products progress further based on unrepresentative datasets and technological infrastructure that is not available to all, the represented populations will gain benefits and advantages over the marginalised, and in the process create even more value and wealth. While in the short term marginalised populations may be provided with temporary resources in the form of low-pay and precarious work, ultimately, this can result in an exacerbation of inequality and marginalization in the longer run where the marginalised end up in an even less equitable position. In short, AI can further marginalise the marginalised, and a more meaningful way of examining and understanding such risks is necessary in order to manage this dynamic of complex, inter-related benefit and harm.

7. Theoretical developments & implications for practice

7.1. A relational view of risk and the future of AI

We propose a relational perspective to envisioning the future of AI as an emerging technology, and in conceptualizing risk as a duality that integrates both benefit and harm (Bednarek, Chalkias, & Jarzabkowski, 2021). Earlier literature has highlighted the importance of understanding the imbricating relationship of digital technology and risk (Ciborra, 2006). For example, in financial markets, digital technology can be beneficial in serving as a risk management tool that identifies, assesses, and mitigates risk while also being a source of new risk providing potential harm to market participants. Therefore, the extent to which technology is a "risk object" depends on perception, and the act of designating riskiness to technology is ultimately a creative act situated within a social and cultural context (Boholm & Corvellec, 2011). Hilgartner's (1992) concept of emplacement highlights actors' roles in defining and constructing an object risk within sociotechnical networks. Here, while an entity can be assigned (or "emplaced") as a risk object that has the ability to influence the future of the network, the reverse (or "displacement") can also occur, where an object's risks are neutralized or its capacity to influence the network removed. This relational understanding of risk continues to be prevalent, and scholars such as Boholm and Corvellec (2011) have called for further research on the relationships and semantic associations between risk objects and objects at risk. The rise of (generative) AI and its simultaneous implications for marginalization has highlighted the necessity to understand better the dynamic nature and relationships of the technology within an ongoing ecology of risks.

AI as a future emerging technology can be conceptualized as a risk object (Barrett et al. 2023), whose meaning is fluid and evolves over time (Bauman, 2005) rather than fixed. While emerging technologies such as AI are often portrayed as beneficial for the future, through their ongoing development and use, new harms are often simultaneously identified, or emphasized. As "new" risks emerge, complex dynamics within the technology's wider ecology of risk develop; and what was once potentially perceived as mostly a transformative technology can evolve to become a potential catalyst for harm and destruction.

It is through this duality of risk (Hilgartner, 1992), consisting of interdependent yet contradictory elements of benefit and harm, that we understand how emerging technologies as risk objects are constructed and unfold over time. That is, beyond the identification of the risks and benefits that any technology can pose, it is a relational, situated understanding of how this duality and relationship is socially constructed, that allows for actions to be taken to displace identified risks and retain benefits for the future. In the ongoing dynamic construction of AI and its potential futures as a risk object, organizations retain significant agency through their prioritization of either harm or benefit, and crucially how the prioritization of one can be a way to access the other (Bednarek et al., 2021).

Take, for example, AI futures. Over the last decade, AI, now in its third generation as a technology, has become increasingly important and pervasive. We are constantly aware of it being a part of our daily lives, whether it be in feeding us social media videos, or increasingly being the front line of customer service helping us to resolve customer service issues as it seems surprisingly convincing and human like while displaying unexpected novelty. However, the recent launch of ChatGPT, has had escalating attention and heightened promise for an AI future (Logue & Grimes, 2022). Generative AI, is widely viewed to be a new type of breakthrough AI technology which could disrupt a wider range of economic tasks in the domain of creativity that were previously seen to be at low risk of automation. Unlike past waves of automation which have focused on the use of AI and robotics to perform an increasing scope of routine tasks, creative tasks are now open to being displaced (e.g. Hollywood scriptwriters, artists). With generative AI technologies, changes in work and social fabric associated with AI and its future is happening in ways unimaginable.

