The Responsible Adoption of (Highly) Automated Decision-Making Systems

Camille Grange
HEC Montreal
camille.grange@hec.ca

Alain Pinsonneault
McGill University
alain.pinsonneault@mcgill.ca

Abstract

The next-generation technological era will be marked by the prevalence of highly automated decision-making systems (ADMS), which promote technological autonomy at the expense of human agency. In this paper, we examine the role and importance of socio-ethical factors in the responsible adoption of ADMS by organizations. In doing so, we draw on the unique characteristics of ADMS and leverage the literature on social responsibility to conceptualize what a responsible adoption process and a responsible adoption decision involve. The resulting framework makes a much-needed connection between technology adoption and social responsibility and offers a progressive foundation to study ADMS adoption.

1. Introduction

A number of pundits have started to observe a role reversal whereby humans are becoming used and shaped by technologies. For example, Demetis and Lee [1] claim that "human agency is becoming subordinate to automatic executions…" (p. 930) and that "technology is overtaking not only human decisions and the context of their embeddedness, but also entire subsystems of society" (p. 932). We refer to this development as the rise of highly automated decision-making systems (ADMS). ADMS have several benefits that make their adoption by organizations attractive. They are consistent, efficient, scalable, and can manage a much greater level of complexity than humans can. Yet, they are fallible and can yield (intentionally or not) harmful consequences such as privacy violations, unwarranted surveillance, uninformed control, or unfair discrimination [2], [3]. They can also have second-hand effects, often detrimental, on those subject to their use by others. In sum, ADMS have become quite controversial especially given that the activities and contexts to which they are increasingly being applied involve socially sensitive situations that have high ethical content; for example, social ranking [4], crime prediction [5], [6], and bail, parole, and criminal sentencing [7].

Our goal in this paper is to develop a theoretical framework that emphasizes the socio-ethical factors surrounding the decision made by organizations to deploy an ADMS. This endeavour, which echoes contemporary callings like critical technological citizenship [8] and critical digital capitalism [9], as well as early writings on the diffusion of new technology e.g., [10], [11]–[13], is intended to move researchers' attention beyond the rather instrumentalist perspective espoused by existing information technology (IT) adoption theories [14], [15]. This literature focuses on the questions of why, when, and how prospective organizations and individuals adopt or intend to adopt an IT for their own use and benefits, without inquiring whether the adoption decision or the process leading to the decision is *socially responsible* [16], [17].

In developing the framework, we specify a new set of constructs to capture the gist of the phenomenon and we introduce the theoretical logic underlying their relationships. We follow a deductive approach whereby we develop a conceptualization of ADMS that considers their unique characteristics and we combine it with contextualized insights from social responsibility theory. The resulting theoretical framework aims to make a three-fold contribution: clarifying our understanding of what the responsible adoption of ADMS by organizations entails, offering a platform for future research on the topic, and opening the door to a paradigm change in the study of IT adoption that is likely to yield important new insights into how organisations engage with the decision to deploy (or to restrain from deploying) increasingly autonomous and value-laden technologies. In the remainder of this paper, we specify the theoretical boundaries of our inquiry, we present the theoretical framework, and we discuss the research's contributions and its limitations.

2. Theoretical Boundaries

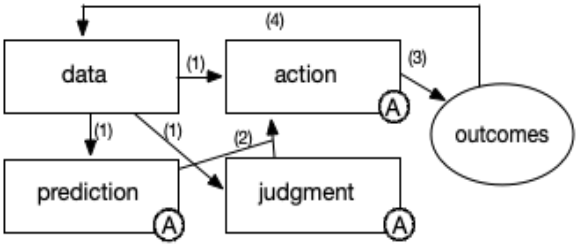
2.1. What is an ADMS?

We define ADMS as software products or software-enabled objects that bring a significant level of automation into decision making processes.

URI: https://hdl.handle.net/10125/71213
978-0-9981331-4-0

Importantly, very high levels of automation pave the way to technological autonomy, a situation where "humans find themselves outside, i.e., cast out to the environment, outside of these decisions, and (…) human agency—acting on behalf of another, or providing a particular service—is being replaced by technologized agency" [1], p. 933. As an example of the link between automation and autonomy, think of motor vehicles. The US Department of Transportation defines six stages of automation ranging from level 0 (no automation, i.e., the driver performs all the driving task) to level 5 (full automation, i.e., the vehicle is autonomous, driverless), with intermediary stages where the driver and the automated-driving system share control and responsibilities.

ADMS are different from decision-support systems (DSS). A DSS is an interactive system aimed at *helping* people (e.g., managers, consumers) make decisions (often characterized as ill-structured problems) [18]. In turn, an expert systems is a computer program (often based on artificial intelligence techniques) that aims to *mimic* experts at making complex, non-algorithmic decisions [19][20]. The goal of an expert systems can thus include automation, making it similar to an ADMS; yet as we will explain, the scope of automation in an ADMS can go much further than that of an expert system. Because the notions of decision-making and automation are complex and central to our conceptualization of ADMS, illustrated in Figure 1, we describe them in more details in the rest of this section.



