Responsible AI Practices: Histories, Definitions, Barriers and Future Directions

Lorenn P. Ruster

School of Cybernetics, Australian National University
Birch Building, 35 Science Road,
Canberra, ACT 2601, Australia
lorenn.ruster@anu.edu.au

Abstract

Responsibility practices in organizations are not new; they far precede the recent introduction of artificial intelligence (AI). At the same time, how to enact responsible AI in practice is immature and emerging. This paper explores and synthesizes literature on responsible AI practices through technical and sociotechnical lenses. Four common barriers to responsible AI practice are identified: (perceived) neutrality, power to act, modularity and the supply chain metaphor, and organizational culture. When these barriers and the current technical and sociotechnical approaches to responsible AI practices are mapped to a framework for systems intervention in responsible AI, a mismatch is revealed: barriers to enacting responsible AI practices operate across all four leverage zones, but technical approaches do not. Six future directions embracing sociotechnical approaches to responsible AI practice are identified from the literature – design practices, ethics-as-a-service, located accountability practices, Indigenous approaches, integrative approaches and organizational culture approaches. All in all, this work exposes a fundamental mismatch between the systemic nature of responsible AI barriers and the narrow focus of existing technical approaches, advocating for renewed focus on sociotechnical interventions alongside technical ones.

Introduction

What responsible technological systems look like is a question that persists throughout human history, and has preoccupied a range of scholars, practitioners and policymakers across disciplines and contexts. While questions of responsibility and ethics relating to computing go back decades (Maner 1978), and have been explored through philosophies and methods spanning value sensitive design (Friedman and Hendry 2019), systems theory (Midgley 2000) and sociotechnical systems design (Mumford 2000), these questions have renewed urgency, as artificial intelligence (AI) systems are more and more embedded in the social and economic fabric of society.

If we take as a given the interactive stance of technologies (Friedman and Hendry 2019) – that humans are shaping technologies and technologies are shaping humans – then it comes as no surprise that responsibility considerations accompany the advent of any new technology, including AI. Indeed, many responsibility concerns regarding new technologies are discussed both in the academic literature as well as in the mainstream media, such as the responsible age for children to be engaging with social media (Long 2024), the impact of AI model development on the environment, particularly in terms of energy and water consumption (Li et al. 2023), and the role of AI in justice, banking, provision of government services etc. (Eubanks 2017; Noble 2018; Benjamin 2020; Felstead, Stockdale, and Scheepers 2023; Marjanovic, Cecez-Kecmanovic, and Vidgen 2021).

Responsible technology practices are not a new concept; they can be found in the fields of legal compliance, privacy, corporate social responsibility, sustainability, labor conditions, diversity, equity and inclusion, anti-slavery etc. (Polonsky and Jevons 2009; Graafland and Smid 2019; Treviño et al. 1999; Mulligan and Elaluf-Calderwood 2021; Bromley and Powell 2012). For example, back in the 1990s, Weaver, Treviño, and Cochran (1999) reviewed the Fortune 1000 firms as listed in 1994, to understand their corporate ethics activity. They found a high degree of adoption of low cost activities, namely ethics policies and codes but a large variability in the extent to which these policies and codes are put into practice, rendering the policies and codes as largely symbolic gestures. They conclude that "corporate ethics programs are not self-sufficient; they depend heavily for their success on support from other organizational systems and informal norms and practices" (p. 293). They go on to specify that understanding the relationships of ethics programs and policies to organizational activities and structures is critical and there needs to be more emphasis on the role of informal practices in the implementation of ethics activities in an organization.

Thirty years later, and now with the advent of AI in so many organizations, the story is a similar one. For example, Bughin (2024)'s more recent quantitative study analyzes the extent to which large global firms have implemented responsible AI and the antecedents to such action. They find that "extensive operationalization remains rare", with a decoupling between design (and a tendency to overengineer principles) and implementation. The challenges associated with the "decoupling" of principles and practices is also referred to as "ethics-shopping" or "ethics washing" (Floridi 2019). As Rakova et al. (2021) summarize:

> All of these domains [legal compliance, privacy, di-

> versity and inclusion] appear to have gone through a process that is mirrored in current algorithmic responsibility discussions: publication of high-level principles and values by a variety of actors, the creation of dedicated roles within organizations, and urgent questions about overcoming challenges and achieving "actual" results in practice and how to avoid investing in processes that are costly but do not deliver beyond cosmetic impact. (p. 4)

Enacting any sort of responsibility in practice is a gnarly, complex and long-standing problem; responsible AI practice is no exception. As AI and its benefits and harms proliferate, understanding and engaging with what responsible AI practices look like is increasingly critical. This paper begins by illustrating the long-standing interconnection between technology and responsible practice, demonstrating the enduring importance of sociotechnical approaches. After providing a working definition of responsible AI practice, it organizes current responsible AI practices according to two approaches: technical and sociotechnical. A synthesis of the academic literature around common barriers to responsible AI practice follows, and identified barriers and approaches are mapped according to Nabavi and Browne (2023)'s 5P leverage zone framework, showing the need for sociotechnical approaches. Future directions that align with the sociotechnical approach are then summarized from the literature.

Responsible Technology Practices Across Time

Responsible technology practices are not a new phenomenon. This section explores approaches and lessons of the past – namely Indigenous perspectives on responsible technology, which have existed for millennia and persist today, and learnings from other significant technologies such as the steam engine and the internet.

Indigenous Perspectives on Responsible Technology

For many Indigenous cultures[1], the creation of technologies is often done with responsibility considerations intertwined from the outset; a need to separately consider responsibility is counterintuitive because "technology is a cultural practice... contingent on social systems and structures" (Harle, Abdilla, and Newman 2018, p. 12). For example, for many Indigenous Australian communities, humans and the environment are inextricably and symbiotically linked, and thereby any technology development a priori considers the impact on the environment and on other humans. As Worimi man Deen Sanders and Kamillaroi man Rick Shaw state:

> An indigenous perspective begins outside the self, and sees all activity as serving the purpose of creating abundance and harmony for all. For us, ethics is wrapped into the core of our being, because we see ourselves as custodians of the land from which all benefit. The interests of the land are inseparable from the interests of its people. The ethics of the system are the intrinsic aim of the system, not an externally imposed restraint on commercial or other outcomes. Our cultures measure knowledge and shared outcomes, not individual accumulation. We revere wisdom, not wealth... the interests of the self are the same thing as the interests (collectively and intergenerationally) of the land and the people. The self is activated by its responsibility to that whole and to each other. (The Ethics Centre and Deloitte Access Economics 2020, p. 22)

An example of an Indigenous technological system is Baiame's Ngunnhu (Brewarrina fish traps) – a millennia-old system situated in the far north west of New South Wales, Australia. This technological system is relevant to AI systems of today, even though it does not contain silicon chips and rare earth metals; its specific arrangement of stones is a technological system that works with the flows of the Barwon river to catch fish in small ponds and keep them fresh and ready to feed a large crowd that would gather there as a meeting place. According to oral history, these fishtraps were inspired by the pelican, "with the traps acting like a pelican's beak to scoop fish out of the water" (Murdi Paaki Regional Assembly 2015). Although it is protected under the Australian Heritage List (Australian Government Department of Climate Change, Energy, the Environment and Water 2021), Western-imposed regulation and historical mismanagement have disrupted the use of this technological system, illustrating the importance of aligned governance.

A sustainable system at its finest, Daniell and Moggridge (2024) share how carefully this aquaculture system was engineered to work with waterflow, ecology, climate change and culture. For example, in addition to the technology itself, is Aboriginal Lore governing the use of the fishtraps which ensured that responsibilities to the environment and to eachother are observed when people gathered for work, trade and consumption; the system is built to reflect the people it is intended to benefit which contributes to the care given to its maintenance and continuation. This is but one example of how "Indigenous epistemologies (or theories of knowledge) provide frameworks for understanding how technology can be developed in ways that integrate it into existing ways of life, support the flourishing of future generations and are optimized for abundance rather than scarcity" (New Frontiers in Research Fund 2023). Today, many of these ideas are applied to ensuring culturally appropriate design, through approaches such as Country-centred design (Abdilla et al. 2021) and the development of Indigenous protocols for AI (Abdilla et al. 2020; Gledhill-Tucker 2024).

The Steam Engine, the Internet and Responsibility Practices: Precursors to Responsible AI

Of course, not all technologies have been built with worldviews that embody the interconnectedness of humans with eachother and with nature. This is particularly true in more

[1] A multiplicity of Indigenous knowledge systems are involved in technologies, including AI (Abdilla et al. 2020). Thus, a single or representative "Indigenous perspective" is impossible to capture and is not the goal here; rather, this section shares illustrative, non-exhaustive perspectives.

recent technological developments such as AI which, as Ali et al. (2023) argues, "participates in the extractive and racist legacies of colonialism, race science and capitalism" (p. 9). A complicated relationship between technology and responsibility can also be seen with the creation of the steam engine in the 18th century. This technological innovation has shaped us as humans and shaped our societies, enabling the transportation of goods across previously unimaginable distances, leading to a more connected trade system, and some would say, facilitating the fruits of globalization (Pascali 2017) and harms of colonization (Buchanan 2024).