However, alongside the anticipated benefit of generative AI, there are often-understated harms that evolve with these applications which are both undefined and uncertain. Viewing AI as a responsible innovation is increasingly recognized as concerns related to inadequate recognition of the potential harms are brought to light (Caulfield & Condit, 2012). We suggest that a relational perspective of risk for understanding AI and its future would provide a lens to unpack the sociotechnical relationships of technology policy makers, AI users, the wider (marginalised) population, and our broader perception of AI as risk objects. What may currently be perceived as unacceptable risks need to be "displaced" or neutralized via policy actions to manage the risks, such that potential harms are mitigated and benefits reaped. We suggest that this recognition of the potential harm as part of the duality of risk associated with AI requires novel policy initiatives and regulatory frameworks.

8. Practical implications for regulating AI risks in the future

8.1. Data protection for AI as an emerging technology

The development of AI raises questions about how existing data protection frameworks apply to this emerging technology in order to manage the widening ecology of risk. AI development fits into existing data protection frameworks in two ways: firstly, to the collection, use, and sharing of personal data used to train AI systems, and secondly, AI systems themselves must comply with data protection law when they process personal data. Each of these levels of data responsibility connect to different forms of risk. The former data related risk is often an underlying reason for some people and subgroups to express new vulnerabilities around their data (Lee, Barrett, Prince, & Oborn, 2022), while the latter around the agency of the algorithm raises different forms of accountability and responsibility around the risk objects, often requiring new categories in the legal domain.

In the European Union and the United Kingdom, for example, the General Data Protection Regulation (GDPR) applies to the collection, use, and sharing of personal data used to train AI algorithms. This means that organizations must obtain valid consent for the collection and use of personal data, and must provide individuals with access to their personal data, the right to rectify inaccurate data, and the right to be forgotten. AI algorithms, themselves, must also comply with data protection law when they process personal

data. This means that AI must be designed to respect the rights of individuals and must be transparent in their agentic decision-making processes. They must also provide individuals with meaningful information about how their personal data is being used and must ensure that appropriate safeguards are in place to protect personal data. Further legislation being developed in this area, increasingly needs to account for the duality of risk associated with the wider ecology of risks discussed in this paper.

8.2. Policy initiatives and regulatory structures

We suggest three important ways in which policy and regulatory structures need to become responsive to emerging AI technologies and their risk dynamics.

First, there is a growing role for policy to become responsive in defining what the emerging AI risks are, as well as demystifying those risks deemed less salient. We suggest policy initiatives need to understand risk as a two-way process, of delineating the harmful risks that are being reckoned with and emplaced, whilst simultaneously, removing and displacing the capacity of harmful risks in influencing work and broader civil society. In delineating risk, both harm and benefit, policy initiatives should locate the identified risks within their broader relations, both social and technical. This entails understanding how the risks are being constructed, how the risks are becoming defined (and by whom), who the risk-bearers are, and the mechanisms that are linking them to potential harm. In so doing, policy can more readily place control mechanisms as it seeks to govern emerging risk, rendering them with less capacity to influence dynamic relations within the wider social fabric. By clarifying and governing risks across a network of risk relations, the risk outcomes can be made more reliable, understood and predictable.

Second, policy processes require a structure that is increasingly responsive to the evolving and dynamic nature of AI as an emerging future. Rather than being a static entity, policy makers need to find mechanisms to become more dynamic, responding to needs as they become apparent. In this sense, policy should be processual in orientation and develop learning capabilities (Lee & Lee, 2022), drawing on the present and past to inform the future. Given the ongoing generativity of AI technologies, and thus the changing risks, policy itself needs to be correspondingly dynamic. In this regard, we suggest that policy initiatives could draw more on real world data to inform on emerging benefit and harm of AI use, responding and adjusting in an iterative and cyclical manner (Lee, Barrett, Prince, Oborn, Li, & Thomas, 2022). To do so, policy decision makers might draw more effectively on real world evidence in informing policy content. For example, social media can point to trends such as alienation, loss of respect or dignity or perspectives on marginalization. While such real-world data needs to be considered and acted on with care, this should be done with awareness of possible data traps and being mindful of unanticipated consequences as they arise.

