Received: 28 September 2020 | Revised: 9 April 2021 | Accepted: 15 April 2021

DOI: 10.1111/padm.12742

ORIGINAL ARTICLE

public administration WILEY

Assessing public value failure in government adoption of artificial intelligence

Daniel S. Schiff[1] | Kaylyn Jackson Schiff[2] | Patrick Pierson[2]

[1]School of Public Policy, Georgia Institute of Technology, Atlanta, Georgia, USA

[2]Department of Political Science, Emory University, Atlanta, Georgia, USA

Correspondence
Daniel S. Schiff, School of Public Policy, Georgia Institute of Technology, 685 Cherry Street, Atlanta, GA 30332, USA.
Email: schiff@gatech.edu

Funding information
CFDE FIT Grant, Emory University

Abstract

In the context of rising delegation of administrative discretion to advanced technologies, this study aims to quantitatively assess key public values that may be at risk when governments employ automated decision systems (ADS). Drawing on the public value failure framework coupled with experimental methodology, we address the need to measure and compare the salience of three such values—fairness, transparency, and human responsiveness. Based on a preregistered design, we administer a survey experiment to 1460 American adults inspired by prominent ADS applications in child welfare and criminal justice. The results provide clear causal evidence that certain public value failures associated with artificial intelligence have significant negative impacts on citizens' evaluations of government. We find substantial negative citizen reactions when fairness and transparency are not realized in the implementation of ADS. These results transcend both policy context and political ideology and persist even when respondents are not themselves personally impacted.

1 | INTRODUCTION

Governments have a special responsibility to realize key values held by the public. Yet scholars of public value failure recognize that this core responsibility may not be satisfied when administrators overemphasize efficiency and other economic goals at the expense of publicly-held values such as equity, transparency, and responsiveness (Bozeman, 2002; Jørgensen & Bozeman, 2007). In this study, we examine public value failure in the context of a phenomenon of increasing importance: the delegation of discretion to automated decision systems (ADS), or artificial intelligence (AI)-based tools that provide predictions and recommendations to government officials. In particular, the adoption of AI by governments to enhance efficiency in service provision may result in supplanting or otherwise

reprioritizing key public values. While scholars have studied the delegation of discretion to technology—as well as potential public value implications of digital governance—this paper is the first to empirically assess potential public value failure in the case of government use of AI. Specifically, we investigate whether members of the public are concerned about prominent values arguably threatened by ADS, whether some of these values are more salient than others, and whether public attitudes vary depending on the policy use case and citizens' political ideology.

While public administration scholars have evaluated the implications of digital discretion for the roles of bureaucrats and for engagement with citizens (Bovens & Zouridis, 2002; Buffat, 2015; Young et al., 2019), there is a need to better understand which public values should guide government implementation of AI systems. As Bullock (2019) argues, "it seems obvious that a clear assessment of public values is needed" to prioritize values, maximize good governance, and minimize "administrative evil." Devising such a framework is especially essential at the present moment when major implementation, procurement, and regulatory decisions are being made that will shape AI's long-term impacts in the public domain. We therefore address this need by drawing on Bozeman's (2002) theory of public value failure and implementing a survey experiment to assess and rank values in the case of government adoption of AI.

Prior methods used to assess public value failure have largely been limited to qualitatively developing inventories of potentially relevant public values. In turn, scholars of public value failure have raised concerns about a lack of concrete empirical testing (Fukumoto & Bozeman, 2019), leading to an inability to identify which values are most salient to the public. To address these concerns, we leverage experimental methodology to quantify concern for—and thus rank—public values. Based on a preregistered design and real-world cases, we administer a set of randomized vignettes describing prominent ADS applications and potential public value failures to 1460 American adults. The factorial design of the experiment incorporates two distinct policy sectors—the child welfare system and the criminal justice system—and considers three specific public value failures—bias, lack of transparency, and lack of human responsiveness—resulting from government use of ADS. We use respondents' evaluations of government to assess and rank the salience of the three public values associated with the experimental treatments, namely fairness, transparency, and human responsiveness.

We find statistically significant and substantial negative citizen reactions when failure emerges along the values of fairness and transparency, with effect sizes on the order of one-third and one-quarter of a standard deviation, respectively. These results transcend both policy context and political ideology and persist even when respondents are not themselves personally impacted, demonstrating the truly public nature of the values held. Our findings suggest that governments should be especially attentive to how citizens rank these values and should work to address fairness and transparency as key priorities when adopting AI systems.

2 | THEORY

2.1 | Public value failure and ADS

How does government adoption of ADS affect the realization of public values? While the concept of public value(s) itself admits to several overlapping definitions (Nabatchi, 2018), here we focus on *public values* as "those providing normative consensus about…the principles on which governments and policies should be based" (Bozeman, 2002).[1] *Public value failure*, then, is the failure to realize these public values. To determine whether public value failure has occurred, Bozeman provides seven evaluative criteria which have been used in domains such as nanomedicine (Slade, 2011) and science policy (Bozeman & Sarewitz, 2011). The public value failure framework can be understood as a challenge to market-based approaches to public administration and policy and as a response to perceived failures of the New Public Management (NPM) paradigm (O'Flynn, 2007; Stoker, 2006; Stone, 1997). This more value-centric orientation toward governance is also reflected in the New Public Service, which departs from NPM by emphasizing the hollowing out of government, the need for citizen participation and public responsiveness, the role

of advanced information and communication technology (ICT), and the incorporation of a broader set of governance objectives and public values beyond efficiency (Bryson et al., 2014; Denhardt & Denhardt, 2000).

This begs the question: how does the adoption and implementation of ADS technologies contribute to the realization—or failure—of public values in public administration? Figure 1 presents our conceptual framework that applies the public value failure literature to the context of AI implementation by government and highlights the pathway to public value failure that we isolate for study. In particular, when implementing government policies, administrators may delegate decision-making authority to technology in ways that fail to express certain public values, resulting in possible public value failure.

While the public value failure framework offers promise to scholars of public administration, recent scholarship in this field identifies several areas for further theoretical and empirical development. First, public value failure can benefit from conceptual refinement (Bozeman, 2009), especially given the plurality of concepts surrounding public value and values (Alford & O'Flynn, 2009; Nabatchi, 2018). Second, there is a need to provide more concrete strategies for empirical testing (Bryson et al., 2014), which can help address the lack of empirical studies to date (Williams & Shearer, 2011). For example, there is a need to distinguish between the mere *identification* of an inventory of public values and an *assessment* of their relative importance or relevance in specific contexts (Jørgensen & Bozeman, 2007). Third, measuring public values or their failure does not yet provide clear linkages to strategies to address identified needs or problems. Therefore, measurement efforts should aim to capture public values and their failure in "some analytically useful form" (Bozeman & Sarewitz, 2011).

We respond to these needs by incorporating experimental methodology into the study of public value failure and demonstrate how resultant rankings of public values can more clearly inform government responses. Coupled with the methodology we use, public value failure offers an actionable framework for considering the ethical implications, trade-offs, and values implicated by important government decisions, especially in controversial domains or in response to paradigm shifts, such as those involving new technology. Our study demonstrates the utility of this

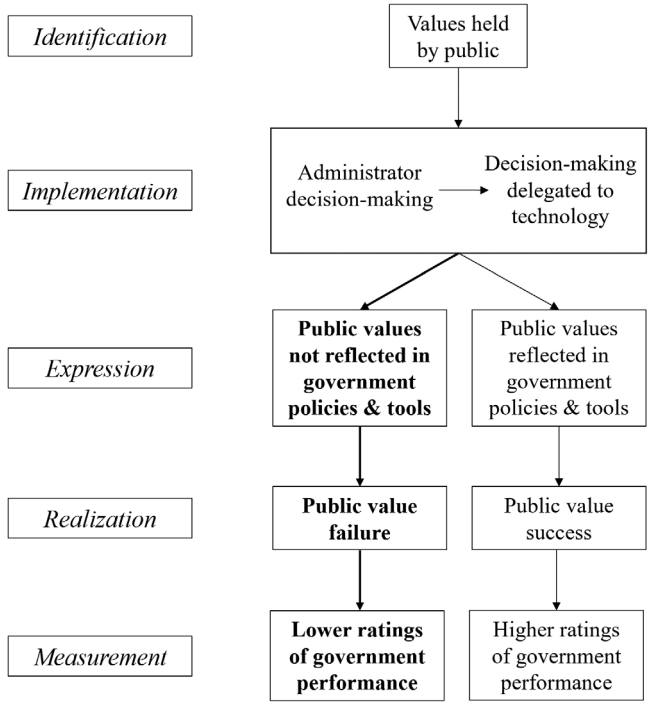

FIGURE 1 Conceptual framework: Public value failure in government adoption of artificial intelligence

approach by evaluating public value failure in the novel context of government use of AI, shedding light on an issue of increasing importance in public administration.

2.2 | The new paradigm of automated government: Impacts on administration and values

Governments around the globe have begun to leverage advanced ICT systems and AI by developing, procuring, and implementing "automated decision systems" (ADS), or predictive algorithms that aim to "automate, aid, or replace" decision-making by government officials (Hidalgo, 2020). Hundreds of ADS are actively deployed across the United States in government domains as diverse as education, child welfare, citizen contact, criminal justice, education, emergency services, healthcare, housing, and immigration (AI Now Institute, 2019; Diakopoulos et al., 2020), with various technical implementation, governance, and ethical challenges coming into focus (Wirtz et al., 2019).