A range of changes in our social lives, that today are just considered part of the fabric of life, were adjusted to incorporate the steam engine, including the creation of standardized time so that trains could run to a schedule that was shared across distances and a variety of laws to accommodate new railways (Kostal and Kostal 1997). Arguably, this technology was not shaped with a responsibility-first mindset, instead it neglected a holistic understanding of the real and potential impacts of the steam engine on the planet. As Bell (2020) laments, imagine how different our trajectory would be if someone urged the engineer to redesign the steam engine system to ensure less environmental degradation! With today's climate emergency, we are witnessing attempts to retro-fit a responsibility agenda to a series of choices made centuries ago and which have scaled over time to grave effect on our climate, our planet and our lives (Malm 2016). This example alone suggests that the questions we ask now can have a profound impact on the trajectories of technologies and the shaping powers they have over us and our environments.

Norbert Wiener, widely known as one of the major proponents of cybernetics in the 1940s, highlighted the power of questions and one's own responsibility when it comes to technological development. Against the backdrop of the Cold War, Wiener was acutely aware of how his involvement in technological projects could be used irresponsibly for weapons development. In his book, *The Human Use of Human Beings* (1950), Wiener hypothesized about the impact of self-guided systems and information technologies on human values and society, years before many of the futures he forecast seemed possible. He was also a vocal opponent of the increasing militarization of science, and of scientific practice divorced from consequence, calling on his peers to take responsibility for the harms their creations might cause:

> If therefore I do not desire to participate in the bombing or poisoning of defenseless peoples – and I most certainly do not – I must take a serious responsibility as to those to whom I disclose my scientific ideas… I do not expect to publish any future work of mine which may do damage in the hands of irresponsible militarists. I am taking the liberty of calling this letter to the attention of other people in scientific work. I believe it is only proper that they should know of it in order to make their own independent decisions, if similar situations should confront them. (Wiener 1947, p. 31)

Responsibility conversations were also central in engineering circles, especially from the 1960s and early 1970s where engineers' identification as responsible professionals was under threat from new autonomous technologies and the problematic technologists associated with them. At the time, member associations such as the American Society of Mechanical Engineers (ASME) played an important role in setting aspirations for engineers as a profession and their role in taming out-of-control technology. The Overriding ASME Goal in 1970 was:

> To move vigorously from what is now essentially a technical society to a truly professional society sensitive to the engineer's responsibility to the public, and dedicated to a leadership role in making technology a true servant of man (Goal Report Committee, as cited in Wisnioski 2012, p. 84).

Of course, the increasing influence of autonomous technologies and militarization of science cannot be seen as all bad or all good, but rather a system with a range of complexities. For example, it also led to the creation of the internet, a major technological innovation of the 20th Century and one that carries numerous responsibility considerations. Initially designed to facilitate communications between the military and universities during the Cold War in the 1960s, the Advanced Research Projects Agency Network (ARPANET) was the pre-cursor to the World Wide Web (WWW) and embodies a core underlying design principle of open architecture networking (Leiner et al. 2009).

Internet responsibility conversations continue today in many forms, some flowing from the openness design principle, for example the role and responsibility of Internet intermediaries for the third party content that they host (Thompson 2015/2016), the shared responsibility of internet governance (Cerf, Ryan, and Senges 2014), various concerns around privacy and cybersecurity (see for example Muggah 2021; Pearlson 2024) and the political power and dominance that Big Tech companies wield on citizens' rights (Lindman, Makinen, and Kasanen 2023). Further, as Imran (2023) argues, we face a digital inequality paradox between, on the one hand, the narrative of a ubiquitous internet and recognition of internet access as a human right (Reglitz 2020), and on the other hand, the current reality that, as at 2024, over 30% of the world's 8 billion people still do not have basic internet access (Statistica 2024).

All of these experiences point towards the importance of considering technologies as part of a wider system, also referred to as adopting a sociotechnical approach.

What Are Responsible AI Practices?

Responsible AI is a contentious term, often entangled with other terms like ethical AI, trustworthy AI and AI for good (Lu et al. 2024). In this paper, I draw upon cybernetics, organization studies, systems sciences and science and technology studies to examine responsible AI practices from a sociotechnical perspective. Leveraging definitions provided by Lu et al. (2024) and Vassilakopoulou et al. (2022), I define responsible AI as follows:

> Responsible AI is practices undertaken by humans to design, develop, deploy and govern AI systems in

ways that adhere to fundamental values and benefit individuals, groups, wider society and the environment, while minimising the risk of negative consequences.

Cybernetician and systems scholar Ray Ison (2017), describes practices as:

> comprising a practitioner (P) with a history, a tradition of understanding, possibly a chosen framework of ideas (F), a chosen method (M) and a situation (S) in which they practise...practice is concerned with understanding, discovering, describing or changing some aspect of a situation. (p. 49)

In the context of responsible AI practices, the situation is about ensuring that AI systems are developed responsibly (which has different meanings in different contexts). Similarly, responsibility practices more generally are about ensuring the organization and its individuals are enacting a commitment to being responsible in what they do and how they operate. In organization studies, for example, Orlikowski (1992)'s structurational model of technology acknowledges the interactional interplay between humans and technologies and discusses the role of the organization as an additional and indispensable dimension.

Cybernetics compels us to rethink the boundary between how we think and how we act, that is, between theory and practice. According to Sweeting (2015) second-order cybernetics would consider theory something that we construct, and thereby a form of practice in and of itself. In addition, cybernetics considers the relationship between theory and practice as integrated, reflective and circular, similar to the steering of a ship "where the steersman's understanding of the effects of his or her action [theory] informs how he or she continues to act [practice]" (Sweeting 2015, p. 1398). He argues that:

> ...practice is therefore always cybernetic in structure, consisting of a circular relation between how we act and how we explain or understand that action... our actions are interdependent with how we understand them. (Sweeting 2015, p. 1400)

Understanding practices from the perspective of Science and Technology Studies (STS) is also very relevant given that STS scholars claim that "STS attends to practices" – they look at "how theories, methods, and materials are used in practice in specific social, organizational, cultural, and national contexts – and they look at the effects of those practices" (Law 2017, p. 31). STS focuses on studying method (scientific and otherwise) in practice, and in doing so acknowledges that theory, method and the empirical are woven together with social institutions and sometimes objects (such as technologies). As such, it embraces a sociotechnical perspective.

Two Approaches to Responsible AI Practices: Technical and Sociotechnical

Although my interpretation of responsible AI practices is grounded in cybernetic, sociotechnical and adaptive perspectives, this is not necessarily the norm in how responsible AI practices are conceived of today. This section describes two approaches to responsible AI practices – technical and sociotechnical – recognizing that there may also be others. To assist in delineating the differences between these two approaches, I draw upon Nabavi and Browne (2023)'s 5P framework – a systems-based approach to responsible AI development. The first P refers to the problem space and then the remaining 4Ps specify four leverage zones or places to intervene in a system to move it towards responsible AI:

1. the **Purpose zone** – focused on changing intent, mental models and paradigms
2. the **Pathway zone** – focused on redefining system design and structures that govern processes and parameters, such as rules, information flows, incentives and constraints
3. the **Process zone** – focused on changing the societal and technical processes and feedback
4. the **Parameter zone** – focused on tweaking algorithms and modifying parameters

This framework will be used to help describe how current responsible AI approaches intervene.

Technical Approach to Responsible AI Practices

Technical approaches are an extension of the technosolutionism that operates more generally in technological development, especially with AI (Pham and Davies 2024; Hollanek 2024; Lindgren and Dignum 2023; Morozov 2013). Technical approaches consider responsible AI as a technical problem, namely one that has

> ...known solutions that can be implemented by current know-how...resolved through the application of authoritative expertise and through the organization's current structures, procedures and ways of doing things. (Heifetz, Grashow, and Linsky 2009, p. 7)

Technical responsible AI practices are numerous and often take the form of technical fixes, toolboxes and toolkits (Hollanek 2024). Examples of technical responsible AI practices include gradient-based methods for explainable AI (Bughin 2024), anti-classification, classification parity and calibration standards in fair machine learning (see Corbett-Davies et al. 2023, for a review) or industry-led open-source solutions such as IBM Research (n.d.)'s "AI Fairness 360" toolkit which assists in finding bias in datasets or Microsoft (2024)'s "Responsible AI Toolbox" which provides a range of tools for model and data exploration and assessment, shared through the popular developer platform, GitHub.

Across some of the academic literature too (see for example Scantamburlo, Cortés, and Schacht 2020; Qiang, Rhim, and Moon 2024; Economou-Zavlanos et al. 2023), the framing of responsible AI practices assumes that responsible AI is a technical problem; that is, that there are known fixes or solutions. In this vein, Scantamburlo, Cortés, and Schacht (2020) describe five types of responsible AI practices: 1) Assessments, questionnaires and checklists, 2) End-to-end frameworks, 3) Strategy guides and canvasses, 4) Design guides and 5) Software toolkits. Similarly, Bughin (2024) provides an overview of responsible AI practices, across three domains – design, procedures & toolkits, and scaling.

Further, Ayling and Chapman (2021)'s review yielded three types of responsible AI practices – impact assessments, audits, and technical / design tools. Despite the naming used, all of these responsible AI practices would be considered "technical" in nature and, to use the 5P framework, operate at the *parameter* and *process* zones.

