Employee Perceptions of Effective AI Principle Adoption

Stephanie Kelley

Smith School of Business, Queen's University, Kingston, ON, Canada, K7L 3N6,
stephanie.kelley@queensu.ca

Draft: March 12th, 2021

Abstract

This study examines employee perceptions on the adoption of artificial intelligence (AI) principles in their organizations. 49 interviews were conducted with employees of 24 organizations across 11 countries. Participants worked directly with AI across a range of positions, from junior data scientist to Chief Analytics Officer. The study found that there are eleven components that could impact the effective adoption of AI principles in organizations: communication, management support, training, an ethics office(r), a reporting mechanism, enforcement, measurement, accompanying technical processes, a sufficient technical infrastructure, organizational structure, and an interdisciplinary approach. The components are discussed in the context of business code adoption theory. The findings offer a first step in understanding potential methods for effective AI principle adoption in organizations.

Key words: AI principles, AI ethics, artificial intelligence, adoption

Introduction

Reports of organizations proliferating bias and discrimination, jeopardizing customer privacy, generating decisions from customer data without informed consent, and making other ethical mistakes in their use of artificial intelligence (AI) are increasing (Whittaker et al. 2018). For example, the Apple Card, a joint venture between Apple and Goldman Sachs, was recently accused of discriminating against women in their use of an AI-based credit approval model[1]. Not long before that, IBM, Microsoft, Amazon, and Megvii were accused of proliferating racial and gender discrimination in their AI-based facial recognition technologies (Buolamwini and Gebru 2018). In response, several of the firms altered their technologies to reduce the bias (Raji and Buolamwini 2019), and others called for moratoriums[2]. While these are all examples of AI ethics issues that were noticed and investigated, the use of AI in organizations is only projected to grow and with it, the number of ethical issues are likely to increase (Kaplan and Haenlein 2019a), many of which could go noticed or investigated. These, and other reports

suggest that adhering to the existing laws may not be enough to prevent unethical AI outcomes, a concern shared by legal (Barocas and Selbst 2016) and management scholars (Martin 2019).

Increasingly, organizations are turning to self-regulatory initiatives (Schwartz 2001) such as principles (Mittelstadt 2019), partnerships (Cath et al. 2018), and oversight boards (Bietti 2020) for AI ethics issues. While these initiatives may indicate the private sector is attempting to institutionalize AI ethics (Cath et al. 2018), or accept responsibility for their AI outcomes (Martin et al. 2019), organizations may be putting them in place for less altruistic reasons including ethics washing, creating an ethical façade (Bietti 2020), evading formal regulation (Rességuier and Rodrigues 2020), or avoiding the need to operationalize their initiatives (Mittelstadt 2019). Despite the potentially complex motivations for the creation of self-regulatory AI ethics initiatives, preliminary regulatory recommendations[3,4] support their use. Self-regulatory initiatives by themselves may, however, be deficient in affecting the ethical use of AI (Rességuier and Rodrigues 2020) and, as such, additional proposals have been put forth to develop formal regulations (e.g., The Algorithmic Accountability Act in the United States) and industry-led standards (e.g., IEEE Standards Association P7000) as complements. At the same time, self-regulatory initiatives continue to be developed, with many organizations generating AI principles, the most common self-regulatory initiative, and the focus of this study. Several examples of AI principles have been gathered by the AlgorithmWatch, available here: https://inventory.algorithmwatch.org/.

Before proceeding to the research question, a definition of 'AI' and 'principles' is proposed. It is important to define a technology as its study is shaped by its definition (Martin and Freeman 2004). 'AI' refers to 'artificial intelligence,' "a system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation" (Kaplan and Haenlein 2019a). This study focuses on AI, as opposed to automation (or autonomous agents) which are "computational entities that makes decisions and executive actions in response to environmental conditions, without direct control by humans" (Wellman and Rajan 2017). 'Principles' describes the contents of documents which primarily contain stakeholder principles (Kaptein 2004), including transparency, fairness, and accountability (Fjeld et al. 2020; Jobin et al. 2019). The word 'principles' is used in the title of most of these document (e.g. Google[5], Microsoft[6], Telefonica[7]), and in the literature (e.g., Fjeld et al. 2020; Floridi 2019).

Artificial intelligence principles (AIPs) are defined herein as *a formal document developed* (Kaptein and Schwartz 2008) *or selected by an organization that states normative declarations* (Fjeld et

al. 2020; Hagendorff 2020) *about how artificial intelligence ought to be used by its managers and employees.*

Although less common, AI principles are also referred to as guidelines (Jobin et al. 2019), tenets (Mittelstadt 2019), codes of ethics (McNamara et al. 2018), declarations, ground rules, frameworks, strategies, and statements (Fjeld et al. 2020). In this study AIPs are treated as a mutation of business codes (BCs). They share a similar overall function, and some overlap in content, but differ in their audience, use of compliance or values-based language, and development. BCs and AIPs share a similar function: BCs "provide a set of prescriptions…on multiple issues…" (Kaptein and Schwartz 2008), while AIPs similarly state normative declarations about how AI ought to be used. Content overlap also exists between the two types of documents: some principles present in AIPs are discussed in BCs (transparency [55% of BCs], fairness [45% of BCs], and accountability [18% of BCs]) (Kaptein 2004), but many of the topics (e.g., data privacy, cyber safety and security, transparency, and explainability (Fjeld et al. 2020)) are unique. AIPs also differ in their audience compared to BCs: they target employees that interact with AI, as opposed to all employees. AIPs also rely heavily on "normative declarations," (Fjeld et al. 2020; Hagendorff 2020), which use values-based language (Spiekermann 2016) as opposed to the mix of compliance-based and values-based language of BCs (Weaver et al. 1999a). Their development also differs: an AIP can be developed or selected by an organization, whereas a BC must be an internal document "developed by and for a company" (Kaptein and Schwartz 2008). In practice, organizations have been found to generate their own AIPs (e.g., HSBC[8]) or declare their adherence to principles developed by another party, such as an industry consortium (e.g., Partnership on AI[9], The Toronto Declaration[10], The Montreal Declaration for Responsible AI Development[11]), an intergovernmental organization (e.g., The Organisation for Economic Co-operation and Development[12]) or a governing body (e.g., Monetary Authority of Singapore[13]).

Although the study of AIPs is nascent, the study of AI ethics in organizations has been of interest since the technology was developed, and was first highlighted as a management concern by Khalil (1993), who argued that managers must remain legally and ethically responsible when using AI in decision making due to the technology's possible incorporation of intentional or accidental bias; and lack of human intelligence, emotions and values. Since that time, only a handful of scholars have studied AI ethics in organizations (e.g., Huang and Rust 2018; Kaplan and Haenlein 2019b; Martin 2019; Martin et al. 2019; Morley et al. 2020), but Khalil's (1993) notion, that AI ethics is a management concern, remains valid.

Empirical studies on AIPs are limited, as companies only recently started adopting them, with the first AIP thought to have been developed in 2016 by the Partnership on AI (Fjeld et al. 2020; Jobin et al. 2019). The nascent adoption of AIPs has guided the research to date with the majority being "*content* oriented" studies (what is in the actual principles), as opposed to "*transformation* oriented*"* studies (how the principles are adopted or not in an organization), or "*outcome* oriented" studies (what effects the principles have) (Helin and Sandström 2007), which cannot occur before AIPs are developed and adopted.

Several *content* oriented studies have occurred in recent years analyzing the ever-growing list of AI principles. Jobin et al (2019) review 84 ethical principles and guidelines for AI and conclude that there exists a degree of convergence around five ethics principles: transparency, justice and fairness, non-maleficence, responsibility, and privacy (Jobin et al. 2019). Fjeld et al. (2020) review 36 AI ethics principles and find there are eight key themes: privacy, accountability, safety and security, transparency and explainability, fairness and non-discrimination, human control of technology, professional responsibility, and promotion of human values (Fjeld et al. 2020). Schiff et al. (2020) review 88 AIPs and discuss the overarching challenges of the existing principles, motivations behind their creation, and their potential governance impact. Although primarily *content* focused, Schiff et al. (2020) propose five factors that could impact AIP adoption, extending the work into the *transformation* oriented realm: engagement with law and governance, specificity of the document, document reach, enforceability and monitoring, and iteration and follow-up. Schiff et al. (2021) review a more-recent AIP list of 112 documents, and compare differences across public, private, and NGO sectors, and conclude that organization type impacts AIP content, with the private sector focused on client and customer-related issues, the public sector focused on economic growth and unemployment issues, and NGOs focused on more nuisanced issues.

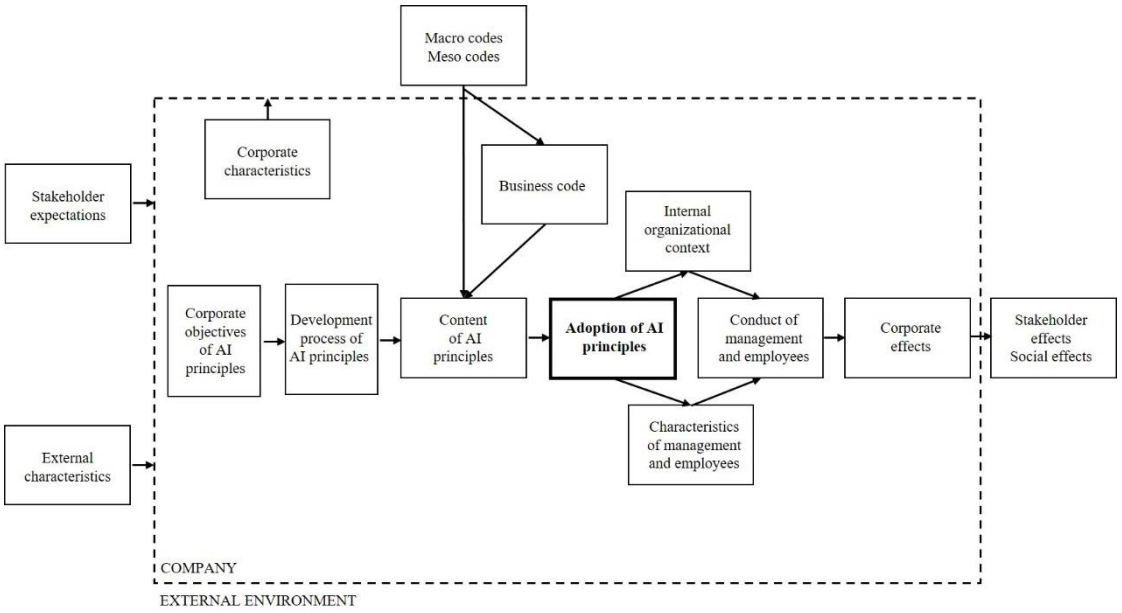
To date there has been a single *output* oriented study on AIPs by McNamara et al. (2018), which uses ethical vignettes to measure the response of software students and developers to the Association of Computer Machinery's Code of Ethics. They conclude that the AIP has no effect on ethical decision making when compared to a control group that did not read the code (McNamara et al. 2018). The finding of this *output* oriented study contrasts the large body of literature on the efficacy of BCs, where the majority of studies find BCs to be effective in changing behaviour (Babri et al. 2019). It has been suggested that the use of ethical vignettes may not provide a strong enough manipulation in ethics behavioural research (Kaptein and Schwartz 2008), which could explain the inefficacy found by

McNamara et al. (2018). Additional *output* oriented studies focused on answering the question "are AI principles effective?" have not yet occurred given the limited development and adoption of AIPs.

