

2022

Research Perspectives: Toward Theoretical Rigor in Ethical Analysis: The Case of Algorithmic Decision-Making Systems

Uri Gal

University of Sydney, uri.gal@sydney.edu.au

Sean Hansen

Rochester Institute of Technology, shansen@saunders.rit.edu

Allen S. Lee

Virginia Commonwealth University, AllenSLee@alum.mit.edu

Follow this and additional works at: <https://aisel.aisnet.org/jais>

Recommended Citation

Gal, Uri; Hansen, Sean; and Lee, Allen S. (2022) "Research Perspectives: Toward Theoretical Rigor in Ethical Analysis: The Case of Algorithmic Decision-Making Systems," *Journal of the Association for Information Systems*, 23(6), 1634-1661.

DOI: 10.17705/1jais.00784

Available at: <https://aisel.aisnet.org/jais/vol23/iss6/1>

This material is brought to you by the AIS Journals at AIS Electronic Library (AISeL). It has been accepted for inclusion in Journal of the Association for Information Systems by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Research Perspectives: Toward Theoretical Rigor in Ethical Analysis: The Case of Algorithmic Decision-Making Systems

Uri Gal,¹ Sean Hansen,² Allen S. Lee³

¹University of Sydney, Sydney, Australia, uri.gal@sydney.edu.au

²Rochester Institute of Technology, USA, shansen@saunders.rit.edu

³Virginia Commonwealth University, USA, allenslee@alum.mit.edu

Abstract

Algorithmic decision-making systems (ADMS) are increasingly being used by public and private organizations to enact decisions traditionally made by human beings across a broad range of domains, including business, law enforcement, education, and healthcare. Their growing prevalence engenders profound ethical challenges, which, we maintain, should be examined in a structured and theoretically informed fashion. However, much of the ethical exploration of ADMS within the IS field draws upon an atheoretical application of ethics. In this paper, we argue that the “big three” ethical theories of consequentialism, deontology, and virtue ethics can inform a structured comparative analysis of the ethical significance of ADMS. We demonstrate the value of such an approach through an illustrative case study of an ADMS in use by an Australian bank. Building upon this analysis, we address four characteristics of ADMS from the three theoretical perspectives, provide guidance on the contexts within which the application of each theory might be particularly fruitful, and highlight the advantages of theoretically grounded ethical analyses of ADMS.

Keywords: Ethical Analysis, Algorithmic Decision-Making Systems, Theoretical Rigor, Consequentialism, Deontology, Virtue Ethics

Sutirtha Chatterjee was the accepting senior editor. This research perspectives article was submitted on January 19, 2021 and underwent three revisions.

1 Introduction: Structured Ethical Analysis and the Rise of Algorithmic Decision-Making

Ethics is one the oldest domains of philosophical inquiry, focusing on questions of right and wrong action or how individuals within a society ought to behave (Finnis, 1983; Sidgwick, 1907). Given their sweeping implications for enabling and constraining individual and collective action, information systems (IS) present manifold avenues for ethical analysis. Indeed, a consideration of the ethical implications of various IS elements is a regular feature of research in the discipline (e.g., Couger, 1989; Mason, 1986; Mingers & Walsham, 2010).

One class of technologies that poses particularly challenging ethical questions is algorithmic decision-making systems (ADMS) (Lo Piano, 2020; Mittelstadt et al., 2016; Neyland, 2019). ADMS are autonomous self-learning systems that make judgments with little or no direct human intervention (Mökander et al., 2021). They are increasingly utilized by governments, public agencies, and private corporations, making their way into our cities, workplaces, and homes. They enact decisions traditionally made by humans across a range of domains, including business, law enforcement, education, and healthcare (Crawford, 2021). Vendors of ADMS offer visions of data analytics that are “speedy, accessible, revealing, panoramic, prophetic, and smart” (Beer, 2018, p. 469). We are told that they

are objective and rational, far exceeding human analytical abilities. Accordingly, they can help us make evidence-based decisions, get to know our customers, manage employees, and lead healthy lives.

Despite this promise, the spread of ADMS has engendered significant concerns, including their potential to detrimentally reshape the relationships between countries and their citizens, organizations and their employees, and among individuals (Zuboff, 2015). Some fear that a small number of powerful corporations, controlling vast amounts of data, can shape people's interpretations of reality and alter their behavior (Beer, 2018; Ward, 2018). The expanding scope of data collection required for ADMS can erode our privacy (e.g., Stahl and Wright 2018) and create the foundations of a surveillance state (Liang et al, 2018). The utilization of ADMS also has the potential to introduce systematic bias (Kordzadeh & Ghasemaghaei, 2022; O'Neil, 2016), for instance based on gender (Rai, Constantinides, & Sarker, 2019), and produce unfair outcomes (Mittelstadt et al. 2016; Wachter et al, 2020). This is especially worrying because many ADMS lack accountability (Pasquale, 2015), being protected by corporate or government confidentiality rules, and are too complex for even their designers to understand (Faraj et al., 2018).

In light of the multifaceted implications of ADMS, much of the research on this class of technologies has considered its ethical ramifications (e.g., Bostrom & Yudkowsky, 2014; Gal et al., 2020; Marabelli et al., 2017). Several researchers have questioned the quality of the data used to inform algorithmic analysis, including its potential insufficiency and inconclusiveness (Garcia, 2016). Observers have similarly scrutinized the analytical processes themselves, with particular concern for the unpredictability and inscrutability of ADMS (Zerilli et al. 2019). Still others have emphasized the potential of ADMS to condition the way we see the world (D'Agostino and Durante 2018; Floridi 2014). With the recognition of these diverse ethical facets, many place the onus on ADMS developers to create tools that are auditable, predictable, and in line with prevailing ethical principles (Bostrom & Yudkowsky, 2014; Marabelli et al., 2017).

These studies offer valuable contributions in highlighting ethical challenges emanating from ADMS use. However, the application of ethical concepts in this literature is generally not couched in ethical theory (Stahl, 2012). While a handful of studies explicitly apply ethical theory to ADMS (e.g., Ananny, 2016; Gal et al., 2020; Kraemer et al., 2011; Sandvig et al., 2016), most ethical treatments of ADMS focus instead on broad classes of ethical concerns, such as privacy, bias, and accountability, without grounding their

examinations in formal theory (Jobin et al., 2019). This approach may be problematic because it accepts as a given that such concepts are ethically salient without interrogating the foundational theoretical principles on which that determination is made.

Since the growing prevalence of ADMS, and their new and embedded forms of material agency (Berente et al., 2021), can be viewed as raising the ethical stakes, we believe it behooves us to examine the ethical challenges that they present in a structured and theoretically informed fashion. Such an examination is required to develop a cumulative body of research about the ethical aspects of ADMS. Moreover, it can help researchers generate theoretically substantiated and nuanced discussions about the nature and effect of ADMS, mitigate the risk of ethnocentric ethical evaluations that can lead to moral tribalism and polarization (Markowitz & Shariff, 2012), and clarify issues of theoretical uniformity and moral relativism.

Accordingly, we argue for enhanced theoretical rigor in the ethical evaluation of ADMS by drawing on the “big three” ethical theories: consequentialism, deontology, and virtue ethics (Arnold et al. 2010). Each of these ethical traditions has a rich history marked by distinct foundational principles. In brief, *consequentialism* assesses the ethicality of acts based on their outcomes, *deontology* on adherence to universal moral rules, and *virtue ethics* on the personal disposition of the actor that performs them. These ethical theories and their associated bases for adjudicating right action have significant implications for any assessments of a technology and its use. Moreover, their juxtaposition can help enrich the ethical vocabulary applied in such assessments, thereby improving the quality of ethical reflection. Such a juxtaposition can also help researchers assess the conditions or contexts under which one or another of the theories is analytically efficacious.

The paper proceeds as follows: In the next section, we explore the need for theoretically grounded ethical analyses in assessments of ADMS. After that, we provide an overview of the “big three” ethical theories as a foundation to support such an analysis. We then demonstrate the practical ramifications of the three theories through their application to an illustrative case of ADMS use. Building upon this analysis, we consider how the theories can help address salient ethical issues that are borne out of the defining characteristics of ADMS. We further offer an assessment of the applicability of the three theories in various contexts of ADMS use, and discuss key benefits conferred by the application of theory to the ethical analysis of ADMS. We conclude by outlining the implications of our research and proposing avenues for future research.

2 Ethical Studies in IS and the Need for a Theoretically Grounded Ethical Analysis of ADMS

The IS field has a long history of research on ethics. In their extensive review of leading IS journals, Paradise et al. (2018) identify 145 papers that examine ethical issues in relation to various aspects of technology, such as design and innovation, privacy and security, and use and planning. While valuable for highlighting the significant ethical implications of IS, much of this work either applies no formal ethical theory or does so only latently.

To substantiate this observation, we draw on a framework developed by Stahl (2012), which distinguishes between four levels of normativity that can be used in research. The first level, *moral intuition*, includes studies that implicitly assume that certain things are good or bad, right or wrong. However, the reasoning for classifying things into a given category is not specified, explained, or justified. Studies that engage in the second level of normativity, *explicit morality*, openly assert the appropriateness of things or actions and provide criteria for making such designations. Research that draws on professional codes of conduct entails explicit morality because these codes overtly express moral standards. While the two first levels of normativity are a-theoretical, the third level, *ethical theory*, explicitly utilizes ethical theoretical frameworks to interrogate the moral nature of things or events. IS studies that employ ethical theory have applied such theories as utilitarianism, deontology, and virtue ethics to examine various technology-related phenomena. The final level, *reflection and meta-ethics*, acknowledges that different ethical theories may diverge in their evaluations of a given situation or action, applying multiple theories to examine complex moral situations from diverse perspectives.

Much IS research falls within Stahl's first two categories (i.e., moral intuition and explicit morality): Scholars assume that various IS issues or outcomes are ethically significant and desirable or not, but they do not detail the criteria for these evaluations or draw on ethical theory to substantiate them (Stahl, 2012). For instance, Mason's (1986) seminal PAPA framework identifies privacy, accuracy, property, and accessibility as core ethical concerns without justifying this choice vis-à-vis existing literature or theory (Stahl, 2012). Similarly, Banerjee et al. (1998) and Leonard and Cronan (2001) equate unethical behavior with misuse of IS that can lead to business losses but do not explain the rationale for this classification. Other studies discuss the need to protect people's privacy (Culnan & Williams, 2009) in the digital age (Bélanger & Crossler, 2011) or in the face of big data (Günther et al., 2017). While all of these studies explicitly

assert that privacy is ethically desirable, they do not explain why. Such research does not reflect on the possible contributions of ethical theory to current research and practice (Stahl, 2012).

Some research from IS-adjacent fields (most notably, computer science) does draw on ethical theory to inform the design of autonomous moral agents (Berberich & Diepold, 2018) and hence falls within Stahl's third and fourth levels of normativity. For instance, Bench-Capon (2020) builds on consequentialism, deontology, and virtue ethics to design autonomous agents. Thornton et al. (2017) apply the "big three" theories to design different aspects of an automated vehicle control algorithm. Govindarajulu et al. (2019) build on virtue ethics to derive formalized rules, which they build into the code of an autonomous agent. These studies share the assumption that ethics can be encoded into algorithms and that the burden of doing so falls on developers. While this research explicitly draws on ethical theory, it largely elides the complex organizational and social contexts within which algorithms are used, as well as the competing demands and inherent equivocality (Daft & Weick, 1984) that characterize the problem domains of human decision-makers using these algorithms.

Another stream of research that draws on ethical theory discusses its relevance for IS research. For example, Laudon (1995) offers a typology of rule-based and consequentialist theories and reviews their significance for IS research. Mingers and Walsham (2010) discuss the relevance of discourse ethics to IS research, and Chatterjee and Sarker (2013) draw on consequentialist, deontological, and virtue ethics to propose research questions about knowledge management programs. Other IS research emphasizes how the application of ethical theory can promote effective managerial and organizational practice. For instance, Smith and Hasnas (1999) explore how stockholder, stakeholder, and social contract theories can help managers navigate IS dilemmas. Siponen and Iivari (2006) show how different ethical theories can be used to inform IS security policies. Chatterjee et al. (2009b) draw on deontological ethics to derive metarequirements in order to design tools that support ethical collaboration. Chatterjee et al. (2009a) demonstrate how the application of postmodern ethics can enhance a sense of moral responsibility among IS development project stakeholders. Sojer et al. (2014) incorporate principles from teleological and deontological ethics to examine individuals' unethical reuse of code. Chatterjee et al. (2015) examine how certain IT affordances can help generate organizational virtues that are positively associated with organizational improvisation and innovativeness.

While these IS studies examine various technologies and phenomena, only a handful of IS studies have applied ethical theory to study ADMS. In our review of the IS "basket of eight" journals, as well as *Information &*

Organization, we found 13 papers that discuss the ethical aspects of ADMS. Of these, only two explicitly apply ethical theory (Stahl's third level of normativity), and none applied a multi-lens ethical approach (Stahl's fourth level of normativity). We describe this review of the literature in more detail in Appendix 1. This finding highlights the need for IS research that engages in structured ethical analysis of ADMS. To this end, we next describe three ethical theories that can provide a theoretical framework to facilitate such investigations.

3 Three Theories of Ethical Reasoning

The philosophical analysis of ethics is dominated by three core perspectives, each with a rich tradition—consequentialism, deontology, and virtue ethics. Given the abundance of literature on the three ethical traditions, our review is not meant to be exhaustive. Rather, we seek only to introduce the theories and the core principles underlying them, emphasizing critical points of comparison and contrast. On this basis, we wish to provide an integrative analytical lens to assess the ethical nature of ADMS. A summary is provided in Table 1.

