The Gaps that Never Were: Reconsidering Responsible AI's Principle-Practice Problem

Lorenn P Ruster*
lorenn.ruster@anu.edu.au
School of Cybernetics, Australian National University
Canberra, Australia

Jenny L Davis
jennifer.davis@vanderbilt.edu
Department of Sociology, Vanderbilt University
Nashville, USA; School of Sociology, Australian National University
Canberra, Australia; Center for Democracy and Technology
Washington D.C., USA

Abstract

Metaphors shape perception and action. Across domains of responsible AI, the 'gap-bridge' metaphor proliferates. Troubled by an ostensible gap between principles and practices, the field has produced an array of bridging instruments in the form of toolkits, guidelines, and frameworks. The corpus of these instruments is expansive and robust. Yet, the principle-practice problem persists. Rather than propose new and better bridging devices, we step back to interrogate the metaphor itself. Reading metaphors as artifacts, we breakdown the elements of bridges and gaps, abstracting these into identifiable characteristics: *separation, linearity,* and *stasis*. We then draw on fieldwork with three startup organizations to understand whether and how gap-bridge characteristics manifest. The organizations under study all have AI in their portfolios and are actively working to (re)construct their principles and practices, thus showcasing the principle-practice problem in action. Findings show the gap-bridge metaphor misfits with these organizations' in-situ efforts, in which principles and practices are integrated, nonlinear, and subject to dynamic values that transform across time and circumstance. With these findings in mind, we consider alternative metaphors and invite collective reconsideration of principles, practices, and their relationship to responsible AI.

CCS Concepts

• **Human-centered computing → Interaction design theory, concepts and paradigms**.

Keywords

Affordances, Principle-Practice Relation, Responsible AI

ACM Reference Format:
Lorenn P Ruster and Jenny L Davis. 2025. The Gaps that Never Were: Reconsidering Responsible AI's Principle-Practice Problem. In *The 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25), June 23–26, 2025, Athens, Greece.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3715275.3732024

*Corresponding Author

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.
FAccT '25, June 23–26, 2025, Athens, Greece
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1482-5/2025/06
https://doi.org/10.1145/3715275.3732024

1 Introduction

An ostensive 'gap' between principles and practices has long plagued the field of responsible AI[1]. This gap bears blame for a bevy of AI related harms, while apparently undermining efforts to realize AI's potential for social good. Researchers, activists, government, and industry have converged on the problem, devising guidelines, toolkits, and frameworks to bridge the divide between principles and practices such that community values materialize in code.

Morley and colleagues' 2020 review of the AI literature identified 425 'bridging' resources in academic publications alone, acknowledging the multitudes that eluded or fell outside of their search parameters [38]. These resources have since been joined, appended, revised, and displaced by scholarly and corporate outputs that together constitute a billowing and ever-evolving corpus [4, 6]. Yet these various instruments—numerous in quantity and diverse in kind—are not living up to their promise[2]. Each new entrant indicates the inefficacy of those that came before, as the need for more bridges evinces gaps that endure and persist.

In the present work we do *not* present a more optimal bridging tool. Instead, we step back to question the gap-bridge metaphor itself. We approach this question both analytically and empirically, unpacking the assumptions and characteristics of a gap-bridge metaphor and using those assumptions as a guidepost for fieldwork with three organizations in Australia's technology startup sector. Each organization works with AI, expresses affirmative commitments to social responsibility, and is in a building and/or rebuilding phase focused on principle-practice development. These cases instantiate the principle-practice problem in action, fostering insight about the gap-bridge metaphor as an object of inquiry and orienting device

[1] Responsible AI refers to a field of scholarship, activism, and practice committed to social and political considerations in the production and deployment of AI technologies. Responsible AI is often used interchangeably with AI ethics, and we move between the two terms in the present work. We note, however, that responsible AI and AI ethics are distinct from the now dated term 'ethical AI' due to the latter's problematic implication that technology itself has an ethics or can act ethically.

[2] Many valid reasons explain the persistence of AI's principle-practice problem—a misplaced focus on AI instead of its production environments [3, 8, 51], a lack of care and time allocated to the process early enough and with the right stakeholders [7], prescription-based and technosolutionist approaches that lack ethical reflection [26], an emphasis on useability instead of effectiveness [9], and a framing around problem/solution and rationality instead of relationality [10], amongst others

2 Metaphors As Artifacts

Metaphors are consequential, configuring both perception and action [11, 19, 29, 30]. The power of metaphor is clear across sociotechnical domains. For instance, the metaphor of an 'information superhighway' emphasizes speed while treating information as homogenous cargo rather than diverse and generative co-productions [23] the data-as-oil and cloud metaphors convey data as both natural and immanently extractable, unmoored from physical infrastructures and human labor [1, 52, 56]; and framing people as quantified cases, as with insurance and other actuarial algorithms, fosters comparative justice ethics while foreclosing holistic accounts of complex human subjects and the intrinsic dignity those subjects warrant [28].

With their social origins and societal effects, metaphors have an artifactual quality, reflecting and structuring material and symbolic culture. We thus read the gap-bridge metaphor as an analyzable object, with affordances that enable and constrain [28]. 'Affordance' is an analytic device used to delineate how the properties of an artifact shape that artifact's uses and effects [22, 39]. Although traditionally applied to material technologies, affordances are relevant to *any* artifact of interest. Following Kiviat, we consider the affordances of a sociotechnical metaphor [28], relying in particular on the mechanisms and conditions framework of affordances (M&C framework). The M&C framework is an operational model that specifies how artifacts *request, demand, encourage, discourage, refuse,* and *allow* across subjects and circumstances, capturing the push and pull of artifacts in action [13–15]. With this theoretical framing, we investigate the implications of a gap-bridge metaphor—how it affords—and consider alternative framings that may afford otherwise. In doing so, we at once analyze an axiomatic construct (i.e., the gap-bridge metaphor), level a critique, and lay ground for alternative framings of a pervasive problem space.

3 The Gap-Bridge Metaphor in Responsible AI

Framing responsible AI in terms of a gap-bridge metaphor is commonplace. Here we document the scale and scope of this metaphor across academic and grey literature (i.e., government and industry reports). This background sets the stage for analytic scrutiny and empirical investigation of principle-practice development.

In 2019, Mittelstadt's influential paper in the journal *Nature* forewarned that AI ethics principles alone cannot guarantee 'ethical AI' [36]. Here, he spoke of various challenges associated with a principled approach to AI ethics, one of which was 'the absence of proven methods to translate principles into practice' [[36], p.503]. This notion of translation from principles to practice has stuck; some refer to it as 'operationalizing AI ethics' [27, 37] or the 'what' (principles) and 'how' (practice) of AI ethics [18]. Schiff and colleagues provide six explanations for why responsible AI principles may never make it into practice – mismatched organizational incentives, an underestimation of the complexity involved in working through the impacts, the number of disciplines that need to be involved, organizational structures that hamper shared responsibilities, governance challenges, and the fact that there are too many tools and frameworks, leading to overwhelm and paralysis [49].

