Science and Public Policy, 2023, **50**, 161–176
DOI: https://doi.org/10.1093/scipol/scac056
Advance Access Publication Date: 14 October 2022
Article

What governs attitudes toward artificial intelligence adoption and governance?

Matthew R. O'Shaughnessy[1,*], Daniel S. Schiff[2], Lav R. Varshney[3], Christopher J. Rozell[1,†] and Mark A. Davenport[1,†]

[1]School of Electrical and Computer Engineering, Georgia Institute of Technology, 777 Atlantic Dr NW, Atlanta, GA 30332, USA, [2]Department of Political Science, Purdue University, 110 North University St, West Lafayette, IN 47907, USA and [3]Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 306 N Wright St MC 702, Urbana, IL 61801, USA

[†]These authors contributed equally.
*Corresponding author. E-mail: moshaughnessy6@gatech.edu

Abstract

Designing effective and inclusive governance and public communication strategies for artificial intelligence (AI) requires understanding how stakeholders reason about its use and governance. We examine underlying factors and mechanisms that drive attitudes toward the use and governance of AI across six policy-relevant applications using structural equation modeling and surveys of both US adults ($N = 3{,}524$) and technology workers enrolled in an online computer science master's degree program ($N = 425$). We find that the cultural values of individualism, egalitarianism, general risk aversion, and techno-skepticism are important drivers of AI attitudes. Perceived benefit drives attitudes toward AI use but not its governance. Experts hold more nuanced views than the public and are more supportive of AI use but not its regulation. Drawing on these findings, we discuss challenges and opportunities for participatory AI governance, and we recommend that *trustworthy AI governance* be emphasized as strongly as trustworthy AI.

Key words: artificial intelligence policy; public opinion; public engagement.

1. Introduction

Artificial intelligence (AI) may fundamentally reshape our economy and society, but across wide variety of application areas, its prospective benefits are accompanied by potential harms. For example, AI's impact on economic growth may be felt unevenly across the labor market. The use of AI in new medical systems raises questions about trust, fairness, and privacy even as it enables new treatments. AI-based systems provide new tools for free expression while simultaneously powering authoritarian crackdowns and the spread of disinformation.

Realizing the benefits of emerging technologies like AI while mitigating their accompanying harms requires governance strategies that are respectful of the diverse values and beliefs held by the public (Stirling 2008; Macnaghten and Chilvers 2014; Ulnicane et al. 2020; Stix 2021). Inclusive and participatory governance is a central pillar of AI development frameworks released by academic, industry, government, and international groups (Organisation for Economic Cooperation and Development 2019; IEEE 2019; European Group on Ethics in Science and New Technologies 2018; United States Office of Management and Budget 2020). In representative suggestions, IEEE's framework suggests that developers and regulators of AI should remain aware of the 'diversity of cultural norms among users' (IEEE 2019) while the AI Now Institute stresses the importance of expanding 'cultural, disciplinary, and ethnic diversity' in the development and governance of AI (Campolo et al. 2017).

However, the technical complexity of AI makes it difficult to design governance structures that the public can participate in effectively. As a result, discourse about AI governance can become opaque and expert-based, making the policy process ineffective at representing diverse viewpoints, vulnerable to capture by vested interests (Ulnicane et al. 2020), and liable to 'ethics-washing' (Stix 2021; Sloane et al. 2020). Moreover, while recent opinion surveys have found that the US public is generally supportive of AI (Morning Consult 2017; European Commission 2017; Smith and Anderson 2017; The Harris Poll 2017; Gallup, Inc 2018; Morning Consult 2018; Smith 2018a, b; Zhang and Dafoe 2019; United Kingdom Government 2019; Johnson and Tyson 2020), their awareness of it is limited (DeCario and Etzioni 2021): even as AI is pervasive in applications like resume screening and credit scoring, surveys have found little public support for AI in these 'sensitive' settings (Smith 2018a). These seemingly contradictory views suggest that public opinion may change rapidly as AI's capabilities, limitations, and societal impacts become more apparent.

Ensuring that diverse public opinion is respected in AI governance processes thus requires that AI developers and policymakers better understand the underlying values and motivations that shape how public attitudes toward AI could evolve. This understanding is also critical for equipping the public to meaningfully engage with AI governance: science communication literature suggests that processes for public outreach and dialogue are most effective when they are

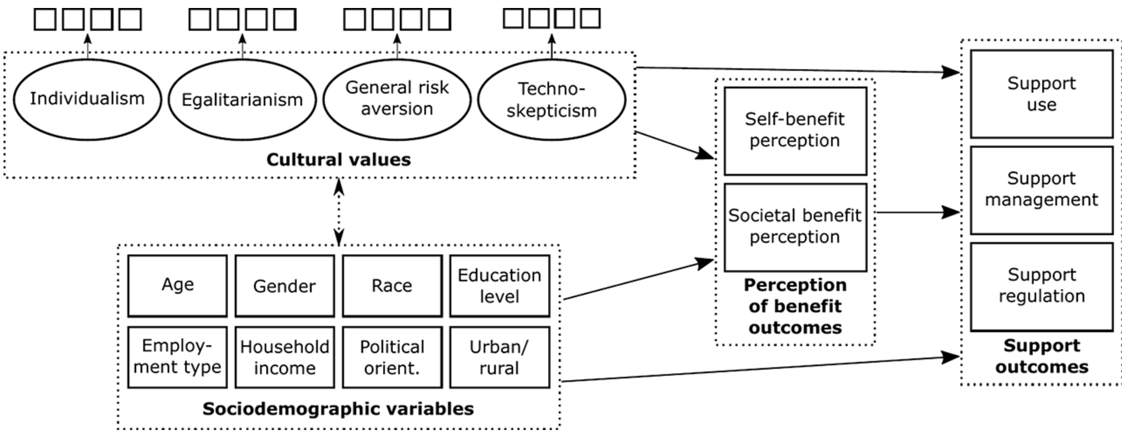

Figure 1. SEM used for analysis. The full SEM, S, allows variables within each group (denoted by dashed boxes) as well as cultural values and sociodemographic variables to covary; we treat demographic variables as exogenous. Two nested models are used in our analysis: $S_{\setminus C}$, which constrains paths from cultural value constructs to outcome variables to zero, and $S_{\setminus B}$, which constrains paths from perceived benefit outcome variables to support outcome variables to zero.

tailored to the public's values, beliefs, and motivations (Kahan et al. 2011; Lupia 2013). Although previous work has explored how attitudes of AI professionals (Zhang et al. 2021; Aiken et al. 2020) and the public (Morning Consult 2017; European Commission 2017; Smith and Anderson 2017; The Harris Poll 2017; Gallup, Inc 2018; Morning Consult 2018; Smith 2018a, b; Zhang and Dafoe 2019; United Kingdom Government 2019; Johnson and Tyson 2020) differ across sociodemographic groups, little existing work has explored the *underlying values and mechanisms that drive attitudes toward AI*.

In this paper, we take a step toward better understanding what shapes attitudes toward AI by looking at factors and mechanisms beyond sociodemographic characteristics. We explore the following questions, which are key to designing effective AI governance and science communication strategies:

(1) How do sociodemographic factors, cultural values, and perceived benefit influence attitudes toward AI?
(2) How do these attitudes—and the factors that inform them—differ between experts and the public?
(3) How do these attitudes—and the factors that inform them—differ across common contexts of AI use?

To explore these questions, we conducted two online surveys in April and May 2021. The first survey sampled $N = 3{,}524$ US adults recruited and compensated through the Lucid Theorem platform, which uses quota sampling to obtain participants representative of adult US residents on age, gender, race, and region. The second survey sampled $N = 425$ students who had recently completed a graduate AI course at Georgia Tech. Most (93.9 per cent) of these students had undergraduate degrees in technical subjects, and 93.5 per cent previously or concurrently worked in computer science or another Science, Technology, Engineering, and Math (STEM) field.[1] In addition to standard sociodemographic variables, we consider the impact on attitudes of perceived self- and societal benefits and of the cultural values of individualism, egalitarianism, general risk aversion, and techno-skepticism—constructs found to inform the perception of many other technological risks (Kahan et al. 2007; Renn and Benighaus 2013; Tennant et al. 2019).

The main contribution of this work is to increase understanding of attitudes toward AI use and governance by (1) exploring a set of attitudinal drivers that is broader than the typically-considered sociodemographic variables, including both perceptions of benefit, and cultural values inspired by the cultural theory literature; (2) directly comparing the attitudes of experts and the public; and (3) considering attitudes across a range of policy-relevant contexts of AI use. Our preregistered analysis strategy uses the structural equation model (SEM) shown in Fig. 1 (described in more detail in the next section), which allows us to naturally address the three key research questions defined above. Our results provide insights that can aid policymakers in crafting governance strategies that are respectful of diverse beliefs and assist AI developers in effectively communicating the broader implications of their work to the public. Drawing on these results, we offer recommendations for engaging the public in dialogue about AI governance and offer suggestions for future research.

2. Background and theory

2.1 Underlying factors governing attitudes toward technology

Prior work has found that race, gender, and political ideology (Wildavsky and Dake 1990) are highly predictive of attitudes toward issues such as nuclear power (Slovic et al. 1991), climate change (Flynn et al. 1994), genetically engineered food (Finucane et al. 2000), and radiation (Peters et al. 2004). Similar sociodemographic divides have been found in attitudes toward AI. Those reporting familiarity and comfort with AI are more likely to be young, be male, be educated, live in urban areas, and have higher incomes (Morning Consult 2017, 2018; Zhang and Dafoe 2019; United Kingdom Government 2019; Johnson and Tyson 2020; Morning Consult 2021). Sociodemographic divides also shape perceptions of AI's impact on society. Those in urban areas, blue-collar workers, and political liberals are more likely to believe that AI will deepen inequality and reduce employment (Morning Consult 2017; Gallup, Inc 2018), while those with more education, white-collar jobs, and higher incomes are more likely to believe that AI will be beneficial to society and the economy

(Morning Consult 2017; Smith and Anderson 2017; Gallup, Inc 2018; Morning Consult 2018; Zhang and Dafoe 2019).