Selbst (2021) describes five traps involved in separating the technical from the social – assuming technology *must* be involved in a solution (the solutionism trap), only choosing certain technical parts of the system to model and manage (the framing trap), assuming generalisability of solutions to different contexts (the portability trap), defining terms such as fairness in relation to what can be solved technically (the formalism trap) and failing to understand the flow-on impacts of a new technology on wider social systems (the ripple effect trap). In its place, a sociotechnical approach is proposed.

Sociotechnical Approaches to Responsible AI Practices

There are many ways to consider sociotechnical approaches to responsible AI practices (see for example Wang, Xiong, and Olya 2020; Smit and Eybers 2022; Sloane and Zakrzewski 2022; Rakova et al. 2021; Stahl et al. 2022; Madaio et al. 2024; Prem 2023; Selbst et al. 2019). At its core, sociotechnical approaches take a much broader, systems view and consider the interactions between technology, humans and the environment (Sarker et al. 2019; Bell 2021) and value instrumental and humanistic outcomes.

Gutierrez Lopez and Halford (2024) describe a sociotechnical perspective as follows:

> Instead of asking how to explain the model (or algorithm) at the centre of a wider sociotechnical network, we ask how to explain the wider sociotechnical network of practice within which the model is conceived, developed and used. It is in this wider network of 'machine learning practice' that decisions are made about how ML [Machine Learning] will come into use, how it is used in everyday work and consequently how outcomes are derived. (p. 5)

Importantly, a sociotechnical approach often operates across different leverage zones, moving beyond parameters and processes to also consider pathways and purpose. For example, Davis, Williams, and Yang (2021) operate at the *purpose* and *pathway* zones in proposing algorithmic reparation in place of an algorithmic idealism mindset. Grounded in a sociotechnical approach, algorithmic reparation is presented as an alternative to current technical fixes dominating fair machine learning. Similarly, Allen et al. (2025) also begin by framing their suggestions in terms of a paradigm shift (*purpose* zone) required for AI governance. They argue for a new framework of "Power-Sharing Liberalism" which, in place of reactive, punitive approaches to AI governance, embraces "the expression of an expansive, proactive vision for technology – to advance human flourishing". Both of these proposals also discuss practices targeting process and parameter zones, but do not stop there.

Responsible AI practices focused on awareness building, learning and reflection are also sociotechnical in nature. Rakova et al. (2021) highlighted the importance of the practices of responsible research teams that contribute to internal awareness and education. Similarly, Stahl et al. (2022) reiterated the role of organizational awareness and reflection as a responsible AI practice, whilst also observing relatively little attention paid to these aspects in the literature, despite its focus in practice. Further, Madaio et al. (2024) investigate the wider systems structure of how practitioners learn about responsible AI and how their learning pathways influence how they use responsible AI resources and thereby the responsible AI practices taken. Moreover, Prem (2023) identified education as one of seven categories of approaches for addressing ethical AI issues. In addition to training, they suggested the potential future role of coaching and consulting. These practices intervene in the *purpose* zone by intending to change intent and mental models and often translate into changes in information flows (*pathway* zone) as well as *processes* and *parameters* as learners integrate new knowledges.

Other sociotechnical approaches to responsible AI practice focus on organizational structures (*pathway* zone) and processes (*process* zone). For example, Rakova et al. (2021) interviewed industry practitioners in large companies developing or using AI and identified organizational structures that currently support or hinder responsible AI. Governance practices that assisted in responsible AI included internal ethics review boards and distributed accountability integrated into the product development lifecycle, including escalation pathways and processes to hold teams accountable and internal review boards. Similarly, Stahl et al. (2022) identified ethics review boards and following standards and codes of ethics as important responsible AI practices.

Finally, practices targeting incentive structures are also often sociotechnical in nature. Rakova et al. (2021) identified a range of organizational practices that were thwarting responsible AI efforts such as misalignment between taking responsible AI action and incentive structures resulting in a lot of unrecognized or volunteer work, ill-informed performance trade-offs based on problematic and sometimes misleading metrics (or a lack thereof) and decision-making structures that are reactive to external pressures (such as catastrophic media attention). Rakova et al. (2021)'s study also highlighted activities aimed at tackling misaligned incentives, but overcorrected by creating overly rigid incentives, which were demotivating. These practices focused on incentive structures are broadly aligned with intervening in the *pathway* zone.

Barriers to Enacting Responsible AI Practices

Many responsible AI practices are never fully realized, and even if they are "There is, as of yet, little evidence that the use of any of these translational tools / methods has an impact on the governability of algorithmic systems." (Morley et al. 2021, p. 241). Four barriers to enacting responsible AI practices, synthesized from the literature, are explored below.

Barrier 1: (Perceived) Neutrality

Technological neutrality – the idea that if one makes a pipe that eventually gets turned into a gun, you cannot be responsible for making the gun – has often been applied to the case of AI development. Widder and Nafus (2023) discuss how this sentiment stems from uncritical technical practice where there is little attempt at understanding who may use the tool being created and the social relations created through the use of the tool. In stark contrast to the responsibility questions Wiener was posing to himself in 1947 (and explored in an earlier section), Widder and Nafus (2023) describe many machine learning (ML) developers' approaches today as follows:

> He only imagines other inert containers of software, enabling him to normalize harmful ML practices as a general matter of course, or theoretical possibility, and not question his participation in it or his choices about who he allows to access his technology. (Widder and Nafus 2023, p. 4)

For Widder and Nafus (2023), this idea of perceived neutrality is nurtured within the neoliberal economic context that many technologists operate in, which emphasizes not having relations as a result of the technology creation and assumes that there are no social ties following the economic exchange.

The interactive stance of technology presented in Friedman and Hendry (2019)'s value sensitive design approach rejects technological neutrality. Instead, it emphasizes that humans shape technologies, which in turn shape humans. This interactive stance is also at the heart of Winner (1980)'s argument that artifacts have political qualities either through the potential flexibility of their design and needing to understand the social actors who can influence the designs and arrangements chosen, or because there are certain intractable properties of some technologies that are reflective of institutionalized power and authority. There is a recognition that practitioners involved in the design, development, implementation, monitoring and regulation of technologies are non-neutral actors and the technologies created reify these non-neutral influences into a material artefact that extends beyond the original decisions made. Failing to recognize the interactive stance of technology, for example by hiding behind the myth of objectivity or neutrality, thwarts responsible AI efforts, enabling an abdication of responsibility. As Heinz von Foerster (2003) notes:

> I mentioned objectivity before, and I mention it here again as a popular device for avoiding responsibility. . . .With the essence of observing (namely the processes of cognition) having been removed, the observer is reduced to a copying machine with the notion of responsibility successfully juggled away. (p. 293)

Barrier 2: Power To Act

Widder et al. (2023) found that software engineers were capable and willing to identify a wide range of ethical concerns, even without the use of responsible AI instruments such as checklists or principles. The main challenge to enacting responsible AI practices lies in their power to resolve the issues that they have identified. Widder et al. (2023) share the strategies that practitioners could see to resolve the ethical issues identified – proposing technical software solutions, creating a business case for why harms caused warranted action, refusing (overtly or covertly) to work on unethical tasks, legal action, boycotting and collective bargaining – and the challenges that they faced to act upon these strategies. All of these challenges stem from power imbalances in the organization context. For example, many practitioners decided not to act because of financial precarity of doing so, fear of being blacklisted for future jobs or the need to stay for immigration visa purposes. Similarly, workplace culture played an important role in whether they felt they had power to act. For some, authoritarian and hierarchical structures made raising concerns difficult, but difficulties were also experienced for some in a "friendly" workplace culture. Remote working and fewer social ties were seen for some as enabling an escalation of concerns.

Rakova et al. (2021), in her interviews with AI practitioners on organizational practices, also found instances where there was difficulty to act. They reflected on the fragility of the current practices because they often relied on the individual's resources and power to act, rather than a more scalable or systemic organizational structure. In addition, they contend that having reputational risk as a core motivator ties accountability to individual incentives instead of a more stable approach of ensuring accountability through policies and processes.

Barrier 3: Modularity, the "Many Hands" Issue and Supply Chain Metaphor

The challenge of collective responsibility for particular actions – also referred to as the "many hands" issue – is not unique to computing and has been discussed in a range of other contexts over time such as business, engineering and politics (Thompson 1998; De George 1981). The "many hands" issue is rife in any organizational context where there are multiple people involved. This is especially true when it comes to the development of technology products because it generally happens in an institutional setting and suffers from a modularity approach where, consequently, no one person grasps the whole system (Nissenbaum 1994). Unsurprisingly, the many hands issue has been referred to as one of six explanations for the principles-to-practice gap in responsible AI (Schiff et al. 2021). Ironically, there is a disconnect between the instruments created, which often assume full control by a few people, and the reality of decentralized and modularized design of AI systems, where there are a range of people involved, with little oversight or understanding of the system as a whole (Prem 2023; Nissenbaum 1994; Widder and Nafus 2023).