Direct use of gapped language is rife, exemplified through the turn of phrase 'bridging the gap' (see for example [26, 50]) and derivatives thereof. For example, Ibáñez and Olmeda explore the 'gap between practice and principles in AI ethics' [27]; Schiff et al.'s explanations try to account for the unbridged gap between principles and practice [49]; a 2023 whitepaper from Australia's National Artificial Intelligence Centre (NAIC) promises to 'support bridging the gap between the Australian AI Ethics Principles and the business practice of responsible AI' [[47], p.iii]; and an amendment to the EU AI Act indicates that:

> ...in order to foster the development of AI systems in line with Union values, the Union needs to address the main *gaps* and barriers blocking the potential of the digital transformation.... (emphasis added) [[43], Amendment 17]

Indeed, scholars and practitioners are occupied by 'addressing ethical gaps' [44], 'closing the gap in operationalizing responsible AI' [33], and 'closing the AI accountability gap' [45], all ostensibly achieved with bridging tools[3].

We can conjecture about why this gap-bridge metaphor has found traction in the responsible AI field—such as its presence in an early and influential paper [36] and its established use in other domains. The gap-bridge metaphor is common not only in the context of responsible AI but resonates more broadly with academic audiences – for example 'gaps in the literature', 'research gaps', the 'research-practice gap' [34, 48] – and also with industry. In an industry context, it is common vernacular to speak of 'market gaps' [40], 'skill or capability gaps' [35], to undertake 'a gap analysis' [32] etc. Gap metaphors also pervade policy rhetoric. In Australia (our context of study), 'Closing the Gap' is a national strategy to reduce the disparity between Indigenous and non-Indigenous communities on issues such as life expectancy, health, justice, digital inclusion and education outcomes [2] and the 'gender pay gap' is a key metric in the Australian Bureau of Statistics national and business surveys [41]. Accordingly, the gap metaphor acts as a bounded frame within which different audiences – academic, industry and policy – can participate, problem-solve, innovate and collaborate [31].

Further investigating the metaphor's origins and influence is backward looking and outside the scope of the present work. We focus instead on the fact that this gap-bridge metaphor pervades, yet never seems to resolve—see Schiff et al.'s point about the stifling quantity of bridging tools [49]. The persistence of a problem amidst so many fixes indicates that the problem space may not be configured as it seems. Our purpose is thus to understand how a defining metaphor comports with and/or diverges from professional praxis.

3.1 Unpacking the Gap-Bridge Metaphor

Metaphors are inextricable from their physical representation. In this case, these representations are of bridges and the gaps those bridges close. Bridges are weightbearing structures with polemic ends, linked by a tightly bounded pathway. The bridge elevates above an air-filled cavity, under which water, cars, wind and/or earth sit below. This cavity may be still or active, with flowing rivers, for example, or stagnant rocks. Bridges serve the function of transport, supporting people and things that move from one pole to the other and from the other pole back to the first. That

[3]Even one of the authors of this paper has used a gap-bridge metaphor to motivate a novel approach to AI assessment and design in a recent publication [14]

which traverses a bridge is channeled by guardrails erected on lateral sides. The bridge has clear start and end points onto which travelers enter and from which they exit. Bridges are conduits rather than destinations, meant for temporary occupancy on the way from one place to the next.

Bridges can take many forms. They may be one lane or many, rickety or strong, suspended or trussed and so forth. So too, have responsible AI bridging devices come in multiple varieties, with a near infinite collection of variables that may be selected to create new implements that are different from old, and ostensibly better ones than those that came before. Yet all bridges share core characteristics which will always be present in their physical manifestation and thus, in their metaphoric application. These characteristics are *polemic separation, linear trajectories,* and *static construction.* Bridges are marked by polar entrance and exit points, always separate and distinct from the other, connected only through a third element that touches both simultaneously. The pathway between entrance and exit is linear, following a straight path along a singular bounded channel. The linear pathway between polemic thresholds is stable and enduring, as are the thresholds themselves. Bridges are not modular and moldable, but static structures that hold strong as the world moves over, under, through, and around them.

As applied to responsible AI and shown in Table 1, the core elements of bridges imply a hard distinction between principles and practices, singular pathways from one to the other or from the other to the first, and a stable set of standards or values, to be achieved through the bridge's narrow parameters, and then left behind once traversal is complete. Through our fieldwork, we examine whether and how these characteristics, or other ones, emerge in-situ.

4 Methods

Our study is based on participatory research with three organizations in the Australian startup sector. Each organization is committed to responsible technology, has AI as part of their portfolio, and is in a common context of early-stage startups – two (OrgB and OrgT) are Australian-founded, early-stage startups and one (OrgH) is an ecosystem-enabling 'Think+Do tank' that supports early-stage startups. All three organizations are engaged with AI— either building AI-enabled products now, planning to do so in the near future, and/or supporting organizations that build and use AI in their businesses. All three organizations have an expressed commitment to 'being responsible' in the technologies they develop (or assist others to develop) and were in a development or rebuilding phase at the time of study, focused on establishing and/or refining organizational principles and practices. These cases thus offer relevant contexts to interrogate principles and practices in-situ, as they interplay with responsible technology in general, and responsible AI in particular.

The interaction between the case organizations and the primary researcher is grounded in intervention research [5, 25, 53] – that is, the researcher is not an independent outsider or a passive participant, but actively involved with the organizations themselves. As such, this research holds mutual benefit to both researcher and case organizations, underpinned by trust and collaboration [46]. Data were gleaned from these cases using a variety of methods, with different foci depending on organizational context and the relevant goals for both researchers and participants. These included a combination of participatory workshops and document analysis, with varying levels of emphasis that we describe in Section 4.1.

Participatory workshops were designed according to the pragmatic goals of each organization. The researcher often provided a broad structure for activities in the participatory workshops which was then iterated upon and co-designed with participants. The collaboration looked different in each case. For OrgB, the shared purpose was to reconstitute their organizational values so that those values would reflect a growing leadership team; OrgT was focused on generating a multifaceted responsible AI pledge; and for OrgH there was mutual value in understanding the genesis and evolution of the organization and reflecting on the uniqueness of its organizational design. Across cases, interactions between the primary researcher and the target organization lasted a year or more.

We analyzed the full body of data through abductive methods, moving between theory, literature, and the grounded insights particular to our sites of study [57]. The first author, who conducted the fieldwork, also led the analysis, with the second author providing insight and perspective from a more distanced vantage. This entailed a dialogic process of identifying how principles and practices figured into participants' efforts towards social responsibility, sensitized by the 'gaps' and 'bridges' that heretofore occupy responsible AI literature. Here we incorporated reflexive thematic analysis [12], exploring how the data reflect principle-practice relations through thematic codes such as "iterative nature", "operationalization" etc. Utilizing the M&C framework as a theory-method scaffold, we overlay the data with an affordance vocabulary, whereby the varied renderings of principles and practices request, demand, encourage, discourage, *refuse,* and *allow* various outcomes. The primary researcher comes to this work with a decade of experience as a consultant predating academic pursuits, as well as experiences in building both entrepreneurial and intrapreneurial ventures. This was vital for gaining access to and trust from tech startup organizations. The second author is an academic with experience studying the AI industry, offering a theoretically grounded but practically distant perspective on both study design and data analysis.