3.1 Consequentialist/Utilitarian Ethics

Consequentialist theory refers to a variety of ethical perspectives that assess the ethicality of acts on the basis of their consequences. These perspectives identify certain conditions as a “good state of affairs,” with the appropriateness of actions being assessed based on the degree to which they lead to such conditions (Foot, 1985). The quintessential example of a consequentialist theory is *utilitarianism*. Formalized in the work of Bentham (Burns & Hart, 1996) and Mill (1859), utilitarianism holds that the appropriate measure of right action is the greatest well-being, good, or happiness for the greatest number of people.

Consequentialist theories hold that for an act to be morally correct, it is not enough that it *improves* the state of affairs in the world; rather, there must be no other available act that would have had better consequences (Shaw, 2006). At the same time, there may be situations where multiple acts have equally positive outcomes, in which case there can be multiple ethically appropriate actions. Consequentialist theories presume the possibility of an impartial vantage point from which to judge the rightness of actions and their consequences (Portmore, 2001). This agent-neutral approach implies that it is feasible to compare the consequences of different outcomes to determine which is preferred (e.g., an action resulting in the death of one person is objectively better than one resulting in the death of two).

This point illustrates the consequentialist preference for the “good” over the “right”: We can determine the appropriateness of any act by its consequences, without engaging in a normative assessment of the act itself. This can be done prior to, and independent of, the proposed act through a calculative process—weighing the expected pros and cons, aggregating them, and ranking actions accordingly. These calculations are expected to converge across individual sensible actors. Thus, consequentialism instructs us to do what is likely to have the best results, as judged by what a reasonable person in our circumstances can be expected to know (Shaw, 2006).

While consequentialism guides us to take actions that bring about the best possible results, it does not explicitly define what the best results are. The notion of “good” remains vague. As noted above, utilitarianism identifies the best state of affairs as that which maximizes happiness, pleasure, or well-being. Such well-being is additive in the sense that it is possible to measure the net amount of total well-being by adding up the good or bad experienced by every individual. This principle presumes that we can compare people's levels of well-being despite cultural differences, personal preferences, and dissimilarities.

Critics of utilitarianism hold that this presumption is overly simplistic and impossible to implement (Williams, 1973). Further, even if one could effectively calculate the sum of good and bad outcomes, doing so necessitates a finite temporal window (Garcia & Nelson, 1994). Over an unbounded timeframe, any action could have infinite outcomes, both good and bad. Another critique leveled at utilitarianism is that it ignores everything other than happiness or well-being which may have intrinsic value to individuals, such as autonomy, friendship, or knowledge. The utilitarian response to such claims is that those things only have derivative value to individuals. In other words, we deem autonomy, friendship, or knowledge as valuable only because they engender happiness or pleasure (Shaw, 2006).

Ultimately, many maintain that utilitarian ethics are devoid of something morally essential and often conflict with our moral intuitions (Macdonald & Beck-Dudley, 1994). Since acts are not judged by their inherent moral value, everything is permissible, as long as it can be argued to increase well-being. Nothing is entirely forbidden; an intuitively abhorrent act, such as the murder of innocents, may still be rationalized and justified (e.g., if such an act would save the lives of more people than were murdered). This purported “blind spot” for utilitarian ethics is a central concern for the second of the frameworks we consider—deontological ethics.

Table 1. A Summary of the Three Ethical Theories (adapted from Dignum, 2019)

Dimensions	Ethical theories		
	Consequentialism/ utilitarianism	Deontology	Virtue
Primary focus	Consequences of act	Moral obligations	Character development
Guiding value	Good (seen as maximum happiness or well-being)	Right (in accordance with moral rules)	Virtue (to develop human character)
Evaluative measure	Maximizing well-being and happiness	Satisfying categorical imperatives	Pursuing eudaimonia; developing practical wisdom
Locus of ethical evaluation	People impacted by act	Action and its compliance with moral rules	Individual motivation for action

3.2 Deontological Ethics

The theory of deontological ethics is concerned with the inherent morality of acts. Any act can be judged on its own merit and assessed for its ethical nature. The basic criterion for such assessment is whether the act conforms to a moral norm. When acts do not conform to moral norms, they are deemed wrong, regardless of how positive their consequences may be. Similarly, acts that do conform to moral norms are deemed right, even if they have negative consequences. Thus, in contrast to consequentialist theory, deontological ethics prioritizes the “right” over the “good.” The theory of deontological ethics is often traced to the work of Immanuel Kant, with his focus on the notions of good will and duty. For Kant, there is nothing that is unreservedly good in the world, apart from good will itself (Kant, 1785/2002). Good will implies acting out of respect for moral law in service to moral duty. A good person is therefore committed to taking moral considerations as ultimate reasons to guide their behavior.

To act morally requires that one deliberately act for the sake of moral duty itself (Kant, 1785/2002). For Kant, a moral duty is one that satisfies the test of a *categorical imperative*—an obligation that is not conditional on a given circumstance, but universal across all contexts. Categorical imperatives are derived from pure practical reason (Kant, 1785/2002), which determines what ought to be done without reference to situational factors, the individuals involved, or their subjective desires. The rational person would recognize categorical imperatives and realize that they are bound by them. Thus, to act against a categorical imperative is immoral *and* irrational (Kerstein, 2002).

Kant offered two essential formulations for the categorical imperative. The first states that one should only act if one could will that the principle, or maxim, underlying the action would become a universal law, without creating a logical or practical contradiction (Kant, 1785/2002; Korsgaard, 1985). Logical contradiction means that the maxim’s universalization

would render the proposed action inconceivable. Practical contradiction means that the maxim’s universalization would render the proposed action less effective (Korsgaard, 1985).

For example, the principle of lying in order to get something one wants is not *logically* warranted, because its universalization implies a world where everyone lies to get what they want and everyone knows that everyone else lies. In such a world, truthfulness does not exist and no one expects to be told the truth. This would logically contradict the maxim of lying to get what one wants, since it makes no sense in a world without the assumption of truthfulness. It is therefore strictly immoral. Similarly, committing an act of violence for the purpose of gaining some advantage would lead to a *practical* contradiction when universalized. Because the advantage gained through the act of violence would make oneself a target of violence, the principle would thus reduce or eliminate the advantage obtained through the original act.

Kant’s second formulation of the categorical imperative states that we should never treat the humanity of others as a means to our own ends, but always as an end in itself (Kant, 1785/2002). This requires that we not interact with them exclusively for our own benefit, without consideration of their needs and interests. We are self-governed—able to set our own ends and make free decisions based on our rational wills. This imbues us with absolute moral worth, such that we should not be manipulated or manipulate others for our own benefit (Kant, 1785/2002). The humanity formulation means that engaging in lying or deception is morally wrong, because a deceived person cannot make autonomous decisions about how to act since this decision would be based on false information (Paton, 1971).

Kant’s approach addresses some of the main concerns voiced by critics of utilitarianism, which permit all acts as long as they maximize happiness, even if they seem objectionable. Kant’s ethical theory sets strict

guidelines that prohibit engaging in such acts. However, the deontological approach is critiqued on other grounds. Hegel famously noted that, in setting up a conflict between reason and human desire, the Kantian perspective undermines an individual's motivation to act ethically (Geiger, 2007). Additionally, the professed rational and objective basis for the categorical imperative may be dogmatic and restrictive, leading to the denial of the moral worth of individuals' subjective motivations (Campbell & Christopher, 1996). These motivations, and the way they relate to one's character, are the focus of a third ethical perspective—virtue ethics.

3.3 Virtue Ethics

Of the ethical theories explored here, *virtue ethics* has the oldest pedigree, being traced to the works of Plato and Aristotle (Hursthouse, 1999). Aristotle, in particular, highlighted the role of personal character traits—i.e., virtues—in determining the ethical nature of individuals and their actions (Aristotle, ca. 340 B.C.E./1987). Unlike utilitarian and deontological ethics, virtue ethics focus on the virtuous agent rather than on right actions or what anyone should do in a given situation. Aristotle's starting point was to ask what makes life worth living and, in answering this question, to construct an account of how a moral person should live. His answer was that one must live a life characterized by *eudaimonia*, a state of profound fulfillment that comes from achieving one's ultimate purpose, or *telos* (Hughes et al., 2003). The *telos* is a final end toward which individual actions are performed and from which they derive meaning (MacIntyre, 2007). It must be chosen for its own sake, rather than as a means to some other end. For example, the *telos* of an architect is to design structures, of a composer to create music, of a scientist to generate knowledge. It is with respect to *telos* that human traits emerge as virtues (MacIntyre, 2007).

Aristotle's virtue ethics are different from the other two ethical theories in that they do not offer concrete moral guidelines. While virtue ethics state a general requirement to act virtuously, they do not describe virtue as a matter of following rules. Rather, to live a virtuous life, one should engage in a process of cultivating excellence in character, which will naturally lead to moral action (Fowers, 2003) through the development of “practical wisdom” (*phronesis*)—the capacity to recognize and respond to morally salient elements in any given situation. Doing the right thing cannot be determined in a formulaic fashion by adhering to predetermined moral principles; rather, it is a matter for the individual actor to identify right actions within a particular circumstance. Therefore, “practical wisdom includes a moral sensibility that highlights what is at stake in a given situation for the ends that we seek” (Fowers, 2003, p. 417). This

practical wisdom, acquired through habituation and repetition, is the consequence of training that involves performing virtuous acts (MacIntyre, 1998). To be virtuous, therefore, is not situational, but a life choice: a life dedicated to understanding the virtues and developing one's character. This process involves socialization into a community; interacting with and observing others who possess practical wisdom, learning the moral code of a community's collective life, and developing the capacity for virtue (Tsoukas, 2018).

For an act to be rightly called virtuous, three conditions must be met (Aristotle, ca. 340 B.C.E./1987). First, the act must originate from an agent and be undertaken in accordance with the agent's wants or beliefs. It cannot be coerced. Second, the agent must not act for the wrong reasons, such as out of ignorance or to avoid punishment. If someone acts honestly because they are afraid that they will be caught lying, it is not a virtuous act. Third, the agent must have sufficient relevant knowledge concerning the act. If we ask the agent why they acted in a certain way, they can articulate an honest answer that allows us to appreciate why they believed the action was right. These conditions highlight the centrality of volition in virtue ethics. Virtues can only manifest in voluntary actions because we cannot reasonably assign praise or blame to actions not willfully undertaken (MacIntyre, 2013).

The role of agency in virtue ethics raises the question of whether this perspective can be applied beyond the behavior of individual ethical actors. While the traditional approach to virtue ethics focuses on individuals, a number of researchers (e.g., Beggs, 2003; Chatterjee et al., 2015; Chun, 2005; Fricker, 2010) have advanced a collectivist perspective on virtue ethics, whereby “some standardly individual attributes have collective counterparts ... similar enough to play the same role at both the collective and individual levels” (Cordell, 2017, p. 44). Collectivist theorists argue that this perspective is regularly reflected in natural language when we attribute virtues or vices to a group (e.g., “The research team is tenacious”; Fricker, 2010). They further note that the virtues of a group may differ from those of any individual members, defying a summative perspective that merely “adds up” the virtue of all members. Importantly, this argument remains a contentious philosophical principle, and we make no attempt to adjudicate the debate here. Rather, in our subsequent analysis, we will endeavor to consider the focal case from both individualist and collectivist perspectives.

As with consequentialist and deontological theories, virtue ethics face criticisms from those advocating for other perspectives. From the perspective of *action-based* consequentialist and deontological schools, the *agent-based* approach (Klein, 1989) of virtue ethics provides limited guidance for concrete action in the

face of an ethical quandary. Similarly, the “situationist critique” of virtue ethics argues that the theory fails to account for the idiosyncrasies of a situation that can frame the options available to an actor, regardless of their virtuous traits (Sreenivasan, 2013). Virtue ethics have also been critiqued for the vagueness that often surrounds descriptions of virtues themselves and how they can be assessed.

This summary of the “big three” ethical theories highlights some of the key differences between them with respect to foundational assumptions, core principles, and areas of focus in assessing the ethicality of human behavior. Before applying each theory to an illustrative ADMS case, it is worth underscoring why we focus on these three theories, in particular. We concur with earlier scholars within the IS community, most notably Mingers and Walsham (2010) and Chatterjee and Sarker (2013), who have argued at length that the big three frameworks incorporate the vast breadth of ethical reasoning with regard to organizational environments and that a variety of applied ethical perspectives in the IS domain, including disclosive computer ethics (Brey, 2000; Introna, 2005), information ethics (Floridi, 1999), and the ethics of care (Adam, 2001; Gilligan, 1993), have their principle foundations in one or more of the big three theories. Apropos of our present focus, there is also a rich body of research on AI ethics (Coeckelbergh, 2020; Dignum, 2019), yet this research has tended to focus on the articulation of design principles and practical guidelines for the ethical development of AI resources (Hagendorff, 2020; Jobin et al., 2019), with little exploration of the theoretical foundations upon which such normative arguments are constructed. Indeed, in a systematic analysis of AI ethics literature, Jobin et al. (2019) found that the leading ethical principles invoked within the domain, including transparency, fairness, non-maleficence, privacy, and responsibility, remain undefined in the relevant works, much less grounded in broader ethical theory. In light of these observations, and in keeping with both Mingers and Walsham (2010) and Chatterjee and Sarker (2013), we determined that the big three ethical frameworks provide the most well-established and inclusive foundation for a comparative theoretical analysis of ADMS. Next, we describe the illustrative case to which we apply the three ethical theories.