As Widder and Nafus (2023) argue, a modularity approach reinforces a sense that technology creation is part of a supply chain, and in the process, this limits a sense of agency and responsibility. Modularity and "many hands" also leads to conceiving AI development as a supply chain (Hockenberry 2021; Widder and Nafus 2023). The logistic metaphor

describes the relations between actors as distanced, where code created in one module serves as a container that is passed on to the next part of the chain as a supply, without looking inside (Hockenberry 2021), forming a chain of actions that are near or far to the end user. Widder and Nafus (2023) describe these conditions as leading to a dislocated accountability – "acknowledgement of harms was consistent, but nevertheless another person's job to address, always elsewhere" (p. 1).

Some contend that open-source communities mitigate some of this challenge by providing transparency in the codes used and also a visible community response to whether it is accepted. Widder et al. (2022) confirm that the traceability involved in open-source approaches assists with identifying and rectifying implementation-based harms (such as code quality), but does little for use-based harms because "harm can be wrought not only from parts of code which may malfunction or be ethically inadequate in some way, but from the whole software package operating as its creators intend, but for a harmful use they did not intend" (p. 2043). Further, it can be argued that the principles of unconstrained use, circulation and modification common to open-source communities, actually proliferate such harms by increasing the ease of unintended use by others (Widder et al. 2022).

Barrier 4: Organizational Culture

Orr and Davis (2020) specify that the organizational context within which AI is developed is essential because the speed of regulatory codes is too slow to match the speed of technological development, leaving organizations to often set their own standards, both officially and as informal organizational norms. Lauer (2021)'s article *You cannot have AI ethics without ethics* sums up the challenge:

> AI ethics simply cannot exist without a broader culture of ethics…only organizations with a firm grounding in ethics, and an appreciation for the way complex systems behave can succeed at ethical deployment of AI. (p. 21)

From this perspective, organizations without a culture committed to responsibility cannot band-aid some responsible AI checklists or frameworks and expect to achieve responsible AI; these activities are likely to lead (or are potentially intended to lead to) ethics washing activities (Floridi 2019). For example, Madaio et al. (2020) found that AI fairness checklists, whilst being helpful in formalising ad-hoc processes and empowering individuals to act, is most effective when practically aligned with existing processes and "supported by organizational culture" (p. 10). Similarly, Widder and Nafus (2023) reflect on the places in which responsible AI practices can exist and have trouble existing. Places where responsible AI practices can exist include those outside of the supply chain of AI development, for example motivated by an organizational culture that values customer-centricity or is concerned about the impact on reputation. On the flip side of this argument, is that organizational cultures that do not value responsibility practices are unlikely to truly embrace responsible AI.

Organizational culture is also influenced by organizational size. Vakkuri et al. (2020) note that developers within startup environments may have ethical concerns regarding their systems but did not actively consider them because they were just "developing a prototype". Despite the fact that initial versions of a product actually do influence the final one developed (Duc and Abrahamsson 2016) and thus are valid sites for responsible AI consideration, the fail fast mantras associated with startup culture works in opposition to such responsible AI considerations.

Further, organizational culture can be influenced and shaped by the characteristics and incentives of the economic system in which they operate, such as a need for growth and scale, agile working that spurs rapid release of minimum viable products and incentives focused on revenue generation (Rakova et al. 2021). For example, Widder and Nafus (2023) reflect that practitioners spoke of "embodied moments when scale becomes indifference" and how these dynamics are counterproductive to a responsibility-focused organizational culture. In addition, results from Rakova et al. (2021)'s survey of responsible AI practitioners suggests that "what individuals working on responsible AI need is for the organizational structures around them to adapt in order to support rather than hinder their work" (p. 18). All of these insights point towards the critical role organizational culture, structure and processes play in the implementation of responsible AI.

Future Directions for Responsible AI Practices

Figure 1 maps the different approaches to responsible AI practices and the four common barriers to responsible AI practice synthesized from the literature to Nabavi and Browne (2023)'s four intervention zones. Doing so demonstrates a profound mismatch: barriers to enacting responsible AI practices operate across all four leverage zones, but technical approaches do not. In contrast, sociotechnical approaches are also intervening at various leverage zones in the system.

Having established the need for sociotechnical approaches to responsible AI practice, I now turn to synthesizing the literature that discusses sociotechnical approaches to responsible AI practice, the known barriers that each approach (at least partially) addresses, as well as the leverage zones they target. See Table 1 for an overview.

Design Practices

Instead of seeing responsible AI as an end goal, it can also be positioned as a design process. Morley et al. (2021) argue that a potential solution in the face of not-yet actionable responsible AI principles is "to bring ethical guidance down to the Design level, by providing tools and methods that translate the 'what' of AI ethics into the 'how' of technical specifications" (p. 242). This style of approach would be akin to what Dignum (2019) characterizes as "ethics in design" – where designers and developers consider the purpose and values embedded in the AI system (*purpose* zone) and the potential impacts of its use (*pathway* and *process* zones). Other design approaches proposed by Dignum (2019) includes "ethics by design" which looks at the behaviour of AI

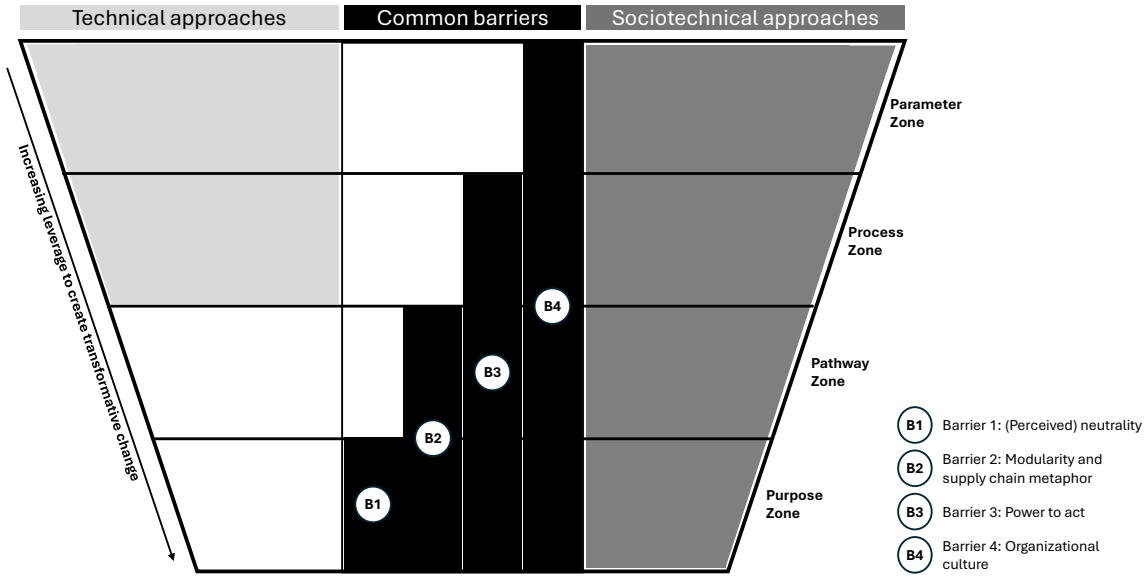

Figure 1: Mapping technical and sociotechnical approaches to responsible AI practices and barriers to enacting responsible AI to 5P Framework. Adapted from Nabavi and Browne (2023). The first 'P' is Problem, for example operationalising responsible AI. The remaining 4Ps – Parameter, Process, Pathway, Purpose – describe the 'leverage zones' or places to intervene in the system to address the problem. The purpose zone has the greatest potential for transformation change in of the system, but is the hardest to intervene in.

systems once deployed and what actions are permissable in the context of use, and "ethics for designers" which may include the professional codes of conduct, principles and regulations that motivate and govern AI practitioners (*pathway* zone). All of these design practices lean in to recognition of the interactive stance of technology and reject any sense of being neutral actors and empower action.

Ethics-as-a-Service

Morley et al. (2021) explore Digital Catapult's AI Ethics Framework and their own theoretical exploration, and conclude that translational tools and methods assist with lowering the level of abstraction seen in many responsible AI instruments, but only offer a partial solution. They suggest moving beyond translational tools in isolation and towards an AI Ethics-as-a-Service approach, drawing an analogy to a Platform-as-a-Service model which operates in the middle ground between methods that are too flexible and too strict, and between devolved and centralized governance. This is achieved through several components operating together. For example, an independent multi-disciplinary ethics advisory board, which provides some of the ethical infrastructure – an ethical code, a process to follow for ethical design and a regular audit processes – alongside the AI practitioners responsible for "customising" the approach – contextualising the principles, identifying appropriate tools while designing an AI system and documenting their processes. Ethics-as-a-service would thus target *pathway* and *process* zones and assist in ensuring the contextualization that is often absent from modularity and supply chain metaphors.

Located Accountability Practices

In the face of the challenges of a modularity approach to AI development and a felt lack of power to act, Widder and Nafus (2023) propose "located accountability" which begins from an assumption that personal relations between actors exist and are important within AI development. It also acknowledges that there is partial, situated knowledge within those relations. They admit that asking AI practitioners to accept located accountability wholesale is a tough request; taking this into account, Widder and Nafus (2023) propose three interventions.