4.1 Cases

OrgB is an Australian-founded, early-stage startup which began with two co-founders and currently has four co-founders. OrgB's mission is to demystify sustainability and make it easy for Small Medium Enterprises (SMEs) to take sustainability actions. This is done by providing customized recommendations regarding sustainability actions that SMEs can take and connecting SMEs with service providers they can use. Org B also keeps track of and rewards SMEs' progress through the creation and monitoring of an SME's Environmental Social Governance (ESG) score. OrgB's current technology platform does not yet involve AI, however this is a core component of their near-term growth and scale plans, and a strong motivator for refining their values. At the time of data collection, OrgB was testing a Minimum Viable Product (MVP) with a range of SMEs, and hoping to raise seed capital in the next six months. Believing that organizational values are important to align on before undertaking significant scale and growth, OrgB engaged with the primary researcher to adapt the initial value set developed by the first two co-founders. This adaptation would reflect all four

Table 1: Gap-Bridge Characteristics Applied to Responsible AI

Separation	Principles and practices are fundamentally separate and thereby need 'bridging'
Linearity	Linear movement from principles to practices *or* practices to principles
Stasis	Static start and/or end points that can be left once principles are achieved through practice

co-founders of the now expanded team. The primary researcher conducted 15 interactions with the four co-founders of OrgB, individually and together. We draw in particular on four individual interviews, two reflection-style interviews, and one participatory workshop.

OrgT is an Australian-founded, early-stage startup that at the time of data collection, had yet to raise substantial capital and was bootstrapping its current operations. OrgT had built an AI-enabled browser extension (with an app forthcoming) that aims to burst narrow news bubbles and expose users to a variety of information sources. The product suite is based on a recommender algorithm fueled by a Generative Pre-trained Transformer (GPT-3). OrgT engaged with the primary researcher because they wanted to strengthen social responsibility as a foundational element of their organizational ethic and business model. Over the course of 18-months, the primary researcher conducted 15 participatory workshops and three reflection interviews with the three co-founders of OrgT to develop a set of responsible AI pledges that would guide further development of their product suite. Finally, OrgH is an ecosystem-enabling 'Think+Do tank' that supports several startups creating responsible technology for humanitarian purposes. OrgH is affiliated with a large international humanitarian organization and is philanthropically funded. There are several important components of OrgH. First, it runs an accelerator-like program called The Lab. The Lab provides funding, resources, and mentorship to technology startups, many of which are using AI in their products. Second, OrgH established a 'prototype' – a resource to assist and guide responsible technology design, development, and implementation. The prototype represents what is important to OrgH and to the startups they support. Third, OrgH runs a yearly Summit that connects technologists, community leaders, corporate innovators, policy makers, academics, and the humanitarian sector. Data collected and analyzed include 6 individual interviews with members of OrgH and analysis of internal and publicly available documents.

5 Analysis

The argument to this point stipulates that (1) metaphors shape how we think and act; and (2) current responsible AI discourse describes a relationship between principles and practices that is 'gapped' and thereby needs 'bridging'. In this section, we analyze fieldwork with three tech organizations, each engaging with AI and committed to social responsibility. Our analysis is premised on a comparison between the characteristics of a gap-bridge metaphor—separate, linear, and static (see Table 1)—and the in-situ realities for the three organizations under study. Across all three cases, we examined the gap-bridge metaphor and its component parts. For the sake of analytic clarity, however, we present each component part on its own, highlighted through a single case study that elucidates the processes at play.

5.1 Observing the Gap-Bridge Metaphor

5.1.1 Separate. The first metaphoric characteristic is that of separation, where two poles sit opposite each other, linked by a bridging conduit. Here we would expect principles to develop separate from practices, and practices separate from principles, only affecting each other once fully formed. That is, a gap-bridge metaphor *demands* principle-practice distinction. Such distinction does not feature in our data, which is instead, marked by entanglement and integration. To illustrate this point, we draw upon the case of OrgH, focusing on their organizational design, linguistic choices, and the tools this organization developed. These organizational, linguistic, and technical elements together *request* and *encourage* the integration of principles and practices, resisting a born-separate ontology.

Think+Do tank model

OrgH was established as a Think+Do tank. This structural choice—to integrate principles (think) and practices (do)—was deliberate and, from the perspective of the co-founders, critical to its credibility as an organization advancing responsible technology. Initially this view was not shared by their executive sponsors:

> They [the executives] wanted us to be a think tank…and our point was 'that's great, however unless we can actually gather some evidence from our own engagement with technology and test some of our thinking through that engagement, people won't take us seriously'…it's not necessarily something that we can just do as a third-party observer. (OrgH initiator)

Eventually the executives were convinced, and it was agreed that a Lab – a place where startups would test and learn with responsible technology principles - would be central to OrgH's remit.

Over five years since its inception, OrgH's decision to integrate principles and practices is considered a defining element of the organization, which prides itself and the startups it supports, on being pragmatic and principled at the same time:

> I think one of the unique components is having a Lab where we are working closely with people who are designing, testing, developing technologies. So it wasn't in a vacuum… we were trying to be very practical and pragmatic in understanding the limitations of these principles, rather than going 'here you go, here's these lofty ambitions but actually the reality is just impossible to implement'…. Making sure that they [principles and practices] were a bit more symbiotic. (OrgH initiator)

Such integration is further reflected in how the Lab operates on an ongoing basis:

> We wanted to make sure that we were not just telling people about the principles, but we were also working with the principles... the principles led our way

> of working as well… reporting… program design… every learning workshop [in the Lab] had to be in some way related or concerned with the principles… everything is one big virtuous circle. (OrgH Lab Lead)

'Prototype' language

One of OrgH's core activities was the co-creation of a 'prototype' – a resource to support and guide responsible technology design, development and implementation. OrgH was intentional in its labeling of this resource as a 'protoype'. They wanted to signal a co-existing practicality or tangibility to principles as they materialize in practice. According to one co-founder of OrgH, 'a prototype has a sense that you can *do* something with it' [emphasis added]. For another, engaged as a volunteer, tangibility was important: 'A prototype tends to have associations in my eyes, not just with concepts or guidelines, but to a prototype of something'.

They also chose the word 'prototype' because it gestures towards an iterative approach. One of the co-founders reflects, 'this was the first time that we'd actually published something externally that was unfinished and acknowledged it and said this is part of our iterative process'. This indicates a sense that principles and practices co-exist in cycles of mutual and ongoing development. It also speaks to the continual nature of principles and practices, a topic we touch on more fully in Section 5.1.3.