Third, current policies on data protection need to be revised, accounting for the harm and benefit, while holding AI technologies and their development to account within its framework. While existing data protection frameworks provide a reasonable foundation for the responsible development of AI, issues of transparency and confidentiality remain inadequately answered and pose unanticipated risks. For example, as foundational models such as Chat-GPT or DALL-E2 begin to mass-produce text, software code and images that are then used to train further models, the issue of data ownership becomes increasingly controversial. In the context of healthcare, it has already been suggested that generating synthetic medical records, imaging or tabular data could be used as a mechanism of evading data privacy legislation (Arora & Arora, 2022). Halting these machine learning systems is undesirable because they do carry positive benefits. The same algorithms that could be used to generate synthetic records to evade GDPR may be applied to rebalance or augment datasets, anonymise data for legitimate purposes, and in so doing enable the pooling of data between silos and enhance medical education.

In line with our emphasis of a relational perspective to AI risks, it is important to note how countries and regions may take disparate positionings when seeking to engage with AI through policy. For instance, some such as the UK may choose to foreground the benefits of AI innovation with a "pro-innovation approach" (DSIT, 2023), while other regions such as the EU may choose to initially focus on displacing AI risks as much as possible (Engler, 2023). Therefore, when considering technology policies, it is pertinent to be cognizant of specific regions' positioning towards innovation and the technology, the perceptions of the policymakers, and how together this dynamic influences the ultimate future of emerging technologies such as AI.

Finally, future research could usefully address other key implications for risk and AI futures. For example, identifying new and reviewing existing marginalised groups facing increasing risks with the wider implementation of AI. This includes marginalised groups within the Global North that are unable to voice concerns through a lack of knowledge or an appropriate platform. Research into mechanisms for maintaining visibility of both harms and benefits could be useful for encouraging change in AI design decisions and ensuring marginalised groups can engage with their own futures.

Additionally, as AI technologies become more effective, there may be less incentivisation for individuals and organizations to increasingly rely on their own professional judgement. This approach can lead to enfeeblement of rights (especially of the marginalised) and over dependence on AI in defining core values, judgments thereby shaping social boundaries. There is also growing concerns about misinformation and the targeted disinformation campaigns that create division and unrest while hindering collective decision making. Concerns about the representation of the marginalised and/or minority groups and the Global South both in the development and uptake of AI systems are increasingly gaining attention. New systems of power and domination are arising in society, as AI systems are created by fewer and fewer, and thus values become narrower, and interests may become less aligned to the public good. These influences require ongoing policy consideration and leadership. Such instances have led to the calls for more equitable data governance frameworks, that would protect underrepresented individuals and communities in the global data landscape through policy and regulation.

9. Conclusions

Our paper has highlighted the future of AI as an emerging technology which is reshaping diversity, inclusion and marginalization through its algorithmic design and supported by contentious practices of data colonialism. We have proposed a relational risk perspective to understand how AI, as a risk object, evolves in relation to a dynamic risk ecology with implications for its emergent future.

We conclude by reflecting on the implications of these data practices and associated marginalization for AI ethics. Algorithmic bias can have adverse implications for inclusivity and diversity, which raises important ethical issues. To help address these challenges, ethics-by-design approaches can, for example, use synthetic data as highlighted in the abovementioned example for addressing algorithmic bias in ophthalmology. However, there is an urgent need for research that investigates the ethical challenges associated with practices that are global in scale and influence the design and development of AI. Our discussion of data colonialism, particularly in relation to AI, showcases the marginalised in the Global South becoming increasingly marginalised through increasing digital inequality over time. We contend that current ethics by design approaches, while important, may not be sufficient in addressing the consequences of these ethical issues and dilemmas for society.