The cultural theory of risk perception posits that 'cultural' worldviews can be more concise and informative predictors of attitudes toward technological risk than sociodemographic factors alone (Kahan et al. 2007; Weber and Hsee 2000; Johnson and Swedlow 2021). These cultural values have been hypothesized to define identity groups, imbue potential risks with affective qualities (Peters et al. 2004), and encourage biased information processing (Lord et al. 1979). Indeed, literature has found that successfully communicating scientific topics to the public benefits from careful attention to how messages may interact with the cultural values held by the public (Kahan et al. 2011; Lupia 2013; Druckman and McGrath 2019). For policymakers seeking to design inclusive governance and communication strategies, it is critical to understand how cultural values relate to views on AI and whether this relationship differs across specific AI use cases.

We use two cultural values that originate with the grid-group cultural theory of Douglas and Wildavsky (1982), were operationalized for survey research by Kahan et al.'s 'cultural cognition theory' (Kahan 2012), and have been identified as salient to technological risk perception (Kahan et al. 2007; Dake 1991). The first represents attitudes toward the role of individuals in society: *individualists* favor social orderings in which individuals are responsible for 'securing their own well-being without assistance or interference from society' and thus prefer to minimize the role of government when ensuring collective welfare comes into tension with individual preferences (Kahan et al. 2011). The second cultural value represents attitudes toward well-defined social hierarchies: *egalitarians* favor greater equality between groups defined by race, gender, wealth, and political power; they spurn stratified social orderings based on fixed characteristics. Literature on risk analysis and related disciplines have used cultural theory generally—and the conceptions of individualism and egalitarianism we borrow from cultural cognition theory in particular—to explain differences in opinion between environmentalists and the public (Ellis and Thompson 1997), disagreements on controversial issues such as gun control and global warming (Kahan et al. 2007), and divides in acceptance of scientific consensus (Kahan et al. 2011).

We also consider two cultural values that describe general attitudes toward risk and technology. First, many individuals tend to avoid small risks even at the cost of foregoing larger benefits; general *risk aversion* has been found to be a powerful predictor of attitudes toward technology (Renn and Benighaus 2013). Here we use the risk aversion construct of Sharma (Sharma 2010), which assesses attitudes toward general lifestyle risks. Second, *techno-skeptics* are uncomfortable with the use of new technology, cynical about the intentions of groups developing new technological advancements, and opposed to the use of technology to solve social problems (Meadows et al. 1972; Krier and Gillette 1985). Techno-skepticism has been found to partially explain divides in opinion on topics such as nuclear waste (Barke et al. 1997), climate change adaptation (Gardezi and Arbuckle 2020), and autonomous vehicles (Tennant et al. 2019). In the context of AI, techno-optimism and techno-skepticism are well-reflected in popular narratives about utopian and dystopian scenarios driven by AI (Cave et al. 2018).

2.2 Perceived benefit and hypothesized model

In contrast to technologies whose benefits are perceived as broadly shared, popular narratives about AI often feature clear losers (Fast and Horvitz 2017): workers who lose their jobs to automation, for example, or minorities who suffer discrimination at the hands of automated decision systems. These narratives may make views about AI governance—perhaps more so than views about other technological risks—subject to perceptions of who stands to benefit and lose from the continually increasing adoption of AI. However, while there is some evidence that perceived self-interest informs support for AI-based technologies (Morning Consult 2018; Liu et al. 2019; Dixon et al. 2020), other literature has suggested that perceived benefit does not always eclipse affective and value-based concerns (Sears and Funk 1991; Chong et al. 2001). To evaluate how perceived benefit influences attitudes toward AI (and understand how it is influenced by sociodemographic variables and cultural values), we use an SEM (Kline 2016) analysis framework.

The SEM that forms the core of our analysis describes hypothesized relationships between demographic variables, cultural values, perceived individual and societal benefits from AI, and support for AI use and governance. The SEM also mathematically defines how each variable is measured. In SEM analysis, model parameters (e.g. path coefficients and (co)variances) are estimated by minimizing the difference between the observed covariance matrix and the model-implied covariance matrix according to a certain statistical criterion (Kline 2016).

Our model, shown in Fig. 1, assumes that demographic variables and cultural values drive both categories of outcome measures (perceived benefit, support for AI adoption, and governance), but that the reverse driving relationships do not exist. This reflects the assumption that cultural values are broad concepts likely to integrate beliefs and experiences from a wide variety of sources and that views about AI are unlikely to be sufficiently present in the public discourse to fundamentally alter cultural values.[2] Each cultural value construct was measured by four survey items. While the cultural value constructs were allowed to covary in our SEM, each survey response item was modeled as independent (i.e. each survey items are independent of each other when conditioned on their parent construct). Our SEM also assumes that perceived self- and societal benefits drive support for AI use and governance, but that the support outcomes do not drive perceived benefit.

The relationship between sociodemographic variables and cultural values is a more subtle question. For example, it seems likely that age and gender drive cultural values, and conversely, literature has suggested that cultural values drive political orientation (Wildavsky and Dake 1990). Our model includes sociodemographic variables as exogenous variables, allowing unmodeled covariance between them and between sociodemographic variables and cultural values. This represents the possibility that there exist causal relationships between these variables, or that unmodeled confounding is present. These covariances are denoted by the bidirectional dotted line in Fig. 1. Similarly, variables within each group may be causally related or be jointly affected by unmodeled variables. For example, techno-skepticism and risk aversion may be driven by individualism and egalitarianism, rather than existing as discrete constructs.[3] We model this by allowing variables within each group (sociodemographic variables, cultural values, perception of benefit, and AI support) to covary.

Our SEM bears some similarities to popular models of technology acceptance and adoption used in psychology and marketing research literature. The theory of reasoned action (Fishbein and Ajzen 1975) focuses on the relationship between behavior and behavioral intention, which is modeled as being shaped by attitudes and subjective norms. The Multi-Attribute Attitude Model (Fishbein and Funke 1976) models an individual's attitude toward a brand or product as a weighted linear combination of attributes. Unlike this model, in which each individual is modeled by a unique set of weights, our SEM models all respondents collectively with a single set of inferred parameters. The influential Technology Acceptance Model (Davis 1985) posits that attitudes toward technology use are governed by perceived usefulness and ease of use, which are in turn governed by a set of 'external factors'. While extensions of this model use more extensive sets of external factors (including culturally-relevant variables such as gender (Venkatesh and Morris 2000)), the set of sociodemographic and cultural variables we use in our SEM is broader than typically considered in this literature.

2.3 Differences between experts and the public

It is particularly important to understand the ways in which public and expert attitudes diverge when discourse about policy is dominated by experts. Research on other emerging technologies has suggested that technical experience often negatively associates with risk perception, with experts tending to be particularly tolerant of risks stemming from technology aligned with their discipline (Barke and Jenkins-Smith 1993; Sjöberg and Drottz-Sjöberg 1993). Restricting policy discourse to those who are most knowledgeable therefore threatens to limit the influence of the very people who may perceive the most risk. Previous work has also found that scientists' views on risk vary based on gender, institutional affiliation, and cultural and political values (Barke et al. 1997, 2006; Funk et al. 2015). AI experts differ from the public along each of these dimensions; failure to appreciate how these factors influence attitudes toward AI may hinder the creation of inclusive policy dialogue.

Indeed, prior surveys comparing the attitudes of AI experts and the public have found major differences in the trust placed in government, technology companies, the US military, and international organizations (Zhang et al. 2021; Aiken et al. 2020), suggesting a potentially wide gulf in attitudes toward who should be responsible for governing AI. AI professionals also differ from the public on many sociodemographic variables that typically predict regulatory preferences: compared to the public, AI practitioners tend to be better educated, be more racially diverse but overwhelmingly male, have higher income, and live in more urban areas (Zhang et al. 2021). Understanding expert attitudes is particularly relevant in the context of AI because technology workers have demonstrated substantial leverage in determining where and how AI is used and governed (Belfield 2020).

2.4 Differences across use contexts

Further complicating the design of inclusive governance and science communication strategies is the diversity of contexts in which AI can be used. This diversity makes it difficult to know how findings relevant to AI's impact on labor automation, for example, generalize to AI used in medical research or automated weapons systems. To better understand these differences, in addition to examining attitudes toward AI in general, we explore attitudes toward AI used in six policy-relevant contexts: predictive policing, labor automation, medical diagnosis, automated vehicles, personalization, and weapon systems (see Section 3.3 and Supplement Section B for more details on these contexts).

The use of AI in each of these contexts raises different questions about risks, distribution of impacts, and ethical questions like fairness. Modeling each of these contexts allows us to understand how the factors we study—sociodemographic variables, cultural values, and perceived benefit—impact attitudes differently across application areas.

3. Methods

Our survey and analysis procedure were preregistered at https://osf.io/pcsvf/. Supplement Section E contains results from the complete analysis procedure specified in the preregistration; Supplement Section G describes minor deviations from the preregistration. The research was approved by the Georgia Tech institutional review board under the protocol number H21112.

3.1 Data

Our first sample consisted of $N = 3{,}524$ US adult participants recruited and compensated online through the Lucid Theorem platform, which uses quota sampling to match the US census marginal distributions on age, gender, ethnicity, and region. Previous research has found that samples provided by Lucid provide results generally similar to US probability samples or samples provided by Amazon Mechanical Turk (Coppock and McClellan 2019). However, this sample may not generalize to US adults on dimensions such as comfort with technology. Recent studies have found decreased participant attention on Lucid and other online survey platforms coinciding with the Covid-19 pandemic (Aronow et al. 2020; Peyton et al. 2021); we expected that this would reduce effect sizes. As a robustness check, we replicated our results with inattentive respondents removed (see Section 3.4). The completion rate (defined as the number of participants entering the survey who completed it) for this sample was 86 per cent.

Our second sample consisted of $N = 425$ master's students at the conclusion of a graduate-level AI class in Georgia Tech's Online Master of Science (OMS) in Computer Science (OMSCS) or Analytics programs. OMS students have undergraduate degrees in technical subjects, and in 2020, most of them worked full-time in technical fields in the industry while completing the degree. In their current and post-graduation roles, most will be in a position to have an impact on how AI is used and governed. Recruitment materials for this sample are provided in Supplement Section C. Participants were provided course extra credit, and nonparticipants were offered an alternative method for obtaining the extra credit. The response rate for this sample was 61.7 per cent.