Keys:
(1) Prediction needs data, and so do judgment and actions;
(2) A decision-making agent (human or machine) takes actions relying on both prediction and judgment;
(3) Actions lead to outcome (consequences)
(4) Outcomes generate data that can be used as an input to other/future decisions;
A: Prediction is automated, and so could judgment and action

Figure 1. Conceptual Representation of an ADMS (based on [21])

Decision making plays an equally critical role in everyday life [21] and in organizations [22]. Figure 1 highlights a fundamental assumption in our work: decision making requires applying *judgment* to *predictions* and then *acting* [21], all of which may be conducted by either a human and/or a machine. Prediction is an important input to decision-making "the process of filling in missing information… (taking) information you have, often called *data*, and (using) it to generate information you don't have," ([21], p. 24). Note that some decision-making situations can be formalized, thus automated, fairly easily. These do not rely on prediction and data, but on a set of pre-established rules accounting for known contingencies. Yet, in many complex real-life situations, it is virtually impossible to specify and formalize all the rules that would be necessary to faithfully capture a situation (e.g., hiring a new CEO, driving a vehicle in an urban area). Such situations require prediction, thus data. In sum, a*utomated prediction* is at the core of ADMS

Note that a predictive engine designed into an ADMS may also derive its own rules from data itself (i.e., by "learning")—this is the domain of machine learning (ML), a technology at the core of AI applications. AI is an umbrella term that is commonly used to refer to data-driven algorithms that enable computing engines to learn how to perform tasks which would be virtually impossible to automate based on formal rules (e.g., object identification, language processing). Similar to Demetis and Lee ([1], p. 944), we consider such applications as representative examples of extreme cases of sophisticated ADMS, where predictive models can recalibrate themselves automatically based on data. These applications require different data sets: a training data set is used to create the predictive model, an input data set enables running it, and a feedback data set serves as a basis to improve the model.

In the healthcare context an ADMS could automate *prediction* (e.g., estimate the most likely outcomes of applying treatments A or B based on X-rays, blood tests, and monitoring data) so as to inform a doctor who could then apply his or her *judgment* (e.g., considering other criteria such as patients' age, lifestyle, and sensitivity to potential risks and side effects) and *enact* the decision (e.g., administer treatment A). But as Figure 1 also stresses, prediction is not the only component of an ADMS that might be automated. As with motor vehicles, the trend is toward more extensive automation, thus full technological autonomy [1], which happens when a machine undertakes an entire task, not just the prediction component of it. For example, a fully autonomous vehicle automates *prediction* (e.g., anticipating vehicles' trajectory), *judgment* (e.g., evaluating whether the vehicle should hit what would appear to be a pedestrian or risk killing the driver), and *action* (e.g., braking or changing direction).

2.2 From Adoption to Responsible Adoption

What does adoption refer to in the context of the present research? Information Technology (IT) adoption (or acceptance) research offers a large variety of models explaining how, why, and when individuals or collectives decide to adopt a new technology. Some studies have focused on *users'* adoption decision e.g., [23] while others have examined on *organizations'* decisions e.g., [24]. Altogether, this research suggests that adoption is an information-intensive process wherein potential adopters consider a set of relevant factors (e.g., about the context into which they situate, the innovation per se, its consequences) so as to inform a decision about the commitment (of resources) to deploy a new IT. We follow this overall view of adoption, and we proceed by further asking what *responsible* adoption involves from the perspective of an organization.

In its contemporary usage in a business context, the concept of social responsibility (SR) is often associated with that of corporate social responsibility (CSR), and its application to different business functions such as marketing [26], [27], finance [28], and logistics [29]. The content covered by CSR has varied over the years. For example, in Carroll's [30] view, CSR touches upon economic, legal, ethical, and philanthropical aspects, while the European Commission [31] only points to social and environmental concerns and the triple bottom-line proponents consider economic, social, and environmental targets [25], [32]. Despite divergences in content, all views concur on the fact that CSR aims to promote the accountability of profit-seeking business entities in a market environment. Hence, SR is closely linked to CSR, but it applies more widely outside of a corporate context and in both individual and group capacities. Because ADMS adoption decisions are not restricted to private corporations, SR appears to be a more suitable lens to inform a working definition of responsible adoption. In its most generic portrayal, SR implies caring for societal values and for needs and effects that go beyond narrowly defined indicators and self-centered points of reference. On that basis, we define responsible adoption in terms of both the decision-making process and the decision that take into consideration the socio-ethical implications of putting an ADMS into use.

A couple of points in this definition need to be highlighted. First, similar to Van der Duin [33] in the context of responsible innovation, we choose a broad perspective that enables examining responsible adoption both as a means (the process) and as an end (the decision). Second, while the process and the decisions are our two focal units of analysis, some entity, at the end, makes a decision and is responsible for it. In our context, the entity or entities involved in the adoption decision can be viewed as those having moral agency. That is, any collective making a moral decision, where the later refers to a volitional decision that has a moral component given its possible harm on or benefit to others [34].

3. Framework Development

3.1. Responsible Adoption Decision (RAD)

We start by specifying three constructs to capture the extent of *responsibility* of an ADMS adoption decision. Altogether, these constructs imply that a RAD (i) takes into account the moral implications of the technology on the community concerned by it (*civic-mindedness*), (ii) is sensible to its impacts in the long haul (*foresightedness*), and (iii) is accessible and understandable by the community (*transparency*).

First, social responsibility theory emphasizes the idea that decision-makers have moral obligations. Indeed, the primary concern of SR involves the consideration of a socio-ethical dimension when acting and making decisions; in other words, it requires caring for others who might be impacted by one's actions and decisions [34]. This is especially relevant in the technological context in general [35]–[37] and in the ADMS context in particular, as the range of applications that ADMS enable cover highly socially-sensitive domains. For example, hiring, evaluating, monitoring, censoring, arresting, and judging people are all susceptible to being governed by ADMS, and this triggers complex risks such as the potential for systematic, large-scale discrimination [38] and for geopolitical instability [39]. Balancing the benefits and risks when it comes to making decisions and innovating is not a trivial endeavor, and for ADMS it might require weighting conflicting moral values such as security and privacy, or freedom and control [3] and considering a larger set of ethical principles [40], all of which cannot be done unobservant to situational value standards—we will discuss this further in section 3.2. Thus, we define as first dimension of RAD, ***civic-mindedness***, as the degree to which the adoption decision adheres to the values and moral expectations of the community that is relevant in the considered context.