Firstly, acting within the modules by asking practitioners to include their partial understanding of "flaws, limitations, divergent provenances and contexts of use" in various documentations that are handed on to the next part of the chain of activity (*process* zone). Secondly, by strengthening the interfaces between modules, taking on a value chain approach (instead of a supply chain approach). For example, customers in the chain could ensure items like Model Cards, training data and appropriate pay for data labelers are included in contractual agreements (*pathway* zone). This is aligned with Ruster and Brown (2020) who suggest that including a termination for cultural misalignment clause in AI systems built with and for Indigenous users is one way to ensure the wellbeing of Indigenous communities. Such in-

Proposed Sociotechnical Practices	Barrier(s) (Partially) Addressed	Leverage Zone(s) Targeted
Design practices	Barrier 1: (Perceived) neutrality; Barrier 3: Power to act	Purpose, pathway and process zones
Ethics-as-a-Service	Barrier 2: Modularity and supply chain metaphor	Pathway and process zones
Located accountability	Barrier 2: Modularity and supply chain metaphor; Barrier 3: Power to act	Purpose, pathway and process zones
Indigenous approaches	Barrier 2: Modularity and supply chain metaphor; Barrier 3: Power to act	Purpose and pathway zones
Integrative approaches	Barrier 2: Modularity and supply chain metaphor	Purpose and process zones
Organizational culture approaches	Barrier 4: Organizational culture	Purpose and pathway zones

Table 1: Mapping sociotechnical practices and the barriers they address to the Nabavi and Browne (2023) 5P leverage zone framework.

clusions will allow suppliers to "reframe ethics work as an act of delivering customer value" (Widder and Nafus 2023, p. 9). Thirdly, modularity can be rejected entirely and, in its place, a priority placed on "building good relations as a matter of first order order concern, building code second, and 'scale' as a distant matter at most" (Widder and Nafus 2023, p. 9) (*purpose* zone). This relationship-first approach is aligned with Indigenous data sovereignty principles and notions of design justice (Costanza-Chock 2020) and ultimately rejects the idea of generic tools and the assumption of panoptic control and knowledge over the AI systems.

Indigenous Approaches

Indigenous approaches are currently not very widely evidenced; however, there are some illustrative examples shared in this section. Despite its broad application to data in a variety of fields (not just in AI), the CARE Principles for Indigenous Data Governance (Carroll et al. 2020) are nevertheless relevant to how we might approach responsible AI because ultimately AI is built on data. The impetus for the CARE principles stems from the premise that mainstream values associated with research and data are often inconsistent with collective rights and Indigenous culture. This is also a premise discussed specifically in the context of AI development (see for example Gwagwa, Kazim, and Hilliard 2022). Indigenous approaches advocate for finding ways to recognize Indigenous Data Sovereignty – "an assertion of the rights and interests of Indigenous Peoples in relation to data about them, their territories and their ways of life" (Carroll et al. 2020). Indigenous Data Sovereignty operates in *purpose* and *pathway* zones and specifically addresses a known barrier of power to act.

A complement to existing FAIR principles (Wilkinson et al. 2016), the CARE principles (Carroll et al. 2020; Global Indigenous Data Alliance 2022) are as follows:

- Indigenous data must enable **Collective benefit** for Indigenous Peoples
- Indigenous Peoples must have **Authority to control** and govern data that affects them
- Working with Indigenous data carries a **Responsibility** to nurture relationships with the communities from where the data originates and be accountable for how the data used supports self-determination and collective benefit
- **Ethics** must be upheld to ensure Indigenous rights and wellbeing as the core concern throughout the data lifecycle and across the data ecosystem. This includes not only minimising harm, maximising benefits, promoting justice and allowing for future use, but also ensuring that Indigenous Peoples are the ones to assess these aspects based on their own community values.

The *Indigenous Protocol and Artificial Intelligence Position Paper* (Abdilla et al. 2020) shares some similar ground to the CARE principles, aimed at "any person, group, organization, institute, company and/or political or governmental representative that wishes to undertake responsible and fair development of AI systems with Indigenous communities" (Abdilla et al. 2020, p. 20). It specifies the importance of data sovereignty, how AI systems should be responsible, relevant and accountable to Indigenous communities and how there should be a core ethics of do-no-harm. It rejects modularity and supply chain metaphors, advocating instead for AI systems that are locally grounded yet globally connected, with Indigenous knowledge relationality and human-nonhuman reciprocity integrated throughout the entire AI stack. Further, it describes "all technical systems as cultural and social systems", which has several implications. First, bias in models is thereby nearly impossible to mitigate and is amplified in models when unrecognised (Harle, Abdilla, and Newman 2018). Second, practitioners need to be aware of their own biases and work through ways to accommodate other cultural and social frameworks; this includes Indigenous communities developing their own computational approaches that reflect Indigenous values as a form of cultural resilience and continuity (see for example Jones et al. 2025). Third, criteria for model creation and governance need to focus on relational understanding of the world (Abdilla et al. 2021).

Integrative Approaches

Attard-Frost and Widder (2024) argue that a core limitation to many responsible AI efforts lies in the wide range of concerns involved across different actors, resources, contexts

and scales, and how current approaches fail to integrate these challenges. In response, they propose theorising AI as value chains (*purpose* zone) – "co-creation structures that exist within a network of actors and enable patterned resourcing activities to occur between actors" (p. 5). Despite the tendency to use the two terms – supply chain and value chain – interchangeably, they insist that they are very different with "different ontological, ethical, practical, and policy implications". This is consistent with Widder and Nafus (2023)'s call to move away from the modularity assumed in supply chain metaphors and towards broader forms of co-creativity and relationality embraced by ideas of value chains.

Combining multiple approaches has also been a pathway taken by research ethics and medical ethics. Prem (2023) argues that generating a suite of responsible AI approaches, following the path of medical ethics, is a more realistic aspiration than relying on responsible AI instruments alone:

> Given the enormous breadth of possible approaches to designing AI systems, it is unlikely that principlism alone will achieve their ethicality. Just as medical ethics has evolved to establish best practices, tools such as committees, guidelines, and regulations, AI ethics will require much more research into its practical underpinnings from notions to code, best practices, infrastructure such as described above, education, and communities of practice. (Prem 2023, p. 712)

Organizational Culture Approaches

Rakova et al. (2021) shared practitioners' reflections regarding how to shift organizational culture away from ethics-washing and towards responsible AI practices. They identified four enablers that sit in the *pathway* zone:

1. Reward internal education efforts
2. Reward "risk-taking for the public good"
3. Undertake internal investigations to followup on potential issues
4. Enable cross-functional collaboration

Organizational culture approaches also stem from how individuals see their responsibility, which, as discussed earlier, can be fragmented due to the challenge of many hands being involved in the development of technologies. Fahlquist (2015) argues that embracing responsibility-as-virtue may assist in avoiding the many hands issue because it puts emotions at the forefront, where people feel responsible, stemming from a deeper place of care (*purpose* zone). She goes on to specify three parts to acting responsibly in this vein – firstly caring about the impact of their activities on other people and the environment, secondly having the emotional ability to morally imagine the impacts and risks of the activities (moral imagination) and thirdly, the cognitive ability to move from concern into practice and action (practical wisdom). This style of responsibility is grounded in responsibility-as-virtue, where responsibility is forward-looking, focused on relations to others, requires seeing oneself as part of a wider context and requires action in certain ways over time (Fahlquist 2015).

Conclusion

Unlike Indigenous approaches to technology development, which have endured millennia, not all technologies have been built with worldviews that embody the interconnectedness of technology, humans and nature across time and space. Much can be learned from these Indigenous approaches and from the checkered histories of recent technological breakthroughs, the steam engine and the internet: namely, that sociotechnical approaches are needed for responsible AI practice.

Many claim that the advent of AI is as disruptive and society-shaping as the introduction of the steam engine (World Economic Forum 2024) or the internet (Olson 2024) before it. But despite lessons learned from these contexts, we see a burgeoning number of responsible AI instruments – frameworks, tools, guidelines, toolkits etc. – that adopt a technical approach. That is, they assume that responsible AI is something that can be "solved", intervening primarily in the *parameters* and *processes* leverage zones. It is unsurprising then, that in 2020, McLennan et al. observed:

> It remains unclear whether the influx of guidelines has actually made any impact on improving the ethical development and implementation of AI. (p. 21)

Arguably, little has changed since then. Accordingly, whilst current instruments may be helpful for identifying ethical issues or providing a technical fix, they are less convincing when it comes to assisting with holistic responsible AI practice (Prem 2023; Mittelstadt 2019; Morley et al. 2020; Munn 2023; Pant et al. 2024) and, as this paper has shown, addressing the common barriers experienced by practitioners, which operate at within all four leverage zones. Figure 1 illustrates the profound mismatch between technical approaches and barriers faced, providing further evidence for the importance of sociotechnical approaches.

Pant et al. (2024) claims that the challenges faced with responsible AI practice are not because of a lack of awareness of the relevance and importance of ethical challenges when building AI systems in general; rather they suggest practitioners associate responsible AI with certain, limited (technical) approaches such as transparency (Christodoulou and Iordanou 2021), fairness (Holstein et al. 2019), accountability and privacy (Ibáñez and Olmeda 2021), and may be unaware of the wider sphere available. This paper contributes to addressing this gap by synthesizing the literature available on sociotechnical approaches to responsible AI practice, providing examples of what it looks like and future directions it could take.