Differences between these two samples go beyond academic and professional AI-related experience. In 2020, 81 per cent of OMSCS students were male and over one-third were not US citizens or permanent residents. While the OMSCS program has enrolled students from 122 countries and 53 US

Table 1. Means, standard deviations, 95 per cent confidence intervals for differences in means, and *P*-value (Welch's two-tailed *t*-test) for each variable in the US public (Lucid) and expert (OMS) samples. Gender was coded as a binary variable (male, female or other gender), and age was coded using Pew's classification of generational groups (18–25, 26–40, 41–56, 57–75, and 76+). Race was coded as White, Black, Asian, or other, as we anticipated that only these groups would be large enough in both samples to detect effects. We used four-level scales each for education, household income, and urban/rural residence. Political orientation was collected using a five-point Likert scale with end points 'strong liberal' and 'strong conservative'.

	$\bar{x}_{\text{Lucid}}$	$\bar{x}_{\text{OMS}}$	$\bar{x}_{\text{Lucid}} - \bar{x}_{\text{OMS}}$	*P*-value
Age group (0–4)	1.75 (1.12)	0.89 (0.56)	(0.79, 0.92)	<0.001
Gender = Male	0.49 (0.50)	0.81 (0.39)	(−0.37, −0.29)	<0.001
Ethn = White	0.75 (0.43)	0.41 (0.49)	(0.29, 0.39)	<0.001
Ethn = Black	0.13 (0.33)	0.03 (0.16)	(0.08, 0.12)	<0.001
Ethn = Asian	0.05 (0.22)	0.48 (0.50)	(−0.48, −0.38)	<0.001
Education (0–3)	1.36 (1.08)	2.24 (0.43)	(−0.94, −0.83)	<0.001
Cognitive employment	0.25 (0.43)	0.97 (0.17)	(−0.75, −0.70)	<0.001
Manual employment	0.14 (0.35)	0.00 (0.05)	(0.12, 0.15)	<0.001
Social employment	0.22 (0.42)	0.01 (0.10)	(0.20, 0.23)	<0.001
Household income (0–3)	1.23 (1.06)	2.19 (0.91)	(−1.05, −0.87)	<0.001
Political orientation (−2 to +2)	−0.01 (1.23)	−0.52 (0.93)	(0.42, 0.61)	<0.001
Urban (0–3)	1.59 (1.05)	2.16 (0.79)	(−0.66, −0.49)	<0.001
Individualism (standardized)	0.06 (0.92)	−0.47 (0.73)	(0.46, 0.61)	<0.001
Egalitarianism (standardized)	−0.05 (0.90)	0.21 (0.78)	(−0.34, −0.18)	<0.001
Techno-skepticism (standardized)	0.06 (0.92)	−0.45 (0.81)	(0.43, 0.59)	<0.001
Risk aversion (standardized)	0.03 (0.89)	−0.24 (0.69)	(0.20, 0.34)	<0.001

states/territories, most work full-time in computing-related jobs and are therefore more likely to be geographically concentrated than our nationally representative US sample. They also tend to be younger and have higher incomes than the US public. Table 1 shows summary statistics comparing sociodemographic variables and cultural values in our two samples.

Previous research has revealed differences in opinion between distinct groups of AI and computer science practitioners, such as between AI-skilled professionals at US technology companies (Aiken et al. 2020) and active researchers who publish at machine learning conferences (Zhang et al. 2021). Our graduate student expert sample adds an additional perspective to this literature; OMS students may differ from previously-surveyed expert samples in their propensity to work in industry versus academia, their level of experience with AI, and their sociodemographic and cultural factors. Respondents in our OMS sample completed undergraduate degrees largely in North America (66.1 per cent) or Asia (25.6 per cent), primarily in computer science (43.1 per cent) or other STEM fields (50.8 per cent). Most of them concurrently or recently worked in computer science or software engineering but not specifically in AI (63.8 per cent); 18.1 per cent reported working in another field of science or engineering; and 11.8 per cent reported working directly in AI (see Supplement Section A).

3.2 Survey design

Our survey consisted of two parts. The first portion assessed sociodemographic information, cultural values, opinion on risks posed by technologies other than AI, and self-reported familiarity with AI. We included standard sociodemographic factors that have been found to associate with the opinion on questions related to AI in previous surveys (Morning Consult 2017; European Commission 2017; Smith and Anderson 2017; The Harris Poll 2017; Gallup, Inc 2018; Morning Consult 2018; Smith 2018a, 2018b; Zhang and Dafoe 2019; United Kingdom Government 2019; Johnson and Tyson 2020; Deeney 2019; Boyon 2019; ARM 2020; Selwyn et al. 2020): gender, age group, race/ethnicity, job type (cognitive/analytical, manual/physical, social/people-oriented, or other), education level, household income, urban/rural residence, and political orientation (see Table 1 for coding details). We also included questions assessing attitudes toward other technologies for which expert and public risk perception has been well-studied. Participants were asked, on a five-point Likert scale ('risks significantly outweigh benefits' to 'benefits significantly outweigh risks'), about their perception of genetically modified foods, nuclear power, coal-burning power plants, vaccines, and synthetic biology.[4]

The cultural values of individualism and egalitarianism, described in Section 2.1, were adapted from Kahan et al.'s operationalization of grid-group cultural theory for survey research (Kahan et al. 2007).[5] Two clarifications are needed to position our use of these constructs within the broader cultural theory literature. First, Kahan et al.'s cultural cognition theory differs from the broader cultural theory literature by constructing survey items directly from the 'grid' and 'group' axes of Douglasian cultural theory (Douglas and Wildavsky 1982). These survey items for individualism and egalitarianism improve on conceptual issues with other cultural theory measurement strategies (Kahan 2012), have demonstrated high predictive validity in studies of other technological risks, and are perhaps the most popular measurement approach in cultural theory (Johnson and Swedlow 2021). However, they have been shown to be facially and empirically limited, particularly because they do not incorporate the cultural values of hierarchy and fatalism (Swedlow et al. 2020).[6] Second, we depart slightly from the 'cultural cognition' hypothesis of Kahan et al. (2007) by analyzing the effects of individualism and egalitarianism as *individual* constructs rather than analyzing their intersection.[7]

The techno-skepticism construct was created from items previously used in the literature and modified after testing in two small pilot surveys (see Supplement Section D); the final construct consisted of the following four items: 'new technologies are more about making profits rather than making peoples' lives better', 'I am worried about where all this technology is leading', 'technology has become dangerous and unmanageable', and 'I feel uncomfortable about new technologies'. The general risk aversion construct was adapted directly from Sharma (2010).

The second portion of the survey assessed opinion about AI. We first provided respondents with a brief definition of AI adapted from Zhang and Dafoe (2019): 'Artificial intelligence (AI) refers to computer systems that perform tasks or make decisions that usually require human intelligence. AI can perform these tasks or make these decisions without explicit human instructions. Today, AI has been used in the following applications: identifying people from their photos, diagnosing diseases like skin cancer and common illnesses, blocking spam email, helping run factories and warehouses, and predicting what one is likely to buy online'. We then assessed five outcome measures separated into two groups. The first two outcomes assessed whether respondents believed that (a) they personally and (b) society more generally would benefit from AI. These outcome measures (self- and societal benefit) were intended to disambiguate respondents who were supportive or apprehensive about AI use because of its perceived effect on their own lives from respondents were excited or concerned about its effects on society at large. The remaining three outcomes assessed, again on five-point Likert scales, support for whether AI should be (a) 'use[d]', (b) 'carefully managed', and (c) 'regulated by the government', language adapted from Zhang and Dafoe (2019). The differentiation of management and regulation was intended to better disambiguate opinion on *whether* some form of AI governance should occur from opinion on *who* is best suited to perform this governance. This distinction is particularly salient in light of impending regulatory efforts and ongoing debates on the comparative merits of self-regulation, soft law, and formal government regulation.

These five outcome measures, which assessed opinion of AI in general, were repeated for each of the six AI application contexts described below. Before answering survey items for each application, respondents were provided with two-sentence vignettes describing the potential benefits and harms of AI use in that context (see below). To reduce participant fatigue in the US public (Lucid) sample, each respondent was provided with only three of the six contexts, so that the sample size for each of the six specific AI contexts in the US public sample was $N \approx 3{,}524/2$. The expert respondents, who we anticipated would suffer less fatigue, each provided data for all six contexts. The full survey instrument is contained in Supplement Section C.

3.3 AI application contexts

We assessed our five outcome variables (perceived self-benefit, perceived societal benefit, and support for use, 'careful management', and 'regulat[ion] by the government') for AI in general and in the context of six policy-relevant application contexts. Before being asked about AI in general, participants were provided a brief definition of AI adapted from Zhang and Dafoe (2019) (see above). Before being asked about each context, participants were provided a two-sentence vignette describing both potential benefits and concerns about the use of AI in that context. The points highlighted in each vignette were chosen in an attempt to reflect arguments present in typical discourse about AI, particularly those that may associate affective qualities with the application:

- Predictive policing: 'Some police departments use AI to predict where crime is likely to occur, helping them decide where to deploy their resources. But civil rights groups and some researchers argue that these AI systems simply increase arrests in minority neighborhoods without actually reducing crime.'
- Economic/labor impact: 'AI systems are likely to automate many tasks. Some think that these AI systems will make work less tedious and produce higher standards of living. Others believe that these AI systems will increase unemployment and inequality.'
- Medical systems: 'AI-powered medical systems can detect diseases earlier and more accurately than human doctors. But some fear that these AI systems could occasionally produce incorrect results without doctors understanding why.'
- Autonomous vehicles: 'AI-powered self-driving cars could save lives by reducing traffic accidents caused by human error. But some are concerned that the AI systems in self-driving cars are vulnerable to malfunctioning or being hacked.'
- Personalization: 'AI systems can provide personalized news, social media content, and product recommendations using data collected from users. But some worry that this can undermine individual privacy and lead to misinformation and political polarization.'
- Autonomous weapons: 'Lethal autonomous weapons controlled by AI systems could improve our national security while putting fewer service members in danger. But some worry that AI-powered weapons could be dangerous or lead to a reckless arms race.'

Supplement Section B contains a more detailed discussion of each application context along with tables summarizing the impact of sociodemographic and cultural factors on support for AI in each context.