In turn, over the last two decades, scholars have devoted increased attention to the impacts of ICT on government, noting how the adoption of technology by governments fundamentally alters "the way in which public offices organize and deliver services" with "political and administrative consequences" (Cordella & Bonina, 2012). This scholarship has considered the impact of technology on the roles of human administrators, the organizational and legal regimes that underpin administration, and the diminution of human discretion. Lenses used to describe this transformational shift include e-government and "digital-era governance" (Dunleavy, 2005), "screen-level" and "system-level" bureaucracies (Bovens & Zouridis, 2002), and "artificial discretion" (Young et al., 2019). For example, Bovens and Zouridis (2002) argue that the discretion of public administrators, particularly street-level bureaucrats, is decreasing under a new screen-level paradigm in which ICT plays a leading rather than supportive role and human interference in decisions is minimal. They further conceptualize the notion of a system-level bureaucracy, in which the role of ICT is decisive and human discretion is almost eliminated, similar to what Young et al. (2019) describe as a growing norm of "artificial discretion" under which AI systems are predominant.

Under this paradigm, AI systems are now capable of automating complex cognitive and analytical tasks previously thought not automatable. Importantly, AI systems may be more appropriate in some contexts than others, depending on "task complexity, quality and availability of data, technical requirements, [and] limitations" (Young et al., 2019). Bullock (2019) argues that AI systems are best suited to tasks that are both low in complexity and low in uncertainty. Yet, because AI has clear advantages with low-cost scalability, it may be the case that there is "no economic rationale for the public sector to favor human labor" (Young et al., 2019) and, as such, AI may be adopted even in cases when it is less appropriate.

Critically, the increasing adoption of technological systems also has implications for the realization of values, potentially reflecting and imparting some while impeding the expression of others. For example, Busch and Henriksen (2018) analyze 44 studies of digital discretion and find mixed impacts on democratic, professional, and relational values. Young et al. (2019) apply Salamon and Elliott's (2002) framework for evaluating governance tools to the case of AI, considering possible implications for effectiveness, efficiency, equity, manageability, and legitimacy and political feasibility. These studies exemplify a growing body of scholarship that recognizes the ethical importance and potential trade-offs inherent in use of AI systems, and seeks to better understand these issues (Andrews, 2019). However, scholars have also recognized that this line of research does not yet have a sufficiently nuanced and analytically tractable framework to answer such questions (Bullock, 2019).

We argue that the public value failure framework, coupled with the experimental methodology we present, is suited for this job. Importantly, as government use of ADS is rapidly becoming the new norm (Diakopoulos et al., 2020; Shrum et al., 2019), our study acknowledges this context as a reality rather than a contingency, and thus assumes AI adoption as a given. Therefore, we do not seek to contrast human versus AI decision-making as our primary focus.[2] Instead, we seek to evaluate ethical implications (or public value failures) when AI has already been

adopted and is in use. The following sections argue that efficiency is the predominant value that underlies government use of AI and that, in this context, public values of fairness, transparency, and responsiveness may be at risk.

2.3 | Values prioritized by AI in government: Efficiency

What are the implicit values associated with ADS? We suggest that AI and other automating technologies are significantly characterized by goals of efficiency. For example, the National Academy of Public Administration (NAPA) (Shrum et al., 2019) highlights that AI's implementation in work processes brings "promises of efficiency" and "the potential for increased productivity," due in part to its capacity to automate and replace human labor. Expressions like these are often upfront and prominent in documents concerned with AI adoption. Concerns about social and ethical harms, as well as broader values that AI may foster, are often only secondarily addressed and arguably in service to the efficiency goals. These efficiency goals seem well suited to AI, given that the algorithms which constitute ADS are based on optimization and quantification, lending them to economistic thinking.

Adoption of ADS thus may reflect a political and managerial logic echoing a private-sector and technological "mythology" of new and better government (Bekkers & Homburg, 2007), one which hearkens back to a machine metaphor (Gulick, 1984). This is not dissimilar from earlier ICT reforms (Andersen et al., 2010; Cordelia, 2006) which were "largely conceived…along the basic principles of efficiency gains and cost savings," but sometimes failed to promote democratic values (Cordella & Bonina, 2012). The intentional logic behind ADS use therefore further elevates AI's inherent efficiency goals into definitive administrative values. Yet it does so at a potential cost to other public values (Anderson, 1995). In fact, the delegation of decision-making to ADS is arguably efficient in part *because* it fails to ensure certain public values, for example, by facilitating the downsizing and hollowing out of government (Gruening, 2001).

2.4 | Values at risk: Fairness, transparency, and responsiveness

What other values may be traded off when efficiency is explicitly, or implicitly, prioritized? Scholars have considered a variety of means by which to identify public values, including government documents, scholarly literature, opinion polls (Jørgensen & Bozeman, 2007), values statements (Slade, 2011), and other formal and informal sources. Scholars have also noted that the public values considered most pressing vary by context, such as the centrality of the value of responsiveness in e-government (Karunasena & Deng, 2012) and values related to security and defense in the nanotechnology sector (Fisher et al., 2010). In the case of ADS, we aim to identify candidate values that might be particularly threatened under automated government. Based on emerging scholarly literature, media discourse, and a comprehensive study of AI values statements and ethics documents (Schiff et al., 2021), fairness, transparency, and responsiveness emerge as three public values deemed most at risk under government adoption of ADS.

Not only is *fairness* a long-established value in the public administration and public value literature (Bryson et al., 2014; Jørgensen & Bozeman, 2007; Salamon & Elliott, 2002), it is also one with clear significance for AI and ADS in particular. For example, research on AI that identified biases in mainstream commercial facial recognition software (Buolamwini & Gebru, 2018) has spurred increased attention to issues of algorithmic bias, discrimination, and fairness. In the public sector, scholars have identified racial biases in AI use in criminal justice (Chouldechova, 2017), healthcare (Howard & Borenstein, 2017), and more (Eubanks, 2017). NAPA argues that bias in AI "is arguably the largest ethical issue and the one that has received the most attention," given that fairness and consistency are crucial to due process, and among "the cornerstones of administrative decisionmaking" (Shrum et al., 2019). This risk of bias against marginalized groups is "particularly critical in areas of administration where individuals may be awarded or denied a benefit, a consequence such as a sentence in a criminal case, access to a service or treatment, a grant or waiver, etc." (Shrum et al., 2019). As an example, experimental evidence suggests that racial

subgroups may, in turn, respond differently in judging the fairness of ADS used in policing (Miller & Keiser, 2021). Therefore, our first candidate value is fairness, conceived of in this study as the absence of biases that could harm vulnerable groups.

A second value of great importance is *transparency*. According to the AI Now Institute, which has produced recommendations on ADS use by government agencies, many AI systems operate as so-called "'black boxes'—opaque software tools working outside the scope of meaningful scrutiny and accountability" (Reisman et al., 2018). Put differently, the algorithms are thought to be so complex that even the designers cannot explain how they arrived at their recommendations or predictions (Adadi & Berrada, 2018; Castelvecchi, 2016). While AI ethics researchers typically emphasize technical transparency, concerns about policy and process transparency have also fueled ongoing debate regarding e-government and more recent and advanced forms of ICT in government (Andersen et al., 2010; Cordella & Bonina, 2012; Wirtz et al., 2019). Thus, transparency involves not only understanding the inner workings of an algorithm, but also informing the public about an algorithm's existence and explaining how decisions or predictions are made (Reisman et al., 2018). In our experimental vignettes, we describe situations in which the processes behind ADS-provided recommendations are not transparent to the government officials who rely on them, much less to the broader public.

A third value is the role of *human responsiveness* in government services, a value closely related to the delegation of discretion and a longstanding concern associated with AI use (Barth & Arnold, 1999). Citizens value human responsiveness in government services because personal interactions with government officials can convey empathy and promote trust, while "digital rigidity" and a "zero-touch logic" may not allow for sufficient flexibility and customization (Bovens & Zouridis, 2002; Dunleavy, 2005). For example, an ADS adopted in Indiana to determine welfare eligibility deemphasized face-to-face contact with caseworkers in favor of automatic electronic systems, contributing to a $437 million lawsuit in 2010 (Eubanks, 2017). In this case, citizens may have been rightly concerned that government services were "applied rigidly and insensitively" (Jørgensen & Bozeman, 2007). Yet, citizens may also prefer ADS if they decrease the exercise of arbitrary power and bias by bureaucrats (Bovens & Zouridis, 2002). In contrast to the previous two values then, it is less clear whether the public would perceive a loss of human responsiveness as an unambiguous value failure.

2.5 | Assessing public value failure through public opinion

To assess the importance and relative ranking of these public values, we use a survey experiment to measure citizens' reactions to hypothetical scenarios in which public values are not realized. We hypothesize that *citizens will exhibit reduced evaluations of government when use of ADS fails to express salient public values*. Therefore, we use citizens' ratings and evaluations of government as the outcome measures in our survey experiment. Measures of public opinion are useful in identifying and assessing public value success or failure, as citizen evaluations of government legitimacy and performance are often based on their perceptions of whether core public values, such as efficiency, accountability, and fairness, are realized by government (Weatherford, 1992). Indeed, public opinion is arguably the most important arbiter of whether public value failure has occurred, and our methodology reflects this theoretical position.

As reflected in our conceptual framework in Figure 1, we expect that particularly concerning public value failures will lead to decreased evaluations of government, helping to identify which public values are most salient (or relevant) for a given policy context. We thus rely on common measures of government performance, including (1) citizen support for government actions, (2) trust, (3) beliefs about government service quality, and (4) expectations of personal impact (Cordella & Bonina, 2012; DeLone & McLean, 1992; Karunasena & Deng, 2012; Kelly et al., 2002; Nye, 1997; Omar et al., 2011). We discuss each measure of government performance in more detail in the Methods section. Note that it is important to assess not only which values are deemed salient, but also how they are ranked.