Similarly, there are only a handful of *transformation* oriented studies on AIPs, very few of which are empirical studies given the limited adoption of AIPs to date. Mittelstadt (2019) suggests there is an absence of a proven method to translate AIPs into practice. Vakkuri et al. (2019) perform a case study of five organizations, and find there is a gap between AIPs and their adoption, which they argue is primarily due to a lack of tools for practitioners. Raji et al. (2020) propose an algorithmic auditing framework aimed to help organizations assess the fit of their AI decisions given their AIPs. Madaio et al. (2020) suggest an AI fairness checklist as one method to operationalize a specific subset of AIPs related to bias, discrimination, and fairness. Schiff et al. (2020) argue there is a lack of clarity for how organizations should implement AIPs in practice; and suggest a single impact assessment framework to aid in adoption. While these studies make several suggestions as to why AIP adoption is lacking, and propose recommendations, they do not empirically investigate the effective adoption of AIPs. This paper addresses this gap by asking: "according to the perceptions of employees who work with AI, what components might relate to the effective adoption of AIPs?" The research question is timely given the recent growth in AIP adoption, prior to which empirical studies would not have been possible.

As AIPs are proposed as a mutation of BCs in this study, Kaptein and Schwartz's (2008) integrated research model for the effectiveness of BCs is proposed as a foundation for the study of AIPs. An adapted version of the model is proposed in Figure 1. The integrated research model acts as a starting place to determine which components from the BC adoption literature might impact effective AIP adoption. The model also helps to conceptualize which components are distinct from effective AIP adoption (e.g., AIP content), and should therefore be studied separately. Lastly, the hope is that by positioning this study on as part of a larger theoretical framework, it could help lay the foundation for a consistent body of research on the broader study of AIPs and their effectiveness in the future.

Figure 1. An integrated research model for the effectiveness of AI principles, adapted from Kaptein and Schwartz (2008)



Location Figure 1

Several components have been empirically found to impact effective BC adoption; they act as a starting point for this study of effective AIP adoption, and are summarized below in Table 1. At the outset of this study, it was unclear what the impact of these components could be on effective AIP adoption.

Table 1. Summary of components that impact effective business code adoption

Adoption components	References
(1) Communication	
Reach	(Weaver et al. 1999b)
Distribution channel	(Adam and Rachman-Moore 2004; Schwartz 2004; Stevens 2008)
Sign-off process	(Schwartz 2004; Singh et al. 2011; Weaver et al. 1999b)
Reinforcement	(Kaptein 2011; Schwartz 2004; Smith-Crowe et al. 2015; Weaver et al. 1999b)
Communication quality	(Kaptein 2011; Schwartz 2004)
External communication	(Singh 2011)
(2) Management support	
Local management support	(Kaptein 2011; Petersen and Krings 2009)
Senior management support	(Kaptein 2011; Schwartz 2004; Singh et al. 2011; Trevino et al. 1999)
(3) Training	
Existence of training	(Adam and Rachman-Moore 2004; Schwartz 2004; Singh 2011; Weaver et al. 1999b)
Preferred trainers	(Schwartz 2004; Trevino et al. 1999)
(4) Ethics Office(r)	(Kaptein 2015; Singh 2011; Weaver et al. 1999b)
(5) Reporting Mechanism	
Existence of a reporting mechanism	(Kaptein 2015; Schwartz 2004; Singh 2011; Trevino et al. 1999; Weaver et al. 1999b)
Existence of a standardized procedures	(Weaver et al. 1999b)
(6) Enforcement	
Audits	(Kaptein 2015; Singh et al. 2011)

Location Table 1

Penalties	(Adam and Rachman-Moore 2004; Singh 2011; Singh et al. 2011; Trevino et al. 1999)
Communicating violations	(Schwartz 2004)
Incentive policies	(Kaptein 2015; Schwartz 2004; Trevino et al. 1999)
(7) Measurement	(Schwartz 2004; Weaver et al. 1999b)

It is important to note that the study does not attempt to directly measure adoption, instead it uses the perceptions of employees working with AI to assess the potential components that could impact the adoption of AIPs. This presumes that AIP adoption is intrinsically tied to employee comprehension and compliance to the normative declarations included in the AIP, and therefore employee perceptions on adoption are relevant. BC studies also use employee perceptions in a similar manner (Schwartz 2004).

After discussing the methodology in the next section, the paper presents evidence, based on qualitative interviews, for the existence of components (e.g., Table 1) that could impact the effective adoption of AIPs. The paper concludes with a discussion of the practical implications, limitations, and future avenues of study.

Methodology

To explore the research question, 49 in-depth, semi-structured interviews were conducted with individuals employed in financial services organizations who work with AI. Financial services was chosen due to its wide-spread development of AIPs to date (DeutscheBank et al. 2019), a prerequisite to the study of their adoption. AI is used extensively across financial services organizations (e.g., fraud detection, credit lending, customer service chat bots, talent acquisition) (Financial Stability Board 2017), with current adoption rates and projected growth rates second only to the technology industry (Bughin et al. 2017). The financial services industry is highly regulated, potentially making effective AIP adoption more likely than other less regulated industries (e.g., technology) and, as such, serves as a strong case to study employee perceptions on effective AIP adoption.

Organizations were first pre-screened for the existence of an AIP via an online search. The researcher then contacted AI ethics leaders at these firms, as it was assumed they would have the best knowledge of the AIPs and their adoption. The researcher then asked these leaders to suggest other employees in their organizations, or other organizations who had AI principles, to be interviewed. The snow-ball sampling technique was used because the desired interviewees were in a hard-to-reach group. Moreover, it was difficult to determine from outside the organization which employees were in the small target audience of the company's AIP (only employees working with AI), and knew about its adoption (Miles and Huberman 1994). Of those contacted, 35% agreed to be interviewed, of which ~69% (34/49)

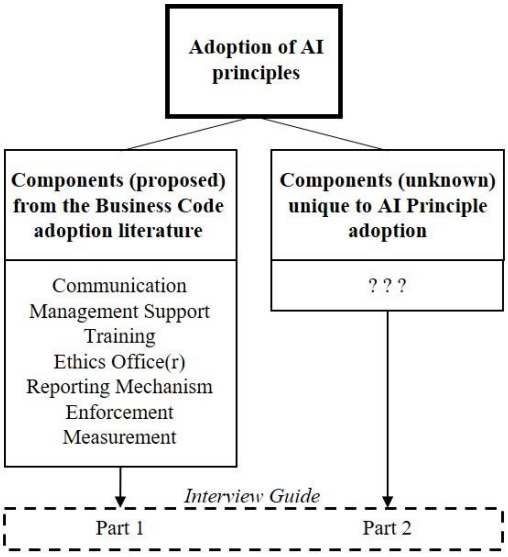
identified as men, and ~31% (15/49) as women. ~39% (19/49) were executives or vice presidents, ~55% (27/49) were managers, and ~6% (3/49) were non-managers. Interviews spanned 24 organizations, with an average of two employees interviewed at each organization. The 49 participants were in the United States (3), Singapore (9), Canada (23), the United Kingdom (4), Australia (1), Sweden (2), China (1), Thailand (1), Mexico (2), Brazil (1), and South Africa (2).

21 of the participants worked directly with technical teams on AI; 12 of whom helped develop and implement the AIPs. These "technical" participants were primarily managers and non-managers, with only 4 self-identified executives, whom at the time of the interviews lead the AI group at their organization. The other 28 participants were not directly involved in the technical aspects of AI; 16 of them were however directly involved in developing and driving adoption for AIPs, while the other 12 were tasked AIP adoption but not the original development. These "non-technical" participants were split between executives (15), and non-executives (11 managers and 2 non-managers who were both lawyers). Of the 49 participants, 28 were directly involved in AIP development and adoption, 15 of whom were executives, 11 managers, and 2 non-managers (lawyers). The other 21 participants were involved in AIP adoption but not AIP development. Participants had worked at their company from 1 to 17 years, with an average of 4.5 years at their organization.

The interview-based methodology was selected because it has been suggested as a better fit than quantitative methods to understand "the relationship between codes and behaviour" (Schwartz 2001) and "how codes work" (Babri et al. 2019). Furthermore, it allows for the investigation of AIPs in "an actual corporate setting involving actual users" (Schwartz 2001), which extends beyond the laboratory study of AIPs to date (e.g., Fjeld et al. 2020; McNamara et al. 2018).

The semi-structured interview was split into two parts: first, semi-structured questions were asked based on the components of effective BC adoption from the literature (per Table 1). Second, two open-ended questions were asked to explore other components of effective AIP adoption, not found in effective BC adoption theory. The interview guide was reviewed by two subject matter experts, and two qualitative research experts prior to the study's commencement for clarity and validity. A summary of the conceptual framework used to develop the interview guide is provided in Figure 2, and the full interview guide is provided in Appendix A.