Principle-practice tools

The responsible tech prototype was tested with different cohorts of startups who were a part of the Lab. In doing so, a range of additional tools were developed, also intertwining principles and practices. The first was a diagnostic tool that assisted Lab participants in moving beyond siloed responsible tech principles to instead, actively integrate them into their own practices. The second tool developed was a set of cards. These cards featured each of the responsible tech principles, a short statement and then a series of questions that enabled the Lab cohort to surface what the principles meant for them in practice. Both tools facilitated ways for the principles and practices to co-exist and enabled discussion of principles and practices in a mutually reinforcing relation:

> [Once we] showed the cards, I felt like the conversations got more rich… the questions are really great because people were actually understanding the principles better once they had the prompts…the most resounding response was 'they're really useful…I hadn't thought of that. I will now think about that more in my work'. (OrgH Lab Lead).

Here we observe principles animating practices, and practices (re)conceived in light of principles. This integration between principles and practices resists the notion of separate or polemic ends, highlighting instead a merging of ideals and pragmatics.

5.1.2 Linear. Moving in a linear fashion *from* principles *to* practices, or the reverse, would align with a gap-bridge metaphor, which *requests* continuity and narrowly confined routes. Instead, we find evidence of nonlinearity; that is, moving continuously between principles and practices in a multidirectional and multidimensional fashion. To demonstrate a nonlinear relation between principles and practices, we turn to the responsible AI pledge creation process undertaken by OrgT. Here, we observe twists, turns, offshoots and spirals that exceed the strictures of linear paths.

Generating responsible AI pledges

OrgT was focused on building responsible AI pledges, and they approached this task from a cybernetic foundation premised on interconnected feedback loops between principles and practices. For example, in one phase of the process, the co-founders explored what other organizations are doing in terms of their responsible AI principles. The idea was to assist OrgT in choosing responsible AI principles for their own context; however, that was not what happened. What transpired was a real fear that they may adopt principles that they cannot uphold in practice and thereby set themselves up for failure, become perpetrators of ethics-washing, and ultimately erode user trust. Reflecting on the experience of studying other organizations' principles, OrgT's co-founder talked about toggling between common ethical priorities expressed by these organizations and the realities OrgT faced as practitioners in their daily work.

> A very common question for us [became], 'but how do we know you're not biased?' 'What are your algorithms and how do they work?' I think that this [work of creating responsible AI pledges] is essential, methodologically…and commercially because of the nature of our audience…I think that intentionality really is the word because we're trying to cover every angle while we build rather than just go 'let's see what happens, whatever!'. (OrgT co-founder)

OrgT's movement was not a straight line from ideals to implementation. Rather, they paused, explored various side-roads (other organizations' principles), circled back to themselves, dove deep into technical reflections, and resurfaced from a new starting place.

Consequently, a phase followed that the co-founders called 'developing an intuition' for what responsible AI principles mean in practice. This looked like a series of discussions about what they had already done and how that reflected what they would or could practically do if they were to adopt a certain principle. This conversation was oriented by a combination of pragmatism, discernment of feasibility, and thought experimentation, by which principles, practices, and their relation shifted in tone, depth, and dimension:

> Let's say we want to use all the pillars [of the responsible AI pledges] and do hundreds of features…How should we talk about it when we're planning it? How should we code it up? What checks and balances should we have in place every time we write a line of code…or propose some funding for this particular feature…what are the extra questions we should be asking…about the particular pillar of ethics that we've picked? (OrgT co-founder)

At one point, the team faced a tension between the principles that they aspired to adopt and the realities of their current product in practice. A core stated value within the organization was that of 'autonomy and freedom'. They worked to achieve this through the content of their news delivery algorithm. However, they came to understand that their algorithm drew from a limited data repository, due to economic and technical limitations of GPT-3 at the time. This realization sparked a reconsideration among the organization's

leadership about what freedom and autonomy mean in practice, and if and how that can be achievable. In considering limitations and costs, the team adjusted both the product itself—optimizing for plurality of perspectives—and shifted their operationalization of autonomy and freedom as a driving value. What was once broad and absolute would now be narrower and tightly scoped—reshaped and repositioned to account for the realities of data availability and concurrent values of economic prudency. The trajectory was therefore not a singular line from principles to practices, nor from ill-conceived principles and practices to better ones. Rather, we observed an irregular movement whereby a practice (data processing) gave pause in relation to a principle (freedom and autonomy), prompting the reshaping of that principle, a cyclical return to practice, and a lateral consideration of economic responsibility as a concomitant value that inflects the others.

The conversations about both product choices and value sets flowed between principles and practices at different levels. For example, the co-founders reflected amongst themselves how their current product choices already demonstrated what was important to them and how a particular principle could help them prioritize certain product features. Following this exploration of what the principles mean or could mean in practice, they consolidated their 'intuition' back into a higher-level principles discussion, ultimately creating pledges as living documents that evolved with product development.

For OrgT, the general format of each pledge consisted of three parts. First, a heading that expressed the general principle; second, the first sentence that summarized the pledge at a high level, in aspirational language; and third, a sentence or two that explains what the aspiration looks like in practice, beginning with 'For us, this means…'. Even once OrgT landed on a set of responsible AI pledges that they determined were sufficiently developed for publication on their company website, the *nonlinearity* of the principle-practice relation continued; they made it clear that the pledges will be perpetually tested and refined as they learn by doing—shifting, revisiting, and reconfiguring both values and practices through starts and stops, cycles, mergers, and outgrowths.

5.1.3 Static. Bridges are static, with clear stop and end points. Applied to responsible AI, this implies that stable values can be identified, achieved, and then left. Instead, we observe value-dynamism, in which the relationship between principles and practices remains unsettled, in-flux, and always situationally contingent. We draw on OrgB to exemplify value-dynamism in two ways. First, OrgB was a growing operation, and worked with the primary researcher to reconsider organizational values in light of new members joining their leadership team. Second, value-dynamism is present in the way one co-founder initiated efforts to reconstitute the organization's values – using ChatGPT as a reflective assistant. In these ways, the gap-bridge metaphor, which *demands* rigidity and *encourages* finality, is resisted by a fluid movement that *requests* flexibility and *encourages* ongoing reconsideration.

Values adjustment

OrgB began with two co-founders, later adding two more to complete a four-person team. The original co-founders established a set of initial organizational values, representing their own inclinations. They used these values to sense-check the third co-founder's 'culture fit' for the organization, at which point the values began to evolve:

> [The values came up] very early on… [initial co-founder] was like these are our values… Any deal-breakers for you? Anything missing?...Talk to me now because there's no point having the conversation that happens next if you're not aligned with what we're trying to do…and that's where I started to say, well, yeah, inclusion is one thing that I want explicitly written in. (Third co-founder of OrgB)

As one of the initial co-founders reflected: 'there's nothing there [in the original OrgB value set] you can be fundamentally opposed to, but it's not necessarily reflective of what they [the other co-founders] truly need'. Consequently, OrgB underwent a process with the primary researcher to review and rework their organizational values to represent the full co-founding team. There was an acknowledgement that the values need to be representative of all four co-founders for them to feel ownership in the organization, which was deemed critical to their individual and collective success. Explicitly adding 'inclusion' was a direct aim, given the new co-founders expressed prioritization of this value. One of the original co-founders explained:

> I think you can argue that inclusivity should be mentioned somewhere more than it is currently. It's definitely a value that we all hold high and that has been driving [new co-founder] for the last couple of years even more than us. (initial co-founder of OrgB)

As a result of the process, 'inclusion' was proposed as part of OrgB's core values, evolving through conversation, reflection, interaction, and iteration, reflecting a principle-practice relation that remains open and pliant.