Second, the more recent applications of SR to the technological innovation context emphasize the need to be prepared and care for the future, that is, make decisions that are sensitive to time and forward-looking. Technological changes are inherently uncertain, and their social consequences are powerful but also hard to predict beforehand [41], [42]. However, the difficulty to examine an uncertain future, with its multiple causal chains and dependencies, does not preclude from taking

decisions that are mindful of it [33], [43]. On the other hand, it may stress "our ability to act responsibly in the present on behalf of future generations" [44], p. 1880. ADMS adoption decisions that are responsible should be particularly sensitive to this need for anticipation because they have substantial societal ramifications that make adopters morally accountable to others now but also in the future [16], [17]. Thus, we define as second dimension of RAP, ***foresightedness***, as the degree to which the adoption decision is sensible to the long-run implications of using the ADMS on the community that is relevant in the considered context.

A third important notion that links social responsibility to ADMS adoption is transparency [45]–[47]. In general terms, transparency can be viewed as a "right to know", which is a principle that individuals have the right to know about the decisions (and possible ensuing risks) to which they may be subject [48]. In a technology context, transparency has two key components, accessibility and comprehensibility [49], and has sometimes been translated into regulation. For example, the new European General Data Protection Regulation ("GDPR") mentions that "the data subject shall have the right to [know about] the existence of automated decision-making." Accessibility and comprehensibility are particularly salient characteristics in the context of ADMS adoption because ADMS tend to be concealed and have indirect, second-hand effects. Thus, the socio-ethical quandaries that they produce may not be easily resolved, but at minimum, they could be open to scrutiny. As reminded by Florini ([48], p.viii), it is sometimes the case that "sunshine is the strongest antiseptic." Thus, we define as third dimension of RAP, ***transparency***, as the degree to which the adoption decision is inspectable and understandable by the community that is relevant in the considered context.

3.2. Responsible Adoption Process (RAP)

Because ADMS are quite unique, a responsible adoption *process* will need to account for their idiosyncratic characteristics. Given the absence of theory on ADMS, we specify RAP by two means. First, we draw on explicit premises about the differentiating characteristics of ADMS (i.e., they are opaque, value-laden, and can have side and systemic effects) to identify four constructs that capture the responsibility of an adoption process in the context of ADMS. Second, we specify (for each four constructs) how they would manifest with respect to each conceptual building block of an ADMS (data, prediction, judgment, and action – see Figure 1)

A starting premise is that the functioning of an ADMS can be quite opaque [50][46][45][50]–[52]. Burrel [50] explains that software code can generate three types of opacity. One (intentional secrecy) results from a decision to keep software code proprietary (e.g., the google search engine algorithm). Another (technical illiteracy) is a consequence of the specialized skills required to write (and read) code and of the fact that such skills are not widespread among the population. A third type (interpretability) is more specific to ADMS as it is a consequence of the complexification of algorithmic, often data-driven, systems and the resulting difficulty of "understanding the algorithms in action, operating on data" (p. 5). Despite some progress e.g., [53], [54], explaining the decisions generated by highly-sophisticated ADMS remains difficult, an issue sometimes known as the interpretability problem [55]. ADMS can be quite complex and more or less autonomous based on whether and how prediction, judgment, and action are automatized. To be able to evaluate the possibilities and risks associated with adopting an ADMS, a clear understanding of its functioning mechanisms and assumptions is needed. Thus, we define a first dimension of RAP, ***functional scrutiny***, as the degree to which the adoption process includes a thorough analysis of the key components and operational logic of a focal ADMS. A more specific coverage of functional scrutiny can be derived from the key components of an ADMS (see Figure 1) and is presented in Table 1.

Table 1. The Scope of Functional Scrutiny

DATA. Analyzing the data that is being used in the prediction: its origin, its quality (e.g., accuracy, reliability, completeness), and its role (e.g., training, feedback, operation) in the ADMS.
PREDICTION. Analyzing what is being predicted by the ADMS, what factors are used to derive the prediction as well as their weight in the prediction, and how accurate the result of the prediction is expected to be.
JUDGMENT. Analyzing the potential outcomes (what may happen as a result of a potential action), the value that is assigned to each outcome, their prioritization, and whether the ADMS automatizes judgment or leaves it to humans' responsibility.
ACTION. Analyzing the decision that is at stake, the possible actions associated with this decision, and whether the ADMS automatizes actions or leaves it to humans' responsibility.

A second premise is that ADMS embed moral values [3], [49], [52], [56]–[58]. The idea that information technologies are not value-neutral dates back several decades e.g., [59] and is at the core of a

number of fields of research such as value-sensitive design [35], [60]–[62] and engineering and computer ethics [37], [63], [64]. As Van den Hoven [17] puts it, our social world is "shaped by the algorithms that determine how far our messages reach into our networks, what is recommended to us on the basis of what the system has learned about our search history and preferences, what is filtered out and how our reputation is built." (p. 66) Although all technological design decisions may be viewed as morally loaded as they represent (intentionally or not) the value system of designers, the automatization and scalability of data-driven predictions make that statement much more salient in the ADMS context. In essence, ADMS rate and rank things, events, and also individuals [65], [66], and this can be particularly problematic when predictions are based on sensitive categorizations such as race or gender [67]. For O'Neil [38], such algorithmic intelligence is nothing else than "opinions embedded in math" (p. 19). To be able to examine what is at stake from a moral standpoint if an ADMS is to be deployed, a clear understanding of the nature and magnitude of the ethical challenges it involves is needed. Thus, we define a second dimension of RAP, ***ethical reflexivity***, as the degree to which the adoption process involves the careful deciphering of the moral values embedded in or promoted by a focal ADMS. The key components of an ADMS (see Figure 1) help further circumscribe the scope of ethical reflexivity (Table 2).