Cybernetician W. Ross Ashby (1968)'s law of "requisite variety" suggests that in order to respond to the challenges that emerge from complex systems (like responsible AI development), a repertoire of responses at least as nuanced as the problem faced is required. From this perspective, a complex system of approaches are required, including sociotechnical ones. This paper therefore urges researchers and practitioners to further develop, demonstrate and communicate sociotechnical approaches to responsible AI, harnessing lessons from the past and expanding the variety of responses needed for effective responsible AI practice.

Acknowledgments

I am especially grateful to feedback received from Professor Katherine A. Daniell and Professor Angie Abdilla. This work was supported by a Florence Violet McKenzie Scholarship and an Australian Government Research Training Program Scholarship.

References

Abdilla, A.; Arista, N.; Baker, K.; Benesiinaabandan, S.; Brown, M.; Cheung, M.; Coleman, M.; Cordes, A.; Davison, J.; Duncan, K.; Garzon, S.; Harrell, D. F.; Jones, P.-L.; Kealiikanakaoleohaililani, K.; Kelleher, M.; Kite, S.; Lagon, O.; Leigh, J.; Levesque, M.; Lewis, J. E.; Mahelona, K.; Moses, C.; Nahuewai, I. I.; Noe, K.; Olson, D.; Parker Jones, Ō.; Running Wolf, C.; Running Wolf, M.; Silva, M.; Fragnito, S.; and Whaanga, H. 2020. Indigenous Protocol and Artificial Intelligence Position Paper. Technical report, Concordia University Library.

Abdilla, A.; Kelleher, M.; Shaw, R.; and Yunkaporta, T. 2021. Out of the Black Box: Indigenous Protocols for AI. Technical report, Old Ways, New.

Ali, S. M.; Dick, S.; Dillon, S.; Jones, M. L.; Penn, J.; and Staley, R. 2023. Histories of Artificial Intelligence: A Genealogy of Power. *BJHS Themes*, 8: 1–18.

Allen, D.; Hubbard, S.; Lim, W.; Stanger, A.; Wagman, S.; Zalesne, K.; and Omoakhalen, O. 2025. A Roadmap for Governing AI: Technology Governance and Power-Sharing Liberalism. *AI and Ethics*, 5: 3355–3377.

Ashby, W. R. 1968. Variety, Constraint, and the Law of Requisite Variety. In Buckley, W., ed., *Systems Research for Behavioral Science*, 129–136. New York: Routledge, 1st edition. ISBN 978-1-315-13056-9.

Attard-Frost, B.; and Widder, D. G. 2024. The Ethics of AI Value Chains. arXiv:2307.16787.

Australian Government Department of Climate Change, Energy, the Environment and Water. 2021. National Heritage Places - Brewarrina Aboriginal Fish Traps (Baiame's Ngunnhu). https://www.dcceew.gov.au/parks-heritage/heritage/places/national/brewarrina (accessed: 2024-20-09).

Ayling, J.; and Chapman, A. 2021. Putting AI Ethics to Work: Are the Tools Fit for Purpose? *AI and Ethics*, 2: 405–429.

Bell, G. 2020. Anthropology, Cybernetics, and Establishing a New Branch of Engineering at ANU. https://www.thisishcd.com/episode/genevieve-bell-anthropology-cybernetics-and-establishing-a-new-branch-of-engineering-at-anu (accessed: 2024-20-09).

Bell, G. 2021. Talking To Ai: An Anthropological Encounter with Artificial Intelligence. In Pedersen, L.; and Cliggett, L., eds., *The SAGE Handbook of Cultural Anthropology*, 442–458. London, UK: SAGE Publications Ltd. ISBN 978-1-5297-0387-0.

Benjamin, R. 2020. *Race after Technology: Abolitionist Tools for the New Jim Code*. Cambridge, UK: Polity. ISBN 978-1-5095-2640-6.

Bromley, P.; and Powell, W. 2012. From Smoke and Mirrors to Walking the Talk: Decoupling in the Contemporary World. *The Academy of Management Annals*, 6(1): 483–530.

Buchanan, R. A. 2024. History of Technology - Renaissance, Industrial Revolution, Enlightenment. https://www.britannica.com/technology/history-of-technology/The-emergence-of-Western-technology-1500-1750 (accessed: 2024-23-09).

Bughin, J. 2024. Doing versus Saying: Responsible AI among Large Firms. *AI & Society*, 40(4): 2751–2763.

Carroll, S. R.; Garba, I.; Figueroa-Rodríguez, O. L.; Holbrook, J.; Lovett, R.; Materechera, S.; Parsons, M.; Raseroka, K.; Rodriguez-Lonebear, D.; Rowe, R.; Sara, R.; Walker, J. D.; Anderson, J.; and Hudson, M. 2020. The CARE Principles for Indigenous Data Governance. *Data Science Journal*, 19: Article 43.

Cerf, V.; Ryan, P. I.; and Senges, M. 2014. Internet Governance Is Our Shared Responsibility. *I/S: A Journal of Law and Policy for the Information Society*, 10(1): 1–42.

Christodoulou, E.; and Iordanou, K. 2021. Democracy Under Attack: Challenges of Addressing Ethical Issues of AI and Big Data for More Democratic Digital Media and Societies. *Frontiers in Political Science*, 3: Article 682945.

Corbett-Davies, S.; Gaebler, J. D.; Nilforoshan, H.; Shroff, R.; and Goel, S. 2023. The Measure and Mismeasure of Fairness. *Journal of Machine Learning Research*, 24(1): 14730–14846.

Costanza-Chock, S. 2020. *Design Justice: Community-Led Practices to Build the Worlds We Need*. Cambridge, MA: The MIT Press. ISBN 978-0-262-35686-2.

Daniell, K. A.; and Moggridge, B. 2024. Indigenous Water Engineering and Aquaculture Systems in Australia: The Budj Bim Cultural Landscape and Baiame's Ngunnhu (the Brewarrina Aboriginal Fish Traps). *Blue Papers*, 3(1): 18–29.

Davis, J. L.; Williams, A.; and Yang, M. W. 2021. Algorithmic Reparation. *Big Data & Society*, 8(2): e20539517211044808.

De George, R. T. 1981. Ethical Responsibilities of Engineers in Large Organizations: The Pinto Case. *Business & Professional Ethics Journal*, 1(1): 1–14.

Dignum, V. 2019. *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Artificial Intelligence: Foundations, Theory, and Algorithms. Cham: Springer International Publishing. ISBN 978-3-030-30370-9.

Duc, A. N.; and Abrahamsson, P. 2016. Minimum Viable Product or Multiple Facet Product? The Role of MVP in Software Startups. In Sharp, H.; and Hall, T., eds., *Agile Processes, in Software Engineering, and Extreme Programming*, 118–130. Cham, Switzerland: Springer International Publishing. ISBN 978-3-319-33515-5.

Economou-Zavlanos, N. J.; Bessias, S.; Cary, M. P.; Bedoya, A. D.; Goldstein, B. A.; Jelovsek, J. E.; O'Brien, C. L.;

Walden, N.; Elmore, M.; Parrish, A. B.; Elengold, S.; Lytle, K. S.; Balu, S.; Lipkin, M. E.; Shariff, A. I.; Gao, M.; Leverenz, D.; Henao, R.; Ming, D. Y.; Gallagher, D. M.; Pencina, M. J.; and Poon, E. G. 2023. Translating Ethical and Quality Principles for the Effective, Safe and Fair Development, Deployment and Use of Artificial Intelligence Technologies in Healthcare. *Journal of the American Medical Informatics Association*, 31(3): 705–713.

Eubanks, V. 2017. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: Macmillan Publishers, first edition. ISBN 978-1-250-07431-7.

Fahlquist, J. N. 2015. Responsibility as a Virtue and the Problem of Many Hands. In van de Poel, I.; Royakkers, L.; and Zwart, S. D., eds., *Moral Responsibility and the Problem of Many Hands*, Routledge Studies in Ethics and Moral Theory, 187–208. Hoboken: Taylor and Francis, 1st edition. ISBN 978-1-317-56030-2.

Felstead, C.; Stockdale, R.; and Scheepers, H. 2023. A Dignity Perspective on the Potential Harm of AI Technologies: The Case of Robodebt. In *Australasian Conference of Information Systems (ACIS) 2023 Proceedings*, Article 43.

Floridi, L. 2019. Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical. SSRN Scholarly Paper 3835010, Social Science Research Network, Rochester, NY.

Friedman, B.; and Hendry, D. G. 2019. *Value Sensitive Design: Shaping Technology with Moral Imagination*. Cambridge, UK: MIT Press. ISBN 978-0-262-35170-6.

Gledhill-Tucker, K. 2024. Reflections: Indigenous Protocols for Artificial Intelligence (IP//AI) Workshop #3. Technical report, Old Ways, New; The Australian Network for Art & Technology; Australian National University.

Global Indigenous Data Alliance. 2022. Indigenous Data Sovereignty and Governance. Prepared by Stephanie R. Carroll, Jewel Cummins, Andrew Martinez. https://static1.squarespace.com/static/5d3799de845604000199cd24/t/640792a43ba5c11a1073bbc8/1678217895508/TheCAREPrinciples.pdf (accessed: 2025-02-05).