In addition, OrgB recognized that values are active and context dependent; that is, the organizational values may mean different things to each co-founder and those meanings might change across time and circumstance. One of the original values for OrgB was 'dare to challenge'. For one of the newer co-founders, who is particularly focused on partnerships with SMEs as part of their role, this was confusing:

> Not really sure about that [dare to challenge]…we are trying to be innovative. We are trying to help and support SMEs, push them along the Environmental Social Governance (ESG) journey, but I don't really see it as challenging because something we really try to do is meet the SME's where they're at.… The whole point is we want to make it easy and digestible and manageable for them. (Fourth co-founder of OrgB)

As part of rethinking the value set, it was proposed that 'dare to challenge' be removed, and some of its meaning encompassed in the idea of 'staying bold'. This captured the value of 'innovation' without insinuating that the product was going to challenge users in a way that dissuaded adoption. Overall, we observed values that are dependent on interpretation and context, and that shift dynamically in the collective hands of an evolving organization.

Reflecting with ChatGPT

In the process of engaging with each co-founder regarding their views on OrgB's values, an unanticipated action occurred. The developer co-founder revealed that they had undertaken their own process of understanding the values of the organization in context: namely, they engaged ChatGPT to make suggestions about implementing organizational values at different phases of the user experience journey (their primary domain of activity in OrgB at that time). They asked ChatGPT to generate a table of suggestions organized by value and by phase of the user's experience. They reviewed and made minor adjustments to the outputs and then assessed current activities against the ChatGPT-created matrix.

This activity reflects the need for this co-founder to understand values within a tangible context in order to think more broadly about organizational principles in the abstract. The value-dynamism embedded in the approach is evident in how they explicitly prompted ChatGPT to consider the way values manifest across distinct phases and under various conditions. By undertaking this activity, the developer asked 'What are the things we could do to demonstrate that value in that phase?'. The context specificity of the query is an important framing, recognizing that values in the discovery phase, for example, differ to those of the reward phase.

The way the co-founder assessed current activities against ChatGPT outputs (by shading them in different status colors) also points towards the dynamic nature of principles and practices, reflecting an ongoing project of (re)consideration:

> The green-coded items we've kind of done… yellow-coded items we've got but could be better at, so it's not quite fully developed. The blue-coded items will be happening as we start to do testing… the gray-coded items we haven't yet included… the orange-coded items are in the roadmap for later… [this] identified some… things we should be trying to do through the iteration of the product. (Third / developer co-founder of OrgB)

For this co-founder in particular, the approach of the interview – to discuss the meaning of the organizational values – necessitated a consideration of these values in changing contexts (value dynamism), for which they later engaged ChatGPT to help map a moving perspective.

5.2 Rethinking the Gap-Bridge Metaphor

Interrogating the gap-bridge metaphor, we can begin to understand why instruments meant to unite principles with practices have remained ineffective—and why new and improved bridging tools are not likely to do better. The metaphor itself is misfitted to the realities of technology production and deployment. Bridges are sclerotic, heavy, linear, and stable, connecting two fixed points through a set and singular structure. This diverges from the dynamic, fluid, reflexive and multifaceted environments from which AI emerges and in which practitioners operate. The core characteristics of a gap-bridge metaphor might then be reconceived as the metaphor's fallacies. In their place, we offer correctives (see Table 2).

5.2.1 From Separate to Integrated. A gap between two or more things implies those things are siloed and distinct, hence the inevitable 'bridge' metaphor as a means of infrastructural linkage. Bridges close the apparent distance between two bounded sites, enabling transfer and connection between one enclosed thing and another. With this logic, principles are contained objects disconnected from practices—an arrangement that *demands* polemic binaries, while also *discouraging*, perhaps *refusing* recognition of the ways that principles and practices entangle and co-constitute each other. Such co-constitution is axiomatic in studies of social psychology, cultural sociology, and social studies of design, whereby values at once shape and emerge from actions and doings [16, 19, 54, 55]. Naïve to these sociological insights, the gap-bridge metaphor instead precipitates tools, resources, frameworks and so forth that inevitably exist outside of principles and practices themselves, offering instruments that assume two separate objects in need of an effective go-between.

The first corrective we propose is thus from separation to integration, reflecting how principles and practices operate with and through each other. Integration refutes the premise that principles and practices are born separate and instead emphasizes their symbiosis and intertwining. Far from siloed, integrated principles and practices co-constitute and interconnect with each other, as highlighted especially in OrgH's Think+Do Tank model, prototyping process, and the tools they developed for weaving principles *with* practices.

5.2.2 From Linear to Nonlinear. Wrapped-up with the presumed separation between principles and practices is a second fallacy, that of linearity. Here the gap metaphor brings principles to practices or practices to principles in a one-way trip that *refuses* lateral, horizontal, irregular, or circular movement. Correcting for linearity, we propose nonlinearity, whereby movements between principles and practices are continuous, refracted, fractured, and reciprocal. We observed this with particular clarity when considering OrgT's construction of responsible AI pledges. This was an ongoing process in which the organization broke off to consider others' principles, circled back to their own, burrowed into technical considerations, and zoomed out to reflect on higher level questions

5.2.3 From Static to Dynamic. Gap-bridge metaphors *request* value-stasis, whereby values are either start or end points, to be practically achieved and then let alone. Yet, values are socially embedded processes that shift and evolve. Normative standards change over time and across contexts, as abundantly clear during the Black Lives Matter (BLM) and #MeToo movements, in which sensitivities to race and gender recast past behaviors in new and incriminating light. Such is also the case for the problem spaces of 'tech ethics' and 'responsible AI'. Cultural norms reconfigure in ways that make technical features that were once acceptable, less so, and others that were inconceivable, a standard requirement. With these value shifts, and at times instigating them, are material and practical changes in the ways things are made [20, 21]. For example, de-identification was once an adequate form of data privacy protection, whereas now, attention to proxy variables, along with regulations about aggregate and metadata shape how data are, and legally can be, collected, stored, and used [17, 24, 42]. The rigidity of a gap-bridge metaphor encumbers ethical agility as values, technologies, and values about technologies change over time and across contexts. Moreover, this rigidity instills a false belief that ethics can be 'done' or complete, eliding the ongoing project of constant (re)consideration.

Within our fieldwork, values were unstable and unsettled, leading to our final correction—a shift from stasis to dynamism. Such dynamism was acute for OrgB, whose changing team generated significant efforts to reconsider the organization's driving values and how those values would constitute in practice. Not only did the organizational principles transform to reflect new co-founders, but all co-founders rethought their own value hierarchies through emergent relationships and reflective activities. This value-dynamism indicates that, unlike bridges, there is no off ramp or end point, but a continuous reconsideration that *refuses* finalities in responsible AI fields.