4 Illustrative Case—Curbing Gambling Through ADMS¹

Gambling is widespread in Australia. More than 80% of the country’s adult population gambles—the highest rate in the world.² Opportunities for gambling are never far away, as Australia has more poker machines per person than any other country—nearly 200,000 machines or roughly one for every 114 people. Advertisements encouraging people to gamble are common on primetime television and the options are varied—e.g., horse/dog racing, electronic gaming, casinos, lotteries, and sports. From 2016 to 2017 Australians bet more than \$208b.³ Averaged across the adult population, this is more than \$11,000 per person. Nationally, total losses come to almost \$24b, placing Australia first globally in gambling losses per capita.⁴ Gambling is particularly common among males, people aged 50 and older, those with limited schooling, retirees, and those who draw their main source of income from welfare payments.⁵ Regular gamblers can suffer not only financial disaster and bankruptcy, but divorce, family violence, mental and physical illness, crime, drug and alcohol abuse, and self-harm. The estimated annual social cost associated with gambling in Australia is over \$7b. Recognizing gambling behavior as a public health issue, the Australian Medical Association advises general practitioners to conduct screening for problem gambling as part of a lifestyle risk assessment.

Banks have faced public criticism for their role in enabling gambling behavior. The four major Australian banks, accounting for almost 75% of all lending in the country, allow most gambling transactions on credit cards. In addition, banks regularly provide credit cards, personal loans, and overdrafts to known gamblers (Swanton et al., 2019), with several prominent cases featured in mainstream media.⁶ Against this backdrop, one of the four major banks recently implemented a gambling-reduction program—a variety of measures aimed at curbing gambling behavior among a small group of “problem gamblers.” To identify “problem gamblers,” the bank developed a machine-learning algorithm that was trained on datasets (acquired from the Federal Department of Social Services and the Australian Bureau of Statistics) concerning known gamblers’

¹ The case is based on actual events. The details provided about the scope of gambling in Australia are factual and the description of the bank’s initiative is elaborated from a real intervention by an Australian bank.

² <http://news.bbc.co.uk/2/hi/asia-pacific/6313083.stm>

³ <https://www.qgso.qld.gov.au/issues/2646/australian-gambling-statistics-34th-edn-1991-92-2016-17-summary-tables.pdf>

⁴ <https://www.casino.org/gambling-statistics/>

⁵ <https://aifs.gov.au/agrc/publications/gambling-activity-australia/2-gambling-participation>

⁶ <https://www.abc.net.au/news/2018-03-22/commonwealth-bank-offered-gambler-credit-card-limit-increases/9577654>

behavior, including their financial activity, alcohol consumption, and socioeconomic and demographic information. With these datasets, the algorithm learned to identify problem gamblers and was subsequently applied to the bank's customer database.

Customers algorithmically identified as problematic are subjected to a range of interventions aimed at mitigating their gambling behavior. Using a mix of location data, available cash balances, and previous cash withdrawal histories, the bank creates "friction" in any decision to spend or withdraw money that may be used for gambling. This friction comes in a variety of "hard" and "soft" measures. Hard measures include daily withdrawal limits to cap spending at a level that only allows people to cover basic needs, canceling credit cards, and rejecting loan applications. Soft measures include using the bank's mobile app to promote products that the bank considers to be safer and more conservative (e.g., low-interest credit cards and loans) than those promoted to other customers. To access less-conservative products, problem gamblers must proactively engage the bank and specify the products that they wish to acquire. The bank also sends text messages to problematic customers reminding them of the adverse effects of gambling, citing various real-life examples of individuals who suffered from the effects of chronic gambling. This is done when customers are located in or around a gambling venue. Additionally, the bank partnered with mobile advertising platforms, alcohol producers, and the Ministry of Health to ensure that problematic customers are not targeted with ads for alcoholic drinks and gambling and instead receive public service announcements on the detrimental effects of drinking and gambling.

After implementing these measures for a period of six months, the bank declared them a success. In an internal statement, bank executives reported a 40% average decline in gambling among targeted customers, reflecting thousands of dollars in annual savings per customer. A follow-up survey found that many customers reported improved mental/physical health, including reduced stress, better sleep, and more quality time with family. Following the initial rollout of the program, the bank made it available to its entire customer base on an opt-in basis by promoting it and creating sign-in options on its website and mobile app. In promoting the program, the bank emphasized its positive outcomes but did not specify the measures employed to ensure effectiveness. Early indications show that the bank's program is similarly beneficial to self-identified problem gamblers who volunteer to participate in the initiative.

With the expansion of the program came public scrutiny from social advocacy groups and public policy institutions. Some of the concerns centered on the fact that a minority of targeted customers continued to gamble regularly, despite the bank's efforts. Some of these customers reported increased levels of stress because their disposable income shrank due to the restrictions imposed. Other concerns were raised about the invasion of customers' privacy caused by the bank's data collection activities and tracking of its customers' locations without their consent. In addition, an independent inquiry found instances of misidentification of "problematic customers" by the bank's algorithm. Several customers were wrongly identified, because of high correlations in the data between socioeconomic status and gambling behavior. Thus, the algorithm identified several nongambling customers from low socioeconomic backgrounds as gamblers. Finally, several observers worried that the program limits participants' freedom of choice. If customers are not aware that they are in the program, they do not know their mix of ads has been deliberately altered to change their behavior.

5 An Ethical Analysis of the Anti-Gambling ADMS

Next, we consider the ethical nature of the bank's initiative from the viewpoint of the big three ethical theories. As a prefatory point, we note that the bank's ADMS initiative itself reflects a preliminary judgment regarding the ethical evaluation of gambling and the desirability of institutional action to curb it. That is, the mere suggestion of the program reflects an implied determination that gambling (or at least excessive gambling) is ethically problematic and should be prevented. Of course, this perspective itself could be subject to significant debate and analysis through examination of the appropriate relationship between the state, private sector institutions, and the individual. While we intentionally limit our focus to the ethical evaluation of the ADMS initiative itself, the case provides an illustration of the inherent nested-ness of ethical analyses.

In the current analysis, we apply the big three ethical theories to examine the use of ADMS by the bank. In doing so, we allow the language and assumptions of each theory to inform our analysis of the case, the different questions raised, and the ethical dynamics at play. These elements are summarized in Table 2.

Table 2. Application of the Ethical Theories to the Case Study

	Utilitarian ethics	Deontological ethics	Virtue ethics
Root question	Does the bank's initiative increase overall well-being?	Is the bank's initiative consistent with its moral duties?	Does the bank's initiative enhance customers' capacity to exhibit and develop virtue? Does the bank's initiative reflect a virtuous collective character?
Analytical approach	Identify relevant stakeholders. Identify elements in the bank's initiative that affect stakeholders. Evaluate the likely effects of initiative on stakeholders' well-being.	Clarify the principle underlying the bank's initiative. Determine if the principle results in logical/practical contradictions when universalized. Determine if the principle leads to treating others as a mere means to an end.	Evaluate the effect of the bank's initiative on its customers' ability to (1) develop practical wisdom and (2) engage in voluntary action. Assess the degree to which the initiative is consistent with and helps realize the bank's <i>telos</i> . Assess whether the bank's actions meet the conditions for virtuous action.
Conclusion	In aggregate, the initiative enhanced customers' well-being and is therefore ethical	The initiative violated the bank's moral obligations, failed the test of the two formulations, and is therefore unethical.	At the individual level, the initiative compromised the ability of customers to develop practical wisdom and act voluntarily and is therefore unethical. At the collectivist level, the ethical standing of the initiative is mixed.

5.1 A Utilitarian Analysis of the Bank's Actions

An analysis of the case from a utilitarian perspective focuses on the consequences of the bank's initiatives and whether they increased overall well-being. To do so requires that we: (1) Identify the stakeholders impacted by the initiatives, (2) identify the elements of the initiative that affect the well-being of impacted stakeholders, and (3) evaluate the likely effect of these elements on the identified stakeholders' well-being to reach a conclusion regarding the ethicality of the gambling-reduction program.

We define the relevant stakeholders within the scope of our analysis as those individuals that stand to be significantly and directly impacted by the bank's initiatives. The case describes gambling in Australia as a serious public health issue with broad societal impacts that can potentially cut across entire families, communities, and the population as a whole. For the present assessment, we restrict our focus to the bank's customers and their immediate family members. In this regard, we focus on two specific groups of customers—those who were identified by the bank's algorithm to be part of its gambling-reduction program and those who opted to participate in the program during its second wave of implementation. While a wider associated network of individuals (and, indeed, society as a whole) may be impacted by the bank's initiative, such impacts are likely to be milder and experienced indirectly and over a longer time period

compared to those felt by the customers and their family members. Accordingly, we kept individuals without close contact with the focal customers outside the scope of our analysis.

Next, we identify elements in the bank's initiatives that are likely to have an impact on the well-being of the identified stakeholders. Several elements should be taken into consideration, including hard and soft friction-inducing measures. In addition, the bank's collection of customer data from external sources can be potentially detrimental to individuals and their families (Zuboff, 2015) when it is done without their consent and when it violates their privacy. Finally, the misclassification of some customers as problem gamblers can be damaging when information is used to inform and motivate further actions.

The third step in the analysis involves examining the impact of the banks' measures on the well-being of customers and their families. Importantly, while some aspects of the ADMS initiative seem unpalatable to liberal notions of individual agency and free choice, a utilitarian examination elides debates on the inherent moral nature of acts, focusing strictly on outcomes. In utilitarian ethics, the only valid evaluative measure of the ethicality of actions is their outcomes; no action is *inherently* immoral and actions are ethical if they enhance well-being.

In this respect, the illustrative case provides clear evidence that the cumulative effect of the actions taken by the bank is decidedly positive. Specifically, the case

indicates that most customers benefited not only financially from the bank's interventions (e.g., 40% decline in gambling behavior, thousands in annual savings per household), but also experienced improved physical and mental health (e.g., reduced stress, better sleep, and more quality time with family). Moreover, early indications are that the later opt-in customers experienced similar benefits. In both of these conditions, the positive outcomes achieved also benefit the family members of the focal customers. While there was a small group of customers whose well-being decreased as a result of the initiative, this negative effect was eclipsed by the benefits gained by most customers and their families. Therefore, we conclude that, from a utilitarian standpoint, the bank's actions are ethically justified.

5.2 A Deontological Analysis of the Bank's Actions

In applying a deontological lens, we are interested in the inherent ethical quality of the bank's actions, rather than their outcomes. Specifically, we ask whether the bank's actions are consistent with its moral duties and the universality and humanity formulations of the categorical imperative. The first formulation requires that extending the principles underlying our actions to a universal law does not result in a logical or practical contradiction. To examine whether this formulation holds in the bank's case, we must articulate the proposed maxim, which we state as follows: To curb gambling activities among suspected problem gamblers, banks will apply hard and soft measures in their engagement with their customers to impede gambling behavior. In this scenario, all banks would utilize algorithmic analyses to identify problem gamblers and change their interactions with these customers to reduce gambling behaviors. When we consider this scenario, we conclude that it may result in a practical contradiction. That is, the universalization of the maxim may reduce its effectiveness.

In a world where all banks publicly implement measures to alter the behavior of their customers, and where people have access to multiple sources of information through the internet, it is reasonable to assume that customers would come to expect such interventions and be able to identify them and consequently counter their effect. This is particularly the case for the soft measures used by the banks. For instance, when specific products are shown by default through the bank's mobile app to an unsuspecting customer, they may not know that other, less conservative options are available and therefore never ask for them (Gill, 2018). However, if customers are aware of the efforts to change their behavior, they realize that other product options exist and will seek to have their default offerings changed to gain access to less conservative products. This, in turn,

may thwart the bank's efforts to encourage more risk-averse behavior and render its gambling-reduction initiative less effective.

Universalizing the initiative could also lead to a practical contradiction due to the bank's use of text reminders. As described above, the bank sent text messages to identified gamblers to remind them of the adverse effects of gambling. However, doing so may have the unintended consequence of *increasing* gambling behaviors among these customers for a couple of reasons. First, people may get irritated by being sent unsolicited notifications and react by engaging in a contrary act to the one they were encouraged to engage in (Damgaard & Gravert, 2018). Second, when the reminders point out how many people engage in the same harmful behavior, they may inadvertently normalize the behavior and ultimately promote it (Bicchieri & Dimant, 2022). Thus, when extended across all banks, the unintended consequences of the initiative can reduce its effectiveness.

The bank's initiative also violates the humanity formulation of the categorical imperative, particularly in relation to those customers that did not choose to participate. The humanity formulation requires that we recognize others as fully realized beings, acknowledge their wills, and remain honest in our interactions with them to allow them to make decisions based on truthful information. Furthermore, this formulation stipulates that we never treat others as a means to achieving ends, but as ends in themselves. While the initiative was well-intentioned, the bank did not consult its customers, ascertain what changes (if any) they wished to make in their lives, or secure their consent to participate. The bank effectively coerced its customers into participating, robbing them of the ability to pursue their own ends.