Further, understanding how specific policy contexts influence considerations of public value is important for scholarship and for guiding policy and administration (Gormley, 1986; Hartley et al., 2017). To examine how public value salience may vary based on the policy context, we identify two real-world cases of ADS that are prominent in scholarship and media: (1) a predictive risk algorithm used in the child welfare system in Allegheny County, PA, to "determine whether a maltreatment referral is of sufficient concern to warrant an in person investigation" (Chouldechova et al., 2018) and (2) an ADS used in numerous US states to assess eligibility for pretrial detention or release based on predicting whether a defendant will recidivate or not show up for one's court date (Reisman et al., 2018). These cases provide the inspiration for the policy sectors experimentally manipulated in our treatment vignettes, and are described in more detail in Section E in Appendix S1.

Beyond their real-world relevance, both policy sectors represent the implementation of AI in domains with high complexity and high uncertainty that are arguably best suited to human discretion (Bullock, 2019); because of this mismatch, public value failure may be more likely. However, these policy domains differ noticeably regarding the social construction of their target populations. In particular, we expect that citizens will exhibit more concern about the potential burdens of AI policy design in the child welfare scenario, as children are a population of "dependents" generally felt to be especially deserving of protection. In contrast, detainees in the courts scenario may be thought of as "deviants" for whom the general population is more comfortable assigning burdens (Schneider & Ingram, 1993). Therefore, we hypothesize that *citizens will be more concerned about public value failure in the child welfare scenario*.

Finally, scholars of public administration have raised the concern that various groups of citizens may hold different or even competing values, challenging the identification of values that are truly public and complicating their eventual implementation (Fukumoto & Bozeman, 2019). As partisanship is increasingly central to individuals' social identities (West & Iyengar, 2020), assessing public value salience based on individuals' political ideology is particularly pressing. Historically, politically-motivated public values have played an important role in shaping government adoption of technology, such as how conservative values emphasizing private sector superiority informed the early development of the internet (Rogers & Kingsley, 2004). Yet in the case of AI and ADS, it is currently unclear whether and to what extent the public is divided along political lines (Zhang & Dafoe, 2019).

It is plausible that public values surrounding ADS may be contested if AI is implemented in policy sectors for which there are preexisting, politically-divided issue preferences. Relatedly, political differences regarding social constructions of target populations and views about the role of government may lead to distinct public value conceptions pertaining to ADS in given policy sectors. As a starting point along this line of inquiry, we consider how political ideology moderates judgments of public values within the policy sectors that we study. In particular, longstanding conservative attitudes toward desert in the criminal justice system (Caldeira & Cowart, 1980) inform our hypothesis that *conservative respondents may be relatively less concerned about public value failure in the criminal justice scenario, as compared to liberal respondents*. Therefore, our analysis considers reactions to potential public value failure generally, as well as within policy sectors and across political ideology.

3 | METHODS

This section reviews the design of our survey experiment, including the survey's factorial treatments, the informational vignettes used, the outcomes measured, and the regression models used to evaluate the causal effects of interest. The preregistered analysis plan is available through the Open Science Framework.[3] Supporting information, complete replication code, and data are also available through the Public Administration Dataverse.[4] Human subjects approval was granted by Emory University (IRB00111023) and the Georgia Institute of Technology (H19120).

Surveys of the public can capture preferences, emotions, and values, and are therefore particularly well-suited for studying public value failure (Hartley et al., 2017). *Survey experiments* provide the additional benefit of randomization of treatment such that responses can be compared across groups and differences can be causally attributed

solely to variation in treatment rather than to other possible explanations (Bouwman & Grimmelikhuijsen, 2016).[5] Well-designed survey experiments address concerns about endogeneity and question ordering effects, which adds causal rigor and complements alternative approaches to studying public values, such as public value mapping (Bozeman & Sarewitz, 2011; Fisher et al., 2010).

3.1 | Study sample, factorial design, and vignette treatments

We administered an online survey in June 2019 to American adults recruited through the Amazon Mechanical Turk (MTurk) Marketplace.[6] A total of 1460 respondents completed the full survey.[7] The survey randomly presented participants with one of eight vignettes (short text-based informational prompts) about hypothetical value failures associated with government use of AI, and then asked follow-up questions to assess participants' reactions. The vignettes varied across two dimensions. First, the vignettes varied the policy sector in which a hypothetical predictive algorithm was employed. This included (1) a scenario in which one's local Department of Health and Human Services relies on a predictive algorithm to determine a child's risk of abuse or neglect in order to guide government intervention (the "child welfare scenario") and (2) a scenario in which one's local court system relies on a predictive algorithm to determine a detainee's risk of not showing up for trial in order to assess eligibility for pretrial release without bail (the "court system scenario").

Second, within each policy sector, the informational vignettes also vary which public values are invoked. Each vignette presents a hypothetical example of a given public value failure associated with a particular value: (1) *fairness*, (2) *transparency*, and (3) *responsiveness*. That is, these vignettes describe, in the context of ADS, (1) a lack of fairness (i.e., bias), (2) a lack of transparency, and (3) a lack of human responsiveness. There is also a control vignette that includes no description of a potential public value failure. With external validity in mind, the wording of the vignettes is designed to closely mirror how public value failures are discussed in popular media outlets. The descriptions of both the policy sector use cases and the identified public values come from recent news coverage in sources such as *The New York Times*, *TheWashington Post*, *NPR*, and other mainstream news outlets.[8]

Overall, this constitutes a 2 × 4 factorial design (2 policy sectors × 3 public value failures +1 control) and we randomly assigned participants to one of the eight resultant treatment groups.[9] Table 1 shows the factorial design of the experiment, as well as the number of participants randomly assigned to each vignette. Figure 2 shows the broader logic of the experimental design as well as the wording of the vignette treatments.

3.2 | Outcome measures

As outcome measures, we rely on common measures of government performance, which have also been applied to study public values: (1) citizen support, (2) trust, (3) beliefs about service quality, and (4) expectations of personal impact. As each of these measures captures important dimensions of citizen evaluations of government, they are

TABLE 1 Number of participants for each vignette of the 2 × 4 experimental design

	Policy sector	
Public value failures	**Child welfare scenario**	**Court system scenario**
Control	177	184
Bias	184	190
Lack of transparency	187	179
Lack of responsiveness	176	183

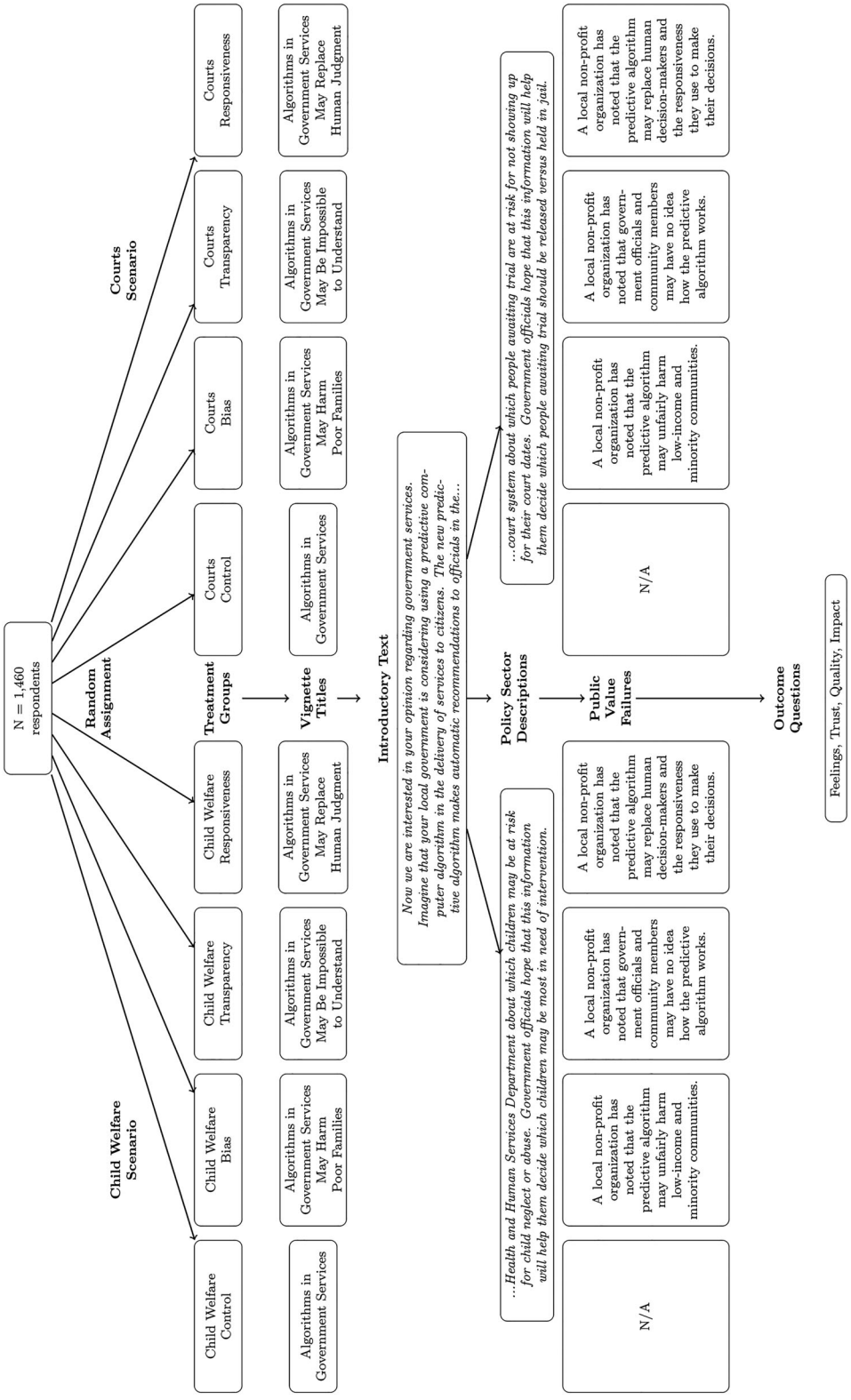

FIGURE 2 Experimental design flow diagram

useful for gauging the extent to which citizens believe that government has failed to realize each of the key public values that we study. Further, each of the government performance measures has been recommended for and applied to public-value-centric evaluations of e-government (Kearns, 2004), a context which closely relates to our study of AI adoption by government.