6 Summary & Conclusions

Across domains, significant energy has gone towards bridging the gap between principles and practices in AI development and implementation. These efforts have materialized in toolkits, guidelines, frameworks, and technical interventions that now saturate the AI landscape. Motivated by a commitment to minimizing AI harms while realizing AI's positive potential, these 'bridging' instruments continue in droves, each ostensibly iterating, displacing, or otherwise improving upon those that came before. Yet the sheer volume of such instruments and their perpetual shortcomings call into question the validity of a gap-bridging approach to the principle-practice problem. In the present work, we thus challenged this prominent metaphor and explored how principles and practices might be otherwise conceived.

Coupling a review of academic and grey literatures with embedded fieldwork, we identified fallacies in gap-bridge framing and proposed correctives that resonate with principles and practices as they relate in-situ. We argued that the gap-bridge metaphor creates a problem space defined by bounded separation between principles and practices, linear movement between principles and practices, and static values that can be forgotten once achieved. This configuration diverges from the principle-practice integration, nonlinearity, and value-dynamism that emerged across the tech organizations we studied, as those organizations strove towards social responsibility. From this, we suggest that gap-bridge metaphors not only misrepresent principle-practice relations, but in doing so, hinder efforts towards 'responsible AI', distracting practitioners with resources that are misfit to purpose. As such, we urge practitioners to stop relying on 'bridging toolkits' because they are borne out of this misfitted metaphor; it is little wonder that they fail to resolve the challenge of practical action on responsible AI. Instead, we encourage practitioners and policy-makers to consider approaches that intertwine and integrate principles and practices in ways that account for iterative developments in norms and products. Further, we hope to see academics suggesting new metaphors and analytic frames that more representatively capture the principle-practice relation. In the next section we suggest some of our own starting points for metaphoric revision.

6.1 If not gaps and bridges, then what?

If gaps and bridges are out, then what will replace them? How do a broad and multifaceted collection of scholars, entrepreneurs, activists, policy makers, and technologists encompass efforts towards responsible AI in ways that allow for a shared vocabulary, collaborative efforts, and coherent dialogues? Indeed, the gap-bridge metaphor may be incorrect, but it has served as a common object and objective for a diverse and loosely connected field, all working towards complex and elusive goals. In this final section, we begin to consider alternative metaphors that *request* and *encourage* integration, nonlinearity, and value-dynamism in the principle-practice relationship. These metaphoric alternatives are not determined or determining. Rather, they are placeholders, conversation starters, and indicative illustrations of how responsible AI might be reconceived. The purpose is not to introduce *the* new responsible AI metaphor, but instead, to invite collective efforts in this regard.

One alternative metaphor is a principle-practice hive. A hive is a living compound made by bees as a place to live, to shelter, to protect themselves from predators, to rear their brood and to store food (honey). A hive is also spurred by humans who create the conditions for the hive (as opposed to a nest which is naturally occurring). The hive shapes bees' individual and communal lives, which in turn reworks the hive which affects the bees, all of which is shaped by and influences the broader environment of trees, flowers, and food systems. In this metaphor, the hive is a manifestation of principle-practice relations. What can be seen in the creation of a hive is an integrated and non-linear relationship between principles and practices: the principles governing how bees work together shape the practices that form the hive and the hive shifts and changes over time, particularly in response to a changing environment. Further, different types of hives form based on their purpose (reflecting values dynamism). For example, in traditional beehives, honeycomb is created within the hive and honey extracted by crushing the honeycomb wax; more beeswax is found in these hives and less honey than in a modern hive. To maximize honey production, modern hives provide frames to encourage productive formation of honeycomb and honey.

Another alternative metaphor is a principle-practice jazz duet. In this metaphor, two individuals (principles and practices) come together to create music (responsible AI). Their relation to each other is non-linear given the spontaneous nature of jazz improvisation. Together, their instruments complement each other to form harmonious sound. In addition, although there are known methods for jazz improvisation – improvising the melody, the harmony or the motives in the arrangement – the values that guide jazz music are dynamic; there are no strict rules, but general social protocols that each member adheres to for the creation of new sounds, with adjustments when disharmony occurs.

6.2 Limitations and Future Work

We concluded by suggesting the displacement of the gap-bridge metaphor with metaphoric alternatives. This conclusion is really a starting point. A limitation of this piece is therefore its disproportionate attention towards deconstructing an ill-fit metaphor, without equal attention towards constructing a better one. This is a function of the gap-bridge metaphor's pervasiveness within the field, and thus the need for concerted and careful work to undo it.

Proceeding with deeper exploration will entail further empirical study as well as expansive application. Here we focused on the field of responsible AI, scoped empirically within the context of startup

Table 2: Gap-Bridge Fallacies and their in-Situ Correctives

Gap-Bridge Fallacies	In-Situ Correctives
Separation: Principles and practices are fundamentally distinct and thus require bridging	*Integration:* Principles and practices operate with and through each other
Linearity: Straight movement between principles and practices	*Nonlinearity:* Principles and practices move continuously along irregular trajectories
Stasis: Values are static and achievable	*Dynamism:* Principles and their practical implementation are fluid and ever changing, never to be solved, only re-solved through ongoing and evolving efforts

organizations operating in a single country (Australia). However, the relationship between principles and practices holds relevance beyond the context of our limited field sites, with potential bearing on larger organizations working with AI, a range of geographic locales, and efforts towards social responsibility in the design and implementation of all technologies, broadly conceived, including software design, policy design, built environments and so forth. Applying, testing, tweaking, and expanding the principle-practice relation that we set forth herein will not only redefine the problem space of 'responsibility', but also serve as a practical and intellectual advance in the ongoing project to manage societal challenges in an era of rapid technological change.

Acknowledgments

We thank the three organizations that form the cases explored in this work. This research has been supported by a Florence Violet McKenzie scholarship and an Australian Government Research Training Scholarship.

Research Ethics and Social Impact

Ethical Considerations Statement

Research in this paper was subject to a University Ethics Review process and each person involved was a part of free, prior and informed consent processes and agreed to their data being used for research. To avoid any unintentional harm to the fieldwork organizations, every effort has been made to pseudonymize references to the organizations and to the individuals quoted in this paper. Further, each organization was contacted to review the sections where their organization was featured to ensure that there was no breach in confidentiality and to give an opportunity for feedback. The feedback given through this process was minimal and did not substantially change any parts of the argument. All feedback received from the organizations was integrated into the manuscript.

Researcher Positionality Statement

The primary author of this manuscript is a current PhD Candidate following a decade-long career in consulting, much of which was spent as an intrapreneur. Her consulting background has highly influenced the very applied nature of this research. Her understanding of entrepreneurship from being an intrapreneur and social entrepreneur has allowed for a strong rapport to be built with the three organizations. Two of the three organizations used in this paper were recruited through the primary author doing a presentation or being a panelist on topics related to Responsible AI for industry audiences. As such, the primary author is strongly cross-sector and interdisciplinary in her approach. The secondary author of this manuscript is a sociologist working at the intersection of social psychology and technology studies and supervisor of the primary author's PhD. Both authors self-identify as white women, affecting access and acceptance within tech-sector field sites that are predominately (but not exclusively) populated by white men.