Another way in which the bank violated the humanity formulation is by using a machine-learning algorithm to construct data-based profiles of its customers. From the Kantian perspective, this practice is ethically problematic because it equates a set of data points with the phenomenon they are intended to describe and ignores the interpretative context within which this data is generated and becomes meaningful (Tsoukas, 1997). Customer profiles generated in this way obscure the richness, complexity, and ambiguity that are inherent to the lived experience of the bank's customers (Constantiou & Kallinikos, 2015). Therefore, in tailoring its interactions with customers based on their datafied profiles, the bank did not acknowledge their subjective motivations and interpretations of their actions, and instead imposed its own algorithmically generated views on its customers. Based on this analysis, we conclude that the bank violated its moral obligations and failed the test of the two formulations. Therefore, from a deontological viewpoint, the bank's actions are unethical.

5.3 A Virtue Ethics Analysis of the Bank's Actions

Our analysis of the bank's initiative from a virtue ethics perspective considers both an individualist and a collectivist perspective. From the individualist perspective, we ask how the bank's actions influenced the ability of its customers to exhibit and cultivate virtue in the context of their relationship with the bank. First, let us consider what makes a virtuous customer. Such a customer must exercise good judgment, self-control, honesty, prudence, and temperance (Aristotle, ca. 340 B.C.E./1987; MacIntyre, 1998). The pursuit and expression of these qualities demand that the virtuous customer make considered choices when selecting products and services that reflect the customer's actual needs and means. The virtuous customer is forthcoming in dealing with the bank and does not try to conceal any relevant information or seek loopholes in its provision of services to gain an unfair advantage. The virtuous customer also shows restraint and avoids rash decisions and uninformed choices. In other words, the virtuous customer demonstrates farsightedness, patience, and a capacity for deliberation that contribute to the customer's flourishing (Vallor, 2016).

How, then, do the bank's actions influence the ability of its customers to be virtuous? One aspect of the bank's initiative could have a particular impact on its customers' capacity to cultivate their virtue and develop practical wisdom—the use of nudging tactics, exemplified in the bank changing the mix of ads presented to customers and altering their personalized product offers. *Nudging* refers to the covert redesign of people's choice architecture in order to manipulate them into acting in ways they would not have otherwise chosen (Thaler & Sunstein, 2008). In IS design, nudging may take the form of default settings that prod users toward predefined courses of action. Indeed, this is precisely what the bank did in restructuring the product offerings and altering the advertising mix to identify problem-gamblers.

Nudging has an adverse effect on an individual's ability to pursue virtue and develop practical wisdom. It entails automating decision-making and delegating the moral responsibility involved in this process to algorithms. For a virtue ethicist, this is morally questionable for two reasons. First, it manipulatively imposes a single, external view of the “right” action. Second, it deprives people of the opportunity to assess the moral significance of their choices and their inherent messiness and ambiguity, thereby curtailing their capacity for reflection, moral growth, and practical wisdom (Morozov, 2013). When the bank nudges, its customers are unable to assess the moral significance of choosing more, or less, risky products and applying good judgment, self-control, prudence, or

temperance. In fact, customers cannot do the “wrong” thing, because they are steered toward what the bank deems desirable behavior. This can result in morally deficient customers who may not do the right thing unless the “wrong” option is eliminated.

The bank's nudging tactics also conflict with the principle that virtue requires volition. Voluntary acts must originate from the agent and reflect their freely developed motivations. Furthermore, to act voluntarily an agent must have sufficient knowledge about the act and its context. The bank's nudging tactics are at odds with both conditions. The bank deliberately and covertly restructured its customers' informational environment to elicit predictable responses whose expected outcomes reflect goals chosen by the bank rather than its customers. Therefore, when customers chose certain products or responded to an ad, their acts unwittingly reflected the bank's interests and not their own. Moreover, since the bank's customers (esp., those who did *not* opt-in) were unaware that their informational environment was altered, they did not have full and reasoned recognition of the circumstances within which they acted. Thus, from an individualist virtue ethics perspective, we conclude that the initiative compromised customers' ability to develop practical wisdom and act voluntarily and is therefore unethical.

Considering the case from a collectivist virtue ethics perspective, we ask what the purpose (*telos*) of a bank is in order to determine whether it is supported by the ADMS initiative. No shortage of claims can be made regarding a bank's *telos*, including enabling the transmission of value within a community (Caldararo, 2013), promoting economic development (Boushall, 1940), functioning as an intermediary between those who have money and those who need it (Davis & Taylor, 1978), screening and monitoring borrowers (Mayer, 2015), and simply making a profit (Fuller, 2015). The purpose that one settles on will have substantive ramifications for the assessment of the virtue of the ADMS initiative. The initiative is consistent with some claimed purposes (e.g., monitoring borrowers) and markedly inconsistent with others (e.g., enabling transmission of value or making a profit). As such, the determination of fit with the bank's *telos* is mixed.

Additionally, as with individual customers, we might consider the degree to which a bank can exercise such virtues as good judgment, self-control, and honesty (MacIntyre, 1998). While one might claim that the ADMS effort was indicative of good judgment and self-control (e.g., forgoing profit in favor of the perceived well-being of the customer), the implementation of nudging behaviors and the lack of transparency with selected customers certainly calls into question the bank's honesty in the ADMS scenario.

It is similarly difficult to determine if the ADMS effort has met the three virtue ethics conditions for voluntary action. *Was the effort uncoerced?* While the bank may be reacting to negative publicity for enabling gambling behavior, it is not clear that this reflects coercion or that the bank is acting in a way inconsistent with its own “wants and beliefs.” *Was the effort undertaken for the right reasons?* The bank ostensibly introduced the ADMS out of concern for the well-being of their customers, but it is difficult to discern if other motivations, such as the anticipation of regulatory action or the pursuit of favorable press coverage, are also relevant. *Did the bank possess sufficient relevant knowledge regarding the effort?* Here at least, we can anticipate that the bank would be capable of offering a thorough justification for the ADMS effort and explicit reasoning behind the various measures taken. Based on the assessment of these three criteria, we can determine that the bank’s effort is likely voluntary. Overall, the evaluation of the ADMS initiative from a collectivist virtue ethics perspective renders a mixed assessment.

6 Pursuing Theoretically Informed Ethical Analyses of ADMS

Our analysis reveals that conclusions regarding the ethicality of the bank’s initiative vary depending on the ethical theory informing the evaluation. The assumptions undergirding the theories translate into different assessment criteria, which, in turn, result in different conclusions about the appropriateness of the bank’s activities. A utilitarian viewpoint suggests that the initiative was ethical because its overall effect was to enhance the well-being of a majority of the primary stakeholders. Conversely, a deontological analysis indicates that the bank’s program was unethical because it failed on both formulations of the categorical imperative. A virtue ethics analysis offers a more mixed determination. At the individual level, virtue ethics suggests that nudging tactics interfere with customers’ ability to develop practical wisdom and engage in voluntary acts—prerequisites for the cultivation of virtue. Yet a collectivist application of virtue ethics renders a “split decision,” with the determination depending upon the purported *telos* of the bank and its motivations for undertaking the ADMS initiative.

Beyond these general evaluations of the bank’s initiative, our analysis highlights some ethically salient characteristics of ADMS: ADMS are *data-driven*, *autonomous*, *complex*, and *scalable*. Although overlapping and interdependent, each of these characteristics poses unique ethical challenges. While

these characteristics have been previously identified in the literature for their ethical significance (e.g., Bigman & Gray, 2020; Mökander et al., 2021; Wachter, 2019), our multi-lens ethical approach can inform a nuanced examination of each characteristic to help researchers address them in a comprehensive manner (see Table 3).

Data-driven: Algorithms embedded in ADMS can only viably function when they are trained on large datasets. Our case details how the algorithm that the bank used to identify problem gamblers leveraged large amounts of data acquired from external sources. Widely noted ethical implications of using big data for algorithmic analysis include the violation of privacy⁷ (Crawford, 2021), behavioral profiling,⁸ and the transmission of embedded bias (Leicht-Deobald et al., 2019). This last challenge is particularly prominent when training datasets contain a disproportionate representation of the population groups to which the resulting algorithm is applied, or when focal variables are highly, but spuriously, correlated with background variables. In such cases, outcomes can be biased against people on the basis of misplaced classifications. This was evident in our case where individuals from a lower socioeconomic background were erroneously identified by the bank’s algorithm as problem gamblers. This is a common issue with ADMS because algorithms are often trained on data that is inexpensive and readily available, even if it is not comprehensive and representative of the broader population.⁹ The ethical implications of the data-driven nature of ADMS are likely to vary based on the ethical theory applied. That is, different ethical theories will engender different questions, concerns, or phenomena of interest

From a *utilitarian perspective*, researchers and practitioners may ask whether ADMS data inputs are sufficient to support a comprehensive assessment of their likely outcomes: Is limited or low-quality data employed by ADMS sufficient to support a valid assessment of their utility? Does the data used by ADMS capture the various population groups to which it is applied, enabling a thorough assessment of how they are likely to be affected by the system? From a *deontological perspective*, researchers may examine whether bias embedded in ADMS, due to partial or low-quality data, can induce the treatment of individuals as a mere means or lead to categorical or practical contradictions. To illustrate the latter contradiction, ADMS have been used in the American legal system to provide ostensibly nonbiased recommendations for bail-setting, sentencing, and parole.

⁷ <https://reut.rs/3u88Z8c>

⁸ <https://cleanupgambling.com/news/cracked-labs>

⁹ <https://bit.ly/3IPPwNH>

Table 3. A Multi-Perspective Evaluation of the Ethical Issues Associated with ADMS Characteristics

ADMS characteristics	Salient ethical issues	Questions for consideration in a theoretically informed ethical analysis of issues		
		Utilitarian ethics	Deontological ethics	Virtue ethics
Data-driven: The value and applicability of models are conditional on the volume and quality of the input data	Violation of privacy; Behavioral profiling; Transmission of embedded bias	Are data inputs sufficient to support a comprehensive assessment of likely outcomes? Do the data inputs used capture the various population groups to which the ADMS is applied?	Does bias embedded in data induce the treatment of individuals as mere means or lead to categorical or practical contradictions?	Does encroachment on people's privacy undermine individual volition?
Autonomy: Degree of independence from direct human intervention and control	Moral status of ADMS; Reduced accountability	Is the autonomy and adaptation of ADMS oriented toward collective benefits?	Do reduced accountability and lack of recourse violate the categorical imperative by treating individuals as mere means to the ends of the system?	Do autonomous systems compromise individual agency?
Complexity: Diversity of algorithmic elements including computational parameters and hidden neural layers	Opacity (lack of transparency)	Does opacity reduce the ability to assess collective benefits and harm?	Can relevant maxims be formulated if emergent decision rules are opaque? Is it possible to determine what is "right"?	Does opacity undermine the ability of people to understand why decisions about them were made and what acts they can take to change these decisions?
Scalability: Relative ease of application to new domains, questions, or topics	Universality of utility; Datafication	Can assessments of collective good translate across domains?	Does treating individuals based on the data traces they leave across different domains risk reducing their humanity?	Does scaling of functionality across domains impact the voluntary nature of action?

Despite its intention, evidence suggests that rather than reducing bias, such ADMS use reinforces bias against African Americans (Kirkpatrick, 2016), thereby rendering the proposed action less effective. From a *virtue ethics* perspective, researchers can examine to what degree, and by which means, expansive data-collection practices encroach on people's privacy (Manheim & Kaplan, 2019), thereby limiting people's ability to act voluntarily. For instance, Google's search rankings are personalized for each individual user based on the extensive data Google collects about them. Research has shown that Google's search rankings are biased for topics with multiple competing perspectives, such as political candidates (Kulshrestha et al., 2017), and that such bias can shift the voting preferences of undecided voters without their awareness (Epstein & Robertson, 2015). The question therefore arises of whether such implicit influence undermines people's ability to act without external coercion.

Autonomy: ADMS have the capacity to learn, respond to changes in their environment, and make choices in unstructured and dynamic situations without direct human intervention (Baird & Maruping, 2021). This characteristic was demonstrated in our case when the bank's ADMS automatically sent text messages to customers based on their identification as problem gamblers and location. In light of the autonomous nature of ADMS, some scholars claim that ADMS should have a significant moral status (Danaher, 2020) and that, like living animals, they should be legally protected from abuse (Darling, 2016). Others have gone a step further and proposed that they should be granted a "status of electronic persons responsible for making good any damage they may cause" (Delvaux, 2017, p. 18). The autonomy of ADMS also raises questions about who should be held accountable for the recommendations and decisions they make. One

approach is that ADMS satisfy the core conditions of accountability because of their agentic nature. Another approach is that ADMS cannot be said to be aware of what they are doing (Coeckelbergh, 2020) and therefore the responsibility for their consequences should be placed elsewhere, for example with their designers (Martin, 2019).

The autonomy of ADMS can be assessed from different theoretical ethical perspectives. Researchers grounded in *utilitarian ethics* may be concerned with the degree to which the autonomous adaptation of ADMS is driven by collective rather than individual measures of benefit. Presumably, all algorithmic adaptation is oriented toward maximizing certain outcomes, but such outcomes could reflect either broad (e.g., reducing societal losses from gambling) or narrow (e.g., maximizing profits of an individual bank) benefits. Thus, determining whether the autonomy of a given ADMS resource engenders “the greatest good for the greatest number” is critical in a utilitarian lens. A *deontological perspective* would likely orient researchers to questions of ADMS accountability. In its absence, and with limited or no recourse, what protections do people have against being treated as mere means to the ends of the ADMS? For instance, did a debt-collection algorithm used by the Australian government violate the humanity formulation by enrolling citizens into a scheme to which they could not, in principle, consent?¹⁰ Finally, *virtue ethics* can guide researchers to ask how individual agency is challenged by autonomous systems. For example, is people’s ability to freely shape their own lives compromised when they use self-driving vehicles that decide what route to take, trading bots that automatically buy and sell stocks, or internet browsers that automatically complete their text inputs?