The first outcome measure addresses citizen feelings about government use of ADS in the given policy context. Citizen feelings have been used to evaluate reactions to new developments in e-government, such as electronic information systems that may implicate public values (DeLone & McLean, 1992; Omar et al., 2011). The second outcome measure captures the extent to which respondents trust the decisions made by the hypothetical algorithm. Trust in government and its policies is another longstanding measure (Nye, 1997) that has been used to study public values (Kelly et al., 2002; Moore, 1995), and one that is prominent in discussions of AI and ADS (e.g., European Commission, 2019). Third, we measure respondents' attitudes regarding the perceived impact of ADS on the quality of government services. Quality of services has been theorized as important to public value (Kelly et al., 2002; Omar et al., 2011) and has been used to evaluate the assumption that e-government reforms would improve public services (Cordella & Bonina, 2012; Karunasena & Deng, 2012). We measure these three outcomes on a −10 to 10 sliding scale.

The final outcome measure captures respondents' expectations about personal impact—that is, whether they feel they would be personally affected by the ADS. Unlike the other outcome variables, we did not have strong predictions regarding how the treatments would affect the personal impact measure, largely because we did not expect that most respondents are part of the populations affected by the ADS described in our study. Importantly, this measure helps to assess the extent to which citizens care about public value failures *even when they are not personally impacted*—an indication that the identified values are truly of a public nature. We measure personal impact on a 7-point Likert scale.

3.3 | Regression analysis for causal identification

In experiments, random treatment assignment ensures that all treatment and control groups are identical, on average, prior to treatment in terms of both observed and unobserved characteristics. This helps to eliminate threats to inference such as unobserved heterogeneity, common cause confounders, and endogeneity. As the only difference, on average, between the groups is their exposure to treatment, it is straightforward to identify the causal effects of different public value failures on public opinion by comparing outcomes across treatment groups.

We first use ordinary least squares regression to estimate average treatment effects (ATEs) for each of the three public value failures (bias, lack of transparency, and lack of human responsiveness), collapsing across policy sectors and analyzing each outcome variable separately. The general model specification used to address our primary hypothesis is:

$$\text{Outcome} = \beta_0 + \beta_1\,\text{bias} + \beta_2\,\text{transparency} + \beta_3\,\text{responsiveness} + \gamma\mathbf{X} + \varepsilon$$

where the ATEs of interest are the coefficients on the treatment indicators (β_1 through β_3), **X** refers to a vector of nine covariates, and ε refers to the error term. The covariates used include: political affiliation, prior knowledge of AI, gender, age, race, educational attainment, marital status, income, and employment industry.[10] The excluded group is the group of participants in the control condition.

To address the hypotheses related to policy sector and political ideology, we use similar regression models to the one shown above. This allows us to estimate ATEs (1) within each policy sector and (2) conditional on political ideology (conditional ATEs, or CATEs). To estimate ATEs within policy sectors, we use indicator variables for all eight vignettes rather than collapsing them by policy sector. To estimate CATEs by political ideology, we separate our data

into subsets of liberals and conservatives. We use *z*-tests to evaluate whether the treatment effects for bias, transparency, and responsiveness are statistically different across policy sectors and across political ideologies.[11]

4 | RESULTS

4.1 | Reactions to public value failures of ADS

If the candidate values that we identified from our review of scholarly literature and media discourse are indeed important to the public, then we would expect statistically significant negative effects of the public value failure treatments on citizen evaluations of government. Table 2 presents the results from the covariate-adjusted regression specification used to test our primary hypothesis, with separate rows for each public value failure and separate columns for each outcome measure. The coefficients correspond to the main ATEs of interest.

The results demonstrate that members of the public do respond to hypothetical public value failures with lower evaluations of government. The three public value failures associated with ADS produce largely statistically significant reductions in citizen *feelings* toward government use of ADS, *trust*, and assessment of *quality* of government services. For example, a lack of fairness in ADS (or bias) elicits highly significant reductions in citizen support ($p < 0.01$) across all four outcome measures. On average, the bias treatment decreased expectations of quality by about 1.6 points and citizen feelings by about 1.5 points, corresponding to standardized effects of about one-third of a standard deviation.[12] These results constitute compelling evidence that the candidate values identified in this paper—fairness, transparency, and human responsiveness—are indeed important to the public in the emerging context of ADS use by governments.

4.2 | Relative ranking of public values

Yet, it is also evident that some public values are more salient than others. To aid in assessing the magnitude and relative importance of the three public values, Figure 3 depicts standardized treatment effects with 95% confidence intervals for the public value failure treatments.

The results provide evidence of clear and consistent differences in the relative importance of the three public values, with fairness as the foremost public value, followed by transparency, and finally human responsiveness. In contrast to the bias treatment, which produced statistically significant negative effects for all four outcomes, the lack of transparency treatment produced somewhat smaller effects ranging from −0.16 (trust) to −0.21 (feeling) standard

TABLE 2 Regression results: Reactions to public value failures of automated decision systems

	Feelings	Trust	Quality	Impact
Bias	−1.502*** (0.377)	−1.334*** (0.376)	−1.582*** (0.364)	−0.262*** (0.085)
Lack of transparency	−1.086*** (0.380)	−0.835** (0.379)	−1.050*** (0.374)	−0.142* (0.085)
Lack of responsiveness	−0.783** (0.387)	−0.411 (0.387)	−0.733* (0.378)	−0.166* (0.085)
Constant	2.942* (1.596)	1.947 (1.681)	2.713* (1.565)	4.793*** (0.455)
Covariates	Yes	Yes	Yes	Yes
N	1460	1460	1460	1460

Note: Robust SEs in parentheses. Feelings, Trust, and Quality are measured on a −10 to 10 sliding scale, while Impact is measured on a 7-point Likert scale.
*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

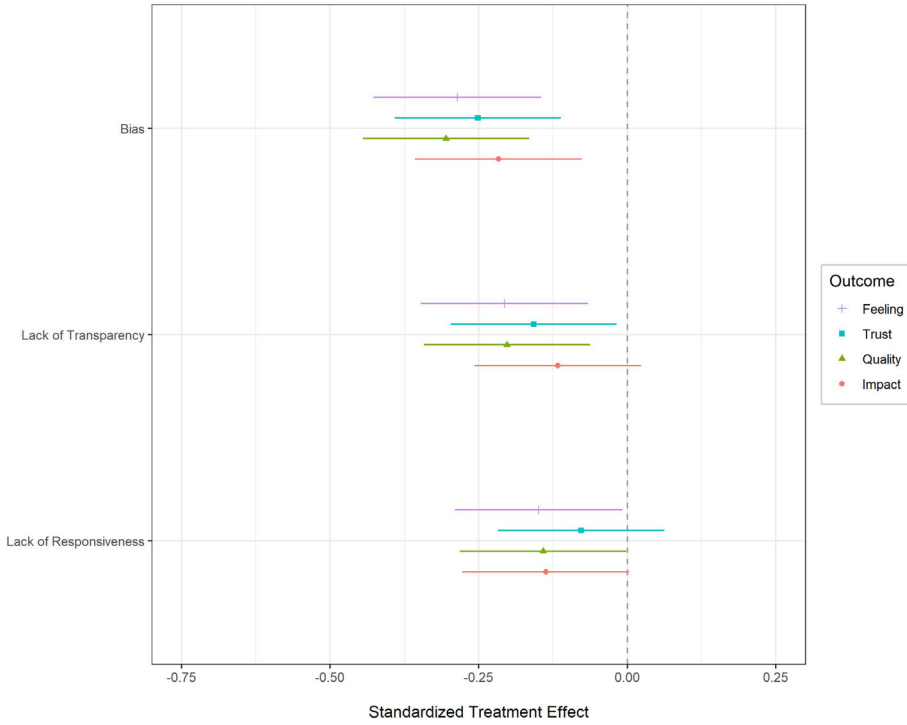

FIGURE 3 Treatment effects: Reactions to public value failures of automated decision systems

deviations, and for only three of the outcome measures. Notably, while the displacement of humans by AI/automation has received much scholarly and media attention, the human responsiveness treatment yielded the smallest impacts on citizen evaluations, impacts which are statistically indistinguishable from zero at the conventional 5% level for three out of four outcomes. This suggests that the public may not be as directly concerned about the loss of human responsiveness itself; yet, the public remains concerned about other public values (i.e., fairness and transparency) that are potentially undermined when human discretion is replaced by technology.

Finally, as per our expectations, there were fewer substantive or significant effects on the personal impact outcome measure. Nonetheless, it is striking that respondents expressed sizable negative reactions to public value failures across the other outcome measures *even when respondents felt less personally impacted*. In particular, in the case of ADS, fairness and transparency seem to be truly public-oriented values that citizens care deeply about and expect governments to uphold for the benefit of all.