Figure 2. A summary of the conceptual development of the interview guide



Location Figure 2

Interviews took place between December 2019 and July 2020. 16 of the interviews were conducted in person, and the remaining 33 were conducted over videoconference technology or the phone due to COVID-19 travel restrictions. Interviews were conducted in English, the primary language of business for all participants and were on average 40 minutes long. 41 of the interviews were recorded, with the remaining eight, at the request of the participants, not recorded, although extensive notes were taken during and after these interviews. Recordings were first transcribed using a natural language processing (NLP) transcription technology (Otter.ai) and reviewed and edited for accuracy.

The transcriptions were then coded using qualitative research software (NVIVO) following a general inductive approach (Thomas 2006) based on a positivist epistemology, and assumed the interviewees were direct holders of accurate information (Johnson and Duberley 2000). Inductive coding was chosen to allow the researcher to uncover components impacting effective AIP adoption without being biased by the effective BC adoption theory (Thomas 2006). During the initial round of open coding, all text was reviewed and coded using interviewee language. Codes were then compared with the existing components from BC effective adoption theory (Table 1). Codes that fit within existing components were grouped and renamed accordingly, while codes unique to AIP adoption were grouped and named using interviewee language. A detailed discussion of the coding approach, along with the final codebook is provided in Appendix B.

Findings and discussion

The findings and discussion are presented per the structure of the interview guide: components impacting effective AIP adoption rooted in business code adoption theory are discussed first, followed by components unique to AIP adoption.

Components impacting AI principle adoption effectiveness from the business code literature

Communication

Communication, the act of making employees aware of a BC is an important first step in its adoption, as simply having an AIP is not enough to ensure to drive implementation (Stevens 2008). Participants were asked about several aspects of communication known to impact BC: (a) *reach*; (b) *distribution channel* (c) *sign-off process*; (d) *reinforcement*; (e) the *communication quality*, and (f) *external communication*.

(a) *Reach.* The reach of BCs, also referred to as their distribution, or the number of employees who receive a copy of the business code (Weaver et al. 1999b), has been found to impact BC adoption (Weaver et al. 1999b), and has also been proposed to impact AIP adoption (Schiff, Biddle, et al. 2020). When asked about AIP reach, there appeared to be two opposing views. About two-thirds of participants suggested that maximum reach among AI employees would be beneficial:

> …very importantly, the actual analytics practitioners in the bank, we've all been involved closely to go through the principles… (Man, Executive, Singapore)

The remaining third of participants questioned whether it was necessary for there to be widespread distribution of the principles to AI employees, and suggested focusing reach on managers and executives.

> … I want to leave my data scientists free to find any possible pairings that they might find interesting or relevant and then we decide whether this is something that is safe to put out in the open. (Man, Manager, Singapore)

A handful of respondents did however suggest with some cynicism that distribution of the AIPs by itself may not be enough to create effective adoption, leaving the importance of it unclear.

> … I think they read it because they wanted to look good – it wasn't genuine – there's a lot of superficiality. (Man, Manager, UK)

> …I think everyone now even at a junior business level is aware of those principles … even though they don't always know what those things mean, they know it's important… (Man, Executive, Singapore)

(b) *Distribution channel*. The distribution channel, or the channel by which an employee receives a BC (Stevens 2008) has been found to impact adoption. The use of informal communication channels, such as managers openly discussing the BC with their employees, or through social norms (Adam and Rachman-Moore 2004), as opposed to formal training, directives, or classes have been found to more beneficial for BC adoption. Participatory design, involving employees in the creation of a BC, thereby communicating its existence prior to the final draft, has also been found to impact effective adoption (Schwartz 2004).

When it comes to AIPs, there appeared to be four channels that participants first found out about their AIP: three informal channels whereby employees were (1) asked to participate in the AIP design, (2) were provided the AIP directly by a manager, or (3) sourced the AIP from an AI ethics office(r); and one formal channel, whereby (4) employees found out about the AIP through internal marketing.

Employees stated no clear preference between the informal and formal distribution channels but suggested a lack of participatory design could hinder adoption, as it does in BCs (Schwartz 2004).

> … it isn't conducive to a cultural mindset shift…there hasn't been a concerted effort to co-create with the practitioners or the developers that need to change their thinking. (Woman, Manager, Canada)

(c) *Sign-off process.* Mandatory sign-off practices, whereby an employee has to acknowledge receipt of a business code (Weaver et al. 1999b) are common practice in several countries (Singh et al. 2011; Weaver et al. 1999b), but are known to receive pushback from employees (Schwartz 2004) which could negate some of the positive impact on adoption.

When asked about AIP sign-off, most participants (30/49) simply noted that it was "too early" in the existence of the AIPs, or that is just was "not a priority" at this point, suggesting it could be potentially important in the future.

(d) *Reinforcement.* BC adoption is known to be impacted by communication reinforcement (Kaptein 2011; Schwartz 2004; Weaver et al. 1999b), the number of communications on the BC received by an employee (Weaver et al. 1999b). Multiple channels are used for communication reinforcement including policies, memos (including email), poster, newsletters, videos, the company intranet (Kaptein 2011), and even meetings pre-employment (Schwartz 2004). A mix of channels has been thought to improve adoption (Smith-Crowe et al. 2015). When it comes to AIPs, participants from just over half of the organizations discussed the importance of AIP communication reinforcement:

> …not just communicate once and say 'this, is out there,' but really have our practitioners and people that support our practitioners understand what it means. (Woman, Manager, Canada)

> …our bosses will send out an email to us to remind us…this is always a reminder and always communicated to us on a very periodic basis. (Woman, Non-Manager, Singapore)

As with BCs, participants noted the use of multiple channels for AIP reinforcement including company intranets, emails, conferences, lunch and learns, executive speeches, annual reports, academic conferences, and public relations. Participants did not explicitly discuss sharing the AIP prior to employment; however, several participants noted they discuss general AI ethics in the hiring process:

> …we asked all the potential candidates to complete a case study – build a model…tell us do you believe the model discriminates against age, or gender? We were trying to get people thinking about that…one way to contain that issue is to discuss it in the hiring process… (Man, Executive, South Africa)

> …it has been discussed right from the beginning…so, it has always been very clear from the start that we shouldn't do anything unethical with out customer's data, anything outside of the framework. (Woman, Non-Manager, Singapore)

(e) *Communication quality.* Communication quality, or the readability, relevance (Schwartz 2004), accessibility, understandability, and usefulness of the communication of a BC has also been found to impact its adoption (Kaptein 2011).

With respect to AIPs, participants suggested that having clear definitions of terms included in the principles (e.g., AI, fairness, explainability) was a way to "quick jump start the awareness campaign," a prerequisite of adoption, while unclear definitions could "hinder" the adoption of AIPs.

Cultural and contextual relevance of the AIP messaging was also suggested as important for adoption, particularly for global banks:

> "When I went to operationalize [the AI principles] I met with representatives…in every country…we spoke at length about how to design a message that is going to be contextually and culturally relevant…that is easily understood in Indonesia and by someone in London…" (Man, Executive, Singapore)

Respondents also suggested that communication understandability could be improved by crafting a clear message with internal marketing. For example, several organizations used either mnemonics or marketing slogans in their AIP communication campaigns. Whether through clear definitions of terms, cultural and contextual relevance, or clear messaging developed with internal marketing, communication quality was noted as an important factor for adoption by 55% (27/49) of the respondents.

(f) *External communication*. Singh (2011) found that communicating a BC externally, for example, sharing the code with customers, suppliers, or displaying it externally on a website, impacts adoption.

When it comes to sharing AIPs, there appears to be four degrees of sharing that participants noted. About 25% of the organizations in the study have not shared their principles externally.

> ….nothing published, which is interesting but deliberate because [the bank] is a privately held company, and overall very, very secretive. (Man, Executive, Australia)

~25% of organizations have shared their AIP through one or more external channels: an annual report, the organization's website, or the media. Another ~25% have created "a white paper," or a summary of their principles to be shared, whilst keeping the AIPs confidential, as a sort of "house view".

> …we've done a few things in the comms, we did a formal publication of the principles… (Man, Executive, China)

> …through the sustainability report we wrote down [the AIPs]…there was also some kind of op ed, [the CTO] and [the CRO] kind of contributed to the public forum, [the newspapers]…more around engagement and awareness. (Man, Manager, Canada)

The fourth group of organizations are those that adopted publicly available AIPs and are therefore exempt from "external communication" given the AIP is already available to the public. For example, the Monetary Authority of Singapore has a set of AI principles [vi] (not formal regulation), which several banks in Singapore have adopted, including five banks that participants in this study are employed at.

Although just over half of organizations in the study share their AIPs externally, a handful of participants at these organizations suggested that sharing them externally was nothing more than 'whitewashing', and that it may not be effective, putting into question its importance in adoption.

> …you can find it, it's open source…. there is a lot of emphases on it – I don't know if it's helpful…it's brought up a lot as an excuse not to do something rather than for legitimate reasons… (Man, Manager, United Kingdom)

> …we're signalling that we're a bank that's very conscious of this, right – we pushed out the training, we advertise it, but it's a signal and it's not effective. (Man, Manager, Canada)

Management Support

Management support, or employees knowing that one or both of (a) *local management support* (Kaptein 2011; Petersen and Krings 2009) or (b) *senior management support* (Kaptein 2011; Schwartz 2004; Singh et al. 2011; Trevino et al. 1999) support the company's BC, has been found to impact BC adoption. Support can include actions such as modelling appropriate ethical behaviour (Kaptein 2011; Schwartz 2004), talking about the code (Schwartz 2004; Singh et al. 2011), knowing and/or understanding the code (Schwartz 2004), or generally taking the code seriously (Trevino et al. 1999).

(a) *Local management support.* BC effectiveness is positively impacted when a local manager, the direct supervisor of an employee (Kaptein 2011), who is not an executive, supports the BC. When it comes to AIPs, just over half of participants (~60%) mentioned the importance of local management support. Local managers show support primarily by speaking about the AIP in meetings, understanding the AIP, and generally being advocates of it.

> …she's super interested in [it]…she's on board, she understands. – that's important. (Man, Manager, Canada)
>
> …the leadership in my region were more progressive on the issue and they became advocates…things have changed a lot in two years… (Woman, Executive, United Kingdom)

(b) *Senior management support.* Seeing senior management, someone more senior than an employee's direct manager, support a BC, impacts adoption (Kaptein 2011). Senior management support is seen as an important factor for AIP adoption; however, given the technical nature of AI, senior managers often do not work directly with it, so managers and non-managers expect them to know about the principles, and to talk about them, but not to model specific behaviour from the AIP.