Complexity: A third critical characteristic of ADMS is their essential complexity, flowing from the dynamic structures of neural networks and other algorithmic approaches (Mökander et al., 2021). One of the central ethical issues raised by such complexity is the opacity of the resulting decision processes (Pasquale, 2015; Rahman, 2021). Human actors interacting with an ADMS often have limited insight into the data processing executed by the system and the criteria on which decisions are based, making it impossible to assess the validity of a system’s determinations (Busuioc, 2021). Indeed, the customers in our case were unaware of the data collected about them, or how the bank’s ADMS processed the data to determine whether they were problem gamblers.

From a *utilitarian perspective*, ADMS opacity poses concerns about the calculus of collective benefit or harm employed within a system. Since human actors cannot “see” the criteria being employed, it becomes difficult to assess potential side effects beyond the focal or ostensible purpose of the ADMS. For instance, if people are unaware that their ride-sharing app only offers rebates to riders that reside in certain zip codes, they may be unable to accurately assess whether it maximizes utility across its user base.¹¹ Employing a *deontological lens*, one might consider whether a maxim for universal application can even be articulated in the absence of known decision rules. In other words, is it possible to determine what is “right”? If we cannot ascertain the criteria employed by an ADMS, how can we assess if their application is appropriate in all like circumstances? Finally, a *virtue ethics* approach might explore whether opacity undermines the ability of individuals to understand why decisions about them were made and consider acts that could facilitate different decisions. For example, can users of Twitter or YouTube truly understand why the algorithmically generated content they are shown is presented to them and what they can do to see content that genuinely reflects their interests or goals?

Scalability: Decisions, assessments, and classifications made by an ADMS in one context can be seamlessly reused to address other questions or problems by an ADMS in a different or broader domain. For example, many online retailers create detailed behavioral profiles of their users to classify them based on their value to the company or specific consumption patterns. Such evaluations can be easily and latently shared across platforms. A recent report showed that one user’s interactions with an online gambling platform were shared with 44 third-party companies.¹² Similarly, the ADMS in our case was trained on data the bank acquired from external domains: The Federal Department of Social Services and the Australian Bureau of Statistics.

Some salient ethical issues raised by scalability are the universality of utility and datafication, which involves treating individuals based on their data traces (Gal et al, 2020). From a *utilitarian perspective*, scalability raises questions about the calculus of utility across domains. Do assessments of collective benefit translate as algorithms are applied to novel or unanticipated domains? For instance, a margin of error of 0.1% may be considered appropriate for a text-recognition system when used to mark student essays, but is it satisfactory

¹⁰ <https://bit.ly/3M7DVM3>

¹¹ This assessment may be equally difficult for other stakeholders such as regulators, policy makers, or municipalities.

¹² <https://bit.ly/3BfTz2U>

when applied to detect online terrorist communications? A *deontological perspective* might ask whether people's humanity is reduced when they are recast as amalgamations of seemingly objective data points in a digitally constructed universe in the service of corporate or state interests (e.g., to increase advertising revenue¹³). A *virtue ethics* perspective might focus on how datafication enabled by scalability impacts people's ability to pursue their freely chosen goals. For instance, how much volition does a person have if one unknowingly receives differential treatment (offered different policies at different prices) from one's insurance company based on one's financial and behavioral profile, which the company sourced from an online broker, which, in turn, received the data from an online gambling platform that the person occasionally uses?

7 Contextual Drivers of Theory Application

As we have argued, the evaluation of the ethical significance of the characteristics of ADMS can suggest different points of focus, depending on the theory applied. We further observe that different domains of ADMS utilization may affect the salience of each of the theoretical perspectives. Since we do not consider any ethical theory to be inherently best-suited to examine ADMS, it is reasonable to ask in what contexts the application of a given theory is likely to be fruitful. Sandvig et al. (2016) note that in the practical ethical reasoning of daily life, "people typically proceed using a hodgepodge of all three approaches" (p. 4981). How then might we assess which frameworks are appropriate to a given aspect of ADMS creation, implementation, or use? Following Marabelli et al. (2021), we suggest an approach to this question based on the choices required at distinct levels of societal action—individual use, organizational adoption, system design, and public policy. These are summarized in Table 4.

At the most granular level of choice, *individual users* may ask how and to what degree they choose to use ADMS in their own lives—e.g., to inform their professional work, lifestyle choices, or interpersonal interactions. While any of the ethical frameworks could support valuable reasoning, individual use choices are the level at which a virtue ethics perspective is most clearly relevant. Individuals may contemplate the degree to which a given ADMS supports or undermines their pursuit of virtue. Can such systems help them to be more judicious, honest, prudential, or temperate (e.g., by highlighting behaviors that are conducive to well-being) and

achieve *eudaimonia*? Conversely, could ADMS support or undermine personal virtue by overriding one's reasoning about virtuous choice? In light of the salience of these types of questions, we argue that the context of individual use choice is one in which a virtue ethics perspective is particularly valuable.

Organizational adoption is a perennial focus for IS research, and it is certainly relevant in studying ADMS impacts. In ethical analyses, organizational research has long explored the duty (fiduciary or otherwise) that organizations have to customers and employees (Eva et al., 2020; Solomon, 1993), particularly in ways that defy a utilitarian analysis (Wiener, 1982). Indeed, this premise underlies the rich stream of research on corporate social responsibility (CSR) (Bowen, 2013; van Marrewijk, 2003). Building on these foundations and our own analysis, we argue that choices of organizational adoption of ADMS are particularly amenable to deontological analyses. The factors surrounding the selection and implementation of ADMS technologies are rife with questions regarding how to ensure that customers and employees are treated as ends in themselves and not mere means to broader organizational objectives. Similarly, the determination of what stakeholder rights are inviolable is critical in the assessment of novel ADMS-based opportunities.

While the deontological perspective appears central for choices of organizational adoption, our analysis suggests that a collectivist virtue ethics approach may also be applicable in this context. Exploring the degree to which a given ADMS enables an organization to pursue its self-identified purpose or demonstrate collective virtues could render an insightful assessment of the technology's appropriateness. Similarly, it would be valuable to evaluate such an initiative with an eye to its impetus (absence of coercion), motivations (reasons), and depth of technical understanding (requisite knowledge).

Design choices have been a central focus in much of the extant research on ADMS. Indeed, designers are key in such ethical reasoning, because they have a significant influence on how ADMS will function. The level of design choice is one at which the deontological perspective may be particularly salient. Since a deontological lens focuses on moral duties to which ADMS should adhere, it can provide practical guidance for system designers (Sandvig et al., 2016): Are there particular principles that ADMS designers should ensure that their designs embody, like the famous "three laws" of robotics (Pasquale, 2020)? To treat others as ends-in-themselves, what ADMS design choices might be required?

¹³ <https://bit.ly/3tbwoU5>

Table 4. Levels of Societal Action and Applicable Ethical Theories

Domain of choice	Relevant ethical theories and questions
Individual use	<i>Individualist virtue ethics</i> : Do ADMS support individuals' pursuit of virtue and help them achieve eudaimonia?
Organizational adoption	<i>Deontology</i> : Can an organization implement ADMS in a way that ensures its customers are treated as ends in themselves? What rights of organizational stakeholders must remain inviolate? <i>Collectivist virtue ethics</i> : Does an ADMS initiative enable an organization to pursue its self-identified purpose or demonstrate collective virtues?
Design	<i>Deontology</i> : Are there general principles that should be reflected in ADMS design, like the famous "three laws" in the domain of robotics? What design choices might be required to ensure users are treated as ends in themselves?
Public policy	<i>Utilitarian</i> : How can ADMS be broadly employed to enhance public well-being? How can policy makers adjust market regulation, taxation, funding, and incentive system to ensure societal welfare from ADMS is maximized?

Finally, ADMS raise the prospect of significant *public policy* choices. In just the past three years, a plethora of research has highlighted the public policy implications of emerging AI capabilities, addressing issues such as market regulation, taxation, R&D funding, and incentive structuring (e.g., Goolsbee, 2019; Misra et al., 2020; Naudé & Dimitri, 2020). As with most public policy research (Hahn, 1982; Mulgan, 2014), this work has a largely utilitarian orientation. This is perhaps unsurprising, because public policy, with its focus on optimizing societal-level outcomes, reflects a natural affinity to pursuing the greatest good for the greatest number. With regard to the ethical analysis of public policy choices around ADMS, we expect this utilitarian perspective to hold.

As this discussion of the contextual drivers of theory application suggests, the determination of which ethical theory (or theories) to employ in a given study is largely contingent on the nature of the choices under analysis. Far from arguing that one particular ethical theory is inherently superior, our analysis underscores the very conditional nature of that selection. Importantly though, that conditionality does not reduce the value of clarity and decisiveness in establishing the theoretical grounding for one's analytic efforts.

8 Benefits of Theoretical Clarity in ADMS Research

In the two preceding sections, we demonstrated how ethical theories can enable a nuanced examination of ADMS characteristics and considered the application of ethical theories in different ADMS domains. Next, we outline four benefits that the IS community can derive from theoretical clarity in the ethical analysis of ADMS.

First, theoretical clarity is required to develop a body of knowledge about ADMS and avoid making unsubstantiated ethical evaluations. As we have noted,

much of the research on the ethics of ADMS has been atheoretical (Stahl, 2012). However, to build a cumulative body of knowledge on the ethical implications of ADMS, conceptual clarity in our definition and grounding of ethics is critical. Simply stating that a certain ADMS, or its effects, are unethical because they erode privacy (Newell & Marabelli, 2015), engender bias (Leicht-Deobald et al., 2019), or are inscrutable (Martin, 2019), may appeal to our common sense, but is not sufficient for establishing a line of scientific research. By turning to theory as a guide, researchers can establish solid grounds on which to reason that the erosion of privacy, the presence of bias, or ADMS inscrutability are unethical. In doing so, researchers would lay out the principles or assumptions upon which these conclusions rest. The requirement for a basis in ethical theory is aligned with acceptable standards in other strands of IS research. For example, we expect scholars who write about IT acceptance, organizational agility, or institutional logics to provide concrete definitions of these terms and couch them in the relevant literature. Research on ethics can and should reflect the same level of theoretical rigor.

Being explicit about the theoretical bases of our ethical evaluations can also enable discussions in which ethical values emerge as questions open to examination rather than conclusions. For example, rather than accepting that privacy is an important ethical value (e.g., Giermendl et al., 2022; Marjanovic et al., 2021) and proposing ways to uphold it by limiting data collection or creating a culture of integrity (Culnan & Williams, 2009), we could employ ethical theories to ask: Is privacy desirable? If so, for what reasons? In what situations and under what conditions? As our analysis demonstrates, consequentialism, deontology, and virtue ethics can provide a theoretical scaffolding on which to construct answers to these questions.

Second, drawing on clearly articulated theoretical frameworks and well-defined concepts can help researchers avoid unjustified claims about the inherent ethical nature and effect of ADMS. Such claims uncritically accept that certain types of technologies, and their organizational and social consequences, are morally preferable to others. Importantly, we are not claiming that all technologies, organizational eventualities, and social arrangements stand on equal ethical ground. Rather, we maintain that in order to make any claims about the ethical significance of a technology, one should apply an ethical framework to provide explicit grounds for such claims.

Similarly, recourse to ethical theory can help researchers avoid conflating their own culturally situated moral evaluations with universal truths. For instance, much of the literature on the ethics of ADMS maintains that group and individual privacy (e.g., Fjeld et al., 2020; Mittelstadt et al., 2016) as well as algorithmic transparency and accountability (e.g., Diakopoulos, 2016; Marabelli et al., 2021) are values to be upheld. However, privacy and transparency are not universally shared ideals. A case in point is the Chinese social credit system, which both lacks transparency and, to Western eyes, could be seen as invading citizens' privacy. The system closely tracks citizens' conduct to ensure ongoing compliance and social control. This system has been widely criticized in the West and characterized as a form of post-panoptic digital surveillance (e.g., Hansen & Weiskopf, 2021). However, Kostka (2019) finds that the system enjoys high approval rates among Chinese citizens, with many viewing it as promoting societal clarity and honesty. Using ethical theory, researchers could make explicit the evaluative criteria underlying judgments of the moral worth of ADMS and demarcate the scope of their applicability—and thereby avoid falling into ethnocentric traps.

Third, and related to the risk of ethnocentric traps, developing a concrete theoretical basis for ethical analysis is important to avoid ethical discussions within and across academic, professional, and social communities that are characterized by crude and superficial ethical characterizations, engendering moral tribalism (Markowitz & Shariff, 2012). An innate receptiveness toward moral tribalism is built into the human mind. The *social intuitionist model* (Haidt, 2001) of moral reasoning suggests that a person's ethical assessment is often an automatic, intuitive process marked by "the sudden appearance in consciousness of a moral judgment, including an affective valence (good-bad, like-dislike), without any conscious awareness of having gone through steps of search, weighing evidence, or inferring a conclusion" (Haidt, 2001, p. 818). While deeper reasoning about ethical issues may occur subsequently, it often entails a search for arguments to support a preformed judgment (Haidt & Bjorklund, 2008).