Graafland, J.; and Smid, H. 2019. Decoupling Among CSR Policies, Programs, and Impacts: An Empirical Study. *Business & Society*, 58(2): 231–267.

Gutierrez Lopez, M.; and Halford, S. 2024. Explaining Machine Learning Practice: Findings from an Engaged Science and Technology Studies Project. *Information, Communication & Society*, 28(4): 1–17.

Gwagwa, A.; Kazim, E.; and Hilliard, A. 2022. The Role of the African Value of Ubuntu in Global AI Inclusion Discourse: A Normative Ethics Perspective. *Patterns*, 3(4): Article 100462.

Harle, J.; Abdilla, A.; and Newman, A., eds. 2018. *Decolonising the Digital: Technology as Cultural Practice*. Sydney: Tactical Space Lab. ISBN 978-0-646-99587-8.

Heifetz, R.; Grashow, A.; and Linsky, M. 2009. *The Practice of Adaptive Leadership: Tools and Tactics for Changing Your Organization and the World*. Boston, MA: Harvard Business Press.

Hockenberry, M. 2021. Redirected Entanglements in the Digital Supply Chain. *Cultural Studies*, 35: 1–22.

Hollanek, T. 2024. The Ethico-Politics of Design Toolkits: Responsible AI Tools, from Big Tech Guidelines to Feminist Ideation Cards. *AI and Ethics*, 5: 2165–2174.

Holstein, K.; Wortman Vaughan, J.; Daumé, H.; Dudik, M.; and Wallach, H. 2019. Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need? In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, 1–16. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-5970-2.

Ibáñez, J. C.; and Olmeda, M. V. 2021. Operationalising AI Ethics: How Are Companies Bridging the Gap between Practice and Principles? An Exploratory Study. *AI & Society*, 37: 1663–1687.

IBM Research. n.d. AI Fairness 360. https://aif360.res.ibm.com/aif360.res.ibm.com (accessed: 2024-10-04).

Imran, A. 2023. Why Addressing Digital Inequality Should Be a Priority. *The Electronic Journal of Information Systems in Developing Countries*, 89(3): e12255.

Ison, R. 2017. *Systems Practice: How to Act*. London: Springer London. ISBN 978-1-4471-7350-2.

Jones, P.-L.; Mahelona, K.; Duncan, S.; and Leoni, G. 2025. Kaitiaki: Closing the Door on Open Indigenous Data. *International Journal on Digital Libraries*, 26(1): 1.

Kostal, R. W.; and Kostal, R. W. 1997. *Law and English Railway Capitalism 1825-1875*. Oxford: Oxford University Press. ISBN 978-0-19-826567-2.

Lauer, D. 2021. You Cannot Have AI Ethics without Ethics. *AI and Ethics*, 1(1): 21–25.

Law, J. 2017. STS as Method. In Felt, U.; Fouché, R.; Miller, C. A.; Smith-Doerr, L.; and Society for Social Studies of Science, eds., *The Handbook of Science and Technology Studies*, 31–57. Cambridge, MA: The MIT Press, fourth edition. ISBN 978-0-262-34599-6.

Leiner, B. M.; Cerf, V. G.; Clark, D. D.; Kahn, R. E.; Kleinrock, L.; Lynch, D. C.; Postel, J.; Roberts, L. G.; and Wolff, S. 2009. A Brief History of the Internet. *ACM SIGCOMM Computer Communication Review*, 39(5): 22–31.

Li, P.; Yang, J.; Islam, M. A.; and Ren, S. 2023. Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models. arXiv:2304.03271.

Lindgren, S.; and Dignum, V. 2023. Beyond AI Solutionism: Toward a Multi-Disciplinary Approach to Artificial Intelligence in Society. In Lindgren, S., ed., *Handbook of Critical Studies of Artificial Intelligence*, 163–172. Cheltenham: Edward Elgar Publishing. ISBN 978-1-80392-856-2.

Lindman, J.; Makinen, J.; and Kasanen, E. 2023. Big Tech's Power, Political Corporate Social Responsibility and Regulation. *Journal of Information Technology*, 38(2): 144–159.

Long, C. 2024. Social Media Ban for Children to Be Introduced before the End of This Year. https://www.abc.net.au/news/2024-09-09/government-plans-social-media-porn-site-age-limit/104329920 (accessed: 2024-23-09).

Lu, Q.; Zhu, L.; Whittle, J.; and Xu, X. 2024. *Responsible AI: Best Practices for Creating Trustworthy AI Systems*. Boston: Addison-Wesley. ISBN 978-0-13-807392-3.

Madaio, M.; Kapania, S.; Qadri, R.; Wang, D.; Zaldivar, A.; Denton, R.; and Wilcox, L. 2024. Learning about Responsible AI On-The-Job: Learning Pathways, Orientations, and Aspirations. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, 1544–1558. Rio de Janeiro Brazil: ACM. ISBN 9798400704505.

Madaio, M. A.; Stark, L.; Wortman Vaughan, J.; and Wallach, H. 2020. Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, 1–14. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-6708-0.

Malm, A. 2016. *Fossil Capital: The Rise of Steam Power and the Roots of Global Warming*. London, UK: Verso. ISBN 978-1-78478-131-6.

Maner, W. 1978. Starter Kit in Computer Ethics. Technical report, Self published in 1978. Republished in 1980 by Helvetia Press in cooperation with the National Information and Resource Center for Teaching Philosophy.

Marjanovic, O.; Cecez-Kecmanovic, D.; and Vidgen, R. 2021. Algorithmic Pollution: Making the Invisible Visible. *Journal of Information Technology*, 36(4): 391–408.

McLennan, S.; Lee, M. M.; Fiske, A.; and Celi, L. A. 2020. AI Ethics Is Not a Panacea. *The American Journal of Bioethics*, 20(11): 20–22.

Microsoft. 2024. Responsible AI Tools and Practices. https://www.microsoft.com/en-us/ai/tools-practices (accessed: 2024-08-05).

Midgley, G. 2000. *Systemic Intervention*. Contemporary Systems Thinking. Boston: Springer US. ISBN 978-1-4615-4201-8.

Mittelstadt, B. 2019. Principles Alone Cannot Guarantee Ethical AI. *Nature Machine Intelligence*, 1(11): 501–507.

Morley, J.; Elhalal, A.; Garcia, F.; Kinsey, L.; Mökander, J.; and Floridi, L. 2021. Ethics as a Service: A Pragmatic Operationalisation of AI Ethics. *Minds and Machines*, 31(2): 239–256.

Morley, J.; Floridi, L.; Kinsey, L.; and Elhalal, A. 2020. From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. *Science and Engineering Ethics*, 26(4): 2141–2168.

Morozov, E. 2013. *To Save Everything, Click Here: The Folly of Technological Solutionism*. New York: PublicAffairs. ISBN 978-1-61039-370-6.

Muggah, R. 2021. Digital Privacy Comes at a Price. Here's How to Protect It. https://www.weforum.org/stories/2021/09/how-to-protect-digital-privacy/ (accessed: 2024-23-09).

Mulligan, C.; and Elaluf-Calderwood, S. 2021. AI Ethics: A Framework for Measuring Embodied Carbon in AI Systems. *AI and Ethics*, 2: 363–375.

Mumford, E. 2000. A Socio-Technical Approach to Systems Design. *Requirements Engineering*, 5(2): 125–133.

Munn, L. 2023. The Uselessness of AI Ethics. *AI and Ethics*, 3(3): 869–877.

Murdi Paaki Regional Assembly. 2015. Brewarrina Fish Traps. https://www.mpra.com.au/brewarrina-fish-traps (accessed: 2024-20-09).

Nabavi, E.; and Browne, C. 2023. Leverage Zones in Responsible AI: Towards a Systems Thinking Conceptualization. *Humanities and Social Sciences Communications*, 10(1): 82.

New Frontiers in Research Fund. 2023. Indigenous-Led AI: How Indigenous Knowledge Systems Could Push AI to Be More Inclusive. https://www.sshrc-crsh.gc.ca/funding-financement/nfrf-fnfr/stories-histoires/2023/inclusive_artificial_intelligence-intelligence_artificielle_inclusive-eng.aspx (accessed: 2024-20-09).

Nissenbaum, H. 1994. Computing and Accountability. *Communications of the ACM*, 37(1): 72–80.

Noble, S. U. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press. ISBN 978-1-4798-3724-3.

Olson, P. 2024. AI's Advances Will Echo the Internet, Not the Steam Engine. https://www.bloomberg.com/opinion/articles/2024-04-09/no-jamie-dimon-ai-won-t-be-just-like-the-steam-engine (accessed: 2022-23-09).

Orlikowski, W. J. 1992. The Duality of Technology: Rethinking the Concept of Technology in Organizations. *Organization Science*, 3(3): 398–427.

Orr, W.; and Davis, J. L. 2020. Attributions of Ethical Responsibility by Artificial Intelligence Practitioners. *Information, Communication & Society*, 23(5): 719–735.