> …what's really helped our bank progress on this is the fact that we have endorsement from executives, very senior executives…that really helps get peoples attention and time. (Woman, Manager, Canada)
>
> One thing I've been really pleased about is the level of interest, commitment at an executive management level… (Man, Executive, China)
>
> … so it's not just our direct managers, it's the CEO. He immediately picks up on how to incorporate ethics which is a good sign of how seriously we think about it. (Man, Executive, South Africa)

Conversely, a lack of senior management support could also be detrimental to AIP adoption:

> …we don't have somebody who frequently talks about this from the executive level, management committee level, no – I wish that would happen. (Man, Executive, United States)

Training

Offering BC training, whereby employees attend a training session or class to educate them on the business code (Adam and Rachman-Moore 2004), has been found to impact its adoption effectiveness (Adam and Rachman-Moore 2004; Schwartz 2004; Weaver et al. 1999b), especially when it is offered to all employees (Singh 2011). Participants were asked their thoughts on the importance of the (a) *existence of training* for the AIP, as well as their (b) *preferred trainers*.

(a) *Existence of training.* Simply the existence of training, regardless of the training content, has been found to improve employee awareness of BCs, and help indicate the importance an organization puts on BCs (Schwartz 2004); two factors which could improve adoption. Close to three quarters of respondents spoke about the "an unfamiliarity with algorithms" which leads to employees "fighting back on the ethics stuff" (Man, Manager, United Kingdom), and suggested training as a key method to close this knowledge gap. Two-thirds of participant organizations have training programs in place, which most participants suggested would be effective.

> … for the few people that I've been able to take through the journey into machine learning…it was literally like night and day when it came to support, discussion about things like ethics…I've seen it as tremendously, tremendously successful. (Man, Executive, Australia)

> …how are we effectively creating awareness and get people to use [the AIPs]…having a checklist is greater, but it no one knows what it is or what they're supposed to do with it, it's not going to be very effective. (Man, Manager, Canada)

Half of the participants also perceived training to be beneficial for data scientists, who they noted are often not trained in AI ethics:

> …the junior data scientist is coming out of school right now – are they learning anything about AI ethics? I am not sure about that – I think probably like 50% or less are, because [they're] still focusing on trying to develop the best model, learn different languages… (Man, Manager, Thailand)
>
> …from the bottom up they need to learn more about this topic (AI ethics)…to train, teach, and educate the people to understand what they're doing and what's a big risk…(Woman, Manager, Canada)

A handful of participants suggested that mandatory training on the AIP may get pushback because "it's like every other corporate training…" (Man, Manager, Canada); which could decrease its positive impact on adoption. One participant even admitted to going to an AIP training but "did not do all the prerequisite

reading…" exemplifying the potential concerns on forced training. Conversely, a couple of participants noted the importance of mandatory training, leaving it unclear as to its impact on adoption.

(b) *Preferred trainers.* Employees most often prefer BC training to be done by someone internal to the organization, preferably a direct manager (Schwartz 2004; Trevino et al. 1999).

When it comes to AIPs, all organizations with a training program (two-thirds of organizations), except one, have their training run by someone internal to the organization. With only one organization using external training, it remains unclear whether the external or internal training could impact AIP adoption differently. Internal training is primarily delivered using a hybrid strategy of "online and in person" training, usually with in-person training for senior leadership and online for more junior employees.

> …we've taken on a training program that's going to be sent out to all analytics practitioners at the bank…and then there's a little bit of in person education… (Man, Manager, Canada)

The in-person sessions are usually delivered by an AI ethics expert in the organizations, sometimes the participant themselves.

> …I've done educations sessions with the compliance leadership…400 plus compliance leadership around the world…I'm going to do one with the risk management leadership team next week. (Man, Executive, Singapore)

This training strategy suggests that it could be more important for the internal trainer to be an AI ethics expert than a direct manager of those being trained, perhaps given the specialized knowledge required to explain the AIP compared to a more general business code. On the other hand, a couple of organizations have a "train the trainer" program, which could suggest the potential importance of direct manager training. Ultimately, internal training appears to be important for adoption, but the impact of direct managers and other trainers on adoption remains unknown.

Ethics Office(r)

Having an ethics office, a specific department or group which deals with ethics and conduct issues (Weaver et al. 1999b); an ethics officer, a designated individual in an organization whom employees can take their ethics concerns to (Kaptein 2015; Singh 2011); or an ethics committee, the group of people in an organization that employees can turn to with their ethics concerns (Kaptein 2015; Singh 2011) have all been found to positively impact BC adoption.

When asked about the involvement of an ethics office or ethics officer, not a single participant said that their AIP was managed by the existing ethics group in their organization. One quarter of participants said their organizations have assigned responsibility for the AIP to a single AI ethics individual, either the "head of analytics and AI" or "an AI ethics officer." The remaining three quarters of participants noted responsibility was assigned to a group of AI ethics experts, such as an "AI ethics committee", an "AI ethics group", or a "panel." Participants spoke about the importance of discussing the "hard decisions," as a group, which may suggest an AI ethics committee may be better than a single AI ethics officer. Regardless, some form of ethics office(r) is important for AIP adoption.

> …we wanted a panel to review the hard decisions, to safeguard the principles, be the governing body if we find the principles don't work, or cases that challenge them… (Man, Executive, China)
>
> …we have a committee of senior leaders…they make decisions on use cases and strategic decisions on what will be on the framework … (Man, Executive, Singapore)
>
> …we have a seminar style presentation of the different projects and everything is questioned – and there's even, within the group, an ethicist. (Man, Executive, Mexico)

Reporting Mechanism

With respect to BC, the (a) *existence of a reporting mechanism*, such as a telephone line, app, or email address that employees can use to report ethical concerns (Kaptein 2015), has been found to impact adoption (Kaptein 2015; Schwartz 2004; Singh 2011; Trevino et al. 1999; Weaver et al. 1999b). The (b) *existence of a standardized procedure*, or a clear routine for reporting any ethics concerns or allegations, has also been found to impact adoption (Weaver et al. 1999b).

(a) *Existence of a reporting mechanism.* Several forms of reporting mechanisms have been found to indirectly impact BC adoption effectiveness (Kaptein 2015), including phone lines, websites, ethics officers, and mail boxes (Weaver et al. 1999b).

When asked about reporting AIP breaches, participants differentiated between malicious and non-malicious acts: "…there's a spectrum – purposely versus accidentally…" (Man, Manager, Canada). When asked about how they would report on malicious breaches, almost every participant said they would call their "whistleblower line", also referred to as an "ethics hotline."

When it came to non-malicious AIP breaches, participants did not think the whistleblower line would work as a solution and were "not aware of any formal ways" to report non-malicious AIP breaches.

> …what is unethical as it refers or applies to analytics and AI maybe isn't as understood...I suspect no, I don't think a whistleblower line will work because I think fairness is such a hard concept to grasp. (Man, Executive, Canada)
>
> …anti-money laundering, or due diligence, or, you know, bribery – those things there are…channels for you to report, but there hasn't been a specific training for AI ethics and a specific channel for reporting AI ethics issues. (Man, Executive, United Kingdom)

Several participants did however provide suggestions as to how they would report non-malicious breaches, all of which included bringing up the issue with someone more senior, pointing to the potential importance of a reporting mechanism. For example, making someone aware in "compliance and model risk management", or the "Chief Privacy Officer and [the head of analytics]." One participant suggested that having a formal mechanism was important for more junior employees who might not "be comfortable" going directly to senior leaders.

(b) *Existence of a standardized procedure*. Having clear procedures for dealing with ethical issues or complaints, creates a sense of procedural justice in an organization which improves BC adoption (Weaver et al. 1999b).

Just under half of participants discussed the importance of "putting some structure around...the governance or ethics in AI and data," (Woman, Manager, Canada) and had done so either through a formal operating procedure or a board approved policy.

> …perhaps the most important one…we have our enterprise risk management framework. (Man, Executive, Singapore)
>
> …we have formalized policies, controls across the organization …people can come from across the enterprise and come and talk about…escalate issues…(Woman, Manager, Canada)

Another quarter of participants noted that they were either in the process of creating a standardized procedure or hoped to do so in the future.

> …we are now finalizing ethics and AI policies…we're going to be building out … the framework and the guidelines and the standards that will feed into those. (Woman, Manager, Canada)
>
> …we are sort of in the early stages of trying to establish where there need to be principles versus practical applications. (Man, Manager, Sweden)

With about three-quarters of participants discussing the importance of a standardized procedure, it is clearly important for effective AIP adoption.

Enforcement

The use of enforcement mechanisms, the methods specified for monitoring, sanctioning or otherwise ensuring compliance with the provision of a code (Singh et al. 2011) have been extensively studied, with four types of mechanisms found to impact BC adoption: (a) *audits* (Kaptein 2015; Singh et al. 2011); (b) *penalties for breaching the code* (Adam and Rachman-Moore 2004; Schwartz 2004; Singh 2011; Singh et al. 2011; Trevino et al. 1999); (c) *communicating violations* (Schwartz 2004); and (d) *incentive policies* (Kaptein 2015; Trevino et al. 1999).

(a) *Audits.* Audits, or the monitoring of an organization's adherence to their BC (Kaptein 2015), by internal and/or external parties are used by organizations to enforce BCs (Singh et al. 2011), and have been found to drive adoption and reduce unethical behaviour (Kaptein 2015).

With respect to AIPs, about a quarter of organizations are using audits to track adoption. Both "internal audit" and "third-party" external auditors are being used. Participants suggested that internal audit would review the "policies and standards and control in place…using the process for right now – business as usual." External auditors are then brought in or are being considered for more technical auditing of algorithms.

> …our audit group have been talking to third parties about how they will audit machine learning… (Man, Executive, Canada)
>
> …we're working with [an external auditor]…it will cover bias and explainability and will help transparency… (Man, Executive, Singapore)

With only a quarter of organizations using audits, it remains unclear as to their importance in effective AIP adoption.