Adverse Impact Statement

Critiquing a metaphor that has arguably enabled cross-sector collaboration around responsible AI could have some adverse impacts. Whilst we argue that the gap-bridge metaphor may be incorrect and is thereby shaping our behaviors in unfavorable ways, we recognize that displacing it may have an adverse effect on the shared space that has been created through the metaphor; there is a risk that an alternative contributes to lost momentum and ultimately thwarts responsible AI efforts. To address this potential concern, we clearly position this work as a starting point for further careful consideration, recognizing that formulating the proposed correctives into a framework with practical and analytic utility is needed.

References

[1] Louise Amoore. 2020. *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others.* Duke University Press. https://doi.org/10.2307/j.ctv11g97wm
[2] Department of the Prime Minister and Cabinet Australian Government. [n. d.]. Closing the Gap. https://www.closingthegap.gov.au/
[3] Agathe Balayn and Seda Gürses. 2024. Misguided: AI regulation needs a shift in focus. *Internet Policy Review* (2024). https://policyreview.info/articles/news/misguided-ai-regulation-needs-shift/1796
[4] Vita Santa Barletta, Danilo Caivano, Domenico Gigante, and Azzurra Ragone. 2023. A Rapid Review of Responsible AI frameworks: How to guide the development of ethical AI. In *Proceedings of the 27th International Conference on Evaluation and Assessment in Software Engineering (EASE '23).* Association for Computing Machinery, New York, NY, USA, 358–367. https://doi.org/10.1145/3593434.3593478
[5] Richard Baskerville and Michael D. Myers. 2004. Special Issue on Action Research in Information Systems: Making IS Research Relevant to Practice: Foreword. *MIS Quarterly* 28, 3 (2004), 329–335. https://doi.org/10.2307/25148642 Publisher: Management Information Systems Research Center, University of Minnesota.
[6] Kathy Baxter. 2022. Ethical AI frameworks, tool kits, principles, and certifications - Oh my! https://blog.salesforceairesearch.com/frameworks-tool-kits-principles-and-oaths-oh-my/
[7] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency.* ACM, Virtual Event Canada, 610–623. https://doi.org/10.1145/