Table 2. The Scope of Ethical Reflexivity

DATA. Inspecting the ADMS in function of ethical considerations associated with the use of data (e.g., privacy, anonymity, security.)
PREDICTION. Inspecting the ADMS in function of ethical considerations associated with operating the predictive engine (e.g., transparency, fairness, justice.)
JUDGMENT. Inspecting the ADMS in function of ethical considerations associated with the making of judgments (e.g., autonomy, accountability.)
ACTION. Inspecting the ADMS in function of ethical considerations associated with the conduct of actions (e.g., control, accountability.)

A third premise is that ADMS have indirect and systemic effects [68]. By this we mean that a (sometimes) large range of people may be involved in operating the ADMS as well as be subject to its consequences. For example, the data used in a prediction engine may not belong to a direct user, and the outcomes may include effects on non-direct users, a phenomenon that Leidner et al. [69] refer to as second-hand effects and that Doorn and Van de Poel [37] characterize as indirect causation. The planned or emergent consequences of ADMS can thus ripple out to a wide range of individuals who may not be directly using the technology. Therefore, in order to examine the set of potential impacts of an ADMS, the individuals who are involved, directly or indirectly, in its operation need to be included in the evaluation process. In fact, the value of an inclusive approach and the need for all relevant agents to exert their influence in a technology-related decision-making process has long been known to facilitate the achievement better results for the relevant community as a whole [11]. Thus, we define a third dimension of RAP, ***stakeholder inclusiveness***, as the degree to which the adoption process involves the participation of a representative set of stakeholders relevant to the particular context of application of a focal ADMS. The key components of an ADMS (see Figure 1) can help specify the scope of stakeholder inclusiveness (Table 3).

Table 3. The Scope of Stakeholder Inclusiveness

DATA. Participation of those whose data is used in the ADMS and/or those who generate, collect, and own the data.
PREDICTION. Participation of those for whom the prediction is made or who are subject to its outcome, as well as those who are involved in designing the prediction engine
JUDGMENT. Participation of those who make the judgments or who are subject to its outcome, and if relevant, those involved in designing the rules underlying the automatized judgment.
ACTION. Participation of those who conduct the actions or who are subject to its outcome, and if relevant, those involved in designing the rules underlying the automatized action.

The fourth construct we propose to capture RAP also relates to the systemic nature of ADMS. To be able to execute the socially responsible practice of making an integrative assessment of the impacts of an ADMS on stakeholders with potentially different interests and concerns, prioritization and conflict resolution is likely needed [3]. This requires a holistic effort in unifying insights and points of view in light of the norms and moral standards relevant to a specific societal context into which the ADMS might be deployed. Technology

engineers are often confronted with moral dilemmas in their design work as they need to solve conflicting (value) requirements [70]. In doing so, they need to resolve difficult trade-offs (e.g., safety vs. efficiency, security vs. privacy, control vs. freedom). As Newell and Marabelli [3, p. 10] illustrate : "when algorithms determine that particular categories of people (e.g., based on race, income, job) are more likely to commit a crime and, as a result, those concerned find difficulty in obtaining a loan or changing job … this clearly violates basic privacy rights, but is justified based on the idea that it will increase security in society." Thus, we define a fourth dimension, ***integrative evaluation***, as the degree to which the adoption process involves a systemic evaluation of the consequences of deploying a focal ADMS in light of the moral standards acceptable in the considered context of application. Similar to the previous three constructs, the key components of an ADMS (see Figure 1) help specify the scope of integrative evaluation (Table 4).

Table 4. The Scope of Integrative Evaluation

DATA. Evaluating the possible consequences of deploying an ADMS in light of the moral dilemmas associated with its use of data.
PREDICTION. Evaluating the possible consequences of deploying an ADMS in light of the moral dilemmas associated with the design and operation of its predictive engine.
JUDGMENT. Evaluating the possible consequences of deploying an ADMS in light of the moral dilemmas associated with the judgment process.
ACTION. Evaluating the possible consequences of deploying an ADMS in light of the moral dilemmas associated with the actions resulting from combining prediction and judgment.

3.3. Relationships between the Components of RAP and RAD

In this section, we propose a set of causal relationships that we suspect to exist between the constructs developed to conceptualize RAP (links 1 to 6) and RAP (links 7 to 8), and those between RAP and RAD (links 9 to 11). They are modeled in the research framework illustrated in Figure 2.

RAP (links 1 to 6). Stakeholder inclusiveness, functional scrutiny, ethical reflexivity, and integrative evaluation define the degree of responsibility of an ADMS' adoption decision-making process. We start with the ultimate outcome in a responsible adoption process: integrative evaluation. The difficult integrative assessments that need to be made during an adoption assessment process, such as those involving the comparison and prioritization of different alternatives, are more effective when they are supported by quality informational inputs [71]. Thus. we expect that stakeholder inclusiveness, functional scrutiny, and ethical reflexivity will improve integrative evaluation, In line with this idea, we propose that integrative evaluation will benefit from functional scrutiny (***link 1***) because it will enable the consideration of more accurate inputs about what a focal ADMS can *afford* and *prevent*. Integrative evaluation will also benefit from ethical reflexivity (***link 2)*** because awareness and knowledge of what is at stake from an ethical perspective with respect to a focal ADMS will help generate more complete inputs to assess its possible impacts and "what is right" in a particular context. Stakeholder inclusiveness will enable integrative evaluation (***link 3***) because including a representative set of stakeholders in the decision-making process is likely to yield inputs that are more nuanced, wide-ranging, and thus more reliable. Next, we expect that functional scrutiny will enable ethical reflexivity (***link 4***) by affording a more accurate representation of the ADMS, which will help detect ethical concerns. This could happen via different means such as the provision of a conceptual map of the key components and mechanisms that will facilitate the ethical examination. For example, knowing that the judgment component in an ADMS is automated and knowing its rules and assumptions will help identify ethical concerns that could be associated with it, such as whether the judgment logic is reliable, fair and transparent. Finally, we expect that stakeholder inclusiveness will improve (i) functional scrutiny (***link 5***)—because a more complete set of domain and technical expertise should contribute to a more thorough analysis of an ADMS' structure and operational logic, and (ii) ethical reflexivity (***link 6***)—by providing a more complete view of the ethical considerations relevant for a focal ADMS.