4.3 | Policy sector differences

We also compare the effects of ADS-related public value failures within two policy sectors: the child welfare system and the criminal justice system. Because certain members of the public likely view children as more deserving of protection than individuals awaiting criminal trial, we expected that respondents would be relatively more concerned about public value failures in the child welfare scenario. To help evaluate this hypothesis, Figure 4 shows the estimated standardized effects with 95% confidence intervals for the public value failure treatments within each policy sector, separated by outcome measure.[13]

We find that members of the public express concern about public value failures of ADS within *both* policy sectors. Moreover, the ranking of values within policy sectors largely mirrors the main results, reinforcing that fairness—

and, to a lesser extent, transparency—are the public values deemed most important in the context of ADS. Yet, based on standard *z*-tests to evaluate differences across policy sectors, we fail to reject the null hypothesis that the treatment effects for each respective public value failure are the same across policy sectors.[14] These results suggest that public values regarding ADS may be held uniformly rather than distinctly in different policy sectors, such that the public believes the same values should apply to everyone from children to detainees.

Nonetheless, some tentative differences between the policy sectors may be worthy of further study. For example, the treatments produce a much smaller effect on trust in the child welfare case, raising questions about citizens' baseline assumptions about the trustworthiness or motivations of child welfare agencies compared to courts. Overall, however, public values seem to largely transcend policy contexts. This may be due to the inherent universal nature of public values. Alternatively, as many members of the public have had limited exposure to ADS, citizens may need more time and information to form more nuanced opinions.

4.4 | Political ideology differences

Finally, we expected that citizens of varying political ideologies might respond differently to the public value failures of ADS, for example by prioritizing certain values or viewing policy sector contexts differently. Figure 5 first presents

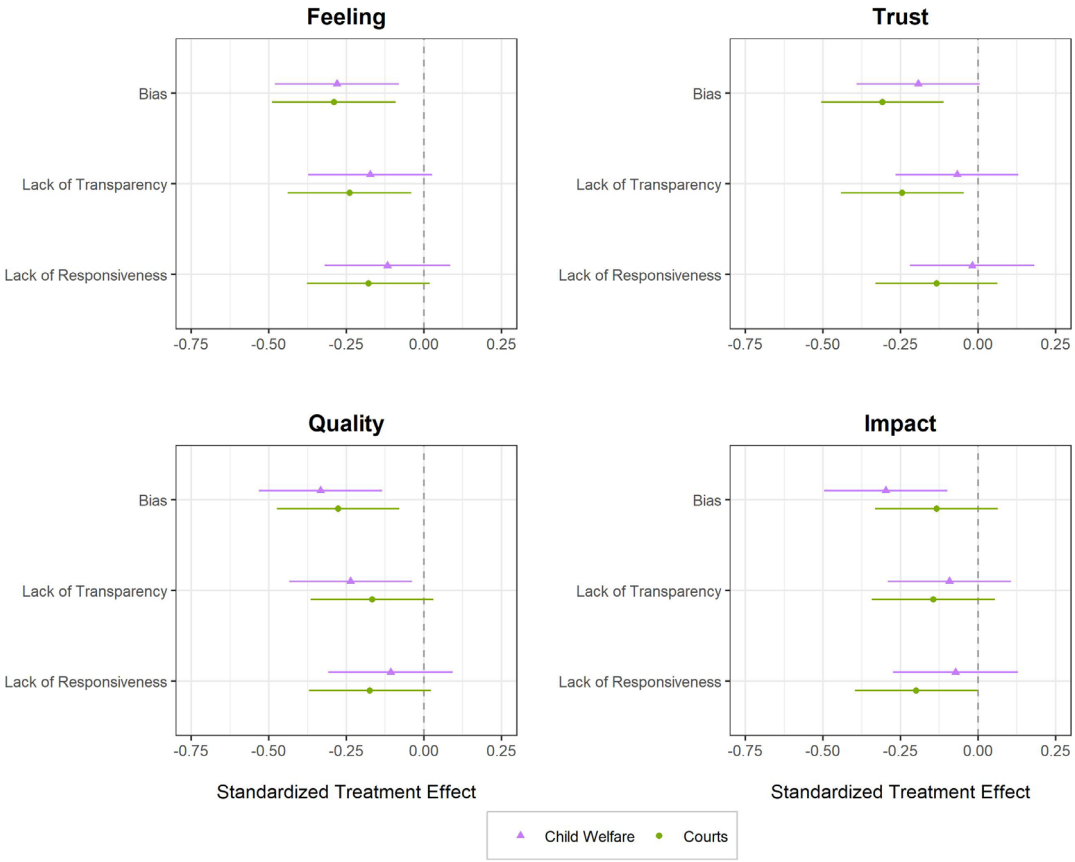

FIGURE 4 Treatment effects: Reactions to public value failures of automated decision systems within policy sectors

descriptive evidence which indicates that conservatives and liberals have different attitudes about ADS. In particular, conservatives (in red) feel more favorably on average toward ADS across almost all experimental conditions.[15]

Yet, unlike surveys that do not use control groups for comparisons, our experimental design allows us to difference out the baseline attitudes represented in the control groups. When we difference out these baseline attitudes and assess the impact of public value failure, we see statistically similar treatment effects across ideological subgroups, as displayed in Figure 6. Overall, this evidence suggests that *public values* regarding ADS are highly similar even when baseline attitudes toward ADS are not. Notably then, the majority of variance regarding attitudes toward public value failure of ADS seems to lie *within* political groups rather than across them. However, the smaller samples available for this subgroup analysis make it more difficult to identify subtle trends that may exist, so further study is warranted.[16]

Most strikingly, the results in both Figures 5 and 6 indicate the same ordinal ranking of values, with fairness first, then transparency, and lastly responsiveness. Liberals and conservatives appear to hold similar public values surrounding ADS—a finding which may be obfuscated in a traditional survey. Overall, our method of assessing public values shows that the candidate values we identified transcend both political ideologies and policy contexts.

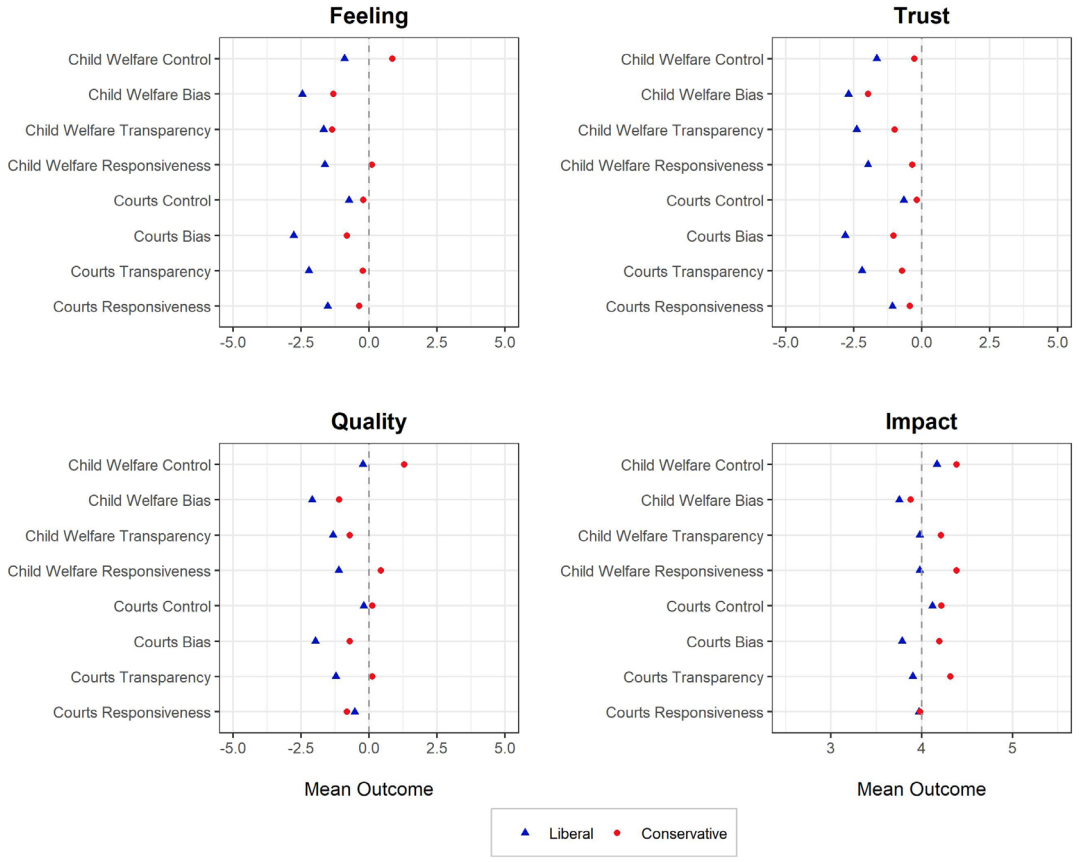

FIGURE 5 Average responses to outcome questions by policy sector and political ideology