3442188.3445922
[8] Glen Berman. 2023. Machine Learning practices and infrastructures. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, Montréal QC Canada, 466–481. https://doi.org/10.1145/3600211.3604689
[9] Glen Berman, Nitesh Goyal, and Michael Madaio. 2024. A Scoping Study of Evaluation Practices for Responsible AI Tools: Steps Towards Effectiveness Evaluations. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–24. https://doi.org/10.1145/3613904.3642398
[10] Abeba Birhane. 2021. Algorithmic injustice: a relational ethics approach. *Patterns* 2, 2 (Feb. 2021), 100205. https://doi.org/10.1016/j.patter.2021.100205
[11] Max Black. 1962. *Models and Metaphors: Studies in Language and Philosophy*. Cornell University Press. https://www.jstor.org/stable/10.7591/j.ctvr6971f
[12] Victoria Clarke, Virginia Braun, and Nikki Hayfield. 2015. Thematic analysis. In *Qualitative psychology: a practical guide to research methods* (3rd edition ed.), Jonathan A. Smith (Ed.). SAGE, London, 222–248.
[13] Jenny L. Davis. 2020. *How artifacts afford: the power and politics of everyday things*. The MIT Press, Cambridge, Massachusetts.
[14] Jenny L Davis. 2023. 'Affordances' for Machine Learning. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 324–332. https://doi.org/10.1145/3593013.3594000
[15] Jenny L. Davis and James B. Chouinard. 2016. Theorizing Affordances: From Request to Refuse. *Bulletin of Science, Technology & Society* 36, 4 (Dec. 2016), 241–248. https://doi.org/10.1177/0270467617714944
[16] Paul DiMaggio. 1997. Culture and Cognition. *Annual Review of Sociology* 23, 1 (Aug. 1997), 263–287. https://doi.org/10.1146/annurev.soc.23.1.263
[17] Heike Felzmann, Eduard Fosch Villaronga, Christoph Lutz, and Aurelia Tamò-Larrieux. 2019. Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society* 6, 1 (Jan. 2019), 2053951719860542. https://doi.org/10.1177/2053951719860542 Publisher: SAGE Publications Ltd.
[18] Luciano Floridi. 2019. Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical. *Philosophy & Technology* 32, 2 (June 2019), 185–193. https://doi.org/10.1007/s13347-019-00354-x
[19] Sonja K. Foss. 2009. *Rhetorical criticism: exploration and practice* (4th ed ed.). Waveland Press, Long Grove, Ill. OCLC: ocn232976511.
[20] Batya Friedman and David G. Hendry. 2019. *Value Sensitive Design: Shaping Technology with Moral Imagination*. MIT Press.
[21] Batya Friedman, Peter H. Kahn, Alan Borning, and Alina Huldtgren. 2013. Value Sensitive Design and Information Systems. In *Early engagement and new technologies: Opening up the laboratory*, Neelke Doorn, Daan Schuurbiers, Ibo Van De Poel, and Michael E. Gorman (Eds.). Vol. 16. Springer Netherlands, Dordrecht, 55–95. https://doi.org/10.1007/978-94-007-7844-3_4 Series Title: Philosophy of Engineering and Technology.
[22] James Gibson. 2017. The Theory of Affordances. In *Perceiving, Acting and Knowing: Toward an Ecological Psychology*, Robert Shaw and John Bransford (Eds.). Psychology library editions. Perception, Vol. 27. Routledge, United Kingdom, 67–82. https://doi.org/10.4324/9781315467931
[23] Raymond Gozzi. 1994. The Information Superhighway as Metaphor. *ETC: A Review of General Semantics* 51, 3 (1994), 321–327. https://www.jstor.org/stable/42579686 Publisher: Institute of General Semantics.
[24] Graham Greenleaf. 2019. Global Data Privacy Laws 2019: 132 National Laws & Many Bills. *Privacy Law & Business International Report* 157 (Feb. 2019), 14–18. https://papers.ssrn.com/abstract=3381593
[25] Armand Hatchuel and Hughues Molet. 1986. Rational modelling in understanding an aiding human decision-making: about two case studies. *European Journal of Operational Research* 24 (1986), 178–186.
[26] Christian Herzog and Sabrina Blank. 2024. A systemic perspective on bridging the principles-to-practice gap in creating ethical artificial intelligence solutions – a critique of dominant narratives and proposal for a collaborative way forward. *Journal of Responsible Innovation* 11, 1 (Dec. 2024), 2431350. https://doi.org/10.1080/23299460.2024.2431350 Publisher: Routledge _eprint: https://doi.org/10.1080/23299460.2024.2431350.
[27] Javier Camacho Ibáñez and Mónica Villas Olmeda. 2021. Operationalising AI ethics: how are companies bridging the gap between practice and principles? An exploratory study. *AI & SOCIETY* (Aug. 2021). https://doi.org/10.1007/s00146-021-01267-0
[28] Barbara Kiviat. 2023. The Moral Affordances of Construing People as Cases: How Algorithms and the Data They Depend on Obscure Narrative and Noncomparative Justice. *Sociological Theory* 41, 3 (Sept. 2023), 175–200. https://doi.org/10.1177/07352751231186797 Publisher: SAGE Publications Inc.
[29] Zoltán Kövecses and Réka Benczes. 2010. *Metaphor: a practical introduction* (2nd ed ed.). Oxford University Press, New York. OCLC: ocn302414610.
[30] George Lakoff and Mark Johnson. 2003. *Metaphors we live by*. University of Chicago Press, Chicago.
[31] Susan Leigh Star. 2010. This is Not a Boundary Object: Reflections on the Origin of a Concept. *Science, Technology, & Human Values* 35, 5 (Sept. 2010), 601–617. https://doi.org/10.1177/0162243910377624
[32] Kimberlee Leonard. 2022. *Conducting A Gap Analysis: A Four-Step Template*. Technical Report. Forbes. https://www.forbes.com/advisor/business/gap-analysis-template/
[33] Qinghua Lu, Liming Zhu, Xiwei Xu, Jon Whittle, and Zhenchang Xing. 2022. Towards a roadmap on software engineering for responsible AI. In *Proceedings of the 1st International Conference on AI Engineering: Software Engineering for AI*. ACM, Pittsburgh Pennsylvania, 101–112. https://doi.org/10.1145/3522664.3528607
[34] Samuel Makin. 2021. The research-practice gap as a pragmatic knowledge boundary. *Information and Organization* 31, 2 (June 2021), 100334. https://doi.org/10.1016/j.infoandorg.2020.100334
[35] McKinsey. [n. d.]. The skillful corporation. http://ceros.mckinsey.com/the-skillful-corporation-desktop
[36] Brent Mittelstadt. 2019. Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence* 1, 11 (2019), 501–507. https://doi.org/10.1038/s42256-019-0114-4
[37] Jessica Morley, Anat Elhalal, Francesca Garcia, Libby Kinsey, Jakob Mökander, and Luciano Floridi. 2021. Ethics as a service: a pragmatic operationalisation of AI ethics. *Minds and Machines* 31, 2 (2021), 239–256. Publisher: Springer.
[38] Jessica Morley, Luciano Floridi, Libby Kinsey, and Anat Elhalal. 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 (Aug. 2020), 2141–2168. https://doi.org/10.1007/s11948-019-00165-5
[39] Donald A. Norman. 1988. *The psychology of everyday things*. Basic Books, New York.
[40] University of Bath. 2021. Entrepreneurial business strategy: how do you spot a gap in the market? https://online.bath.ac.uk/articles/how-do-you-spot-a-gap-in-the-market
[41] Australian Bureau of Statistics. 2023. Gender pay gap guide. https://www.abs.gov.au/statistics/understanding-statistics/guide-labour-statistics/gender-pay-gap-guide
[42] Ahmed El Ouadrhiri and Ahmed Abdelhadi. 2022. Differential privacy for deep and federated learning: A survey. *IEEE access : practical innovations, open solutions* 10 (2022), 22359–22380. https://doi.org/10.1109/ACCESS.2022.3151670
[43] European Parliament. 2023. Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html
[44] Alison B Powell, Funda Ustek-Spilda, Sebastián Lehuedé, and Irina Shklovski. 2022. Addressing ethical gaps in 'Technology for Good': Foregrounding care and capabilities. *Big Data & Society* 9, 2 (July 2022), 205395172211137. https://doi.org/10.1177/20539517221113774
[45] Inioluwa Deborah Raji, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. 2020. Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing. https://arxiv.org/abs/2001.00973v1
[46] Robert N. Rapoport. 1970. Three Dilemmas in Action Research: With Special Reference to the Tavistock Experience. *Human Relations* 23, 6 (Dec. 1970), 499–513. https://doi.org/10.1177/001872677002300601 Publisher: SAGE Publications Ltd.
[47] Alistair Reid, Simon O'Callaghan, and Yaya Lu. 2023. *Implementing Australia's AI Ethics Principles: A selection of Responsible AI practices and resources*. Technical Report. Gradient Institute and CSIRO. https://www.csiro.au/en/work-with-us/industries/technology/National-AI-Centre/Implementing-Australias-AI-Ethics-Principles-report
[48] Jörgen Sandberg and Mats Alvesson. 2011. Ways of constructing research questions: gap-spotting or problematization? *Organization* 18, 1 (Jan. 2011), 23–44. https://doi.org/10.1177/1350508410372151 Publisher: SAGE Publications Ltd.
[49] Daniel Schiff, Bogdana Rakova, Aladdin Ayesh, Anat Fanti, and Michael Lennon. 2021. Explaining the Principles to Practices Gap in AI. *IEEE Technology and Society Magazine* 40, 2 (June 2021), 81–94. https://doi.org/10.1109/MTS.2021.3056286 Conference Name: IEEE Technology and Society Magazine.
[50] Ben Shneiderman. 2020. Bridging the Gap Between Ethics and Practice: Guidelines for Reliable, Safe, and Trustworthy Human-centered AI Systems. *ACM Transactions on Interactive Intelligent Systems* 10, 4 (Dec. 2020), 1–31. https://doi.org/10.1145/3419764
[51] Mona Sloane and Janina Zakrzewski. 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*. ACM, Seoul Republic of Korea, 935–947. https://doi.org/10.1145/3531146.3533156
[52] Luke Stark and Anna Lauren Hoffmann. 2019. Data Is the New What? Popular Metaphors & Professional Ethics in Emerging Data Culture. *Journal of Cultural Analytics* 4, 1 (May 2019). https://doi.org/10.22148/16.036
[53] Gerald I. Susman and Roger D. Evered. 1978. An Assessment of the Scientific Merits of Action Research. *Administrative Science Quarterly* 23, 4 (Dec. 1978), 582. https://doi.org/10.2307/2392581

[54] Ann Swidler. 1986. Culture in Action: Symbols and Strategies. *American Sociological Review* 51, 2 (April 1986), 273. https://doi.org/10.2307/2095521
[55] Ann Swidler. 2005. *Talk of love: how culture matters* (paperback ed., [nachdr.] ed.). Univ. of Chicago Press, Chicago.
[56] Sy Taffel. 2023. Data and oil: Metaphor, materiality and metabolic rifts. *New Media & Society* 25, 5 (May 2023), 980–998. https://doi.org/10.1177/14614448211017887 Publisher: SAGE Publications.
[57] Stefan Timmermans and Iddo Tavory. 2022. *Data Analysis in Qualitative Research: Theorizing with Abductive Analysis.* University of Chicago Press, Chicago, IL. https://press.uchicago.edu/ucp/books/book/chicago/D/bo133273407.html