3.4 Survey administration and attention model

The US public (Lucid Theorem) survey ran from 3 May 2021 to 30 May 2021, with most responses collected from May 3 to 5. Based on recent research on the Lucid platform (Aronow et al. 2020; Peyton et al. 2021), we anticipated that pandemic-induced structural changes in populations completing online surveys might result in reduced effect sizes. The expert (master's student) survey ran from 28 April 2021 to 8 May 2021. Two pilot surveys ($N = 50$ and 150) were administered on 22 March 2021 and 1 April 2021 (see Supplement Section D).

Respondent attention is a concern when using online survey data. Following the recommendations of Berinsky et al. (2019), we assessed participant attention using four attention check questions: three simple grid-type attention checks and one stand-alone attention check. We modeled respondent attention using an item response theory model similar

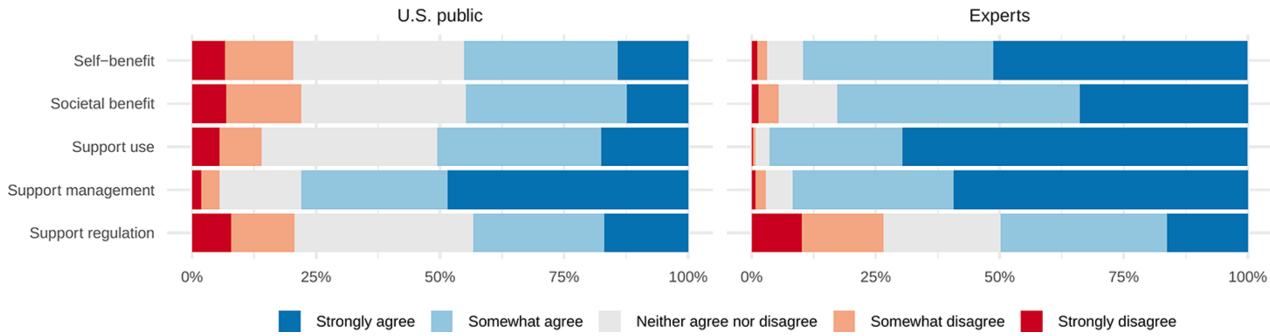

Figure 2. Outcome measures after respondents are presented with a general (context-free) definition of AI.

to that used by Berinsky et al. (2019). Specifically, we used the standard two-parameter Rasch model:

$$p\left(y_{ij}=1\right)=\frac{e^{a_j(\theta_i-b_j)}}{1+e^{a_j(\theta_i-b_j)}} \tag{1}$$

where y_{ij} denotes whether the ith participant correctly answered the jth attention check question, a_j denotes the discriminability of the jth attention check question, θ_i denotes the ith particpant's attention, and b_j denotes the difficulty of the jth attention check question. Inattentive respondents were defined as those in the bottom quartile of attentiveness $\{\theta_i\}$ (computed across the combined US public/expert sample). The US public sample was less attentive overall (two-tailed t-test on mean attention θ_i: $P < 0.001$); 86.1 per cent of the expert sample was retained in the attentive subsample, compared to 73.7 per cent of the US public sample. We expected that including inattentive respondents in our analysis would reduce effect sizes but that excluding them would bias results: respondent attention has been found to associate with characteristics such as age, gender, and education (Berinsky et al. 2014) and may thus influence outcomes. All results reported in this paper are therefore based on analyses that retained the complete sample. As a robustness check, these results are reproduced in Supplement Section F with inattentive respondents removed. Overall, the differences between the full-sample and attentive-subsample results were minor.[8]

3.5 Structural equation model and estimation

We used R version 1.3.9 and lavaan version 0.6–9 (Rosseel 2012) with the default (NLMINB) optimizer to fit the SEMs defined in our analysis. Because outcome measures and cultural values were measured with Likert-scale (ordinal) items, we used the mean- and variance-adjusted weighted least-squares estimator with polychoric correlations (Li 2016) and robust standard errors. Polychoric correlations were also used to compute construct reliabilities. For identifiability, cultural construct variances were fixed to unity and each factor loading was allowed to vary. The only instances of missing data in our survey involved context-specific outcome measures (as only half of the US public sample was asked about each application). The metrics and thresholds we used to assess the quality of fit were preregistered and stemmed from typical recommendations (Kline 2016).

4. Results

4.1 Public and expert attitudes differ in key areas

Compared to the US public, experts were more confident and positive in their attitudes toward AI (Fig. 2). Experts were much more likely to perceive self-benefit (1.04 points on a five-point Likert scale, Welch's unequal variances t-test: $P < 0.001$) and societal benefit (0.82 points, $P < 0.001$). While a plurality of the US public also believed that AI would benefit both them personally (45.2 per cent) and society at large (44.8 per cent), few professed strong opinions. Similarly, our expert sample was much more likely to support the general use of AI than the more ambivalent US public (1.17 points, $P < 0.001$), with almost no experts expressing opposition to AI use. In both samples, support for AI use was strikingly similar to perceived benefit (Supplement Fig. 14), a pattern we explore in more detail below.

Recent surveys have found strong public support for the 'careful management' of AI (European Commission 2017; Smith 2018a; Zhang and Dafoe 2019; Selwyn et al. 2020) but differing opinions on whether this management should be performed by researchers, technology companies, nonprofit groups, or the government (Zhang and Dafoe, 2019). To disentangle attitudes toward AI governance in general from attitudes toward government regulation, we asked respondents both whether AI should be 'carefully managed' and whether AI should be 'regulated by the government', phrasing adopted from Zhang and Dafoe (2019). We found that both experts and the US public were highly supportive of 'careful management' and generally supportive of government regulation (Fig. 2). Notably, we found similar support for government regulation between experts and the public (0.02 point difference, $P = 0.715$), despite experts being more likely to support management (0.28 points, $P < 0.001$). Past surveys have found that, unlike the public, AI experts place more trust in scientific and international organizations than their own government to 'develop and manage' AI (Zhang et al. 2021), suggesting that compared to the public, experts may be more inclined to support soft law governance approaches to governance (see, e.g. Marchant et al. (2020)).

The public's support for the use and governance of AI, shown in Fig. 3, was largely similar across contexts—a notable finding that persisted when the analysis was restricted to only attentive respondents (Supplement Fig. 24; see Section 3.4 for the definition of attentive subsample). By contrast, experts' views were more nuanced, varying much more significantly across contexts. While expert and public attitudes trended in the same direction in many contexts, they featured

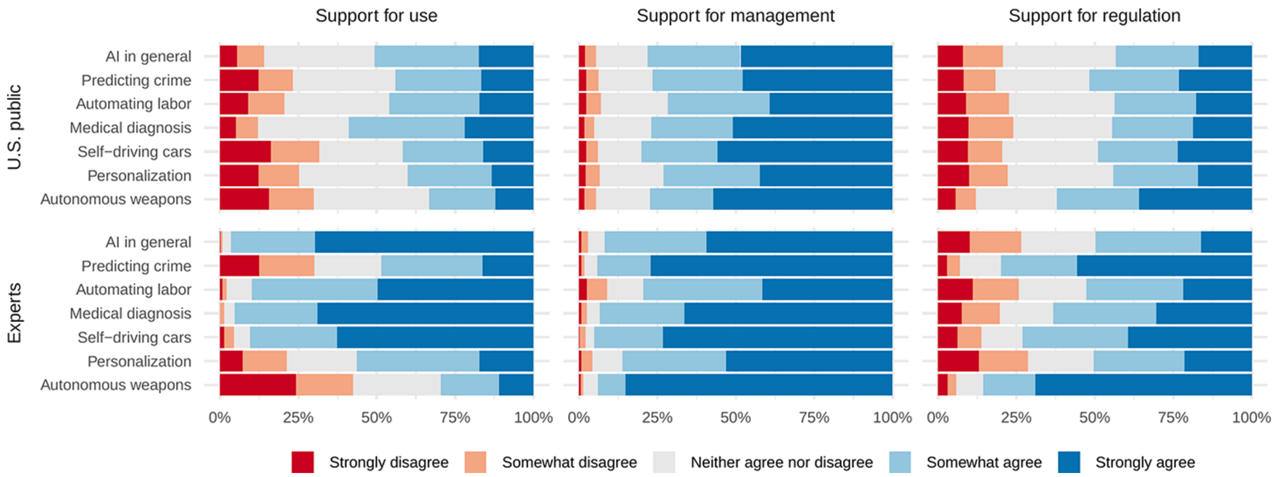

Figure 3. Comparison of support outcome measures between samples and among common AI application areas. Before responding, participants were provided two-sentence vignettes, listed in Section 3.3, describing arguments for and against the use of AI in the context.

distinct splits in others. For example, both experts and the public were wary of AI use in autonomous weapons, recommendation systems, and predictive policing, but experts' overwhelming support for AI use in autonomous vehicles, medical diagnosis, and automating labor stood in stark contrast to the much more divided public.

Our results suggest that greater public awareness about the unique impacts of AI in different applications may be necessary to fully empower the public to share its perspectives on AI use and governance. The cross-context divides we find also suggest that limited support for the regulation of AI *in general* (among both experts and the public) may belie support for tailored government intervention in specific application contexts such as autonomous weapons.

4.2 Cultural factors are strongly informative of attitudes

What drives these expert–public divides that persist across outcome measures and application contexts? These gaps may be due to differences in technical knowledge or due to socialization during AI training. However, they may also be driven by differences in sociodemographics and cultural values. Our expert and US public samples differed significantly on all sociodemographic variables (Table 1) as well as across all four cultural values (P's < 0.001): experts were less individualistic (0.53 points), less techno-skeptical (0.51 points), less risk averse (0.27 points), and more egalitarian (0.26 points) (see Supplement Fig. 7).

To better understand how these factors inform attitudes, we used the preregistered SEM shown in Fig. 1 to explore the relationship between sociodemographic variables, cultural values, perceived benefit, and support for AI use and governance. (The size of our expert sample limited this SEM analysis to the US public.) We first assessed the reliability and fit of the cultural value components of the model. The fit in each sample (evaluated using thresholds defined in our preregistration) was adequate to good, construct reliabilities were satisfactory, constructs loaded appropriately onto each item (with similar loadings in each sample), and model correlation residuals indicated adequate local fit (Supplement Tables 11, 12, and 15).[9] To assess the impact of cultural values on our outcome variables, we compared the fit of S, the full SEM shown in Fig. 1, with $S_{\setminus C}$, the nested model that constrains to zero the paths from cultural values to outcome measures. We found consistent global (Table 2) and local (Supplement Tables 21 and 22) evidence that the inclusion of pathways from cultural values to our outcome variables produced better model fit, indicating that the four cultural values we considered were indeed informative factors in explaining attitudes toward AI.