(b) *Penalties.* Penalties for breaching a BC such as reprimand, fines, demotion, dismissal, or legal prosecution are used by organizations globally (Singh et al. 2011), but have been found to impact adoption in varying degrees (Adam and Rachman-Moore 2004; Singh 2011). With respect to AIPs, Schiff et al. (2020) propose follow-up and enforceability of penalties could have an impact on adoption.

When asked about AIP penalties, participants differentiated between malicious and non-malicious AI ethics issues, as they did when discussing reporting mechanisms. About two thirds of participants suggested that the penalties used for BC breaches (e.g., termination, legal prosecution) could be applied to malicious breaches of AIPs.

> …we have a thing internally where if you break bank principles – effectively employees get conduct points – and if you get negative conduct points then it impacts your bonus at the end of the year, it can impact your ability to get promoted… (Man, Executive, Singapore)
>
> There are clear ethical policies that go beyond data that would envelope that and there's a clear process by which you determine the level of punishment. (Man, Manager, United States)

However, the same penalties would not be applied to non-malicious breaches of the AIP; around half of participants suggested there would be an investigation to understand why the non-malicious breach occurred, but there would be no direct penalties.

> …I don't know that there'll be penalties…in my mind penalties [are] like a financial or job implication – I think there will be repercussions... (Man, Executive, Canada)

(c) *Communicating violations*. Schwartz (2004) found that communicating violations, whereby an organization communicates the details of a BC breach by an employee and the resulting disciplinary action, improve adoption; however, communications should remain anonymous and be reserved for serious violations.

When it comes to AIPs, respondents did not discuss communicating malicious violations; however, a handful of participants suggested that post-mortem reports and sharing of non-malicious breaches would be used, suggesting that it could potentially be important for adoption.

> …they'll be like 'no this is totally the wrong thing to do, why did we ever go down this route?' It would be like a kind of post-mortem if you screw up a project. (Man, Executive, Canada)

(d) *Incentive policies.* Incentive policies, or the act of rewarding good ethical behaviour as defined by a BC (Trevino et al. 1999), are considered important for adoption (Kaptein 2015; Trevino et al. 1999), given they are excluded from performance reviews (Schwartz 2004).

With respect to AIPs, not a single participant knew of an incentive policy at their organization "specific to AI at this point."

Incentive policies for ethics in general were viewed in two very different lights. Some participants in Canada, Singapore, and Europe seemed to think the use of incentives was unnecessary:

> …like a positive or performance-based reward for that ethical behaviour?...I go back to the notion of a code of conduct that all humans that work here sign – we don't walk around and high five each other every year, I think we tick a box on year end to say, I didn't do bad things… (Man, Executive, Canada)

> … I think people would be punished for unethical behaviour, but you wouldn't be rewarded for ethical behaviour. (Man, Executive, Singapore)

Whereas participants from other countries noted the use of ethics incentives, and although not originally designed to include AIPs, suggested these incentives could be used.

> … more broadly there's strong incentives on whistleblowing – there's an annual competition if you report fraudulent or incorrect behaviour you could win [money] as a positive reinforcement. (Man, Executive, South Africa)

This suggests that although they are not in place today, incentive policies for AIPs could be important for adoption; however, the effect could vary across different countries.

Measurement

The use of measurement, some method of evaluating the achievements and/or failures of ethics activities, structures, and personnel may suggest an organization is serious about their BC, increasing its adoption effectiveness (Weaver et al. 1999b). Measurement of employee understanding of the code through testing, however, has been found to be ineffective and potentially patronizing (Schwartz 2004).

Only three participants noted their organization was currently measuring adherence to their AIP. Each participant was from a different firm, and one did not permit the use of direct quotations.

> …every quarter we are measuring how many people have done the [AIP] training, how many use cases are going through the self-assessment at any period of time, the outcomes of the decisions from the [self-assessment] …we'll do some reporting on this. (Man, Executive, Singapore)
>
> …we measure successful education, awareness training of ethics in AI…the number of people certified, number of people formally trained…(Man, Manager, Canada)

Five additional participants noted interest in measuring AIP activities, suggesting its potential importance for adoption. Measurement, whether current or proposed, appeared to be broken into two types: adherence to the AIP, tracking things like "...what percentage of people are following it? How many people have it as part of their formalized procedures?" (Man, Manager, Canada) and technical adherence (i.e., the algorithmic outcomes such as fairness, and explainability).

Novel components impacting AI principle adoption effectiveness

In addition to the above components explored from BC adoption theory, participants were asked "Is there anything else that you feel has led to the effective adoption of the AI principles at your

organization?". Four additional components were uncovered: *accompanying technical processes*, *sufficient technical infrastructure*, *organizational structure*, and using an *interdisciplinary approach.*

Accompanying Technical Processes

Just over half of participants noted the existence, or development, of accompanying technical processes to provide detailed technical guidance on AIPs. In all cases, participants suggested that existing organizational processes could be adapted to fit the AIP. These processes were referred to as checklists, frameworks, assessments, and guidelines.

> …we're looking at ethical checklists or data ethics impact assessments…how to embed that within our existing processes and not create new processes for things. (Woman, Manager, Canada)
>
> …when the [AI principles] were announced and released to the bank internally we started looking at, you know, how that would change what we currently have in place… it's not replacing policies that the bank already has, it's just supplementing it. (Man, Manager, Singapore)

The findings support the idea that there is a lack of technical guidance for AIP adoption, which has been argued by AI developers (Peters 2019) and several researchers (Mittelstadt 2019; Schiff, Biddle, et al. 2020; Vakkuri et al. 2019). Participants opinions also support the suggestion from a recent industry white paper that adapting existing technical processes is sufficient to aid in AIP adoption (DeutscheBank et al. 2019). Although the accompanying technical processes were not explored in detail during the study, the findings suggest that proposed technical solutions such as those from Raji et al. (2020) and Madaio et al. (2020) could be helpful for AIP adoption.

Sufficient Technical Infrastructure

Having sufficient technical infrastructure was suggested as an important factor for AIP adoption, specifically: having a (a) *complete AI inventory* of projects, and ensuring (b) *data and system compatibility.*

(a) *Complete AI inventory.* Having a record of each AI project in the organization was recognized as an important factor for AIP adoption by about half of participants.

> …I can tell you personally every single AI use case in the bank…so it's been helpful that we know where it is… (Man, Executive, Singapore)
>
> …the second step was essentially trying to get an inventory of where AI is being used in the bank. (Woman, Manager, Canada)

Participants noted several benefits of having a complete AI inventory: it helps to "cascade" the AIPs to the relevant individuals, ensures all models are "going through the checklist" and allows for comprehensive measurement. While one quarter of organizations appeared to already have a complete AI inventory, another quarter were actively working to get them completed.

(b) *Data and system compatibility.* About 60% of participants noted that data and system compatibility played a role in AIP adoption. They suggested that technical practitioners would not be able to implement the AIP if data is missing or systems are inaccessible.

> …data is not even that well managed and organized – so that's a task in itself – so then to talk about how you're using that, if it's ethical, yeah it is quite difficult. (Man, Executive, United Kingdom)

Organizational Structure

Participants with centralized AI teams (about half of participants) suggested that this organizational structure helped with AIP adoption. The centralized structure was said to help with general adoption, as well as several specific components of AIP adoption: gathering a complete AI inventory, distributing, and reinforcing the AIP.

> …we've come together as one team… so that makes it a little bit easier to implement something like this…I think it's a very positive move that it's consolidated under one area… (Woman, Manager, Canada)
>
> …there is a central analytical unit for the entire bank…that's helpful. (Man, Manager, Brazil)

Around a quarter of participants suggested decentralized structures could hinder adoption; several challenges with the structure were noted including enforcement, assigning responsibility, AIP distribution, and gathering an AI inventory.

> …right now the structure we still have, for example small data science teams scattered across the bank – while we as the biggest team can set guidelines we can definitely not enforce and take responsibility for other teams and the way they do AI…I think this is the biggest ethics risk... (Man, Manager, Singapore)
>
> …a federated set up – in the current structure it's a bit more difficult to constrain and to make sure the communications are sent to the right people. (Man, Executive, South Africa)
>
> …an AI inventory…it's very, very, tricky, and it's really been proven very difficult to get because it's not contained in any particular part of the organization. (Woman, Manager, Canada).

The importance of a centralized organizational structure noted by participants supports the proposal by Raji et al. (2020) that structural changes could aid in AIP auditing, one aspect of adoption,

and the suggestion by Madaio et al. (2020), who recommend AIP checklists that are designed to fit specific organizational structures.

Interdisciplinary Approach

Throughout the interviews there surfaced a common sentiment from almost every participant, across all levels of seniority: AIP adoption is highly complex problem which no one has come close to solving, but one way to deal with it is to use an interdisciplinary approach for AIP adoption. Participants suggested: (a) *interdisciplinary teams*, (b) *combining AI ethics with data ethics*, (c) *hiring the right people*, and (d) *engaging with third party experts.*

(a) Interdisciplinary teams. The more people involved in the discussion, the better, according to participants, who suggested a wide range of potential roles to involve in AIP adoption: data/analytics/AI practitioners, privacy, legal, compliance, risk, business strategy, HR, and ethics.

> So, we are leveraging cross functional people across the different teams, which includes legal and includes privacy… simply because this is not something that's very clear… (Woman, Manager, Canada)
>
> …an enterprise working group…data and analytics practitioners, legal, privacy…I think there's a better opportunity to have it be successful if there's more people involved... (Woman, Manager, Canada)

The findings support the argument for general increased diversity made by practitioners (DeutscheBank et al. 2019), and the recommendations to include a diverse set of stakeholders in the development of accompanying technical processes, such as checklists (Madaio et al. 2020) and internal algorithmic auditing frameworks (Raji et al. 2020).

(b) *Combining AI ethics with data ethics.* Another interdisciplinary approach just over 60% of participants noted their organizations have embraced is a joint AI and data ethics program. Participants said that AI is highly reliant on data, and therefore AIP adoption should be aligned with data ethics initiatives.