RAD (links 7 and 8). Our conceptualization of RAD indicates that civic-mindedness, foresightedness, and transparency define an ADMS adoption decision' degree of responsibility. We expect that both civic-mindedness and foresightedness will enable transparency because the opening and explaining of a decision is easier and less risky when the content of a decision is more sensitive to socio-ethical matters. Thus, a civic-minded decision is likely to facilitate transparency by making the decision more socially legitimate, thus less risky to communicate (***link 7***). In a similar vein, we expect that foresightedness will encourage transparency because foresightedness

involves anticipation and caution, two attributes that also favor social legitimacy (***link 8***).

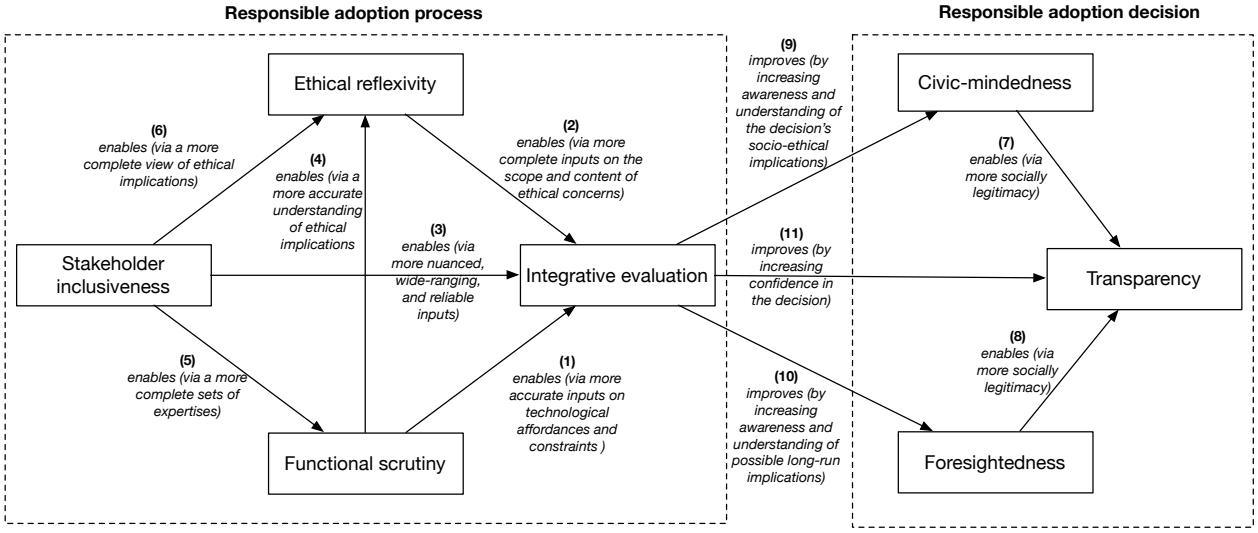


Figure 2. Research Framework

Relationships between RAP and RAD (links 8, 9, and 10). Responsible decisions are not likely to just happen; instead, they are likely to result from a proactive responsible process [35]. Thus, we expect RAP and RAD to be related. Integrative evaluation is a particularly important component of RAP because it involves building on the insights generated by both technological and ethical assessments. As such, we expect that it will be the key driver of RAD. The three resulting relationships can be explained as follows. First, we expect that integrative evaluation will improve civic-mindedness because it will increase decision-makers' awareness and understanding of the socio-ethical implications of adopting a focal ADMS (***link 9***). Second, we expect that integrative evaluation will improve foresightedness because a thorough holistic contextual assessment is likely to help better anticipate longer-run implications of ADMS (***link 10***). Third, we expect that integrative judgement will also improve transparency because it will increase decision-makers' confidence, thus their willingness to make the adoption decision available for public scrutiny (***link 11***).

4. Discussion

The framework that we have introduced in this paper makes some important contributions. First, it introduces a new concept, responsible adoption, and it clarifies what it involves in the context of automated decision-making systems. ADMS constitute an increasingly prevalent type of technology, which is often deployed to automate socially sensitive processes. The adoption of ADMS by public and private organizations, thus their increased presence in society, implies a growing dependence on automated systems in the making of important decisions affecting people in their everyday life (e.g., should parole be given to a particular individual) and managerial decisions alike (e.g., whom to hire or to promote; which product factory to invest in or decommission). This calls for a certain level of caution on behalf of the adopting entity. Thus, a larger implication of our theoretical model is that future research will need to change paradigm when examining ADMS. The existing paradigm in the IT literature espouses a view that is anchored in premises that are not well-aligned with the reality of ADMS and ADMS adoption. This paradigm is technophile, value-neutral, and it focuses on the assessment of benefits and impacts in the immediate surrounding of the decision-maker. In contrast, we propose a broader, responsible-centric view that considers the socio-ethical and systemic aspects that are so inherently tied to the deployment of ADMS.