We next fit the full SEM shown in Fig. 1 to data from the US public sample. Fit statistics are shown in Table 2 along with statistics for the two modified (nested) models used to evaluate the roles of cultural values and perceptions of benefit. The full model achieved the standard thresholds for adequate fit listed in our preregistration.[10] Correlation residuals, shown in Supplement Table 22, generally indicated satisfactory local model fit.[11] Finally, we observed relatively small covariances between support outcomes, consistent with a lack of highly influential unmodeled common causes of these variables. It is important to note that our SEM represents *hypothesized* relationships between variables and that 'equivalent' models with different hypothesized relationships can produce the same covariance structure (MacCallum et al. 1993). Thus, while the fit statistics in Table 2 provide circumstantial evidence in support of our SEM, the primary evidence for the model's correctness is based on our theoretical arguments above.

Inferred SEM path coefficients are shown in Fig. 4. Overall, the results indicated that the cultural values of individualism, egalitarianism, risk aversion, and techno-skepticism were strongly predictive of attitudes toward AI. The influence of sociodemographic variables also contained interesting patterns. Like past surveys (Morning Consult 2017, 2018; Zhang and Dafoe 2019), we found that those who were male, younger, better educated and had higher income both perceived more benefit from AI and were more supportive of its use. Yet we found that support for government regulation was—perhaps surprisingly—often divorced from perceived benefit and support for use and more directly informed by sociodemographic and cultural variables. For example, older and more conservative respondents were more hesitant about AI use. However, despite perceiving less benefit from AI and expressing less support for its use, they were also less

Table 2. Fit statistics for the complete SEM S and two nested models used for analysis. χ^2: model chi-square test, along with model degrees of freedom and P-value, CFI: comparative fit index, RMSEA: root mean squared error of approximation, SRMR: standardized root mean square residual, $\Delta\chi^2$: chi-square difference test (compared to full model S). R^2 values show coefficients of determination for the five endogenous variables in the model. The complete model S achieved adequate-to-good global fit, with CFI and RMSEA indicating adequate fit and SRMR indicating good fit. Reduced models $S_{\setminus C}$ (used to assess the evidence for paths from cultural values to support outcomes) and $S_{\setminus B}$ (used to assess the evidence for paths from perceived benefit to support outcomes) achieved adequate fit on RMSEA and SRMR but poor global fit on CFI.

	Model fit statistics					R^2 (benefit)		R^2 (support)		
	χ^2 (df, P)	CFI	RMSEA (90% CI)	SRMR	$\Delta\chi^2$ (Δdf, P)	Self	Soc.	Use	Mgt.	Reg.
Model S	4650.2 (350, <0.001)	0.903	0.059 (0.058, 0.061)	0.034	–	0.274	0.262	0.470	0.235	0.201
Model $S_{\setminus C}$	8204.2 (370, <0.001)	0.822	0.078 (0.076, 0.079)	0.094	1764.8 (20, <0.001)	0.134	0.110	0.461	0.090	0.084
Model $S_{\setminus B}$	5554.4 (356, <0.001)	0.882	0.064 (0.063, 0.066)	0.047	1173.4 (6, <0.001)	0.552	0.544	0.774	0.220	0.190

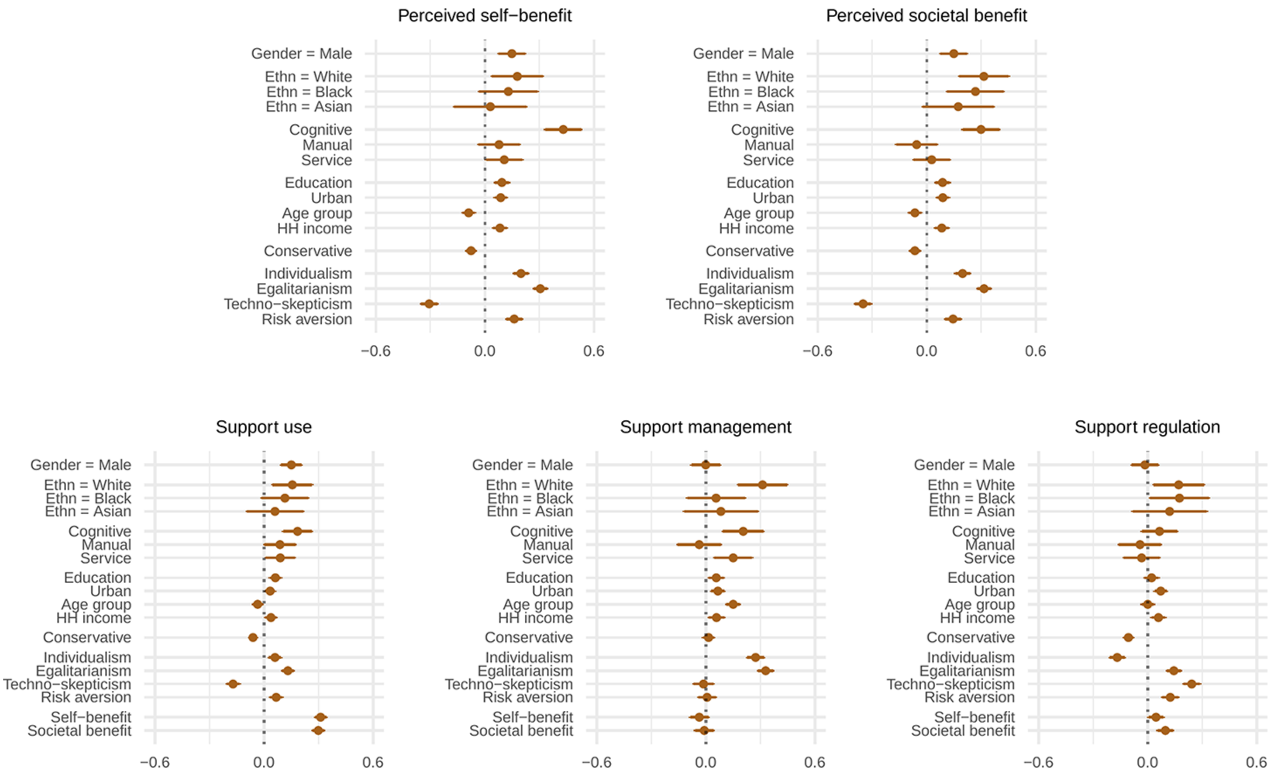

Figure 4. Inferred path coefficients (with 95 per cent confidence intervals) for full SEM S fit with the US public data. Gender, race/ethnicity, and work type were coded as binary; education, household income, and urban residence were coded as four-level variables; age group and political orientation were coded as five-level variables; and cultural constructs and perceived benefit variables were standardized. See Table 2 for fit statistics.

supportive of the government regulating AI. Similarly, those who held cognitive/analytical jobs, lived in urban areas, and had higher incomes perceived greater self-benefit from AI and were more supportive of its use. However, these groups were also more likely to believe that AI should be carefully managed and regulated.

4.3 Cultural determinants of attitudes differed in some applications

Developing effective ‘culturally pluralized’ (Johnson and Swedlow 2021) strategies for science communication and governance requires an understanding of how cultural values affect attitudes toward specific technologies and their applications. While previous research has evaluated how cultural values inform support for other emerging technologies, it is not clear how—or whether—these results generalize to applications of AI.

Notably, our results found that some effects of cultural values (Fig. 4) had reversed directions from the patterns observed for other technologies. For example, both individualism and egalitarianism predicted *increased* perceptions of self-benefit from AI—a contrast with many other technologies, where egalitarianism has been found to associate with lower support.[12] This reversed effect of egalitarianism suggests that AI may be perceived differently from many other technological risks, perhaps due to perceptions that automated systems can temper certain hierarchical social structures that egalitarians perceive as harmful. If this perception does indeed hold among the public, however, it stands in sharp contrast to the increasing realization among AI developers that bias and fairness are significant problems in automated decision-making systems (Mehrabi et al. 2021) and evidence that awareness of these problems negatively affects perceptions of their performance (Schiff et al. 2021).

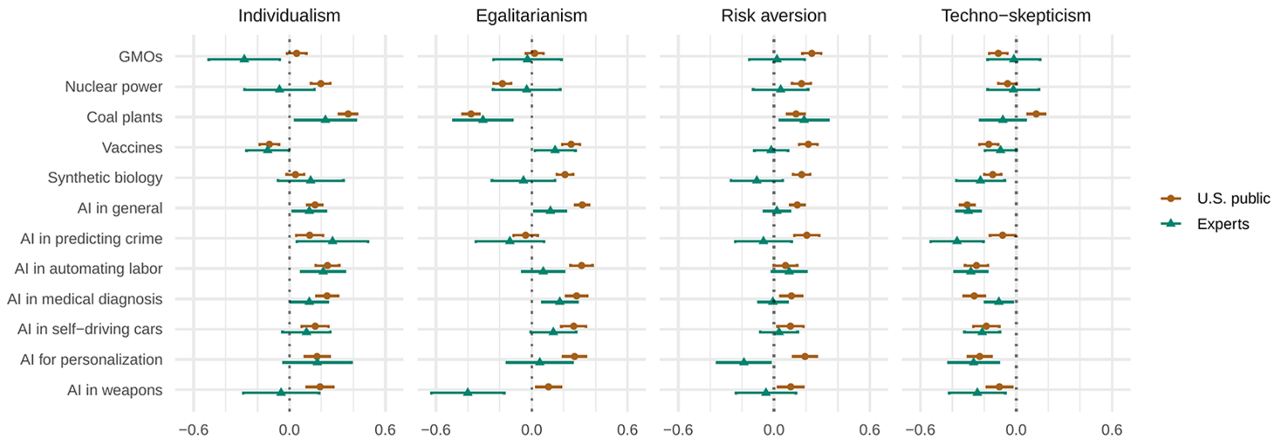

Figure 5. Comparison of cultural values' effects on support for AI contexts and other technologies. Markers show ordinary least-squares regression estimates and 95 per cent confidence intervals when controlling for sociodemographic variables. For support for AI contexts, respondents were asked whether they supported the use of AI in a particular application. For other technologies, respondents were asked whether the technology's benefits outweighed its risks. Each outcome was measured on a five-point Likert scale; cultural value constructs were standardized and inferred from a confirmatory factor analysis model. This analysis was exploratory.