> …we intentionally are doing this under the data management heading because we also see lots of interesting connections between this and privacy… (Man, Executive, Singapore)
>
> …these principles – we use them both on AI, machine learning, privacy and data use – so it's not just about AI anymore…without data there is no machine learning or artificial intelligence that can be done… (Woman, Manager, Canada)

(c) *Hiring the right people*. About one-fifth of participants noted that they don't have the right expertise in their organization to implement AIPs properly, and to rectify this have worked to hire people from outside the organization with the proper skills.

> …on the compliance side what they have done is a lot of self-reflection…and have identified that they don't know enough and started hiring accordingly people that have that skillset. (Man, Executive, Canada)

This finding supports the suggestion that technology vendors can play an important role in implementing AIPs (DeutscheBank et al. 2019). Other participants felt their organizations already had the necessary skills, so hiring people was not a priority, suggesting it is potentially important for effective adoption, depending on an organization's current talent pool.

(d) Engaging with third party experts. Almost 80% of participants noted the importance of engaging with third party AI experts, including technology companies, AI vendors, academia, and AI ethics experts.

> …we've convened a group of experts… a key part of this is Microsoft. Microsoft have an ethics board, and [one member of the board] graciously helped and supported a bunch of the stuff we did… (Man, Executive, Canada)
>
> …asking ourselves all the way through, 'do we need external advice?'…we are already Google and Microsoft customers, so we reached out to speak to the people who run their equivalent panels and principles and brought some of those lessons in. (Man, Executive, China)
>
> ...there was a lot of extra sets of research that were done…in partnership with universities… (Man, Executive, Canada)

(e) *Engaging with regulators*. Around 30% of participants said they were engaging with regulators and suggested it has aided AIP adoption. Engagement includes having regular meetings with the regulator, responding to calls for research or surveys, and inviting the regulator to external AI ethics events. Certain regulators are less active in AI ethics, which may impact the importance of engagement, but there appears to be high levels of engagement in Singapore, Canada, the UK, and Australia.

> …the [regulator's AI principles], we were one of the parties that contributed to the development of that and I think everyone now even at a junior business level is aware of those principles…(Man, Executive, Singapore)
>
> …the whole ethics of AI as well as the AI strategy is really grounded by [our regulator]…we see it as an opportunity…we think's it's in our best interest to kind of forge the discussion around it. (Man, Executive, Canada)

> …a lagged adoption rate in Mexico compared to the US and Canada, which also means a lagged discussion by the regulators. (Man, Executive, Mexico)

Summary of components of AIP adoption

Based on the opinions and perceptions of the employee participants, it appears there are several components that are important, or potentially important for effective adoption of AIPs. Components are identified as "Important" if more than 50% of participants discussed the topic, and "Potentially important" if 15-50% of participants discussed it. A summary of the findings is presented in Table 2.

Table 2. Summary of components that could impact the effective adoption of AI principles

Adoption components	Relationship to AI principle adoption effectiveness	Summary of relationship importance
Components impacting AIP adoption effectiveness from business code adoption theory		
(1) Communication		
Reach	Potentially important	Two groups: "all AI employees should see them," and "only managers need to worry about them"
Distribution Channel	Potentially important	No preference between formal and informal channels; participatory design may help
Sign-off process	Potentially important	Not common practice today, but some organizations may look to implement in future
Reinforcement	Important	Multiple channels are used to increase communication frequency, starting with reinforcement as early as hiring stage
Communication quality	Important	Quality means clear definitions, aligning with marketing, and cultural relevance
External communication	Potentially important	Four groups: not shared, white paper/summary shared, shared, already publicly available; could be considered whitewashing
(2) Management Support		
Local management support	Important	Support shown by speaking about the AIP, understanding the AIP, and general advocacy
Senior management support	Important	Support shown through talking about, knowing AIP; not expected to model behaviour
(3) Training		
Existence of training	Important	Training important to educate non-AI practitioners on AI, and technical team on ethics; mandatory training may get pushback
Preferred trainers	Potentially important	Internal training important, unclear whether external training has different impact No clear preference between internal in-person or online training, direct or senior managers
(4) Ethics Office(r)	Important	Specific AI ethics officer not necessarily important, but responsibility assigned to an individual or ethics panel is vital
(5) Reporting Mechanism		
Existence of a reporting mechanism	Important	Malicious AI principles breaches use existing ethics reporting mechanism; non-malicious acts may not need a reporting mechanism but could benefit junior employees

Location Table 2

Existence of a standardized procedures	Important	Formal operating procedures or board approved policies used, high priority to develop if not currently in place
(6) Enforcement		
Audits	Potentially important	Internal audits used for policy adherence and external audits use for technical adherence
Penalties	Potentially important	Existing penalties used for malicious AIP breaches, but none for non-malicious breaches
Communicating violations	Potentially important	Not important for malicious breaches; important to share non-malicious breaches via post-mortem
Incentive policies	Potentially important	No policies specific to AIPs yet, general ethics incentives cover AIPs in some instances, highly dependent on operating country
(7) Measurement	Potentially important	Future priority for some organizations, however only a handful are measuring today
Novel components impacting AIP adoption		
(8) Accompanying Technical Processes	Important	Helps translate the AIPs into technical guidelines.
(9) Sufficient Technical Infrastructure		
Complete AI inventory	Important	Aids in the distribution and tracking of principles.
Data and system compatibility	Important	Data issues and legacy systems can prevent technical adoption.
(10) Organizational Structure	Important	Centralized AI teams make adoption easier.
(11) Interdisciplinary Approach		
Interdisciplinary teams	Important	Increased diversity of thought, especially from outside the AI team is important.
Combining AI ethics with data ethics	Important	Integration of AI ethics and data ethics and/or privacy given the importance of data to AI.
Hiring the right people	Potentially important	Important if AI ethics talent is not available internally.
Engaging with third party experts	Important	Technology companies, AI vendors, and academia, AI ethics experts.
Engaging with regulators	Potentially important	Dependent on willingness of regulator to engage.

*Components are identified as "Important" if >50% of participants discussed the topic, and "Potentially important" if 15-50% of participants discussed it.

A Brief Discussion of Variation Across Interviewee Groups

In this section, the key differences in the perceptions of AIP adoption between interviewee groups are explored. Perceptions are compared across genders, seniority levels (e.g., executive, manager, non-manager), and cultural dimensions.

With respect to gender, in this study there appears to be no discernable difference in employee perceptions of effective AIP adoption; in line with the limited differences observed between genders across a meta-analysis of several ethics and gender studies (Dalton and Ortegren 2011).

Perceptions of AIP adoption also appear to quite similar between executives, managers, and non-managers, except for the perceptions of training effectiveness. Executives are keen supporters of mandatory AIP training, but managers are more apprehensive, a difference which reflects past research on

general ethics, where the perceptions of senior managers (e.g., executives) were been found to be more positive than the perceptions of lower level employees (e.g., manager, and non-managers) (Treviño et al. 2008). The perceptions of other AIP adoption components did not substantially differ between these groups. Participants did however suggest that certain adoption components could affect certain levels of employees differently (i.e., AIP adoption by managers and non-managers could be greater impacted by sufficient technical infrastructure). While perceptions clearly differ on training, the alignment on the other components suggests differences in seniority have limited impact on AIP adoption perceptions.

The most discernable difference observed in AIP adoption perceptions was between participants from different countries, and accordingly, different cultures. Participants from different cultures explicitly noted differences in their use of incentive policies for good ethical behaviour (i.e., common in South Africa, not common in Canada, or Europe), their country's awareness of inequalities (i.e., racial inequalities are less recognized in Mexico versus the United States), and engagement with regulators (e.g., engagement with the financial regulator is common practice in Canada, Sweden, Australia, and Singapore, but not common in Mexico, Thailand, or Brazil). Variance in cultural dimensions (Hofstede et al. 2010) could be driving these differences as it has been observed that cultures higher in uncertainty avoidance and individuality are better at ethics implementation, whilst higher power distance and masculinity are observed to be negatively associated with ethics implementation (Scholtens and Dam 2007). Differences in the use and perceived effectiveness of incentive policies could therefore be driven by differences in cultural masculinity, whereas varying awareness of inequalities could be driven by power distance differences, and varying levels of engagement with regulators could be driven by variances in uncertainty avoidance (Hofstede et al. 2010). Given the apparent differences in perceptions, further research into the impact of cultural dimensions on AIP adoption is warranted.

Limitations and conclusion

This paper presents the perceptions and opinions of actual AI employees on the effective adoption of AI principles. Eleven components that could impact effective AIP adoption are uncovered through an inductive interview analysis. Seven of the components are from BC adoption theory and were found to impact effective AIP adoption, and four novel components were uncovered: accompanying technical processes, sufficient technical infrastructure, organizational structure, and interdisciplinary approach (see Table 2).

There are, however, limitations to the study. The participants in the study are all employed by financial services organizations, which could limit the generalizability of the findings to other industries.

The financial services industry is highly regulated and given that AI ethics is closely tied to existing and future regulations (e.g., privacy, data, anti-discrimination), the findings may not generalize to less regulated industries. However, it is the high level of regulation that makes financial services so interesting to study as it has resulted in the industry being one of the few to broadly adopt AIPs. Future research could explore AIP adoption in less regulated industries such as technology, or retail. Additionally, the organizations in the financial services industry are large corporations, often with ten of thousands of employees. This could mean the components required for effective AIP adoption uncovered in this study only apply to large corporations. The organizations are also all part of the private sector, and there is potential that the differences in code content across public, private, and NGO sectors discovered by Schiff et al (2021) could also apply to differences in AIP adoption. Future research should therefore explore AIP adoption in small and medium sized enterprises, across the public, private, and NGO sectors. The qualitative nature of the study relies on self-reported data, which could be affected by social desirability bias (Randall and Fernandes 1991). Research on direct behavioural evidence related to code adoption should therefore also be conducted. The use of snowball sampling also has a drawback: the representativeness of the sample is not guaranteed and, as such, there is potential for selection bias.