We suggest that studies could usefully adopt a practice approach to ethics to complement existing ethics-by-design methodologies. This perspective would view ethics as a situated doing, and would focus on understanding how ethics becomes exposed, produced, developed, and integrated in the AI practices comprising data production, synthesizing of data sets, model development, implementation and use of algorithms. These practices would require studies that widen the scope of investigation beyond design and development (Bailey et al., 2022) and include the sociotechnical (eco)system. For example, in studying ethical issues surrounding data colonialism, we could usefully zoom out to examine these emerging practices across space to expose relations of power between Global North promoters and designers of data practices in algorithm development, and those invisible workers for whom the work of cleaning and refining dirty data has been delegated. Our relational risk perspective, a practice approach, would focus on how AI practices are producing harms in the Global South while providing benefit to the Global North.

We suggest that such a relational view to risk can also inform policymakers in their current endeavours to create novel AI legislation. Upcoming AI policies need to be designed to account for the fluidity of this emergent technology and allow for its evolution alongside individual countries and regions' growth and development of digital maturity. Additionally, beyond the consideration of data rights and ethics of current citizens and users of AI within each region, policies also need to account for the rights and ethics of those that are developing such technologies – particularly the marginalised; data colonialism's implications for widening inequalities and furthering marginalization are profound and require more attention from policymakers. To this end, taking a relational view of risks would allow policymakers to consider what the perceived risk objects are, who the objects at risk are, and how policies can influence the varying and complex interplay between the two. Such perspectives will also help to illuminate the difference in approaches to AI regulations across various regions, the recognition of which will aid in furthering global collaboration on AI and related ethical development.

Finally, an effective practice approach to ethics will require multidisciplinary collaboration between important stakeholders. This includes the innovators (academia or industry), policymakers, institutional organizations collecting the data and, importantly, the public. Our prescient viewpoint would suggest that we avoid seeing the future as an eventuality where the Global South outsources AI innovation to the Global North, even if it may perhaps be the easiest path to spreading the benefits of AI. In order to ensure that marginalised populations do not become detached from AI innovation, we must ensure that they are included in the conversation about its development, not just its dissemination.

Acknowledgements

E. Oborn is supported in part by the National Institute for Health Research Applied Research Centre West Midlands.

References

Alba, D., & Metz, R. (2023). Google Opens Access to Bard AI Chatbot, Racing to Catch Up to OpenAI. Bloomberg.com. [Online] 21st March. [Accessed on 27th April 2023] https://www.bloomberg.com/news/articles/2023-03-21/google-chatgpt-rival-bard-now-open-to-public-use.

Arora, A. (2020). Conceptualising artificial intelligence as a digital healthcare innovation: An introductory review. In *Medical devices: Evidence and research* (pp. 223–230). Taylor & Francis.

Arora, A., & Arora, A. (2022). Generative adversarial networks and synthetic patient data: Current challenges and future perspectives. *Future Healthcare Journal. Royal College of Physicians, 9*(2), 190.

Bailey, D. E., Faraj, S., Hinds, P. J., Leonardi, P. M., & von Krogh, G. (2022). We are all theorists of technology now: A relational perspective on emerging technology and organizing. *Organization Science. INFORMS, 33*(1), 1–18.

Barrett, M., & Oborn, E. (2018). Bridging the research-practice divide: Harnessing expertise collaboration in making a wider set of contributions. *Information and Organization. Elsevier, 28*(1), 44–51.

Barrett, M., Oborn, E., Prince, K., & Lee, E. (2023). A relational perspective on digital technologies and risk: The (un)imagined adoption of telemedicine. In *EGOS Conference, July 2023, Sardinia, Italy*.

Bauman, Z. (2005). Liquid life. *Cambridge: Polity Press*.

Bednarek, R., Chalkias, K., & Jarzabkowski, P. (2021). Managing risk as a duality of harm and benefit: A study of organizational risk objects in the global insurance industry. *British Journal of Management. Wiley Online Library, 32*(1), 235–254.