Because the present study has focused on laying out the conceptual basis for the study of the responsible adoption of ADMS, it offers substantial opportunities for future research. One important direction to augment the proposed theoretical framework is to account for *contextual* factors., that is, investigate how the model varies across ADMS adoption circumstances [72]. Researchers could start with examining factors that hinder or facilitate the enabling effect of integrative judgement on the components of responsible adoption decisions. These are particularly interesting contingencies to study because we have reasoned that a responsible adoption process enables but does not

guarantee responsible adoption decisions. In other words, other factors and mechanisms might influence the presence and strength of these links. Contextual factors to consider include those capturing differences in the socio-ethical implications that are at stake, such as the extent of moral intensity [73], as well as those about the domain of application of the ADMS (e.g., type of industry), the decision-making setting (e.g., organizational ethical culture, organization size, subjective norms), and the decision-makers (e.g., agency, decision style, education, cognitive moral development, personal values) [74].

In a similar vein, future work could consider how those factors and others—such as technology/ADMS self-efficacy [75] and stakeholder engagement and participation in IT-related decisions [76]—could make stakeholder inclusiveness, functional scrutiny, and ethical reflexivity more effective so as to better support integrative judgement.

Another path consists in examining how different types of ADMS influence responsible adoption. For example, an ADMS in its preliminary stage of development might be perceived as unstable or highly risky and thus it might be evaluated differently (e.g., ethical reflexivity and functional scrutiny might be more salient) than an ADMS that has been used and tested for some time. Pursuing this line of research would require developing a typology of ADMS and group the types along key characteristics and features that are relevant to RAP and RAD (e.g., degree of complexity/opacity of the ADMS, type of dominant values associated with the ADMS, scale of the system affected by the ADMS).

Another avenue consists in identifying important triggers, drivers, and outcomes of RAP and RAD. In relation to triggers and drivers, it is important to stress that ADMS adoption decisions are rarely made in vacuums and that a variety of factors may hinder the process as well as the decision. This calls for studies about how psychological, social and institutional factors may influence responsible adoption including pressures to engage in such initiatives. Researchers may account for such pressures at different levels including individual (e.g., employees), groups (e.g., advocating groups), country (e.g., governments), and supranational (e.g., intergovernmental organizations) [77].

Finally, while we have assumed responsible adoption to be "good", how good is it really and does goodness vary depending on stakeholder perspectives? In other words, what are the consequences of responsible adoption? These questions call for further investigation into the effects of RAP and RAD on different stakeholders and the community in which they operate. There is some evidence that corporate social/environmental investment can have positive effects on financial performance [78], but we do not know whether similar or different kinds of benefits can be achieved in the case of responsible ADMS adoption.

5. Conclusion

Highly automated decision-making systems differ from most ITs studied in the past, making the traditional adoption paradigm less suitable. In this paper, we have leveraged the literature on social responsibility to propose a new focus on *responsible* adoption, conceptualized in terms of both a process and a decision that take into consideration the socio-ethical implications of putting an ADMS into use. The proposed framework connects technology adoption and social responsibility and offers a progressive foundation to study ADMS adoption.

6. References

[1] D. Demetis and A. Lee, "When Humans Using the IT Artifact Becomes IT Using the Human Artifact," *Journal of the Association for Information Systems*, vol. 19, no. 10, Oct. 2018,

[2] V. Eubanks, *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press, 2018.

[3] S. Newell and M. Marabelli, "Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of 'datification,'" *The Journal of Strategic Information Systems*, vol. 24, no. 1, pp. 3–14, Mar. 2015.

[4] A. Ma, "China has started ranking citizens with a creepy 'social credit' system — here's what you can do wrong, and the embarrassing, demeaning ways they can punish you," *Business Insider*, 2018.

[5] D. Ensign, S. A. Friedler, S. Neville, C. Scheidegger, and S. Venkatasubramanian, "Runaway Feedback Loops in Predictive Policing," *arXiv:1706.09847 [cs, stat]*, Jun. 2017, Accessed: Mar. 03, 2019. [Online].

[6] A. G. Ferguson, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*. New York: NYU Press, 2017.

[7] D. Kehl, P. Guo, and S. Kessler, "Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing," *Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School.*, 2017, Accessed: Feb. 21, 2019. [Online].

[8] V. Eubanks, *Digital Dead End*. MIT Press, 2012.

[9] M. Grimshaw, "Towards a manifesto for a critical digital humanities: critiquing the extractive capitalism of digital society," *Palgrave Communications*, vol. 4, no. 1, p. 21, Feb. 2018.

[10] M. Bunge, "Towards a Technoethics," *The Monist*, vol. 60, no. 1, pp. 96–107, 1977.

[11] E. Mumford, "The ETHICS approach," *Communications of the ACM*, vol. 36, no. 6, p. 82, Jan. 1993.