We used a linear regression model to compare the effects of cultural values on support for AI use between experts and the public. Our use of linear regression rather than SEM was due to the limited size of our expert sample; this portion of the analysis was exploratory (i.e. not preregistered). We found that the direction of cultural values' effects on support for AI use was generally consistent across AI application areas (Fig. 5) and again found that experts' opinions were more nuanced than the public's. Supplement Figs. 15–16 provide additional evidence for this phenomenon, showing that experts' attitudes toward both AI and other technologies varied more than the public's, a pattern that persisted when the analysis was restricted to attentive respondents. This evidence suggests that the public's attitudes toward AI may evolve considerably as they become more informed, underlining the importance of public education on the broader impacts of AI use in specific applications. These results also revealed patterns across the six contexts we explored. For example, attitudes toward the predictive policing and autonomous weapons application contexts were similar, particularly among experts (Supplement Fig. 15).

To examine whether the factors driving attitudes toward these applications were also similar, we fit a multigroup version of the SEM shown in Fig. 1 to data from the US public sample. This multigroup SEM facilitated between-context comparison by allowing path coefficients to differ for each context while constraining the model aspects that defined cultural values to be constant. Some notable patterns emerged from this model, for which inferred parameters are shown in Supplement Section E.4.2. We indeed found key sociodemographic and cultural variables whose impact on attitudes toward predictive policing and autonomous weapons differed from their impact on other contexts. For example, older and politically conservative respondents were *less* supportive of AI, in general, but were *more* supportive of AI use for predictive policing and autonomous weapons. The impact of egalitarianism on support for AI use in these two contexts similarly differed from its impact on most other contexts. More broadly, there were substantial between-context differences in the impact of age on support for AI. For example, older respondents were much less supportive of the use of AI in autonomous vehicles and recommendation systems than they were of the use of AI for medical diagnosis. See Supplement Section B for tables highlighting where these results matched expectations based on prior literature.

Unsurprisingly, AI's impact on labor and the economy was perceived to be more beneficial by respondents with cognitive/analytical jobs and higher education. However, we found that manual/physical employment also predicted greater perceived benefits from AI's impact on labor and the economy. This result is potentially surprising but consistent with the findings that many US workers believe automation is more likely to affect others' jobs than their own (Parker et al. 2019). Interestingly, we also found that perceived societal benefit had a stronger impact on support for labor-automating AI than on AI in general (Supplement Table 32).

Prior work has found that individualism generally predicts higher support for technology, and we found that individualism had a similarly positive impact on support for AI. Less consistent with work on other technologies, however, we found that egalitarianism *also* tended to predict greater support for AI. Perhaps unexpectedly, we found overall positive effects of the general risk perception construct of Sharma (2010) on support for AI across contexts, suggesting that the risk aversion and techno-skepticism constructs used in our survey measured relatively orthogonal aspects of technological risk perception.

That the US public perceived AI as more egalitarian than experts did (Fig. 3) suggests that the public viewed AI as shaping society to be more equitable than experts did. Particularly striking is the positive impact of egalitarianism on support for the use of AI-based weapon systems, suggesting that recent discourse and activism in the AI community opposing autonomous weapons (Belfield 2020) may have been effective in driving experts' opinions but not in breaking through to the general public, who may have been more swayed by our vignette's description of potential safety benefits to service members. It is also notable that egalitarianism drove greater support for labor-automating AI among the public than among experts.

4.4 Perceived benefit substantially informs support for AI use—but not for management and regulation

Our SEM (Fig. 1) hypothesized that perceived self- and societal benefit drove support for AI use and governance. To assess the impact of perceived benefit on these support outcomes, we compared the full SEM S to a nested model $S_{\setminus B}$, in which paths from the perception of benefit outcomes to support outcomes were fixed to zero. Overall, global and local comparisons of $S_{\setminus B}$ and S provided mild support for the existence of an impact of perceived benefit on our support outcomes (Table 2; Supplement Tables 22 and 28).[13]

As shown in Fig. 4, perceived benefit (to both the respondent individually and society at large) predicted substantially greater support for AI use but had much less impact on attitudes toward its governance. Indeed, the total effect of sociodemographic and cultural variables on support for AI *use* was split roughly evenly between direct and indirect effects (Supplement Fig. 10). By contrast, support for AI *management and regulation* was impacted much less by indirect effects. These findings were generally consistent across AI contexts (Supplement Tables 30–36).

Experts' attitudes were again more nuanced than the public's: we found much larger gaps between perceived *self*-benefit and perceived *societal* benefit among experts than among the public. Indeed, in the US public sample, we did not find statistically significant differences between perceived self- and societal benefit in any application (P's > 0.123).

Prior literature has conjectured that AI developers may engage in a form of motivated reasoning that makes them more likely to believe that AI has a positive impact on society (Baum 2016) when it is professionally advantageous for them. We find mixed evidence for this theory. Consistent with this motivated reasoning conjecture, we found that experts were indeed more likely than the US public to believe that AI was beneficial to society (0.82 points on a five-point Likert scale; $P < 0.001$). Our expert sample was also much more likely than the public to believe that AI was beneficial for the society in applications with significant commercial opportunities such as automating labor (0.76 points; $P < 0.001$) and self-driving cars (1.29 points; $P < 0.001$). However, AI experts differed from the public on almost every sociodemographic and cultural trait, typically in ways that our results suggest would predict higher support for AI use (Fig. 2(b)). Moreover, experts were somewhat *less* likely to report that AI-based recommendation systems—a context in which AI experts as a whole have a large commercial interest—were beneficial to society (0.16 points; $P = 0.020$). This counterexample suggests that the AI experts' attitudes might be more substantially driven by underlying sociodemographic and cultural traits rather than by a motivated reasoning mechanism related to their professional orientation, although we would expect that these results may differ in samples of other types of AI experts.

5. Conclusion and discussion

5.1 Summary of key results

The complex and subtle sociotechnical concepts inherent to AI make it challenging to design effective governance and science communication strategies that are informed by and respectful of diverse public views and values. In light of these challenges, this work evaluated underlying factors, values, and mechanisms that influence attitudes toward AI. We explored the role of sociodemographic variables; the impact of the cultural values of egalitarianism, individualism, techno-skepticism, and risk aversion; the potentially moderating effects of perceived self- and societal benefit; differences between experts and the public; and differences across prominent policy-relevant applications of AI.

One consistent finding of our study is that the US public's attitudes toward AI were much less nuanced than experts'. Compared to experts, the public's views on the use, management, and regulation of AI were largely similar across application areas, and the public reported perceiving little distinction between how AI might affect them personally and how it might affect the society more generally. We did, however, find greater support for government regulation in applications such as autonomous weapons and predicting crime, indicating that while recent suggestions for soft law approaches to AI governance (Marchant et al. 2020) may be more likely to find public and expert support in the USA, ambivalence toward broad AI regulation might belie support for 'hard' legally-binding regulatory actions narrowly targeted to certain contexts.

Second, we found that the four cultural values we studied were meaningful predictors of public attitudes toward AI. The relationships between cultural values and attitudes are similar both across application contexts and between experts and the US public (Fig. 5). For example, individualism tended to predict greater support for AI use while techno-skepticism tended to predict reduced support for AI use. These similarities—particularly between experts and the public—advance the hypothesis that cultural values are a useful tool for understanding attitudes toward AI and how these attitudes may evolve. Thus, research on a larger set of cultural values, performed in different regions and with different populations, may be a valuable tool for creating participatory and culturally sensitive AI applications and governance strategies.

A third key finding of our study is that although cultural values had significant impacts on support for AI adoption and governance, these cultural values did not impact attitudes in the same way that they impact attitudes toward many other technologies. For example, egalitarianism and risk aversion are traditionally associated with skepticism toward the use of emerging technologies (Kahan et al. 2007); by contrast, we find that these values predicted *greater* support for AI. This implies that AI's impact on society may be perceived differently from the impacts of other technologies. Governance and public dialogue strategies may be more successful if they take these novel aspects of AI into account. Indeed, previous work has found that science communication is most effective when it tailors its messages to the specific cultural values held by the public (Kahan et al. 2011; Lupia 2013). The relationships we find between specific cultural values and specific AI applications (shown in Fig. 5) suggest which potential dimensions and applications could be emphasized in outreach efforts to more effectively build credibility with the public and honor public values.

5.2 Theoretical implications and contrasts with prior literature

The satisfactory fit of our SEM serves as a proof of concept for the benefits of using a combination of sociodemographic and cultural variables in modeling attitudes toward AI and suggests that a similar approach may be fruitful for studying public attitudes toward other culturally-polarized technologies. In addition, the presence of both strong direct effects and strong indirect effects in our fit model provides tentative (but not conclusive) support for the value of considering self- and societal benefit as mediating variables in understanding attitudes toward technology.

The SEM used in this study shares some features with popular frameworks in the broader technology acceptance literature, such as the Technology Acceptance Model (see Section 2.2). Our work also carries implications for this class of models, providing evidence that factors adopted from cultural theory might also be successfully incorporated as external factors in models of attitudes toward (and use of) other technologies.

Finally, our work provides evidence for cultural theory more broadly, although survey operationalization details (discussed in Section 2.1) suggest that some caution is warranted when interpreting these generalizations. First, the large and statistically significant effects of cultural values on public and expert attitudes toward AI we identified provide evidence in favor of the applicability of cultural theory to attitudes toward AI and toward technology more generally. Enumerating and categorizing values that shape attitudes is particularly valuable for understanding general-purpose technologies such as AI that have multiple overlapping impacts on society, and our work suggests that cultural theory may provide a useful framework for such an effort. Second, our results in Fig. 5—which depict associations of cultural values across multiple technologies and AI application contexts—offer a basis for comparing the impacts of the four cultural values we studied here on a variety of technologies and use cases.