There are several practical implications of the study. First, the four novel components identified in the study for effective AIP adoption indicate that organizations should not treat AIPs as BCs without adjusting for the nuances between the two. Second, the findings suggest a need for AIPs to be studied as their own entity. Third, organizations interested in driving effective AIP adoption now have eleven components of effective AIP adoption to focus their efforts on. These components could, for example, guide the development of internal auditing documents or standard operating procedures around AIPs. Fourth, the unique technical components impacting AIP adoption uncovered in the study, “accompanying technical processes”, and “sufficient technical infrastructure” could also be considered for the study of the effective adoption of other technology-related principles (e.g., Internet of Things technology principles, data ethics principles). Fifth, given the observed differences in employee perceptions on AIP adoption across cultures, it is important for future studies to account for differences to further investigate these variances. Sixth, given the empirical support for the four novel components of AIP adoption, and their alignment with existing research on AIP adoption, it is suggested they are integrated into future studies and theory-building work on AIPs and AIP adoption.

In addition to the future research ideas proposed in response to the study limitations, additional work is warranted. Future research could look to quantitatively measure the adoption components

uncovered in this study, which could provide additional insight into the ranked importance of each of the eleven components. While adoption remains the main focus of *transformation* oriented studies like this one, additional *content* and *output* oriented studies could investigate other aspects of the integrated research model (Figure 1). Of particular value to the literature would be an *output* oriented study that develops a measure of overall AIP effectiveness, per BC studies (Kaptein 2011). This measure would allow future research to answer the ultimate question "are AI principles effective?". In addition to the empirical study, the theoretical study of AI principles is warranted, and could investigate the theoretical underpinnings of the proposed integrated research model, and the BC model that inspired it (Kaptein and Schwartz 2008).

The findings from this study suggest that simply having AI principles will not be enough for organizations to prevent unethical AI outcomes. Participants suggested that several components were important for effective AIP adoption: communication quality and reinforcement; local and senior management support; a training program; an ethics officer or panel; reporting mechanisms with standardized procedures; sufficient technical infrastructure including a complete AI inventory, and data/system compatibility; a centralized organizational structure; an accompanying technical; and an interdisciplinary approach for teams that combines AI and data ethics, with engagement from third party experts. Other potentially important components for AIP adoption are communication reach and distribution channel, a sign-off process, external communication of the AIP, preferred trainers, enforcement mechanisms (i.e., audits, penalties, communicating violations, and incentive policies), measurement, and an interdisciplinary approach (i.e., hiring the right people, and engaging with regulators). Additional research to clarify the potential importance of each of these components is warranted and would be beneficial not only to researchers, but to organizations, to prioritize their AIP efforts and help prevent unethical AI outcomes.

Appendix A

Semi-Structured Interview Protocol

Demographic Questions

To begin, could you please confirm whether you are an executive, manager, or non-manager?

What is your job title?

How long have you worked for the organization that currently employs you?

General AI and AIP Questions

When did your organization first start discussing AI ethics?

Does your organization have their own AI ethics principles in place?

Is your organization impacted or governed by external AI ethics initiatives? This could be from regulators, business groups, or other bodies?

How long have these AI ethics principles been in place?

Part 1: Components (proposed) from the Business Code adoption literature

How did you first hear about the principles?

Do you think all employees know about them?

What did you like about the way you were introduced to the principles?

What didn't you like?

Do you continue to hear about the principles?

How often do you get communication about the principles?

Have you read the entire principles document?

- If you read the entire thing, what made you do that?
- If you haven't read it, what stopped you or prevented you from doing so?

Do you believe the principles are supported by your direct manager?

Do you believe the principles are supported by senior managers?

Do you believe they are supported by the executive team?

Have you seen leaders modelling positive behaviour in line with the principles?

Have you been trained on the principles? What there a specific training? In person? Online?

Is there an ethics office, committee, officer, or group responsible for the principles?

Is there a way to report breaches of the principles?

How does compliance of the principles work?

Are the principles measured? Through KPIs or a dashboard, or other measurement tools?

Are there any penalties for breaching the principles?

In general, does the organization follow-up on reports of unethical use of AI? Or unethical behaviour related to AI?

Are people rewarded for ethical behaviour in line with the AI principles?

Do you think the AI principles at your bank have been adopted effectively? Have they reduced the unethical use of AI from occurring? Have they changed people's behaviours?

Part 2: Components (unknown) unique to effective AI Principle adoption
Is there anything else that you feel has led to the effective adoption of the AI principles at your organization?

Anything else that you feel has prevented or hindered the effective adoption of the AI principles at your organization?

Appendix B

Detailed discussion of the coding approach.

The study was based on a positivist epistemology, and assumed that interviewees were direct holders of accurate information (Johnson and Duberley 2000). The analysis was performed using a general inductive coding approach (Thomas 2006). Specifically, open coding was used, whereby all text in the interview transcripts were coded using labels found within the text. A memo was created for each code keep track of the developing idea, and open coding continued until every line of text was reviewed and sorted from each interview. The open coding process was first applied to 20 randomly chosen interviews, at which point the researcher reviewed the codes and grouped them into components. The components were then compared with the a priori list of components known to impact BC adoption. Components that fit with existing BC-named components were subsumed to establish a link with existing theory, while other components emerged as unique to the effective adoption of AIPs and were named using interviewee language. A stakeholder check (Thomas 2006) was then performed at this stage to evaluate the trustworthiness of the first round of coding; initial components and their accompanying memos were sent to 10 experts in the field of AI ethics; 4 of whom were participants in the study, and 6 whom were not. The experts reviewed the concepts and were asked to comment on whether the components and their descriptions matched their experiences with AI adoption; suggestions were then integrated into the codebook. The remaining 29 interviews were then coded according to the refined list of components, and where necessary additional open coding and corresponding memos were developed when text passages did not fit within the list of concepts from the first 20 interviews. The first 20 interviews were then reviewed again with the complete codebook. A second stakeholder check for finding trustworthiness was performed on the draft manuscript; all 49 participants were provided a draft and were asked to evaluate the components, their descriptions, and the supporting data (quotes). There were no changes to the proposed components during this second stakeholder check.

Codebook.

Final Component (Stage 3)	Grouped Component (Stage 2)	Codes (Stage 1)
Components Impacting AIP Adoption from Business Code Adoption Theory		
Communication	Reach	All AI employees have read AIP
	Distribution Channel	Location of AIPs known
		Participatory design
		First awareness of AIPs
		AIP discussed in hiring
	Sign-off process	Sign-off
	Reinforcement	Communication channel
		Lunch & learns on AI ethics
		Internal conference
		Employee community on AI ethics
	Communication quality	Internal marketing campaign
		Clear definitions
		Healthy AI dialogue
		Cultural relevance of AIP
	External communication	Principles shared externally
Management Support	Local management support	Direct manager prioritizes AIPs
		Direct manager is trained on AI ethics
		Direct manager is aware of AIP
	Senior management support	Top-down communication
		Top management prioritizes AIPs
		Top management is trained on AI ethics
		Top management is aware of AIP
		Executive engagement as barrier
		Funding for more staff as a barrier
Training	Existence of training	Access to training
		Required training
		Basic knowledge on AI
		Onboarding training
		Certification program
		Data scientists aren't trained in ethics
	Preferred trainer	Train the trainer
Ethics Office(r)	Ethics Office(r)	AI ethics contact
		AI ethics panel
		Clear responsibility for AI ethics
Reporting Mechanism	Existence of a reporting mechanism	Reporting AI ethics concerns
	Existence of a standardized procedure	Reporting standardization
		Board of directors approved policy
Enforcement	Audits	Consequences: auditing
	Penalties	Consequences: penalty
	Communicating violations	Consequences: communication
	Incentive policies	Reward for ethical behaviour
Measurement	Measurement	Measure AIP effectiveness
		No measurement mechanism
Novel Components Impacting AIP Adoption		
Accompanying Technical Processes	Accompanying Technical Processes	Consistent data science tool
		Adapting existing processes
		Piloting process
		Automated process
		Integration in product/service development
Sufficient Technical Infrastructure	Complete AI inventory	AI project inventory
	Data and system compatibility	Legacy systems and data
Organizational Structure	Organizational Structure	Organizational structure as barrier
Interdisciplinary Approach	Interdisciplinary teams	Interdisciplinary teams
	Combining AI ethics with data ethics	Data ethics combined with AI ethics
	Hiring the right people	Hiring the right people
	Engaging with third party experts	Leading engagement with regulators
		Leading engagement with academia

End Notes

[1] https://www.bbc.com/news/business-50365609

[2] https://www.wired.com/story/bill-congress-limit-uses-facial-recognition/

[3] https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

[4] https://kammeradvokaten.dk/nyheder-viden/nyheder/2020/06/nu-skal-virksomheder-redegoere-for-dataetik-i-aarsrapporten

[5] https://ai.google/principles/

[6] https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1%3aprimaryr6

[7] https://www.telefonica.com/en/web/responsible-business/our-commitments/ai-principles

[8] https://www.hsbc.com/-/files/hsbc/our-approach/risk-and-responsibility/pdfs/200210-hsbc-principles-for-the-ethical-use-of-big-data-and-ai.pdf?download=1

[9] https://www.partnershiponai.org/partners/

[10] https://www.torontodeclaration.org/declaration-text/english/

[11] https://www.montrealdeclaration-responsibleai.com/the-declaration

[12] https://oecd.ai/ai-principles

[13]https://www.mas.gov.sg/~/media/MAS/News%20and%20Publications/Monographs%20and%20Information%20Papers/FEAT%20Principles%20Final.pdf

References

Adam, A. M., & Rachman-Moore, D. (2004). The methods used to implement an ethical code of conduct and employee attitudes. *Journal of Business Ethics*, *54*(3), 225–244. https://doi.org/10.1007/s10551-004-1774-4

Babri, M., Davidson, B., & Helin, S. (2019). An Updated Inquiry into the Study of Corporate Codes of Ethics: 2005–2016. *Journal of Business Ethics*. https://doi.org/10.1007/s10551-019-04192-x

Barocas, S., & Selbst, A. D. (2016). Big Data's Disparate Impact. *104 California Law Review 671*.

Bietti, E. (2020). From Ethics Washing to Ethics Bashing. In *Proceedings of ACM FAT* Conference (FAT* 2020)*. Barcelona, Spain.