[12] N. Wiener, *Cybernetics: Control and Communication in the Animal and the Machine*. Cambridge, MA: MIT Press, 1948.
[13] L. Winner, *The whale and the reactor: A search for limits in an age of high technology*. University of Chicago Press, 1986.
[14] A. Majchrzak, M. L. Markus, and J. Wareham, "Designing for Digital Transformation: Lessons for Information Systems Research from the Study of ICT and Societal Challenges," *MIS Q.*, vol. 40, no. 2, pp. 267–277, Jun. 2016.
[15] G. Walsham, "Are we making a better world with ICTs? Reflections on a future agenda for the IS field," *J Inf Technol*, vol. 27, no. 2, pp. 87–93, Jun. 2012.
[16] L. Fortunati, "Robotization and the domestic sphere," *New Media & Society*, vol. 20, no. 8, pp. 2673–2690, Aug. 2018.
[17] J. Van den Hoven, "Ethics for the Digital Age: Where Are the Moral Specs?," in *Informatics in the Future*, H. Werthner and F. van Harmelen, Eds. Springer International Publishing, 2017, pp. 65–76.
[18] B. Hosack, D. Hall, D. Paradice, and J. F. Courtney, "A look toward the future: decision support systems research is alive and well," *Journal of the Association for Information Systems*, vol. 13, no. 5, p. 3, 2012.
[19] F. Nelson Ford, "Decision support systems and expert systems: A comparison," *Information & Management*, vol. 8, no. 1, pp. 21–26, Jan. 1985.
[20] J. J. Sviokla, "An Examination of the Impact of Expert Systems on the Firm: The Case of XCON," *MIS Quarterly*, vol. 14, no. 2, pp. 127–140, Jun. 1990.
[21] A. Agrawal, J. Gans, and A. Goldfarb, *Prediction Machines: The simple economics of artificial intelligence*. Harvard Business Press, 2018.
[22] H. Simon, *Administrative Behavior: a Study of Decision-Making Processes in Administrative Organization*. Macmillan Inc., 1947.
[23] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User Acceptance of Information technology: Toward a Unified View," *MIS Quarterly*, vol. 27, no. 3, pp. 425–478, Sep. 2003.
[24] P. Chwelos, I. Benbasat, and A. S. Dexter, "Research Report: Empirical Test of an EDI AdoptionModel," *Information Systems Research*, vol. 12, no. 3, p. 304, 2001.
[25] H. Aguinis and A. Glavas, "What We Know and Don't Know About Corporate Social Responsibility: A Review and Research Agenda," *Journal of Management*, vol. 38, no. 4, pp. 932–968, 2012.
[26] J. Desmond and A. Crane, "Societal marketing and morality," *European Journal of Marketing*, vol. 36, no. 5/6, pp. 548–569, Jun. 2002.
[27] K. Mandal and S. Banerjee, "Responsible Marketing and Its Impact on Business Performance: A Longitudinal Study," *Journal of Nonprofit & Public Sector Marketing*, pp. 1–24, Oct. 2018.
[28] D.-L. Arjaliès, "A social movement perspective on finance: How socially responsible investment mattered," *Journal of business ethics*, vol. 92, no. 1, pp. 57–78, 2010.
[29] P. R. Murphy and R. F. Poist, "Socially responsible logistics: an exploratory study," *Transportation Journal*, pp. 23–35, 2002.
[30] A. B. Carroll, "The pyramid of corporate social responsibility: Toward the moral management of organizational stakeholders," *Business horizons*, vol. 34, no. 4, pp. 39–49, 1991.
[31] European Union, "Corporate Social Responsibility: A business contribution to Sustainable Development," European Commission, Communication from the Commission, 2002. [Online].
[32] D. E. Rupp, P. M. Wright, S. Aryee, and Y. Luo, "Special issue on 'behavioral ethics, organizational justice, and social responsibility across contexts,'" *Management and organization review*, vol. 7, no. 2, pp. 385–387, 2011.
[33] P. van der Duin, "Toward 'Responsible Foresight': Developing Futures that Enable Matching Future Technologies with Societal Demands," *World Futures Review*, p. 1946756718803721, Oct. 2018.
[34] M. G. Velasquez and C. Rostankowski, *Ethics: Theory and Practice*, 1 edition. Englewood Cliffs, N.J: Pearson, 1984.
[35] J. Van den Hoven, "Value Sensitive Design and Responsible Innovation," in *Responsible Innovation*, John Wiley & Sons, Ltd, 2013, pp. 75–83.
[36] P.-P. Verbeek, *Moralizing technology: Understanding and designing the morality of things*. University of Chicago Press, 2011.
[37] N. Doorn and I. van de Poel, "Editors' overview: Moral responsibility in technology and engineering," *Science and engineering ethics*, vol. 18, no. 1, pp. 1–11, 2012.
[38] C. O'Neil, *Weapons of math destruction: how big data increases inequality and threatens democracy*. New York: Crown Books, 2016.
[39] S. Russell, S. Hauert, R. Altman, and M. Veloso, "Ethics of artificial intelligence," *Nature*, vol. 521, no. 7553, pp. 415–416, 2015.
[40] K. Siau and W. Wang, "Artificial Intelligence (AI) Ethics: Ethics of AI and Ethical AI," *Journal of Database Management (JDM)*, vol. 31, no. 2, pp. 74–87, 2020.
[41] D. Collingridge, *The Social Control of Technology*. London: Palgrave Macmillan, 1981.
[42] F. J. Milliken, "Three Types of Perceived Uncertainty about the Environment: State, Effect, and Response Uncertainty," *The Academy of Management Review*, vol. 12, no. 1, pp. 133–143, 1987, doi: 10.2307/257999.
[43] J. Stilgoe, R. Owen, and P. Macnaghten, "Developing a framework for responsible innovation," *Research Policy*, vol. 42, no. 9, pp. 1568–1580, Nov. 2013.
[44] C. Selin, "On not forgetting futures," *Journal of Responsible Innovation*, vol. 1, no. 1, pp. 103–108, 2014.
[45] B. Fernandez-Feijoo, S. Romero, and S. Ruiz, "Effect of Stakeholders' Pressure on Transparency of Sustainability Reports within the GRI Framework," *J Bus Ethics*, vol. 122, no. 1, pp. 53–63, Jun. 2014.