Previous work has found that those who are more comfortable with AI are more likely to be young, male, and educated and to live in urban areas (Morning Consult 2017, 2018; Zhang and Dafoe 2019; United Kingdom Government 2019; Johnson and Tyson 2020; Morning Consult 2021). Our results reflect these divisions. Moreover, we found that across most contexts, these demographic traits had positive and statistically significant effects on support for AI not only directly but also indirectly through paths mediated by perceived self- and societal benefit. Our findings also largely align with prior evidence that individuals with more education, white-collar jobs, and higher incomes are more likely to perceive both self- and societal benefit from AI (Morning Consult 2017; Smith and Anderson 2017; Gallup, Inc 2018; Morning Consult 2018; Zhang and Dafoe 2019).[14]

Our results contrast most sharply with previous findings that blue-collar workers, those in urban areas, and political liberals are most likely to report believing that AI will exacerbate inequality and lower employment (Morning Consult 2017; Gallup, Inc 2018). In seeming contrast, we found that those living in urban areas and political liberals tended to report perceiving a *benefit* to themselves and the society from AI, both in general and in the economic context of labor automation.

5.3 Lessons for public engagement in AI governance

Our study was motivated by the near-universal calls for diverse, interdisciplinary, and public participation in AI governance from global industry, government, and civil society actors. Despite these calls, there are persistent concerns about opaque policy processes vulnerable to industry capture, culturally-insensitive uses of AI techniques, and shallow or ineffectual participatory mechanisms. How can those interested in inclusive governance bridge this gap? Our work both provides insights and suggests challenges that may face even well-intentioned efforts to develop participatory structures.

A first challenge is the sizable gap between the significant public support we find for 'careful management' of AI and the more limited support for 'government regulation' (Fig. 3), a finding that echoes prior research, particularly in the US context (Zhang and Dafoe 2019). However, a growing international expert consensus—including among corporate actors—has articulated a need for AI regulation, and regulatory efforts continue to develop. This reveals a fundamental tension in how public opinion should be respected in AI governance (Dragojlovic 2014). Should regulators take a technocratic approach and base regulatory strategies on the views of experts, even in the face of skepticism from some quarters of the public? Or should regulators, presented with equivocal public support for US government regulation of AI, limit the scope of their involvement even if they believe that public attitudes may evolve significantly as the impacts of AI become more apparent?

One response to this tension that has been embraced by a number of participatory design and governance strategies is to promote public education and genuine public–expert dialogue as part of outreach efforts. In these methods, trained facilitators, researchers, or policymakers may initiate public engagement experiences by providing information about the stakeholders, benefits and costs, policy implications, and trade-offs that can help the public make more informed judgments. The public–expert gaps identified in our study point to the value of these cooperative strategies.

Importantly, these dialogues are not unidirectional; discussion is structured and restructured by the public's situated experiences and values. Examples of relevant approaches can be found both in long-standing participatory design strategies (e.g. Multi-Criteria Decision Analysis (Triantaphyllou 2000) and the Delphi method (Landeta 2006)) and in strategies formulated or adapted specifically for science and technology (e.g. the Citizen Visions on Science, Technology and Innovation method (Gudowsky et al. 2012), the Reflect! platform (Hoffmann 2020), and Deliberative Mapping (Burgess et al. 2007)). These engagement methods can elicit qualitative and quantitative data to inform policy preferences, pointing not only to general values but also guiding specific choices (Mavrommati et al. 2021). Engage2020's Action Catalogue database of participatory strategies (http://actioncatalogue.eu/search) provides one starting point.

Our results also point to specific contexts and value orientations in which further unpacking the complex factors driving attitudes toward AI governance may be particularly useful. We find, for instance, that in AI applications like predictive policing and autonomous weapons, experts are much more likely than the public to support government regulation of AI. Moreover, in these contexts, there are statistically significant differences between experts and the public in how cultural values affect attitudes. For example, our finding that egalitarianism predicts *greater public support* for AI-based weaponry but *less expert support* may suggest that efforts in the AI community to advocate against lethal autonomous weapons (e.g. (Future of Life Institute 2018)) may not have reached the public eye. Similarly, our study's finding that risk aversion predicted greater public support for AI-based recommendation systems but less expert support may suggest that increased public awareness about the potential benefits and harms of these systems could be particularly impactful.

However, our findings also caution that in many application domains, increased public awareness of AI's impacts might not produce major changes in attitudes toward AI governance. We find that public support for AI governance is relatively independent of arguably more malleable factors like perceived self- and societal benefit from AI. Instead, our results suggest that public support for AI governance is more strongly related to factors reflective of broader regulatory preferences such as political orientation and individualism (Fig. 4).

The contrast our study finds between the US public's desire for AI governance and skepticism of government involvement suggests an opportunity for governance strategies. A major focus of AI policy discourse is 'trustworthy AI', an attempt to shape the ways AI is developed and applied in an effort to promote user trust. Our results reveal an additional need for *trustworthy AI governance*. Previous research has indicated that the US public places higher trust in, for instance, military and higher education institutions to manage AI than in the federal government at large (Zhang and Dafoe 2019; Morning Consult 2021). Identifying the aspects that have built trust in these institutions could help government and industry actors demonstrate their own trustworthiness in AI governance. Alternatively, governments could leverage these institutions to develop and implement governance strategies, drawing on trusted local authorities and civil society actors to develop, communicate, and administer aspects of AI governance.

In turn, researchers can help identify participatory strategies, messages, and governance approaches that promote (and deserve) public trust. Little is currently known about which strategies (e.g. third-party conformity assessments, labeling, industry standards, or human rights or well-being impact assessments) are most likely to foster trust. In short, there are many opportunities to promote inclusive AI governance for both AI developers and formal governance bodies. However, the time horizon for doing so is not unlimited. AI systems with major impacts are already commonplace, and a variety of national and international regulatory efforts are currently underway. Understanding effective strategies for trustworthy AI governance—and the role of public views in these efforts—will be a pressing need in the coming years.

5.4 Limitations and future work

Our research has several limitations. The four cultural values used in our model were selected because of their effects in governing public opinion on other technologies, but they may not, of course, be either root or comprehensive causes of differing attitudes toward AI; many other sources of cultural diversity are important to respect when designing AI governance strategies. The broader literature on technology acceptance (e.g. the many variants of the Technology Acceptance Model (Marangunic and Graníc 2015)) describes many examples of factors that may also be influential in the formation of attitudes toward AI. Moreover, while previous work has posited that the *interaction* between cultural values may drive some differences in risk perception (Kahan et al. 2011), our SEM analysis strategy does not analyze interactions between variables.

A second limitation concerns our descriptions of the six AI application contexts used in our survey: while we attempted to faithfully reflect the way each application is framed in public discourse, it is likely that this discourse will evolve in ways that change their associations with particular cultural values. Respondents also differed in their familiarity with AI; knowledgeable participants (and the expert sample in particular) likely considered information from previous knowledge about the AI application contexts beyond what was provided in our vignettes, limiting fair comparisons between samples.

Third, while we believe our graduate student sample provides one informative view on the beliefs of AI experts, this group differs from other samples of AI experts, such as those studied by Aiken et al. (2020) or Zhang et al. (2021). Future work should explore how our findings generalize to groups involved in other aspects of AI development and governance. Our US public sample also suffers from the typical limitations of online surveys: although respondents were representative of US adults on age, gender, race, and region, online samples tend to differ from the general population in ways not captured by these variables.

This work represents the first step toward understanding underlying mechanisms governing expert and public attitudes toward AI. Future research should extend these findings by exploring how attitudes differ in non-US (and non-Western) contexts (IEEE 2019, Sambasivan et al. 2021). Despite their limitations, the four cultural values we used here provide a tool for quantitatively exploring cross-cultural differences in values and attitudes relevant to AI governance; results may help explain emerging transnational political differences in AI governance strategies. It would also be valuable to study other groups of AI experts and practitioners, more fine-grained conceptions of governance than management and regulation, and using other narratives and frames for the application areas we considered.

Supplementary data

Supplementary data is available at *Science and Public Policy Journal* online.

Funding

We gratefully acknowledge funding from the National Defense Science & Engineering Graduate Fellowship (M.R.O.), and from the National Science Foundation via grants CCF-1717530 (L.R.V.), CCF-1350954 (M.R.O. and C.J.R.), and CCF-2107455 (C.J.R. and M.A.D.).

Conflict of interest statement. None declared.

Acknowledgements

We thank M.D. Hunter and R.P. Barke for helpful discussions and feedback on earlier versions of this work, Georgia Tech CS 6601 course staff for assistance in fielding the OMS student survey, and our survey participants. All errors are the responsibility of the authors.

Notes

1. Prior work has studied other samples of AI experts: Zhang et al. (2021) surveys AI researchers publishing in prestigious conferences,

while Aiken et al. (2020) surveys AI professionals in industry. See Section 3.1 and Supplement Section A for more details on the characteristics and limitations of these samples.