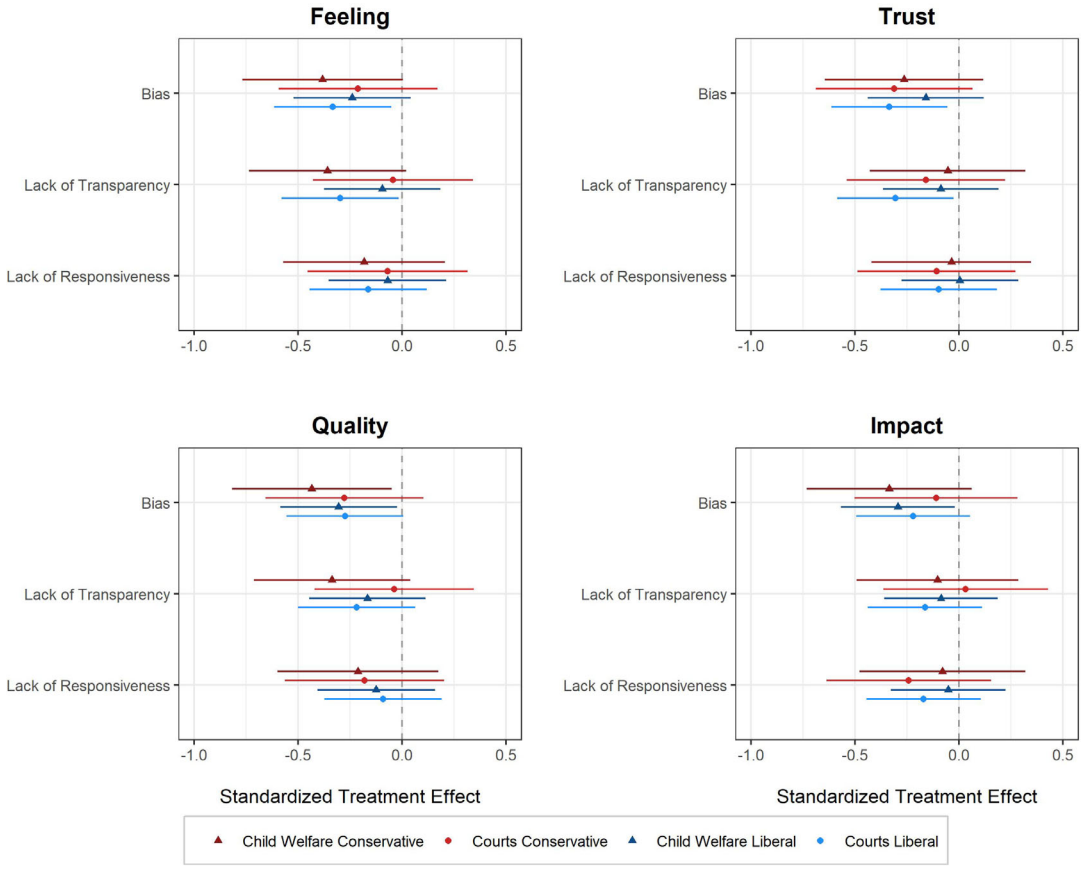

FIGURE 6 Treatment effects: Reactions to public value failure of automated decision systems by political ideology

5 | DISCUSSION

5.1 | Limitations

In any experimental study, external validity is determined by the representativeness of the sample and the authenticity of the treatments, context, and outcomes. Research indicates that the sampling frame of MTurk workers is typically younger, more liberal, and more likely to be male. Yet, the attitudes of both conservative and liberal MTurk workers tend to closely reflect those of their counterparts in the general population (Clifford et al., 2015). Therefore, this study employs separate subgroup analyses by political ideology to reveal any relevant differences, but the sample may not be directly generalizable to the US population in terms of age and education.

A second consideration is the authenticity of our treatments and context. In practice, any given AI application for government services may suffer from all, some, or none of the public value failures explored here, or from other public value failures not assessed in this study. Public value failures are also not necessarily binary (bias or no bias), as they can exist along a complex gradient, and may even interact. Our vignettes thus abstract away to some degree from real-world complexity, presenting each respondent with a single public value failure based on a single associated value. Critically, this approach allows us to isolate and rank the relative importance of individual public values. Moreover, while results based on these vignettes are necessarily dependent upon their particular construction, the

wording and associated public values are valid representations of actual news coverage of prominent government applications of ADS.

5.2 | Implications for scholars

The increasingly prominent role of AI and ADS in public administration presents both new risks and new opportunities (Andrews, 2019; Wirtz et al., 2019). On one hand, it has been argued that digital discretion can advance core ethical and democratic values such as through the standardization of government-citizen interactions (Busch & Henriksen, 2018) as well as key administrative goals like efficiency and productivity (Shrum et al., 2019). On the other hand, increasing public attention and scholarship (Bullock, 2019; Young et al., 2019) raise the specter of ethical and social harms resulting from AI use by governments. Our study addresses these rising tensions by advocating for the application of public value failure theory. The lens of public value failure, particularly when coupled with experimental methodology, can disentangle mixed impacts by revealing trade-offs and public priorities, and in so doing can help to differentiate between appropriate and inappropriate instances of government adoption of AI and ADS.

In particular, our findings suggest a number of potential avenues for future research. First, scholars might devote more attention to citizens' relative rankings of public values. In this study, we find that fairness is a top concern in the context of AI use by government (in relation to transparency and responsiveness), demonstrating that public values are not all prioritized equally. As system- and screen-level bureaucracies proliferate, scholars might consider how these shifts differentially affect those values deemed most important among the public, especially as ICT tools used by governments may impart their own implicit values and impede the expression of others.

Second, the primacy placed on fairness among our respondents invites further evaluation of this particularly salient public value. While recent work suggests delegation of decision-making to ADS may increase perceptions of fairness among some individuals (Araujo et al., 2020), it may erode trust among others, with potential variation across population subgroups and policy domains. In the vein of Miller and Keiser (2021), future work might devote more attention to different conceptualizations of fairness (e.g., procedural vs. distributive) (Marcinkowski et al., 2020), potential variation in assessments of fairness at individual versus group levels (Binns, 2020), and how assessments of public value success or failure may differ across socio-demographic subgroups.

Relatedly, our results highlight an important distinction in how the public views means-oriented versus ends-oriented values, or similarly, procedural versus substantive values (Bruijn & Dicke, 2006). Fairness is arguably an especially ends-oriented value, as bias in ADS can lead to direct impacts on citizens, such as racially disparate outcomes. On the other hand, transparency and responsiveness describe process- or means-oriented values, the neglect of which may produce negative impacts on citizens only indirectly or conditionally. Future work should consider how the transition to screen-level bureaucracy differentially impacts these important "groupings" or "constellations" of public values (Jørgensen & Bozeman, 2007). More generally, scholars of public administration might find our proposed experimental methodology useful in empirically testing whether certain theorized distinctions between classes of public values exist in practice, and which are most salient.

Finally, and in light of our claim that public opinion is the best arbiter of public value success or failure, the survey experiment methodology presented in this paper offers a promising way forward for the assessment of public values. Through priming individuals about hypothetical public value failures, we are able to assess the magnitude and relative importance of public values in particular contexts. Moreover, although our particular experiment did not find such differences, our methodology is able to detect differences across policy sectors, demographic subgroups, administrative implementation choices, or other factors of interest to scholars. These advancements can help to address several needs identified in the public value literature, namely empirical testing, conceptual refinement, ranking of values, and assessment of subgroup differences. In sum, use of the public value failure framework combined with experimental methods will allow scholars to probe the conceptual relationships between values implicated by government actions, consider conditions under which values are made salient, and assess a variety of citizen outcomes beyond those we explore here.

5.3 | Implications for practitioners

Public administrators and policymakers have a special responsibility to protect and promote public values. This is especially true in emerging contexts featuring extensive public expenditures and outstanding implementation decisions, like government use of AI. Importantly, practitioners often have significant choice over the design and implementation of ADS and can exercise this discretion to better realize and protect key public values. For example, when ADS are developed by private entities, governments can choose whether to adopt an ADS in the first place, may be able to select from competing providers, and can shape the solicitations and associated contracts. In addition, governments are increasingly encouraged to develop in-house expertise, which allows for even greater control over the design and implementation of ADS.

To inform these decisions, practitioners concerned with the potential implications of ADS implementation and public value realization should also look to scholarly research in this area, as clearer evidence on public values "could substantially inform policymaking, deliberation, and analysis" (Fisher et al., 2010). As an example, our results highlight the public's concern with the possible failure of ADS on the key values of fairness and transparency, a concern that practitioners can subsequently work to address using a variety of recently developed technical strategies and toolkits (see Mehrabi et al., 2019; Morley et al., 2019). Practitioners will also benefit from emerging AI industry standards and algorithmic impact assessments, such as Canada's Algorithmic Impact Assessment Tool, AI Now's Practical Framework for Public Agency Accountability, and the IEEE 7010-2020 Standard (Government of Canada, 2021; Reisman et al., 2018; Schiff et al., 2020). Finally, to complement research, administrators can foster increased attention to public values in their own practice, such as through forums to solicit public input and participatory design methods.

6 | CONCLUSION

As the adoption of AI by governments to enhance efficiency in service provision may result in supplanting or otherwise reprioritizing key public values, there is a need to better understand which public values should guide government implementation of AI systems. This is especially true at the present moment when major implementation, procurement, and regulatory decisions are being made that will shape AI's long-term impacts in the public domain. In turn, scholars of public administration have identified the need to clearly assess and compare public values in this context. This study demonstrates how survey experiment methodology can be used to quantify and rank public values based on citizen responses to potential public value failures.

Drawing on real-world cases and media discourse surrounding ADS, our survey experiment provides clear causal evidence that public value failures associated with AI have consistent negative impacts on evaluations of government. Our findings show statistically significant and substantial negative citizen reactions when there is failure along the values of fairness and transparency, with effect sizes on the order of one third and one quarter of a standard deviation, respectively. These results transcend both policy context and political ideology, and indicate a ranking of public values that better inform our understanding of the mixed impacts of AI use by governments. Moving forward, this study's methodological and conceptual contributions can serve as a springboard for scholars and practitioners interested in better understanding and realizing key public values in government.