Boholm, Å., & Corvellec, H. (2011). A relational theory of risk. *Journal of Risk Research. Taylor & Francis, 14*(2), 175–190.

Bove, T. (2023). Microsoft Ceo Says A.i. Could Help Create "Utopia," but We Need to Watch Out for the Risks. In *Fortune* [Accessed on 7th August 2023] https://fortune.com/2023/02/09/microsoft-nadella-ai-could-help-humanity-create-utopia/.

Bucknall, B. S., & Dori-Hacohen, S. (2022). Current and near-term AI as a potential existential risk factor. In *Proceedings of the 2022 AAAI/ACM conference on AI, Ethics, and Society* (pp. 119–129).
Burlina, P., Joshi, N., Paul, W., Pacheco, K. D., & Bressler, N. M. (2021). Addressing artificial intelligence Bias in retinal diagnostics. *Translational Vision Science & Technology. The Association for Research in Vision and Ophthalmology, 10*(2), 1–15.
Campbell, J. P., Mathenge, C., Cherwek, H., Balaskas, K., Pasquale, L. R., Keane, P. A., & Chiang, M. F. (2021). Artificial intelligence to reduce ocular health disparities: Moving from concept to implementation. *Translational Vision Science & Technology. The Association for Research in Vision and Ophthalmology, 10*(3), 1-10..
Caulfield, T., & Condit, C. (2012). Science and the sources of hype. In , *Vol. 15 (3–4). Public health genomics* (pp. 209–217). Switzerland: S Karger AG Basel.
Chu, A., Squirrell, D., Phillips, A. M., & Vaghefi, E. (2020). Essentials of a robust deep learning system for diabetic retinopathy screening: A systematic literature review. *Journal of Ophthalmology. Hindawi Limited, 2020*, 1–11.
Ciborra, C. (2006). Imbrication of representations: Risk and digital technologies. *Journal of Management Studies. Wiley Online Library, 43*(6), 1339–1356.
Corley, K. G., & Gioia, D. A. (2011). Building theory about theory building: What constitutes a Theoretical contribution? *Academy of Management Review Academy of Management Briarcliff Manor, NY, 36*(1), 12–32.
Couldry, N., & Mejias, U. A. (2019). Data colonialism: Rethinking big data's relation to the contemporary subject. *Television & New Media. SAGE Publications, 20*(4), 336–349.
Creamer, E. (2023). Authors file a lawsuit against openai for unlawfully "ingesting" their books. In *The Guardian. Books* [Online] 5th July. [Accessed on 7th August 2023] https://www.theguardian.com/books/2023/jul/05/authors-file-a-lawsuit-against-openai-for-unlawfully-ingesting-their-books.
DSIT. (2023). A pro-innovation approach to AI regulation. In *GOV.UK* [Accessed on 7th August 2023] https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper.
Engler, A. (2023). *The EU and U.S. diverge on Ai regulation: A transatlantic comparison and steps to alignment*. Brookings [Accessed on 7th August 2023] https://www.brookings.edu/articles/the-eu-and-us-diverge-on-ai-regulation-a-transatlantic-comparison-and-steps-to-alignment/.
Faraj, S., Renno, W., & Bhardwaj, A. (2021). Unto the breach: What the Covid-19 pandemic exposes about digitalization. *Information and Organization. Elsevier, 31*(1), Article 100337.
Hilgartner, S. (1992). The social construction of risk objects: Or, how to pry open networks of risk. In J. F. Short, & L. Clarke (Eds.), *Organizations, uncertainties, and risk*. Westview Press.
Ibrahim, H., Liu, X., Zariffa, N., Morris, A. D., & Denniston, A. K. (2021). Health data poverty: An assailable barrier to equitable digital health care. *The Lancet Digital Health. Elsevier, 3*(4), e260–e265.
Kwet, M. (2023). Digital colonialism is threatening the global South [Accessed on 8th August 2023] https://www.aljazeera.com/opinions/2019/3/13/digital-colonialism-is-threatening-the-global-south.