Bughin, J., Hazan, E., Ramaswamy, S., Chui, M., Allas, T., Dahlstrom, P., et al. (2017). *Artificial Intelligence - The Next Digital Frontier*. https://doi.org/10.1016/S1353-4858(17)30039-9

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commerical Gender Classification. In *Proceedings of Machine Learning Research* (Vol. 81, pp. 1–15). https://doi.org/10.2147/OTT.S126905

Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). Artificial Intelligence and the 'Good Society': the US, EU, and UK approach. *Science and Engineering Ethics*, *24*(2), 505–528. https://doi.org/10.1007/s11948-017-9901-7

Dalton, D., & Ortegren, M. (2011). Gender Differences in Ethics Research: The Importance of Controlling for the Social Desirability Response Bias. *Journal of Business Ethics*, *103*(1), 73–93. https://doi.org/10.1007/s10551-011-0843-8

DeutscheBank, Linklaters, Microsoft, StandardChartered, & Visa. (2019). *From Principles to Practice: Use Cases for Implementing Responsible AI in Financial Services*.

Financial Stability Board. (2017). Artificial intelligence and machine learning in financial services: Market developments and financial stability implications. *Financial Stability Board*, (November). http://www.fsb.org/2017/11/artificial-intelligence-and-machine-learning-in-financial-service/

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. C., & Srikumar, M. (2020). *Principled Artificial intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI*. *The Berkman Klein Center for Internet & Society Research Publication Series*. https://doi.org/10.1109/MIM.2020.9082795

Floridi, L. (2019). Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical. *Philosophy and Technology*, *32*, 185–193. https://doi.org/10.1007/s13347-019-00354-x

Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines*, *30*, 99–120. https://doi.org/10.1007/s11023-020-09517-8

Helin, S., & Sandström, J. (2007). An inquiry into the study of corporate codes of ethics. *Journal of Business Ethics*, *75*(3), 253–271. https://doi.org/10.1007/s10551-006-9251-x

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and Organizations: Software of the Mind: Intercultural Cooperation and Its Importance for Survival*. https://www.amazon.ca/Cultures-Organizations-Intercultural-Cooperation-Importance/dp/0077074742. Accessed 22 February 2021

Huang, M. H., & Rust, R. T. (2018). Artificial Intelligence in Service. *Journal of Service Research*, *21*(2), 155–172. https://doi.org/10.1177/1094670517752459

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*, 389–399. https://doi.org/10.1038/s42256-019-0088-2

Johnson, P., & Duberley, J. (2000). *Understanding Management Research: An Introduction to Epistemology* (First.). Sage Publications.

Kaplan, A., & Haenlein, M. (2019a). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, *62*, 15–25. https://doi.org/10.1016/j.bushor.2018.08.004

Kaplan, A., & Haenlein, M. (2019b). Rulers of the world, unite! The challenges and opportunities of artificial intelligence. *Business Horizons*. https://doi.org/10.1016/j.bushor.2019.09.003

Kaptein, M. (2004). Business codes of multinational firms: What do they say? *Journal of Business Ethics*, *50*(1), 13–31. https://doi.org/10.1007/978-94-007-4126-3_27

Kaptein, M. (2011). Toward Effective Codes: Testing the Relationship with Unethical Behavior. *Journal of Business Ethics*, *99*(2), 233–251.

Kaptein, M. (2015). The Effectiveness of Ethics Programs: The Role of Scope, Composition, and Sequence. *Journal of Business Ethics*, *132*(2), 415–431. https://doi.org/10.1007/s10551-014-2296-3

Kaptein, M., & Schwartz, M. S. (2008). The effectiveness of business codes: A critical examination of existing studies and the development of an integrated research model. *Journal of Business Ethics*, *77*(2), 111–127. https://doi.org/10.1007/s10551-006-9305-0

Khalil, O. E. M. (1993). Artificial Decision-Making and Artificial Ethics: A Management Concern. *Journal of Business Ethics*, *12*(4), 313–321.

Madaio, M. A., Stark, L., Wortman Vaughan, J., & Wallach, H. (2020). Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI. *Conference on Human Factors in Computing Systems - Proceedings*, (August). https://doi.org/10.1145/3313831.3376445

Martin, K. E. (2019). Ethical Implications and Accountability of Algorithms. *Journal of Business Ethics*, *160*, 835–950. https://doi.org/10.1007/s10551-018-3921-3

Martin, K. E., & Freeman, R. E. (2004). The Separation of Technology and Ethics in Business Ethics. *Journal of Business Ethics*, *53*, 353–364. https://ssrn.com/abstract=1410846

Martin, K. E., Shilton, K., & Smith, J. (2019). Business and the Ethical Implications of Technology: Introduction to the Symposium. *Journal of Business Ethics*, *160*, 307–317. https://doi.org/10.1007/s10551-019-04213-9

McNamara, A., Smith, J., & Murphy-Hill, E. (2018). Does ACM's code of ethics change ethical decision making in software development? In *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering - ESEC/FSE 2018* (pp. 729–733). https://doi.org/10.1145/3236024.3264833

Miles, M. B., & Huberman, A. M. (1994). *Qualitative Data Analysis* (Second Edi.). Sage Publications.

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, *1*(11), 501–507. https://doi.org/10.1038/s42256-019-0114-4

Morley, J., Floridi, L., Kinsey, L., & 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*, 2141–2168. https://doi.org/10.1007/s11948-019-00165-5

Peters, D. (2019). Beyond Principles: A Process for Responsible Tech. *Medium*. https://medium.com/ethics-of-digital-experience/beyond-principles-a-process-for-responsible-tech-aefc921f7317. Accessed 14 September 2020

Petersen, L. E., & Krings, F. (2009). Are ethical codes of conduct toothless tigers for dealing with employment discrimination? *Journal of Business Ethics*, *85*, 501–514. https://doi.org/10.1007/s10551-008-9785-1

Raji, I. D., & Buolamwini, J. (2019). Actionable Auditing: Investingating the Impact of Publicly Naming Biased Performance Results of Commerical AI Products. In *Conference on Artificial Intelligence, Ethics, and Society*. https://doi.org/10.1145/3306618.3314244

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., et al. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. *FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 33–44. https://doi.org/10.1145/3351095.3372873

Randall, D. M., & Fernandes, M. F. (1991). The Social Desirability Response Bias in Ethics Research. *Journal of Business Ethics*, *10*(11), 805–817. https://doi.org/10.1007/978-94-007-4126-3_9

Rességuier, A., & Rodrigues, R. (2020). AI ethics should not remain toothless! A call to bring back the teeth of ethics. *Big Data and Society*, *July-Decem*, 1–5. https://doi.org/10.1177/2053951720942541

Schiff, D., Biddle, J., Borenstein, J., & Laas, K. (2020). What's next for AI ethics, policy, and governance? A global overview. In *AIES 2020 - Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (pp. 153–158). https://doi.org/10.1145/3375627.3375804

Schiff, D., Borenstein, J., Biddle, J., & Laas, K. (2021). *AI Ethics in the Public , Private , and NGO Sectors : A Review of a Global Document Collection*.

Schiff, D., Rakova, B., Ayesh, A., Fanti, A., & Lennon, M. (2020). Principles to Practices for Responsible AI: Closing the Gap. *arXiv*.

Scholtens, B., & Dam, L. (2007). Cultural values and international differences in business ethics. *Journal of Business Ethics*, *75*(3), 273–284. https://doi.org/10.1007/s10551-006-9252-9

Schwartz, M. S. (2001). The nature of the relationship between corporate codes of ethics and behaviour. *Journal of Business Ethics*, *32*(3), 247–262. https://doi.org/10.1023/A:1010787607771

Schwartz, M. S. (2004). Effective corporate codes of ethics: Perceptions of code users. *Journal of Business Ethics*, *55*(4), 323–343.

Singh, J. B. (2011). Determinants of the Effectiveness of Corporate Codes of Ethics: An Empirical Study. *Journal of Business Ethics*, *101*(3), 385–395. https://doi.org/10.1007/s10551-010-0727-3

Singh, J. B., Svensson, G., Wood, G., & Callaghan, M. (2011). A longitudinal and cross-cultural study of the contents of codes of ethics of Australian, Canadian and Swedish corporations. *Business Ethics*, *20*(1), 103–119. https://doi.org/10.1111/j.1467-8608.2010.01612.x

Smith-Crowe, K., Tenbrunsel, A. E., Chan-Serafin, S., Brief, A. P., Umphress, E. E., & Joseph, J. (2015). The Ethics "Fix": When Formal Systems Make a Difference. *Journal of Business Ethics*, *131*(4), 791–801. https://doi.org/10.1007/s10551-013-2022-6

Spiekermann, S. (2016). *Ethical IT Innovation: A Value-Based System Design Approach*. (J. Cantella, Ed.). Boca Raton: Taylor & Francis Group, LLC.

Stevens, B. (2008). Corporate Ethical Codes: Effective Instruments For Influencing Behavior. *Journal of*

Business Ethics, *78*(4), 601–609. https://doi.org/10.1007/sl0551-007-9370-z

Thomas, D. R. (2006). A General Inductive Approach for Analyzing Qualitative Evaluation Data. *American Journal of Evaluation*, *27*(2), 237–246. https://doi.org/10.1177/1098214005283748

Treviño, L. K., Weaver, G. R., & Brown, M. E. (2008). It's lovely at the top: Hierarchical levels, identities, and perceptions of organizational ethics. *Business Ethics Quarterly*, *18*(2), 233–252. https://doi.org/10.1017/s1052150x00010952

Trevino, L. K., Weaver, G. R., Gibson, D. G., & 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., Kultanen, J., Siponen, M., & Abrahamsson, P. (2019). *Ethically Aligned Design of Autonomous Systems: Industry viewpoint and an empirical study*.

Weaver, G. R., Treviño, L. K., & Cochran, P. L. (1999a). Corporate ethics programs as control systems: Influences of executive commitment and environmental factors. *Academy of Management Journal*, *42*(1), 41–57. https://doi.org/10.2307/256873

Weaver, G. R., Treviño, L. K., & Cochran, P. L. (1999b). Corporate ethics practices in the mid-1990s: An empirical study of the fortune 1000. *Journal of Business Ethics*, *18*(3), 283–294. https://doi.org/10.1007/978-94-007-4126-3_31

Wellman, M. P., & Rajan, U. (2017). Ethical Issues for Autonomous Trading Agents. *Minds and Machines*, *27*(4), 609–624. https://doi.org/10.1007/s11023-017-9419-4

Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., et al. (2018). *AI Now Report 2018*. New York. www.ainowinstitute.org