Intuitive ethical reasoning can lead to self-reinforcing loops of moral polarization and tribalism, in which communities become attitudinally homogeneous (Markowitz & Shariff, 2012) and social fragmentation and ideological partisanship proliferate (Sunstein 2007). Mitigating these risks requires that we engage in ethical discussions where the criteria for making judgments are clearly articulated and open for scrutiny, such as with the explicit use of ethical theory. The anchoring of ethical explorations of ADMS in these theories can help strengthen our ability to engage in *reasoned judgment* (Haidt, 2001), drawing conclusions based on the force of logic—by engaging in thought experiments, extrapolating from first principles, or examining an ethical dilemma from multiple theoretical perspectives—to challenge our initial intuitions.

Fourth, the utilization of ethical theory can help to clarify issues of theoretical uniformity and moral relativism. To be sure, our emphasis on the "big three" ethical theories should not be misconstrued as a call for theoretical homogeneity. We wish to emphasize *not* that we should collectively agree on a unified set of moral principles upon which we can categorically determine the ethical nature of ADMS, but that we need to agree on the basis for a debate on the ethical nature of these technologies. The goal in accomplishing this is not a singular moral truth, but rather to develop a robust common denominator upon which we can engage in discussions about the moral significance of ADMS. This common denominator can accommodate different views and evaluative criteria.

What shall we do when different ethical theories lead to contrary conclusions? Is one theory preferable to the others? Is it possible to deduce moral truths from theory and apply them across situations in order to determine their moral significance? Our view is best described as *morally pluralist* (Baghrarian, 2004): There can coexist multiple ethical claims in any given situation and there may legitimately coexist simultaneous assessments of the same ADMS as ethical and unethical. However, this does not undercut the reality of moral truth: Just because there may be many right answers, it does not mean there is no difference between a right and wrong answer. While the application of multiple theories can accommodate multiple moral conclusions, it differs from complete *moral relativism* wherein there is no universally valid morality, only plural, locally valid moralities (Velleman, 2013). Our moral pluralism accepts that there exist diverse locally justified ethical values but avoids an "anything goes" relativism by claiming that these values are ultimately answerable to conditions for human prosperity and other overarching constraints such as the value of human life (Baghrarian & Carter, 2015). The value of explicitly applying different ethical theories is in the exercise of vetting a system

from different ethical standpoints to comprehensively account for the system's ethical significance. Thus, our intention is not to rank order ethical theories. Rather, by juxtaposing different theories, we wish to encourage scholars and practitioners to engage in a judicious process that involves a careful and varied consideration of the principles upon which their analyses are grounded. Of course, the theoretical basis for this process does not have to be restricted to the ethical frameworks we discuss here. Indeed, more work is needed to apply and develop ethical frameworks to examine the use of ADMS. This will foster a rich analytical vocabulary that can be applied to the ethical dilemmas raised by various utilizations of ADMS. As we have sought to illustrate in this analysis, the ethical frameworks we describe here are primary candidates for further adaptation, extension, and elaboration, but other ethical frameworks may be fruitfully employed.

9 Implications and Future Research

The IS discipline endeavors not only to develop scientific knowledge about information technology in organizations and society, but also to support managers and IS professionals in their work with IT, including ADMS. The rapid uptake of ADMS and their ethical aspects means that this is an area where academically based knowledge can be particularly valuable. This paper contributes a multi-lens ethical framework that can support a rigorous, thorough, and transparent analysis of the ethical precepts and ramifications associated with the analysis, design, implementation, and use of ADMS. This study's framework can serve as a tool that managers and IS professionals can use to ensure that ADMS are not only technically feasible but also ethically accounted for.

Indeed, we contend that *not* using an ethical framework could make it difficult to provide useful advice. If we, as researchers and teachers, are unable to say precisely what it is about a certain situation, technology, or outcome that makes it ethically contentious, how can we reliably offer recommendations for action to be implemented by practitioners? Simply stating that something is unethical without articulating the grounds for that conclusion does not provide even the well-intentioned practitioner with sufficient guidance to evaluate an essentially unlimited series of ethical dilemmas posed by emergent ADMS. By contrast, with the application and juxtapositions of ethical theories, clearly elaborated justifications for claims about the moral significance of ADMS can be used to develop evaluative and analytical frameworks that can be applied by practitioners in novel practical settings.

Our aim has been to demonstrate the value in conducting a multi-perspective ethical analysis. We sought to emphasize the contrast among the three ethical theories rather than examine in detail their multiple variants. Our application of the "big three" normative theories can therefore be considered instructive in the spirit of a "first pass." Second and additional passes in future research may "drill down" into the finer points of the three ethical theories, which are umbrella categories that designate diverse approaches. Within consequentialism, for example, one finds "act consequentialism" and "rule consequentialism" (Alexander & Moore, 2012; Sinnott-Armstrong, 2019); within deontology, "agent-centered deontological theories," "patient-centered deontological theories," and "contractarian deontological theories" (Alexander & Moore, 2012); and within virtue ethics, "Eudaimonist virtue ethics," "agent-based and exemplarist virtue ethics," "target-centered virtue ethics," and "Platonistic virtue ethics" (Hursthouse & Pettigrove, 2016). With variance in the theoretical perspective, one may expect variance in the judgment as to whether a given act, situation, or program is ethically justifiable. Just as the research in this paper has endeavored to show the value of conducting a multiperspective analysis across the big three ethical theories, future research could explore the value in conducting a multiperspective analysis *within* a given ethical theory.

Such an analysis could be conducted to investigate different technological applications, such as social media platforms, as well as manifestations of artificial intelligence in addition to ADMS, such as those used in self-driving vehicles, which have been the subject of much ethical discussion (e.g., Awad et al., 2018). IS research applying ethical theory may also fruitfully investigate a wide range of empirical settings; for instance, many private sector organizations other than banks heavily invest in and utilize algorithmic technologies to manage their workforce (Gal et al., 2020). Similarly, governments use algorithmic systems across a range of services, including welfare and benefits (Carney, 2018), public health and safety (Gupta et al., 2020), and law enforcement (Meijer & Wessels, 2019). In being conducted across different technologies and settings, such IS research may serve to test and reveal the limits of any or all of the three ethical theories, thereby supporting further development or adaptation of the theories.

Future ethnographic or action design research could be particularly fruitful. First, ethnographic research mindful of the local culture could provide essential input to the consideration of what is considered ethical. In the example already mentioned regarding the Chinese social credit system, an ethnographic perspective could prevent the Western ethnocentric judgment that the close tracking of citizens' conduct is

an instance of post-panoptic digital surveillance. Clearly, what people themselves value as acceptable is an input to what may be considered ethical or unethical, and ethnographic and other interpretive perspectives can bring out local stakeholders' meanings and values. Second, action design research (ADR) involving the design and implementation of an IT artifact (Sein et al., 2011) represents a natural avenue for research on the ethics of ADMS. According to Sein et al. (2011), the ADR method involves the inseparable "activities of building the IT artifact, intervening in the organization, and evaluating it concurrently" (p. 37). Thus, ADR would be well suited not only to examining how the details of the design and implementation of an ADMS are interwoven with ethical considerations but also to providing research access to the affected people so that an interpretive perspective on their values and meanings regarding ethics can be taken as well. Either an interpretive or ADR approach can be useful for uncovering not only the rich details of situations and processes underlying the ethical aspects of a system but also how stakeholders down the line could be negatively affected.

10 Conclusion

The ethical implications of ADMS are certainly many and varied. As our global societies and organizations wrestle with the diverse applications of these powerful tools, the IS research community is well positioned to guide a thoroughgoing exploration of their ethical facets and the grounds for the limitations on ADMS use that might be warranted. The integration and juxtaposition of multiple ethical frameworks can help us to ensure that such an analysis is theoretically grounded, clearly reasoned, and appropriately nuanced. As the illustrative analysis of the bank's ADMS initiative suggests, the ethical aspects of these emergent technologies may look very different from the perspectives of consequentialist, deontological, and virtue ethical theories. In our roles as both researchers and educators, advocating for structured analyses that clearly articulate foundational assumptions, principles, and the theoretical grounds for the conclusions drawn can foster a cumulative body of knowledge that avoids the very human tendencies toward intuitive reasoning, moral tribalism, and ethnocentric judgment.

References

- Adam, A. (2001). Computer ethics in a different voice. *Information and organization*, 11(4), 235-261.
- Alexander, L., & Moore, M. (2012). Deontological ethics. In E. N. Zalta (Ed.), *Stanford encyclopedia of philosophy*. <https://plato.stanford.edu/>.
- Ananny, M. (2016). Toward an ethics of algorithms: convening, observation, probability, and timeliness. *Science, Technology, & Human Values*, 41(1), 93-117.
- Aristotle. (1987). *The Nicomachean Ethics*. (J. E. C. Welldon, Trans.). Prometheus Books. (Original work published ca. 340 B.C.E.)
- Arnold, D. G., Audi, R., & Zwolinski, M. (2010). Recent work in ethical theory and its implications for business ethics. *Business Ethics Quarterly*, 20(4), 559-581.
- Awad, E., D'Souza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A. F., ... Rahwan, I. (2018). The moral machine experiment. *Nature*, 563(7729), 59-64.
- Baghrmian, M., & Carter, A. (2015). Relativism. In E. N. Zalta (Ed.), *Stanford encyclopedia of philosophy*. <https://plato.stanford.edu/>.
- Baghrmian, M. (2004). *Relativism*. Routledge.
- Baird, A., & Maruping, L. M. (2021). The next generation of research on IS use: A theoretical framework of delegation to and from agentic IS artifacts. *MIS Quarterly*, 45(1), 315-341.
- Banerjee, D., Cronan, T. P., & Jones, T. W. (1998). Modeling IT ethics: A study in situational ethics. *MIS Quarterly*, 22(1), 31-60.
- Beer, D. (2018). Envisioning the power of data analytics. *Information, Communication & Society*, 21(3), 465-479.
- Beggs, D. (2003). The idea of group moral virtue. *Journal of Social Philosophy*, 34(3), 457-474.
- Bélanger, F., & Crossler, R. E. (2011). Privacy in the digital age: A review of information privacy research in information systems. *MIS Quarterly*, 35(4), 1017-1041.
- Bench-Capon, T. J. M. (2020). Ethical approaches and autonomous systems. *Artificial Intelligence*, 281, 1-15.
- Berberich, N., & Diepold, K. (2018). *Virtuous machine: Old ethics for new technology?* Available at <https://arxiv.org/abs/1806.10322>
- Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly*, 45(3) 1433-1450.
- Bicchieri, C., & Dimant, E. (2022). Nudging with care: the risks and benefits of social information. *Public Choice*, 191, 443-464
- Bigman, Y. E., & Gray, K. (2020). Life and death decisions of autonomous vehicles. *Nature*, 579(7797), E1-E2.
- Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. In K. Frankish & W. M. Ramsey (Eds.), *The Cambridge handbook of artificial intelligence* (pp. 316-334). Cambridge University Press.
- Boushall, T. C. (1940). Banks must explore, not exploit. *Bankers' Magazine*, 140(5), 394.
- Bowen, H. R. (2013). *Social responsibilities of the businessman*. University of Iowa Press.
- Brännmark, J. (2009). Ethical theories and the transparency condition. *Ethical Theory and Moral Practice*, 12(5), 449.
- Brey, P. (2000). Disclosive computer ethics. *ACM SIGCAS Computers and Society*, 30(4), 10-16.
- Burns, J. H., & Hart, H. L. A. (Eds.) (1996). *The collected works of Jeremy Bentham: An introduction to the principles of morals and legislation*. Clarendon Press.
- Busuioc, M. (2021). Accountable artificial intelligence: Holding algorithms to account. *Public Administration Review*, 81(5), 825-836.
- Caldararo, N. (2013). The theory of banking: Why banks exist and why we fear them. *International Journal of Sociology and Anthropology*, 5(4), 116-132.
- Campbell, R. L., & Christopher, J. C. (1996). Moral development theory: A critique of its Kantian presuppositions. *Developmental Review*, 16(1), 1-47.
- Carney, T. (2018). The new digital future for welfare: Debts without legal proofs or moral authority? *UNSW Law Journal Forum*, 41(1), 1-16.
- Chatterjee, S., Moody, G., Lowry, P. B., Chakraborty, S., & Hardin, A. (2015). Strategic relevance of organizational virtues enabled by information technology in organizational innovation. *Journal of Management Information Systems*, 32(3), 158-196.
- Chatterjee, S., & Sarker, S. (2013). Infusing ethical considerations in knowledge management scholarship: Toward a research agenda. *Journal*