Lee, E. L. S., Barrett, M., Prince, K., & Oborn, E. (2022). *Developing your digital maturity for competitive advantage: From models to practices in enabling digital transformation*. Centre for Digital Built Britain. September.
Lee, E. L. S., Barrett, M., Prince, K., Oborn, E., Li, O., & Thomas, P. (2022). *Digital twins and service innovation in designing the Hospital of the Future*. Centre for Digital Built Britain. July.
Lee, E. L. S., & Lee, W. O. (2022). Enabling continuous innovation and knowledge creation in organizations: Optimizing informal learning and tacit knowledge. In K. Evans, W. O. Lee, J. Markowitsch, & M. Zukas (Eds.), *Third international handbook of lifelong learning*. Cham: Springer International Publishing (Springer International Handbooks of Education).
Logue, D., & Grimes, M. (2022). Living up to the hype: How new ventures manage the resource and liability of future-oriented visions within the nascent market of impact investing. *Academy of Management Journal. Academy of Management Briarcliff Manor, NY, 65*(3), 1055–1082.
Lovejoy, C. A., Arora, A., Buch, V., & Dayan, I. (2022). Key considerations for the use of artificial intelligence in healthcare and clinical research. *Future Healthcare Journal. Royal College of Physicians, 9*(1), 75–78.
Meta. (2023). Meta and microsoft introduce the next generation of Llama. In *Meta* [Online] 18th July. [Accessed on 7th August 2023] https://about.fb.com/news/2023/07/llama-2/.
Mo Ibrahim Foundation. (2021). COVID-19 in Africa. One year on: Impact and prospects. In *Mo Ibrahim Foundation 2021 Forum Report*. https://mo.ibrahim.foundation/sites/default/files/2021-06/2021-forum-report.pdf.
Panch, T., Mattie, H., & Atun, R. (2019). Artificial intelligence and algorithmic bias: Implications for health systems. *Journal of Globalization and Health International Society for Global Health, 9*(2).
Perrigo, B. (2022). *Inside Facebook's African Sweatshop*. TIME [Online] 14th February. [Accessed on 23rd May 2023] https://time.com/6147458/facebook-africa-content-moderation-employee-treatment/.
Perrigo, B. (2023). *Exclusive: The $2 Per Hour Workers Who Made ChatGPT Safer*. TIME [Accessed on 23rd July 2023] https://time.com/6247678/openai-chatgpt-kenya-workers/.
Roff, H. M. (2019). Artificial intelligence: Power to the people. *Ethics & International Affairs. Cambridge University Press, 33*(2), 127–140.
Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer science. Springer, 2*(3), 160.
Teodorescu, M. H., Morse, L., Awwad, Y., & Kane, G. C. (2021). Failures of fairness in automation require a deeper understanding of human-ML augmentation. *MIS Quarterly, 45*(3).
Ting, D. S. W., Cheung, C. Y.-L., Lim, G., Tan, G. S. W., Quang, N. D., Gan, A., … Lee, S. Y. (2017). Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA. American Medical Association, 318*(22), 2211–2223.
Zheng, Y., & Walsham, G. (2021). Inequality of What? An intersectional approach to digital inequality under COVID-19. *Information and Organization. Elsevier, 31*(1).

A. Arora[a], M. Barrett[b,*], E. Lee[c], E. Oborn[d], K. Prince[c]
[a] *School of Clinical Medicine, Cambridge University, United Kingdom*
[b] *Judge Business School, Cambridge University, United Kingdom*
[c] *CDI, Cambridge University, United Kingdom*
[d] *Warwick Business School, University of Warwick, United Kingdom*

[*] Corresponding author.
E-mail address: m.barrett@jbs.cam.ac.uk (M. Barrett).