ACKNOWLEDGMENTS

The authors wish to thank Natália Bueno, Gordon Kingsley, Jeffrey Ziegler, Maggie Macdonald, and anonymous reviewers for their guidance and helpful suggestions. This study was supported by a CFDE FIT Grant through Emory University.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

OPEN RESEARCH BAGES

This article has been awarded <Open Materials, Open Data, Preregistered Research Designs> badges. All materials and data are publicty accessible via the Open Science Framework at [provided URL]. Learn more about the Open Practices badges from the Center for Open Science: https://osf.io/tvyxz/wiki.

DATA AVAILABILITY STATEMENT

Complete replication code, data, and additional Supporting Information are available through the Public Administration Dataverse at https://doi.org/10.7910/DVN/LIGARA.

ORCID

Daniel S. Schiff https://orcid.org/0000-0002-4376-7303
Kaylyn Jackson Schiff https://orcid.org/0000-0002-4239-5915

ENDNOTES

[1] For example, Moore (1995) emphasizes how public value can be created through the actions of public administrators, akin to the creation of economic value by employees in the firm. Arguably, this is connected to Bozeman's conception of *public values* because their implementation can create *public value* in the Moorean sense.

[2] For readers interested in research contrasting human versus AI/automated decision-making, see, for example: Young et al. (2019), Logg et al. (2019), Araujo et al. (2020), and Miller and Keiser (2021).

[3] Available at https://osf.io/neu3j/

[4] Available at https://doi.org/10.7910/DVN/LIGARA

[5] Our design follows the advice of Bouwman and Grimmelikhuijsen (2016), who advocate for more experimental work based on a systematic review of experimental public administration research. As best practices, they encourage researchers to go beyond basic experimental designs, avoid convenience or student samples, use sources such as MTurk, and offer examination of novel topics relevant for scholars and practitioners alike. We adopt all of these recommendations.

[6] Research supports the validity of MTurk samples regarding political ideology and regarding citizens' attitudes towards public administration (Clifford et al., 2015). For more information about the quality of our survey responses from MTurk workers, see Section F in Appendix S1.

[7] See additional details about survey administration in Section B in Appendix S1, along with power analysis informing our sample size selection.

[8] See Section B in Appendix S1 for a description of the news coverage that motivated the vignette wording.

[9] See Section B in Appendix S1 for a discussion of our block randomization strategy.

[10] As we assigned treatment by blocks (described in Section B in Appendix S1), we control for the covariates used to create the blocks, effectively adding block dummies to the regression specifications. This is necessary for avoiding bias when assignment to treatment varies across blocks. While the probability of assignment to treatment is highly similar across our blocks, we nonetheless control for the blocking covariates and prefer covariate-adjusted regression specifications to produce our main results, as per our preanalysis plan. Additionally, we also calculate block-weighted estimates of the ATEs for our main results (DeclareDesign, 2018). These estimates, presented in Section C in Appendix S1, are nearly identical to covariate-unadjusted and adjusted regression results. Finally, including covariates is generally good practice in the analysis of experiments, as they soak up unexplained variance in the outcomes, thereby reducing standard errors.

[11] Covariate balance information and descriptive information on the covariates and outcome measures are presented in Section B in Appendix S1. Tables and figures for additional analyses, including unadjusted regressions and subgroup-specific findings, are available in Section C in Appendix S1. Section C in Appendix S1 also contains notes on statistical significance and multiple testing. Results for exploratory outcome measures are presented and discussed in Section D in Appendix S1.

[12] On a standardized scale, these effects range from −0.23 (impact) to −0.31 (quality) standard deviations.

[13] A full table of regression results is presented in Section C in Appendix S1.

[14] For example, the largest magnitude difference between policy sectors concerns the effect of the transparency treatment on trust (p-value for z-test = 0.22), not reaching significance at the 5% alpha level. However, this may be due to

insufficient power, as splitting the four primary experimental groups further into two separate policy sectors reduces the sample size by half.

[15] Confidence intervals are not depicted in Figure 5 for purposes of readability, and are largely overlapping.

[16] Nevertheless, some of these trends are explored more in Section C in Appendix S1.

REFERENCES

Adadi, A. & Berrada, M. (2018) Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access*, *6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052.

AI Now Institute. (2019) *Automated decision systems: examples of government use cases*. Available at: https://ainowinstitute.org/nycadschart.pdf

Alford, J. & O'Flynn, J. (2009) Making sense of public value: concepts, critiques and emergent meanings. *International Journal of Public Administration*, *32*(3–4), 171–191. https://doi.org/10.1080/01900690902732731.

Andersen, K.N., Henriksen, H.Z., Medaglia, R., Danziger, J.N., Sannarnes, M.K. & Enemærke, M. (2010) Fads and facts of E-government: a review of impacts of E-government (2003–2009). *International Journal of Public Administration*, *33*(11), 564–579. https://doi.org/10.1080/01900692.2010.517724.

Anderson, E. (1995) *Value in ethics and economics*. Cambridge, Massachusetts: Harvard University Press.

Andrews, L. (2019) Public administration, public leadership and the construction of public value in the age of the algorithm and "big data". *Public Administration*, *97*(2), 296–310. https://doi.org/10.1111/padm.12534.

Araujo, T., Helberger, N., Kruikemeier, S. & de Vreese, C.H. (2020) In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & SOCIETY*, *35*(3), 611–623. https://doi.org/10.1007/s00146-019-00931-w.

Barth, T.J. & Arnold, E. (1999) Artificial intelligence and administrative discretion: implications for public administration. *The American Review of Public Administration*, *29*(4), 332–351. https://doi.org/10.1177/02750749922064463.

Bekkers, V. & Homburg, V. (2007) The myths of E-government: looking beyond the assumptions of a new and better government. *The Information Society*, *23*(5), 373–382. https://doi.org/10.1080/01972240701572913.

Binns, R. (2020) On the apparent conflict between individual and group fairness. In: *Proceedings of the 2020 conference on fairness, accountability, and transparency*, pp. 514–524. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3351095.3372864.

Bouwman, R. & Grimmelikhuijsen, S. (2016) Experimental public administration from 1992 to 2014: a systematic literature review and ways forward. *International Journal of Public Sector Management*, *29*(2), 110–131. https://doi.org/10.1108/IJPSM-07-2015-0129.

Bovens, M. & Zouridis, S. (2002) From street-level to system-level bureaucracies: how information and communication technology is transforming administrative discretion and constitutional control. *Public Administration Review*, *62*(2), 174–184. https://doi.org/10.1111/0033-3352.00168.

Bozeman, B. (2002) Public-value failure: when efficient markets may not do. *Public Administration Review*, *62*(2), 145–161. https://doi.org/10.1111/0033-3352.00165.

Bozeman, B. (2009) Public values theory: three big questions. *International Journal of Public Policy*, *4*(5), 369–375. https://doi.org/10.1504/IJPP.2009.025077.

Bozeman, B. & Sarewitz, D. (2011) Public value mapping and science policy evaluation. *Minerva*, *49*(1), 1–23. https://doi.org/10.1007/s11024-011-9161-7.

Bruijn, H.D. & Dicke, W. (2006) Strategies for safeguarding public values in liberalized utility sectors. *Public Administration*, *84*(3), 717–735. https://doi.org/10.1111/j.1467-9299.2006.00609.x.

Bryson, J.M., Crosby, B.C. & Bloomberg, L. (2014) Public value governance: moving beyond traditional public administration and the new public management. *Public Administration Review*, *74*(4), 445–456. https://doi.org/10.1111/puar.12238.

Buffat, A. (2015) Street-level bureaucracy and E-government. *Public Management Review*, *17*(1), 149–161. https://doi.org/10.1080/14719037.2013.771699.

Bullock, J.B. (2019) Artificial intelligence, discretion, and bureaucracy. *The American Review of Public Administration*, *49*(7), 751–761. https://doi.org/10.1177/0275074019856123.

Buolamwini, J. & Gebru, T. (2018) Gender shades: intersectional accuracy disparities in commercial gender classification. In *Proceedings of Machine Learning Research*, 81:77–91. New York, NY: USA. http://proceedings.mlr.press/v81/buolamwini18a.html.

Busch, P.A. & Henriksen, H.Z. (2018) Digital discretion: a systematic literature review of ICT and street-level discretion. *Information Polity*, *23*(1), 3–28. https://doi.org/10.3233/IP-170050.

Caldeira, G.A. & Cowart, A.T. (1980) Budgets, institutions, and change: criminal justice policy in America. *American Journal of Political Science*, *24*(3), 413–438. https://doi.org/10.2307/2110826.

Castelvecchi, D. (2016) Can we open the black box of AI? *Nature News*, *538*(7623), 20–23. https://doi.org/10.1038/538020a.

Chouldechova, A. (2017) Fair prediction with disparate impact: a study of bias in recidivism prediction instruments. *Big Data*, *5*(2), 153–163. https://doi.org/10.1089/big.2016.0047.

Chouldechova, A., Benavides-Prado, D., Fialko, O. & Vaithianathan, R. (2018) A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions. In *Proceedings of Machine Learning Research*, 134–148. New York, NY, USA. http://proceedings.mlr.press/v81/chouldechova18a.html.

Clifford, S., Jewell, R.M. & Waggoner, P.D. (2015) Are samples drawn from Mechanical Turk valid for research on political ideology? *Research & Politics*, *2*(4), 2053168015622072. https://doi.org/10.1177/2053168015622072.

Cordelia, A. (2006) Transaction costs and information systems: does IT add up? *Journal of Information Technology*, *21*(3), 195–202. https://doi.org/10.1057/palgrave.jit.2000066.

Cordella, A. & Bonina, C.M. (2012) A public value perspective for ICT enabled public sector reforms: a theoretical reflection. *Government Information Quarterly*, *29*(4), 512–520. https://doi.org/10.1016/j.giq.2012.03.004.