- of the Association for Information Systems, 14(8), 452-481.
- Chatterjee, S., Sarker, S., & Fuller, M. (2009a). Ethical information systems development: A Baumanian postmodernist perspective. *Journal of the Association for Information Systems*, 10(11), 787-815.
- Chatterjee, S., Sarker, S., & Fuller, M. A. (2009b). A deontological approach to designing ethical collaboration. *Journal of the Association for Information Systems*, 10(3), 138-169.
- Chen, Y., Deng, S., Kwak, D.-H., Elnoshokaty, A., & Wu, J. (2019). A multi-appeal model of persuasion for online petition success: A linguistic cue-based approach. *Journal of the Association for Information Systems*, 20(2), 105-131.
- Chun, R. (2005). Ethical character and virtue of organizations: An empirical assessment and strategic implications. *Journal of Business Ethics*, 57(3), 269-284.
- Clarke, R. (2016). Big data, big risks. *Information Systems Journal*, 26(1), 77-90.
- Coeckelbergh, M. (2020). *AI ethics*. MIT Press.
- Constantiou, I. D., & Kallinikos, J. (2015). New games, new rules: Big data and the changing context of strategy. *Journal of Information Technology*, 30(1), 44-57.
- Cordell, S. (2017). Group virtues: No great leap forward with collectivism. *Res Publica*, 23(1), 43-59.
- Couger, J. D. (1989). Preparing IS students to deal with ethical issues. *MIS Quarterly*, 13(2), 211-218.
- Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
- Culnan, M. J., & Williams, C. C. (2009). How ethics can enhance organizational privacy: lessons from the choicepoint and TJX data breaches. *MIS Quarterly*, 33(4), 673-687.
- Daft, R. L., & Weick, K. E. (1984). Toward a model of organizations as interpretation systems. *Academy of Management Review*, 9(2), 284-295.
- D'Agostino, M., & Durante, M. (2018). Introduction: The governance of algorithms. *Philosophy & Technology*, 31(4), 499-505.
- Damgaard, M. T., & Gravert, C. (2018). The hidden costs of nudging: Experimental evidence from reminders in fundraising. *Journal of Public Economics*, 157, 15-26.
- Danaher, J. (2020). Welcoming robots into the moral circle: A defence of ethical behaviourism. *Science and Engineering Ethics*, 26(4), 2023-2049.
- Darling, K. (2016). Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behavior towards robotic objects. In R. Calo, A. M. Froomkin, & I. Kerr (Eds.), *Robot law* (pp. 213-232). Edward Elgar Publishing.
- Davis, K. R., & Taylor, B. W. (1978). A Heuristic procedure for determining in-process inventories. *Decision Sciences*, 9(3), 452-466.
- Delvaux, M. (2017). *Report with recommendations to the commission on civil law rules on robotics* European Parliament (Report A8-0005/2017). https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html
- Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM*, 59(2), 56-62.
- Dignum, V. (2019). *Responsible artificial intelligence: How to develop and use AI in a responsible way*. Springer.
- Epstein, R., & Robertson, R. E. (2015). The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences*, 112(33), E4512-E4521.
- Eva, N., Newman, A., Miao, Q., Wang, D., & Cooper, B. (2020). Antecedents of duty orientation and follower work behavior: The interactive effects of perceived organizational support and ethical leadership. *Journal of Business Ethics*, 161(3), 627-639.
- Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. *Information and Organization*, 28(1), 62-70.
- Finnis, J. (1983). *Fundamentals of ethics*. Georgetown University Press.
- Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI* Berkman Klein Center for Internet & Society (No. Research Publication No. 2020-1). <https://cyber.harvard.edu/publication/2020/principled-ai>
- Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, 80(S1), 298-320.

- Floridi, L. (1999). Information ethics: On the philosophical foundation of computer ethics. *Ethics and Information Technology*, 1(1), 33-52.
- Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*. Oxford University Press.
- Foot, P. (1985). Utilitarianism and the virtues. *Mind*, 94(374), 196-209.
- Fowers, B. J. (2003). Reason and human finitude: In praise of practical wisdom. *American Behavioral Scientist*, 47(4), 415-426.
- Fricker, M. (2010). Can there be institutional virtues? In T. S. Gendler & J. Hawthorne (Eds.), *Oxford studies in epistemology* (Vol. 3, pp. 235-252). Oxford University Press.
- Fuller, E. W. (2015). The fractional reserve banking diagram. *Revista Procesos de Mercado*, 12(2) 81-104.
- Gal, U., Jensen, T. B., & Stein, M.-K. (2020). Breaking the vicious cycle of algorithmic management: A virtue ethics approach to people analytics. *Information and Organization*, 30(2), Article 100301.
- Garcia, J. L. A., & Nelson, M. T. (1994). The problem of endless joy: is infinite utility too much for utilitarianism? *Utilitas*, 6(2), 183-192.
- Garcia, M. (2016). Racist in the machine: The disturbing implications of algorithmic bias. *World Policy Journal*, 33(4), 111-117.
- Geiger, I. (2007). *The founding act of modern ethical life: Hegel's critique of Kant's moral and political philosophy*. Stanford University Press.
- Giermindl, L. M., Strich, F., Christ, O., Leicht-Deobald, U., & Redzepi, A. (2022). The dark sides of people analytics: reviewing the perils for organisations and employees. *European Journal of Information Systems*, 31(3), 410-435.
- Gill, D. (2018). When “nudging” is forever: The case of Sweden. *Proceedings of the 130th Annual Meeting of the American Economic Association* (pp. 153-158).
- Gilligan, C. (1993). *In a different voice: Psychological theory and women's development*. Harvard University Press.
- Goolsbee, A. (2019). Public policy in an AI economy. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), *The economics of artificial intelligence: an agenda* (pp. 309-316). University of Chicago Press.
- Govindarajulu, N. S., Bringsjord, S., Ghosh, R., & Sarathy, V. (2019). Govindarajulu, Naveen Sundar, et al. “Toward the engineering of virtuous machines.” *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*.
- Grant, P., Arjoon, S., & McGhee, P. (2017). Reconciling ethical theory and practice: Toward developing a business ethics pedagogical model. *Business and Professional Ethics Journal*, 36(1), 41-65.
- Günther, W. A., Mehrizi, M. H. R., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *Journal of Strategic Information Systems*, 26(3), 191-209.
- Gupta, P. K., Ramachandran, A. T., Keerthi, A. M., Dave, P. S., Giridhar, S., Kallapur, S. S., & Saikia, A. (2020). An overview of clinical decision support system (CDSS) as a computational tool and its applications in public health. In R. Kumar & S. Paiva (Eds.), *Applications in ubiquitous computing* (pp. 81-117). Springer.
- Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, 30(1), 99-120.
- Hahn, F. (1982). On some difficulties of the utilitarian economist. In A. Sen & B. A. Owen (Eds.), *Utilitarianism and beyond* (pp. 187-198). Cambridge University Press.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814-834.
- Haidt, J., & Bjorklund, F. (2008). Social intuitionists answer six questions about morality. In W. Sinnott-Armstrong (Ed.), *Moral psychology: The cognitive science of morality: intuition and diversity* (Vol. 2, pp. 181-217). MIT Press.
- Hansen, H. K., & Weiskopf, R. (2021). From universalizing transparency to the interplay of transparency matrices: Critical insights from the emerging social credit system in China. *Organization Studies*, 42(1), 109-128.
- Hughes, J. A., Sharrock, W., & Martin, P. J. (2003). *Understanding Classical sociology: Marx, Weber, Durkheim*. SAGE.
- Hursthouse, R. (1999). *On virtue ethics*. Oxford University Press.
- Hursthouse, R., & Pettigrove, G. (2016). Virtue ethics. In E. N. Zalta (Ed.), *Stanford encyclopedia of philosophy*. <https://plato.stanford.edu/>.

- Introna, L. D. (2005). Disclosive ethics and information technology: Disclosing facial recognition systems. *Ethics and Information Technology*, 7(2), 75-86.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature machine intelligence*, 1(9), 389-399.
- Kant, I. (2002). *Groundwork for the metaphysics of morals*. (A. W. Wood, Trans. and Ed.). Yale University Press. (Original work published 1785).
- Kerstein, S. J. (2002). *Kant's search for the supreme principle of morality*. Cambridge University Press.
- Kirkpatrick, K. (2016). Battling algorithmic bias: How do we ensure algorithms treat us fairly? *Communications of the ACM*, 59(10), 16-17.
- Klein, S. (1989). Platonic virtue theory and business ethics. *Business & Professional Ethics Journal*, 8(4), 59-82.
- Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: Review, synthesis, and future research directions. *European Journal of Information Systems*, 31(3), 388-409.
- Korsgaard, C. (1985). Kant's formula of universal law. *Pacific Philosophical Quarterly*, 66(1-2), 24-47.
- Kostka, G. (2019). China's social credit systems and public opinion: Explaining high levels of approval. *New Media & Society*, 21(7), 1565-1593.
- Kraemer, F., Van Overveld, K., & Peterson, M. (2011). Is there an ethics of algorithms? *Ethics and Information Technology*, 13(3), 251-260.
- Kulshrestha, J., Eslami, M., Messias, J., Zafar, M. B., Ghosh, S., Gummadi, K. P., & Karahalios, K. (2017). Quantifying search bias: Investigating sources of bias for political searches in social media. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 417-432).
- Laudon, K. C. (1995). Ethical concepts and information technology. *Communications of the ACM*, 38(12), 33-39.
- Leicht-Deobald, U., Busch, T., Schank, C., Weibel, A., Schafheitle, S., Wildhaber, I., & Kasper, G. (2019). The challenges of algorithm-based HR decision-making for personal integrity. *Journal of Business Ethics*, 160(2), 377-392.
- Leonard, L. N., & Cronan, T. P. (2001). Illegal, inappropriate, and unethical behavior in an information technology context: A study to explain influences. *Journal of the Association for Information Systems*, 1(1), 1-31.
- Liang, F., Das, V., Kostyuk, N., & Hussain, M. M. (2018). Constructing a data-driven society: China's social credit system as a state surveillance infrastructure. *Policy & Internet*, 10(4), 415-453.
- Lo Piano, S. (2020). Ethical principles in machine learning and artificial intelligence: Cases from the field and possible ways forward. *Humanities and Social Sciences Communications*, 7(9), 1-7.
- Macdonald, J. E., & Beck-Dudley, C. L. (1994). Are deontology and teleology mutually exclusive? *Journal of Business Ethics*, 13(8), 615-623.
- MacIntyre, A. (1998). *A short history of ethics: A history of moral philosophy from the Homeric age to the twentieth century*. University of Notre Dame Press.
- MacIntyre, A. (2013). *After virtue: A study in moral theory*. Bloomsbury.
- Manheim, K., & Kaplan, L. (2019). Artificial intelligence: Risks to privacy and democracy. *Yale Journal of Law & Technology*, 21, 106-159.
- Marabelli, M., Hansen, S., Newell, S., & Frigerio, C. (2017). The light and dark side of the black box: Sensor-based technology in the automotive industry. *Communications of the Association for Information Systems*, 40(1), 352-374.
- Marabelli, M., Newell, S., & Handunge, V. (2021). The lifecycle of algorithmic decision-making systems: Organizational choices and ethical challenges. *Journal of Strategic Information Systems*, 30(3), 101683.
- Marjanovic, O., Cecez-Kecmanovic, D., & Vidgen, R. (2021). Algorithmic pollution: Making the invisible visible. *Journal of Information Technology*, 36(4), 391-408.
- Markowitz, E. M., & Shariff, A. F. (2012). Climate change and moral judgement. *Nature Climate Change*, 2(4), 243-247.
- Markus, M. L. (2017). Datification, organizational strategy, and is research: what's the score? *Journal of Strategic Information Systems*, 26(3), 233-241.
- Martin, K. (2015). Ethical issues in the big data industry. *MIS Quarterly Executive*, 14(2), 67-85.

- Martin, K. (2019). Ethical implications and accountability of algorithms. *Journal of Business Ethics*, 160(4), 835-850.
- Mason, R. O. (1986). Four Ethical Issues of the Information Age. *MIS Quarterly*, 10(1), 5-12.
- Mayer, C. (2015). Big bang: New beginning or beginning of the end? *Oxford Review of Economic Policy*, 31(2), 186-198.
- Meijer, A., & Wessels, M. (2019). Predictive policing: Review of benefits and drawbacks. *International Journal of Public Administration*, 42(12), 1031-1039.
- Mejia, J., Mankad, S., & Gopal, A. (2019). A for effort? Using the crowd to identify moral hazard in New York City restaurant hygiene inspections. *Information Systems Research*, 30(4), 1363-1386.
- Mill, J. S. (1895). *Utilitarianism*. Longmans, Green & Company.
- Mingers, J., & Walsham, G. (2010). Toward ethical information systems: the contribution of discourse ethics. *MIS Quarterly*, 34(4), 833-854.
- Misra, S. K., Das, S., Gupta, S., & Sharma, S. K. (2020). Public policy and regulatory challenges of artificial intelligence (AI). *Proceedings of the International Working Conference on Transfer and Diffusion of IT* (pp. 100-111).
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 1-21.
- Mökander, J., Morley, J., Taddeo, M., & Floridi, L. (2021). Ethics-based auditing of automated decision-making systems: Nature, scope, and limitations. *Science and Engineering Ethics*, 27(4), 1-30.
- Morozov, E. (2013). *To save everything, click here: The folly of technological solutionism*. Public Affairs.
- Mulgan, T. (2014). *Understanding utilitarianism*. Routledge.
- Naudé, W., & Dimitri, N. (2020). The race for an artificial general intelligence: Implications for public policy. *AI & Society*, 35(2), 367-379.
- Newell, S., & Marabelli, M. (2015). Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of "datification." *Journal of Strategic Information Systems*, 24(1), 3-14.
- Neyland, D. (2019). *The everyday life of an algorithm*. Springer.
- O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books.
- Paradice, D., Freeman, D., Hao, J., Lee, J., & Hall, D. (2018). A review of ethical issue considerations in the information systems research literature. *Foundations and Trends in Information Systems*, 2(2), 117-236.
- Pasquale, F. (2015). *The black box society*. Harvard University Press.
- Pasquale, F. (2020). *New laws of robotics: Defending human expertise in the age of AI*. Harvard University Press.
- Paton, H. J. (1971). *The categorical imperative: A study in Kant's moral philosophy* (Vol. 1023). University of Pennsylvania Press.
- Portmore, D. W. (2001). Can an act-consequentialist theory be agent relative? *American Philosophical Quarterly*, 38(4), 363-377.
- Rahman, H. A. (2021). The invisible cage: Workers' reactivity to opaque algorithmic evaluations. *Administrative Science Quarterly*, 66(4), 945-988.
- Rai, A., Constantinides, P., & Sarker, S. (2019). Next generation digital platforms: Toward human-AI hybrids. *MIS Quarterly*, 43(1), iii-ix.
- Ransbotham, S., Fichman, R. G., Gopal, R., & Gupta, A. (2016). Special section introduction—Ubiquitous IT and digital vulnerabilities. *Information Systems Research*, 27(4), 834-847.
- Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2016). When the algorithm itself is a racist: Diagnosing ethical harm in the basic components of software. *International Journal of Communication*, 10, 4972-4990.
- Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action design research. *MIS Quarterly*, 35(1), 37-56.
- Shaw, J. (2006). Intentions and trolleys. *The Philosophical Quarterly*, 56(222), 63-83.
- Sidgwick, H. (1907). *The methods of ethics* (7th ed.). University of Chicago Press.
- Sinnott-Armstrong, W. (2019). Consequentialism. In E. N. Zalta (Ed.), *Stanford encyclopedia of philosophy*. <https://plato.stanford.edu/>.
- Siponen, M. T., & Iivari, J. (2006). IS security design theory framework and six approaches to the application of ISPs and guidelines. *Journal of*