Pant, A.; Hoda, R.; Spiegler, S. V.; Tantithamthavorn, C.; and Turhan, B. 2024. Ethics in the Age of AI: An Analysis of AI Practitioners' Awareness and Challenges. *ACM Transactions on Software Engineering and Methodology*, 33(3): 1–35.

Pascali, L. 2017. The Wind of Change: Maritime Technology, Trade, and Economic Development. *American Economic Review*, 107(9): 2821–2854.

Pearlson, K. 2024. When Cyberattacks Are Inevitable, Focus on Cyber Resilience. https://hbr.org/2024/07/when-cyberattacks-are-inevitable-focus-on-cyber-resilience (accessed: 2022-23-09).

Pham, B.-C.; and Davies, S. R. 2024. What Problems Is the AI Act Solving? Technological Solutionism, Fundamental Rights, and Trustworthiness in European AI Policy. *Critical Policy Studies*, 19(2): 318–336.

Polonsky, M.; and Jevons, C. 2009. Global Branding and Strategic CSR: An Overview of Three Types of Complexity. *International Marketing Review*, 26(3): 327–347.

Prem, E. 2023. From Ethical AI Frameworks to Tools: A Review of Approaches. *AI and Ethics*, 3(3): 699–716.

Qiang, V.; Rhim, J.; and Moon, Aj. 2024. No Such Thing as One-Size-Fits-All in AI Ethics Frameworks: A Comparative Case Study. *AI & Society*, 39(4): 1975–1994.

Rakova, B.; Yang, J.; Cramer, H.; and Chowdhury, R. 2021. Where Responsible AI Meets Reality: Practitioner Perspectives on Enablers for Shifting Organizational Practices. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1): 7:1–7:23.

Reglitz, M. 2020. The Human Right to Free Internet Access. *Journal of Applied Philosophy*, 37(2): 314–331.

Ruster, L. P.; and Brown, G. 2020. Termination for Cultural Misalignment: Setting up Contract Terms to Ensure Community Well-Being in the Development of AI. *International Journal of Community Well-Being*, 3(4): 523–537.

Sarker, S.; Chatterjee, S.; Xiao, X.; and Elbanna, A. 2019. The Sociotechnical Axis of Cohesion for the IS Discipline: Its Historical Legacy and Its Continued Relevance. *Management Information Systems Quarterly*, 43(3): 695–719.

Scantamburlo, T.; Cortés, A.; and Schacht, M. 2020. Progressing Towards Responsible AI. arXiv:2008.07326.

Schiff, D.; Rakova, B.; Ayesh, A.; Fanti, A.; and Lennon, M. 2021. Explaining the Principles to Practices Gap in AI. *IEEE Technology and Society Magazine*, 40(2): 81–94.

Selbst, A. D. 2021. An Institutional View Of Algorithmic Impact Assessments. *Harvard Journal of Law & Technology*, 35(1): 117–191.

Selbst, A. D.; Boyd, D.; Friedler, S. A.; Venkatasubramanian, S.; and Vertesi, J. 2019. Fairness and Abstraction in Sociotechnical Systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 59–68. Atlanta GA USA: ACM. ISBN 978-1-4503-6125-5.

Sloane, M.; and Zakrzewski, J. 2022. German AI Start-Ups and "AI Ethics": Using A Social Practice Lens for Assessing and Implementing Socio-Technical Innovation. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, 935–947. Seoul Republic of Korea: ACM. ISBN 978-1-4503-9352-2.

Smit, D.; and Eybers, S. 2022. Towards a Socio-Specific Artificial Intelligence Adoption Framework. In *Proceedings of 43rd Conference of the South African Institute of Computer Scientists and Information Technologists*, volume 85, 270–282.

Stahl, B. C.; Antoniou, J.; Ryan, M.; Macnish, K.; and Jiya, T. 2022. Organisational Responses to the Ethical Issues of Artificial Intelligence. *AI & Society*, 37(1): 23–37.

Statistica. 2024. Global Internet Penetration Rate by Region 2024. https://www.statista.com/statistics/269329/penetration-rate-of-the-internet-by-region/ (accessed: 2024-09-12).

Sweeting, B. 2015. Cybernetics of Practice. *Kybernetes*, 44(8/9): 1397–1405.

The Ethics Centre; and Deloitte Access Economics. 2020. The Ethical Advantage: The Economic and Social Benefits of Ethics to Australia. Technical report, The Ethics Centre and Deloitte Access Economics.

Thompson, D. F. 1998. The Moral Responsibility of Many Hands. In *Political Ethics and Public Office*, 40–65. Cambridge, MA: Harvard Univ. Press, 6. print edition. ISBN 978-0-674-68606-9.

Thompson, M. 2015/2016. Beyond Gatekeeping: The Normative Responsibility of Internet Intermediaries. *Vanderbilt Journal of Entertainment & Technology Law*, 18(4): 783–848.

Treviño, L. K.; Weaver, G. R.; Gibson, D. G.; and Toffler, B. L. 1999. Managing Ethics and Legal Compliance: What Works and What Hurts. *California Management Review*, 41(2): 131–151.

Vakkuri, V.; Kemell, K.-K.; Jantunen, M.; and Abrahamsson, P. 2020. "This Is Just a Prototype": How Ethics Are Ignored in Software Startup-Like Environments. In Stray, V.; Hoda, R.; Paasivaara, M.; and Kruchten, P., eds., *Agile Processes in Software Engineering and Extreme Programming*, volume 383, 195–210. Cham: Springer International Publishing. ISBN 978-3-030-49391-2.

Vassilakopoulou, P.; Parmiggiani, E.; Shollo, A.; and Grisot, M. 2022. Responsible AI Concepts, Critical Perspectives and an Information Systems Research Agenda. *Scandinavian Journal of Information Systems*, 34(2): 89–112.

von Foerster, H. 2003. Ethics and Second-Order Cybernetics. In *Understanding Understanding*, 287–304. New York: Springer New York. ISBN 978-0-387-95392-2.

Wang, Y.; Xiong, M.; and Olya, H. 2020. Toward an Understanding of Responsible Artificial Intelligence Practices. In *Proceedings of the 53rd Hawaii International Conference on System Sciences*. Hawaii.

Weaver, G. R.; Treviño, L. K.; and Cochran, P. L. 1999. Corporate Ethics Practices in the Mid-1990's: An Empirical Study of the Fortune 1000. *Journal of Business Ethics*, 18(3): 283–294.

Widder, D. G.; and Nafus, D. 2023. Dislocated Accountabilities in the "AI Supply Chain": Modularity and Developers' Notions of Responsibility. *Big Data & Society*, 10(1): e20539517231177620.

Widder, D. G.; Nafus, D.; Dabbish, L.; and Herbsleb, J. 2022. Limits and Possibilities for "Ethical AI" in Open Source: A Study of Deepfakes. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, 2035–2046. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-9352-2.

Widder, D. G.; Zhen, D.; Dabbish, L.; and Herbsleb, J. 2023. It's about Power: What Ethical Concerns Do Software Engineers Have, and What Do They (Feel They Can) Do about Them? In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, 467–479. New York, NY, USA: Association for Computing Machinery. ISBN 9798400701924.

Wiener, N. 1947. A Scientist Rebels, Letter in Bulletin of the Atomic Scientists. https://cdn.theatlantic.com/media/archives/1947/01/179-1/132381596.pdf (accessed: 2022-20-01).

Wiener, N. 1950. *The Human Use Of Human Beings: Cybernetics And Society*. London: Spottiswoode. ISBN 978-0-7867-5226-3.

Wilkinson, M. D.; Dumontier, M.; Aalbersberg, I. J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; Da Silva Santos, L. B.; Bourne, P. E.; Bouwman, J.; Brookes, A. J.; Clark, T.; Crosas, M.; Dillo, I.; Dumon, O.; Edmunds, S.; Evelo, C. T.; Finkers, R.; Gonzalez-Beltran, A.; Gray, A. J.; Groth, P.; Goble, C.; Grethe, J. S.; Heringa, J.; 'T Hoen, P. A.; Hooft, R.; Kuhn, T.; Kok, R.; Kok, J.; Lusher, S. J.; Martone, M. E.; Mons, A.; Packer, A. L.; Persson, B.; Rocca-Serra, P.; Roos, M.; Van Schaik, R.; Sansone, S.-A.; Schultes, E.; Sengstag, T.; Slater, T.; Strawn, G.; Swertz, M. A.; Thompson, M.; Van Der Lei, J.; Van Mulligen, E.; Velterop, J.; Waagmeester, A.; Wittenburg, P.; Wolstencroft, K.; Zhao, J.; and Mons, B. 2016. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data*, 3(1): Article 160018.

Winner, L. 1980. Do Artifacts Have Politics? *Daedalus*, 109(1): 121–136.

Wisnioski, M. 2012. The Crisis of Technology as a Crisis of Responsibility. In *Engineers for Change: Competing Visions of Technology in 1960s America*, Engineering Studies Series. Cambridge, MA: MIT Press. ISBN 978-0-262-30518-1 978-1-283-70748-0.

World Economic Forum. 2024. Generative AI: Steam Engine of the Fourth Industrial Revolution? https://www.weforum.org/events/world-economic-forum-annual-meeting-2024/sessions/industry-applications-of-generative-ai/ (accessed: 2024-23-09).