DeclareDesign. (2018) *The trouble with "controlling for blocks"*. DeclareDesign, 9 October. Available at: https://declaredesign.org/blog/biased-fixed-effects.html

DeLone, W.H. & McLean, E.R. (1992) Information systems success: the quest for the dependent variable. *Information Systems Research*, *3*(1), 60–95. https://doi.org/10.1287/isre.3.1.60.

Denhardt, R.B., & Denhardt, J.V. (2000) "The New Public Service: Serving Rather Than Steering". *Public Administration Review*. *60*(6), 549–559. http://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=3060584.

Diakopoulos, N., Trielli, D. & Baek, S. (2020) *Algorithm tips—resources and leads for investigating algorithms in society*. Available at: http://algorithmtips.org/

Dunleavy, P. (2005) New public management is dead—long live digital-era governance. *Journal of Public Administration Research and Theory*, *16*(3), 467–494. https://doi.org/10.1093/jopart/mui057.

Eubanks, V. (2017) *Automating inequality: how high-tech tools profile, police, and punish the poor (first edition)*. New York, NY: St. Martin's Press.

European Commission. (2019) *Ethical guidelines for trustworthy AI*. Brussels, Belgium: European Commission, High-Level Expert Group on Artificial Intelligence. Available at: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=58477.

Fisher, E., Slade, C.P., Anderson, D. & Bozeman, B. (2010) The public value of nanotechnology? *Scientometrics*, *85*(1), 29–39. https://doi.org/10.1007/s11192-010-0237-1.

Fukumoto, E. & Bozeman, B. (2019) Public values theory: what is missing? *The American Review of Public Administration*, *49*(6), 635–648. https://doi.org/10.1177/0275074018814244.

Gormley, W.T. (1986) Regulatory issue networks in a federal system. *Polity*, *18*(4), 595–620. https://doi.org/10.2307/3234884.

Government of Canada. (2021) *Algorithmic impact assessment tool*. Available at: https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html

Gruening, G. (2001) Origin and theoretical basis of new public management. *International Public Management Journal*, *4*(1), 1–25. https://doi.org/10.1016/S1096-7494(01)00041-1.

Gulick, L. (1984) The metaphors of public administration. *Public Administration Quarterly*, *8*(3), 369–381.

Hartley, J., Alford, J., Knies, E. & Douglas, S. (2017) Towards an empirical research agenda for public value theory. *Public Management Review*, *19*(5), 670–685. https://doi.org/10.1080/14719037.2016.1192166.

Hidalgo, N. (2020) *Re: Oversight hearing of Local Law 49 of 2018 (Open Algorithms Law) & Int 1806–2019 (aka ADS transparency) & Int 1447–2019 (aka data inventory)*. Available at: https://beta.nyc/2020/01/22/re-oversight-hearing-of-local-law-49-of-2018-open-algorithms-law-int-1806-2019-aka-ads-transparency-int-1447-2019-aka-data-inventory/

Howard, A. & Borenstein, J. (2017) The ugly truth about ourselves and our robot creations: the problem of bias and social inequity. *Science and Engineering Ethics*, *24*(5), 1521–1536. https://doi.org/10.1007/s11948-017-9975-2.

Jørgensen, T.B. & Bozeman, B. (2007) Public values: an inventory. *Administration & Society*, *39*(3), 354–381. https://doi.org/10.1177/0095399707300703.

Karunasena, K. & Deng, H. (2012) Critical factors for evaluating the public value of e-government in Sri Lanka. *Government Information Quarterly*, *29*(1), 76–84. https://doi.org/10.1016/j.giq.2011.04.005.

Kearns, I. (2004) *Public value and e-government*. London: Institute for Public Policy Research.

Kelly, G., Mulgan, G. & Muers, S. (2002) *Creating public value: an analytical framework for public service reform*. London: Strategy Unit, Cabinet Office.

Logg, J.M., Minson, J.A. & Moore, D.A. (2019) Algorithm appreciation: people prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005.

Marcinkowski, F., Kieslich, K., Starke, C. & Lünich, M. (2020) Implications of AI (un-)fairness in higher education admissions: the effects of perceived AI (un-)fairness on exit, voice and organizational reputation. In: *Proceedings of the 2020 conference on fairness, accountability, and transparency*, pp. 122–130. Barcelona, Spain: https://doi.org/10.1145/3351095.3372867.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. & Galstyan, A. (2019) A survey on bias and fairness in machine learning. *ArXiv:1908.09635*. Available at: http://arxiv.org/abs/1908.09635

Miller, S.M. & Keiser, L.R. (2021) Representative bureaucracy and attitudes toward automated decision making. *Journal of Public Administration Research and Theory*, *31*(1), 150–165. https://doi.org/10.1093/jopart/muaa019.

Moore, M.H. (1995) *Creating public value: strategic management in government*. Cambridge, Massachusetts: Harvard University Press.
Morley, J., Floridi, L., Kinsey, L. & Elhalal, A. (2019) From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics*, *26*(4), 2141–2168. https://doi.org/10.1007/s11948-019-00165-5.
Nabatchi, T. (2018) Public values frames in administration and governance. *Perspectives on Public Management and Governance*, *1*(1), 59–72. https://doi.org/10.1093/ppmgov/gvx009.
Nye, J.S. (1997) In government we don't trust. *Foreign Policy*, *108*, 99–111. https://doi.org/10.2307/1149092.
O'Flynn, J. (2007) From new public management to public value: paradigmatic change and managerial implications. *Australian Journal of Public Administration*, *66*(3), 353–366. https://doi.org/10.1111/j.1467-8500.2007.00545.x.
Omar, K., Scheepers, H. & Stockdale, R. (2011) EGovernment service quality assessed through the public value lens. In: Janssen, M., Scholl, H.J., Wimmer, M.A. & Tan, Y. (Eds.) *Electronic government*. Berlin, Heidelberg: Springer, pp. 431–440. https://doi.org/10.1007/978-3-642-22878-0_36.
Reisman, D., Schultz, J., Crawford, K., & Whittaker, M. (2018) Algorithmic impact assessments: a practical framework for public agency accountability. *AI Now*. Available at: https://ainowinstitute.org/aiareport2018.pdf
Rogers, J.D. & Kingsley, G. (2004) Denying public value: the role of the public sector in accounts of the development of the internet. *Journal of Public Administration Research and Theory*, *14*(3), 371–393. https://doi.org/10.1093/jopart/muh024.
Salamon, L.M. & Elliott, O.V. (Eds.). (2002) *The tools of government: a guide to the new governance*. Oxford, England: Oxford University Press.
Schiff, D., Ayesh, A., Musikanski, L. & Havens, J.C. (2020) IEEE 7010: a new standard for assessing the well-being implications of artificial intelligence. In: *2020 IEEE international conference on systems, man, and cybernetics (SMC)*, pp. 2746–2753. Toronto, Canada: IEEE. https://doi.org/10.1109/SMC42975.2020.9283454.
Schiff, D., Borenstein, J., Laas, K. & Biddle, J. (2021) AI ethics in the public, private, and NGO sectors: a review of a global document collection. *IEEE Transactions on Technology and Society*, *2*, 1–12. https://doi.org/10.1109/TTS.2021.3052127.
Schneider, A. & Ingram, H. (1993) Social construction of target populations: implications for politics and policy. *American Political Science Review*, *87*(2), 334–347. https://doi.org/10.2307/2939044.
Shrum, K., Gordon, L., Regan, P., Maschino, K., Shark, A.R. & Shropshire, A. (2019) *AI and its impact on public administration*. Washington, D.C.: National Academy of Public Administration. Available at: https://www.napawash.org/uploads/Academy_Studies/9781733887106.pdf.
Slade, C.P. (2011) Public value mapping of equity in emerging Nanomedicine. *Minerva*, *49*(1), 71–86. https://doi.org/10.1007/s11024-011-9163-5.
Stoker, G. (2006) Public value management: a new narrative for networked governance? *The American Review of Public Administration*, *36*(1), 41–57. https://doi.org/10.1177/0275074005282583.
Stone, D.A. (1997) *Policy paradox: the art of political decision making*. New York, NY: W.W. Norton.
Weatherford, M.S. (1992) Measuring political legitimacy. *The American Political Science Review*, *86*(1), 149–166. https://doi.org/10.2307/1964021.
West, E.A. & Iyengar, S. (2020) Partisanship as a social identity: implications for polarization. *Political Behavior*, 1–32. https://doi.org/10.1007/s11109-020-09637-y.
Williams, I. & Shearer, H. (2011) Appraising public value: past, present and futures. *Public Administration*, *89*(4), 1367–1384. https://doi.org/10.1111/j.1467-9299.2011.01942.x.
Wirtz, B.W., Weyerer, J.C. & Geyer, C. (2019) Artificial intelligence and the public sector—applications and challenges. *International Journal of Public Administration*, *42*(7), 596–615. https://doi.org/10.1080/01900692.2018.1498103.
Young, M.M., Bullock, J.B. & Lecy, J.D. (2019) Artificial discretion as a tool of governance: a framework for understanding the impact of artificial intelligence on public administration. *Perspectives on Public Management and Governance*, *2*(4), 301–303. https://doi.org/10.1093/ppmgov/gvz014.
Zhang, B., & Dafoe, A. (2019) *Artificial intelligence: American attitudes and trends* (SSRN Scholarly Paper ID 3312874). Social Science Research Network. Available at: https://papers.ssrn.com/abstract=3312874

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Schiff D. S., Schiff K. J., & Pierson P. (2021). Assessing public value failure in government adoption of artificial intelligence. *Public Administration*, 1–21. https://doi.org/10.1111/padm.12742