[46] J. Kang and G. Hustvedt, "Building Trust Between Consumers and Corporations: The Role of Consumer Perceptions of Transparency and Social Responsibility," *J Bus Ethics*, vol. 125, no. 2, pp. 253–265, Dec. 2014.
[47] L. Quaak, T. Aalbers, and J. Goedee, "Transparency of Corporate Social Responsibility in Dutch Breweries," *Journal of Business Ethics*, vol. 76, no. 3, pp. 293–308, 2007.
[48] A. Florini, *The Right to Know: Transparency for an Open World*. Columbia University Press, 2007.
[49] B. D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, and L. Floridi, "The ethics of algorithms: Mapping the debate," *Big Data & Society*, vol. 3, no. 2, Dec. 2016.
[50] J. Burrell, "How the machine 'thinks': Understanding opacity in machine learning algorithms," *Big Data & Society*, vol. 3, no. 1, p. 2053951715622512, Jun. 2016.
[51] R. Kitchin, "Thinking critically about and researching algorithms," *Information, Communication & Society*, vol. 20, no. 1, pp. 14–29, Jan. 2017.
[52] T. Zarsky, "The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making," *Science, Technology, & Human Values*, vol. 41, no. 1, pp. 118–132, Jan. 2016.
[53] F. Doshi-Velez and B. Kim, "Towards A Rigorous Science of Interpretable Machine Learning," *arXiv:1702.08608 [cs, stat]*, Feb. 2017, Accessed: Mar. 02, 2019. [Online].
[54] A. Datta, S. Sen, and Y. Zick, "Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems," in *2016 IEEE symposium on security and privacy (SP)*, 2016, pp. 598–617.
[55] P. Voosen, "How AI detectives are cracking open the black box of deep learning," *Science*, Jul. 05, 2017.
[56] I. G. Cohen, R. Amarasingham, A. Shah, B. Xie, and B. Lo, "The legal and ethical concerns that arise from using complex predictive analytics in health care," *Health affairs*, vol. 33, no. 7, pp. 1139–1147, 2014.
[57] B. Lepri, N. Oliver, E. Letouzé, A. Pentland, and P. Vinck, "Fair, Transparent, and Accountable Algorithmic Decision-making Processes," *Philos. Technol.*, vol. 31, no. 4, pp. 611–627, Dec. 2018.
[58] K. Macnish, "Unblinking eyes: the ethics of automating surveillance," *Ethics and information technology*, vol. 14, no. 2, pp. 151–167, 2012.
[59] L. Winner, "Do Artifacts Have Politics?," *Daedalus*, vol. 109, no. 1, pp. 121–136, 1980.
[60] J. Davis and L. P. Nathan, "Value sensitive design: applications, adaptations, and critiques," *Handbook of ethics, values, and technological design: Sources, theory, values and application domains*, pp. 11–40, 2015.
[61] M. Flanagan, D. C. Howe, and H. Nissenbaum, "Embodying Values in Technology: Theory and Practice," in *Information technology and Moral Philosophy*, J. Van den Hiven and J. Weckert, Eds. Cambridge: Cambridge University Press, 2008, pp. 322–353.
[62] B. Friedman, "Value Sensitive Design. Encyclopedia of human-computer interaction," 2004.
[63] H. Jonas, *The imperative of responsibility: In search of an ethics for the technological age*. 1984.
[64] P. Brey, "Disclosive Computer Ethics," *SIGCAS Comput. Soc.*, vol. 30, no. 4, pp. 10–16, Dec. 2000.
[65] D. K. Citron and F. Pasquale, "The scored society: Due process for automated predictions," *Wash. L. Rev.*, vol. 89, p. 1, 2014.
[66] O. Tene and J. Polonetsky, "Taming the Golem: Challenges of ethical algorithmic decision-making," *North Carolina Journal of law and Technology*, vol. 19, no. 1, p. 125, 2017.
[67] O. Tene and J. Polonetsky, "Big data for all: Privacy and user control in the age of analytics," *Nw. J. Tech. & Intell. Prop.*, vol. 11, p. xxvii, 2012.
[68] F. J. Bahamonde-Birke, B. Kickhöfer, D. Heinrichs, and T. Kuhnimhof, "A Systemic View on Autonomous Vehicles," *disP - The Planning Review*, vol. 54, no. 3, pp. 12–25, Jul. 2018.
[69] D. E. Leidner, E. Gonzalez, and H. Koch, "An affordance perspective of enterprise social media and organizational socialization," *The Journal of Strategic Information Systems*, vol. 27, no. 2, pp. 117–138, Jun. 2018.
[70] I. Van de Poel, "Values in Engineering Design," in *Handbook of the Philosophy of Science*, vol. 9, D. Gabbay, P. Thagard, and J. Woods, Eds. 2009, pp. 973–1006.
[71] C. L. Citroen, "The role of information in strategic decision-making," *International Journal of Information Management*, vol. 31, no. 6, pp. 493–501, Dec. 2011.
[72] W. Hong, F. K. Y. Chan, J. Y. L. Thong, L. C. Chasalow, and G. Dhillon, "A Framework and Guidelines for Context-Specific Theorizing in Information Systems Research," *Information Systems Research*, vol. 25, no. 1, 2014.
[73] T. M. Jones, "Ethical decision making by individuals in organizations: An issue-contingent model," *Academy of management review*, vol. 16, no. 2, pp. 366–395, 1991.
[74] J. L. Craft, "A Review of the Empirical Ethical Decision-Making Literature: 2004-2011," *Journal of Business Ethics*, vol. 117, no. 2, p. 221, Oct. 2013.
[75] D. R. Compeau and C. A. Higgins, "Computer Self-Efficacy: Development of a Measure and Initial Test," *Mis Quarterly*, vol. 19, no. 2, pp. 189–211, Jun. 1995.
[76] H. Barki and J. Hartwick, "Measuring User Participation, User Involvement, and User Attitude," *Mis Quart*, vol. 18, no. 1, pp. 59–82, Mar. 1994.
[77] R. V. Aguilera, D. E. Rupp, C. A. Williams, and J. Ganapathi, "Putting the S back in corporate social responsibility: A multilevel theory of social change in organizations," *AMR*, vol. 32, no. 3, pp. 836–863, Jul. 2007.
[78] A. Mackey, T. B. Mackey, and J. B. Barney, "Corporate Social Responsibility and Firm Performance: Investor Preferences and Corporate Strategies," *The Academy of Management Review*, vol. 32, no. 3, pp. 817–835, 2007.