- the Association for Information Systems*, 7(7), 445-472.
- Smith, H. J., & Hasnas, J. (1999). Ethics and information systems: the corporate domain. *MIS Quarterly*, 23(1), 109-127.
- Sojer, M., Alexy, O., Kleinknecht, S., & Henkel, J. (2014). Understanding the drivers of unethical programming behavior: The inappropriate reuse of internet-accessible code. *Journal of Management Information Systems*, 31(3), 287-325.
- Solomon, R. C. (1993). *Ethics and excellence: Cooperation and integrity in business*. Oxford University Press.
- Sreenivasan, G. (2013). The situationist critique of virtue ethics. In D. C. Russell (Ed.), *The Cambridge companion to virtue ethics* (pp. 290-314). Cambridge University Press.
- Stahl, B. C. (2012). Morality, ethics, and reflection: a categorization of normative IS research. *Journal of the Association for Information Systems*, 13(8), 636-656.
- Stahl, B. C., & Markus, M. L. (2021). Let's claim the authority to speak out on the ethics of smart information systems. *MIS Quarterly*, 45(1), 485-488.
- Stahl, B. C., & Wright, D. (2018). Ethics and privacy in AI and big data: Implementing responsible research and innovation. *IEEE Security & Privacy*, 16(3), 26-33.
- Sunstein, C. R. (2007). *Republic.com 2.0*. Princeton University Press.
- Swanton, T. B., Gainsbury, S. M., & Blaszczynski, A. (2019). The role of financial institutions in gambling. *International Gambling Studies*, 19(3), 377-398.
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin Books.
- Thornton, S. M., Pan, S., Erlien, S. M., & Gerdes, J. C. (2017). Incorporating ethical considerations into automated vehicle control. *IEEE Transactions on Intelligent Transportation Systems*, 18(6), 1429-1439.
- Tsoukas, H. (1997). The tyranny of light: The temptations and the paradoxes of the information society. *Futures*, 29(9), 827-843.
- Tsoukas, H. (2018). Strategy and virtue: Developing strategy-as-practice through virtue ethics. *Strategic Organization*, 16(3), 323-351.
- Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Press.
- Van Marrewijk, M. (2003). Concepts and definitions of CSR and corporate sustainability: Between agency and communion. *Journal of Business Ethics*, 44(2), 95-105.
- Velleman, J. D. (2013). *Foundations for moral relativism*. Open Book Publishers.
- Wachter, S. (2019). Data protection in the age of big data. *Nature Electronics*, 2(1), 6-7.
- Ward, K. (2018). Social networks, the 2016 US presidential election, and Kantian ethics: Applying the categorical imperative to Cambridge Analytica's behavioral microtargeting. *Journal of Media Ethics*, 33(3), 133-148.
- Wiener, M., Saunders, C., & Marabelli, M. (2020). Big-data business models: A critical literature review and multiperspective research framework. *Journal of Information Technology*, 35(1), 66-91.
- Wiener, Y. (1982). Commitment in organizations: A normative view. *Academy of management review*, 7(3), 418-428.
- Williams, B. (1973). A critique of utilitarianism. In J. J. C. Smart & B. Williams (Eds.), *Utilitarianism: For and against* (pp. 77-150). Cambridge University Press.
- Zerilli, J., Knott, A., Maclaurin, J., & Gavaghan, C. (2019). Transparency in algorithmic and human decision-making: Is there a double standard? *Philosophy & Technology*, 32(4), 661-683.
- Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1), 75-89.
- Zwitter, A. (2014). Big data ethics. *Big Data & Society*, 1(2), 1-6.

Appendix A: Literature Review for Research on the Ethics of ADMS in the IS Literature.

To scan the existing IS literature on the ethics of ADMS, we used the EBSCO and SCOPUS databases to search the following journals: *Information Systems Research*, *MIS Quarterly*, *Journal of the Association for Information Systems*, *Journal of Management Information Systems*, *Journal of Information Technology*, *Journal of Strategic Information Systems*, *Information Systems Journal*, *European Journal of Information Systems*, and *Information & Organization*.

Within each journal, we looked for the following word combinations in either a paper's title, abstract, or keywords:

ethic* AND algorithm*;

ethic* AND machine learning;

ethic* AND artificial intelligence;

ethic* AND automated decision making;

ethic* AND analytics;

ethic* AND datafication;

moral* AND algorithm*;

moral* AND machine learning;

moral* AND artificial intelligence;

moral* AND automated decision making;

moral* AND analytics;

moral* AND datafication

This search returned 16 papers, which we read in full to determine if they substantively examined the ethical aspects of ADMS. This resulted in excluding three papers from further analysis.¹⁴ We classified the remaining 13 papers according to Stahl's levels of normativity (2012) (see Section 2 for a detailed description of Stahl's framework): Moral intuition, explicit morality, ethical theory, and reflection and meta-ethics. Because the first two levels denote atheoretical engagement with ethical issues, we merged them to arrive at a three-tiered categorization system: atheoretical morality, ethical theory, and meta-ethics. We found that, of the 13 papers, 11 employed atheoretical morality and two ethical theory. None of the studies applied a meta-ethics perspective (Table A1).

Two examples may help to illustrate the classification process employed: Giermindl et al. (2022) list various ethical challenges emanating from the use of people analytics in organizations such as reductively treating employees as data objects or the adverse effects of algorithmic miscalculations on employees. However, because the authors do not draw on any ethical theory to justify these issues as ethically salient, we classified this paper within the atheoretical morality category. On the other hand, Gal et al. (2020) drew on virtue ethics theory to assess the ethical significance of three characteristics of people analytics—opacity, datafication, and nudging. Therefore, we classified this paper in the ethical theory category.

Table A1. A Classification of IS Studies on the Ethical Aspects of ADMS

Level of normativity	Paper	Summary
Atheoretical morality	Wiener et al. (2020). Big-data business models: A critical literature review and multiperspective research framework. (<i>Journal of Information Technology</i>)	The paper discusses the ethical aspects of big-data business models using Mason's PAPA framework (privacy, accuracy, property, and accessibility). However, the reason for embracing these values as ethically desirable is not explained or couched in any ethical theory.
	Baird & Maruping (2021). The next generation of research on IS use: A theoretical framework of delegation	The authors state that the ethical Implications of the use of agential IS artifacts need to be considered. Specifically, transparency is needed, especially when unintended

¹⁴ Markus (2017). Datification, Organizational strategy, and IS research: What's the score?; Mejia et al. (2019). A for effort? Using the crowd to identify moral hazard in New York city restaurant hygiene inspections; Chen et al. (2019). A multi-appeal model of persuasion for online petition success.

	to and from agentic IS artifacts. (<i>MIS Quarterly</i>).	consequences or ethically dubious outcomes occur. However, the importance of transparency is not couched in any ethical theory.
	Stahl & Markus (2021). Let's claim the authority to speak out on the ethics of smart IS. (<i>MIS Quarterly</i>)	The authors call on IS scholars to engage in research on the ethics of smart technologies. In doing so, they mention a range of ethical theories; however, they do not apply or substantively engage with them.
	Berente et al. (2021). Managing artificial intelligence. (<i>MIS Quarterly</i>)	In this editorial, the authors discuss current and future frontiers of AI management. They describe several ethical challenges (e.g., overautomation, fairness), but they do not suggest how addressing these challenges may be informed by ethical theory.
	Newell & Marabelli (2015). Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of "datification." (<i>Journal of Strategic Information Systems</i>).	The authors discuss discrimination, inequality, monitoring, and control as ethical concerns arising from the use of ADMS. However, they do not explain the reasoning for selecting these values as morally significant or couch their selection in ethical theory.
	Günther et al. (2017). Debating big data: A literature review on realizing value from big data. (<i>Journal of Strategic Information Systems</i>)	The authors describe risks to privacy and autonomy, and unjust classification as ethical concerns that arise from the application of Big Data and algorithms. However, they do not draw on ethical theory to explain why these concerns are ethically salient.
	Marabelli et al. (2021). The lifecycle of algorithmic decision-making systems: Organizational choices and ethical challenges (<i>Journal of Strategic Information Systems</i>)	The authors outline strategic choices that arise during the design, implementation and use of ADMS, including ethical considerations involving bias, surveillance, and control. The authors do not draw on ethical theory to explain the ethical significance of these issues or to inform strategies for addressing them.
	Giermindl et al. (2022). The dark sides of people analytics: Reviewing the perils for organisations and employees. (<i>European Journal of Information Systems</i>)	The authors outline the various risks and ethical concerns associated with the use of <i>people analytics</i> , such as surveillance, reductionism, and control. However, they do not refer to ethical theory to explain why these issues are morally significant.
	Faraj et al. (2018). Working and organizing in the age of the learning algorithm. (<i>Information & Organization</i>)	The authors discuss algorithmic technologies as artifacts that both reflect ethical standards and raise complex ethical issues. However, the discussion of various aspects of algorithms (e.g., black-boxed, reductionist) and their implications is not couched in any ethical theory.
	Ransbotham et al. (2016). Ubiquitous IT and digital vulnerabilities. (<i>Information Systems Research</i>)	The special issue editors outline an agenda for research on digital vulnerabilities, identifying one research area as algorithmic ethics and bias. However, they stay within societally or organizationally defined ethical problems, not mentioning any ethical theory.
	Clarke (2016). Big data, big risks. (<i>Information Systems Journal</i>)	The author outlines the risks associated with the growing prevalence of big data analytics and points to the moral responsibility of researchers and professionals in mitigating them. However, he does not draw on ethical theory to inform this discussion.
Ethical theory	Gal et al. (2020). Breaking the vicious cycle of algorithmic management: A virtue ethics approach to people analytics. (<i>Information & Organization</i>)	The authors draw on Aristotle's theory of virtue ethics to outline ethical challenges arising from the use of people analytics (opacity, datafication, nudging) and discuss their implications for people's capacity to pursue internal goods, acquire practical knowledge, and act voluntarily.
	Kordzadeh & Ghasemaghahi (2022). Algorithmic bias: Review, synthesis, and future research directions. (<i>European Journal of Information Systems</i>)	The paper reviews the literature on algorithmic bias to propose a theoretical model that captures the prominent concepts in this literature. The authors rely on various moral theories on justice and fairness to elaborate their position on algorithmic bias.

About the Authors

Uri Gal is a professor of business information systems at the University of Sydney Business School. His research focuses on the organizational and ethical aspects of digital technologies. He is particularly interested in the relationships between people and technology, and in the changes in the nature of work associated with the introduction of algorithmic technologies. Professor Gal received his PhD from Case Western Reserve University. His research has been published in journals such *Organization Science*, *Journal of the Association for Information Systems*, and the *European Journal of Information Systems*. Professor Gal is an associate editor at the *Information Systems Journal*.

Sean Hansen is a professor of management information systems and chair of the Department of MIS, Marketing, & Analytics with the Saunders College of Business at Rochester Institute of Technology (Rochester, New York, USA). He earned his PhD from the Weatherhead School of Management at Case Western Reserve University. His research interests include information systems development, health IT, requirements engineering, distributed cognition, and the ethics of IS design and use. His research has been published in the leading scholarly journals in the IS field, including *MIS Quarterly*, *Information Systems Research*, *Decision Sciences*, *Journal of the Association for Information Systems*, and others.

Allen S. Lee is professor emeritus of information systems at Virginia Commonwealth University. He served as editor-in-chief of *MIS Quarterly* and as a founding senior editor of *MIS Quarterly Executive*. His research program has involved showing not only how qualitative research can be done rigorously and scientifically, but also how quantitative research equally needs to live up to the requirements of science. He is a Fellow of the Association for Information Systems, a LEO Award recipient, and a member of the Circle of Compadres of the Information Systems Doctoral Students Association of the KPMG PhD Project.

Copyright © 2022 by the Association for Information Systems. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and full citation on the first page. Copyright for components of this work owned by others than the Association for Information Systems must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or fee. Request permission to publish from: AIS Administrative Office, P.O. Box 2712 Atlanta, GA, 30301-2712 Attn: Reprints, or via email from publications@aisnet.org.