2. Our survey assessed cultural values before AI was introduced to avoid attitudes toward AI influencing cultural values through priming effects.
3. We consider this alternate model in Supplement Section E.1.4.
4. Because we anticipated that synthetic biology was likely to be less familiar to respondents, this technology featured a one-sentence description.
5. In two small pilot surveys using the Lucid Theorem platform ($N=50$ and $N=150$; see Supplementary Materials for details), we found that the condensed cultural cognition worldview scale of Kahan et al. (2007), which used both positively- and negatively-worded items for each construct, had poor reliability. Based on the results of other recent studies that found reliability issues with the negatively-worded cultural cognition theory items (Hornsey et al. 2018; Johnson et al. 2020), we followed the strategy of Hornsey et al. (2018), by restricting our preregistered final survey to four positively-worded items each for individualism and egalitarianism. The resulting scales had satisfactory reliability in both the full samples and attentive subsamples (Supplement Tables 11 and 39). Our results, however, may not be directly comparable to other work that used the full scale of Kahan et al. (2007).
6. Specifically, cultural theory inspired by the work of Douglas and Wildavsky (1982) posits that the intersection of two axes, 'grid' and 'group,' define quadrants corresponding to four distinct cultural biases: individualism, egalitarianism, hierarchy, and fatalism. Initial attempts to operationalize cultural theory for survey research using these four scales found that many participants did not uniquely belong to a single cultural bias. The cultural cognition theory scales of Kahan et al. (2007) that we use directly measure the 'grid' and 'group' axes as hierarchy–egalitarianism and individualism–communitarianism. This approach sidesteps the issues of participants scoring highly on multiple cultural biases, and is argued to improve on the scale reliability and predictive validity of other approaches (Kahan 2012), but has been criticized for its lack of inclusion of discrete hierarchy and fatalism factors (Johnson et al. 2020; Ripberger et al. 2012; Van der Linden and Conceptual 2016). See Johnson and Swedlow (2021) for a review of cultural theory's development and its relationship to the cultural cognition theory of Kahan et al. (2007).
7. The 'cultural cognition' hypothesis of Kahan et al. (2007) posits that the *intersection* of individualism and egalitarianism define identity groups that imbue attitudes toward risk with affective qualities and lead to directionally motivated reasoning. Other work (e.g. (Johnson et al. 2020)) has also treated these factors as discrete.
8. One notable exception was the covariance between individualism and egalitarianism constructs. In the full results, we found that this negative covariance had much larger magnitude in the expert sample than the US public sample; when restricting the sample to attentive respondents, we found the inferred covariance for the US public sample was much closer to the inferred value in the expert sample.
9. Although cultural construct loadings were similar between samples, there were some notable between-sample differences in the cultural construct covariances between cultural constructs (Supplement Tables 13 and 14). In the US public sample, techno-skepticism was more highly correlated with risk aversion and individualism, suggesting that experts separate their views of technology from their overall risk preferences and individualism somewhat more than the general public does. There was also a much larger negative covariance between egalitarianism and individualism in the expert sample. These differences, however, were much smaller when analysis was restricted to the attentive subsample (Supplement Tables 41 and 42; see Section 3.4).
10. While the model χ^2 statistic indicated a statistically significant difference between the observed and model-implied covariance matrix (a potential indication of inadequate fit), this test is known to be sensitive to large sample sizes such as ours; concluding that a model achieves adequate fit despite a statistically significant result from this test is consistent with standard SEM practice and our preregistration (Kline 2016).
11. One notable exception was the residual variance of the support for use variable (−0.16), whose relatively large magnitude suggested some caution when interpreting results such as the coefficient of determination for this variable.
12. Recall from Section 3.2 that two divides in related literature limit direct comparison of our results to some other work on the impact of cultural values on public attitudes toward technology. First, the constructs of individualism and egalitarianism that we adapt from Kahan et al. (2009) do not model hierarchy and fatalism; cultural elements argued to be important by the broader cultural theory literature (Johnson et al. 2020; Ripberger et al. 2012). Second, like some other literature but unlike Kahan et al. (2009), we model individualism and egalitarianism as discrete constructs rather than examining effects of their intersection.
13. The evidence in support of accepting S over $S_{\backslash B}$ was more equivocal than the evidence in support of accepting S over $S_{\backslash C}$. For example, while overall we found evidence in support of retaining model S over $S_{\backslash B}$, one piece of evidence supported retaining $S_{\backslash B}$: the large residual variance on support for AI use in models $S_{\backslash C}$ and S vanished in $S_{\backslash B}$.
14. For example, we found that employment in a 'cognitive' role had a particularly strong positive effect on perceived self-benefit, perceived societal benefit, and support for use for both AI in general and for AI used in labor automation, perhaps the most economically-oriented application we considered.

References

Aiken, C., Kagan, R., and Page, M. (2020) '*Cool Projects' or 'Expanding the Efficiency of the Murderous American War Machine?': AI Professionals' Views on Working with the Department of Defense*, Washington, DC: Center for Security and Emerging Technology.

ARM. 'AI Today, AI Tomorrow: The ARM 2020 Global AI Survey' (2020).

Aronow, P. M., Kalla, J., Orr, L., et al. (2020) '*Evidence of Rising Rates of Inattentiveness on Lucid in 2020*' preprint *SocArXiv Papers*. <https://osf.io/8sbe4> accessed 28 Jul 2021.

Barke, R. P., Jenkins-Smith, H., and Slovic, P. (1997) 'Risk Perceptions of Men and Women Scientists', *Social Science Quarterly*, 78: 167–76.

Barke, R. P. and Jenkins-Smith, H. C. (1993) 'Politics and Scientific Expertise: Scientists, Risk Perception, and Nuclear Waste Policy', *Risk Analysis*, 13: 425–39.

Barke, R. P., Silva, C. L., and Jenkins-Smith, H. C. (2006) *From Experts' Beliefs to Safety Standards: Explaining Preferred Radiation Protection Standards in Polarized Technical Communities*, Atlanta: Working Paper 14 Georgia Institute of Technology School of Public Policy.

Baum, S. D. (2016) 'On the Promotion of Safe and Socially Beneficial Artificial Intelligence', *AI & Society*, 32: 543–51.

Belfield, H. (2020) 'Activism by the AI Community: Analysing Recent Achievements and Future Prospects' in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pp. 15–21. New York: ACM.

Berinsky, A. J., Margolis, M. F., and Sances, M. W. (2014) 'Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys', *American Journal of Political Science*, 58: 739–53.

Berinsky, A. J., Margolis, M. F., Sances, M. W., et al. (2019) 'Using Screeners to Measure Respondent Attention on Self-Administered Surveys: Which Items and How Many?', *Political Science Research and Methods*, 9: 1–8.

Boyan, N, 'Ipos global poll for the world economic forum' (2019).

Burgess, J., Stirling, A., Clark, J., et al. (2007) 'Deliberative Mapping: A Novel Analytic-Deliberative Methodology to Support Contested Science-Policy Decisions', *Public Understanding of Science*, 16: 299–322.

Campolo, A., Sanfilippo, M., Whittaker, M., et al. (2017) *AI Now 2017 Report*. New York: AI Now Institute.
Cave, S., Craig, C. and Dihal, K. (2018) *Portrayals and Perceptions of AI and Why They Matter*. London: The Royal Society.
Chong, D., Citrin, J., and Conley, P. (2001) 'When Self-Interest Matters', *Political Psychology*, 22: 541–70.
Coppock, A. and McClellan, O. A. (2019) 'Validating the Demographic, Political, Psychological, and Experimental Results Obtained from a New Source of Online Survey Respondents', *Research and Politics*, 6.
Dake, K. (1991) 'Orienting Dispositions in the Perception of Risk: An Analysis of Contemporary Worldviews and Cultural Biases', *Journal of Cross-Cultural Psychology*, 22: 61–82.
Davis, F. D. (1985) 'A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results', PhD thesis Massachusetts Institute of Technology.
DeCario, N. and Etzioni, O. (2021) *America Needs AI Literacy Now*. pnw.ai. <https://pnw.ai/article/america-needsai-literacy-now/72515409> accessed 13 Dec 2021.
Deeney, C. 'Six in ten (61%) respondents across 26 countries oppose the use of lethal autonomous weapons systems' (21 Jan. 2019).
Dixon, G., Hart, P. S., Clarke, C., et al. (2020) 'What Drives Support for Self-Driving Car Technology in the United States?', *Journal of Risk Research*, 23: 275–87.
Douglas, M. and Wildavsky, A. (1982) *Risk and Culture: An Essay on the Selection of Technological and Environmental Dangers*. Berkeley: University of California Press.
Dragojlovic, N. (2014) '*How Meaningful Are Public Attitudes Towards Stem Cell Research?*', *OUPblog*. <https://blog.oup.com/2014/01/how-meaningful-are-public-attitudes-towards-stem-cell-research/> accessed 9 Dec 2021.
Druckman, J. N. and McGrath, M. C. (2019) 'The Evidence for Motivated Reasoning in Climate Change Preference Formation', *Nature Climate Change*, 9: 111–9.
Ellis, R. J. and Thompson, F. (1997) 'Culture and the Environment in the Pacific Northwest', *American Political Science Review*, 91: 885–97.
European Commission, 'Attitudes towards the impact of digitisation and automation on daily life', Special Eurobarometer 460 (10 May 2017).
European Group on Ethics in Science and New Technologies, 'Statement on artificial intelligence, robotics and "autonomous" systems' (2018).
Fast, E. and Horvitz, E. (2017) 'Long-Term Trends in the Public Perception of Artificial Intelligence'. In: *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence AAAI Conference on Artificial Intelligence*. San Francisco: Association for Computing Machinery.
Finucane, M. L., Slovic, P., Mertz, C., et al. (2000) 'Gender, Race, and Perceived Risk: The "White Male" Effect', *Health, Risk & Society*, 2: 159–72.
Fishbein, M. and Ajzen, I. (1975) *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. Reading: Addison-Wesley.
Fishbein, M. and Funke, U. H. (1976) 'A Behavior Theory Approach to the Relations between Beliefs about an Object and the Attitude Toward the Object'. In: M. Beckmann and H. P. Kunzi (eds) *Mathematical Models in Marketing* pp. 87–8. Heidelberg: Springer.
Flynn, J., Slovic, P., and Mertz, C. K. (1994) 'Gender, Race, and Perception of Environmental Health Risks', *Risk Analysis*, 14: 1101–8.
Funk, C., Rainie, L., Kennedy, B., et al. (2015) 'Public and scientists' Views on science and society, (Pew Research Center, 29 Jan. 2015).
Future of Life Institute. (2018) *Lethal Autonomous Weapons Pledge*. Cambridge.
Gallup, Inc. 'Optimism and anxiety: views on the impact of artificial intelligence and higher education's response NortheasternGallup Poll' (Jan. 2018).
Gardezi, M. and Arbuckle, J. G. (2020) 'Techno-Optimism and Farmers' Attitudes toward Climate Change Adaptation', *Environment and Behavior*, 52: 82–105.
Gudowsky, N., Peissl, W., Sotoudeh, M., et al. (2012) 'Forward-Looking Activities: Incorporating Citizens' Visions: A Critical Analysis of the CIVISTI Method', *Poiesis & Praxis*, 9: 101–23.
The Harris Poll. (2017) '*Monthly Harvard-Harris Poll*' <https://harvardharrispoll.com/wp-content/uploads/2017/11/HCAPS-October_Topline-Memo_with-banners_Registered-Voters-Social-Media.pdf> accessed 13 Oct 2020.
Hoffmann, M. H. G. (2020) 'Reflective Consensus Building on Wicked Problems with the Reflect! Platform', *Science and Engineering Ethics*, 26: 793–819.
Hornsey, M. J., Harris, E. A., and Fielding, K. S. (2018) 'Relationships among Conspiratorial Beliefs, Conservatism and Climate Scepticism across Nations', *Nature Climate Change*, 8: 614–20.
Institute of Electrical and Electronics Engineers (IEEE). (2019) *Ethically Aligned Design Version 2: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems*. Piscataway: IEEE.
Johnson, B. B. and Swedlow, B. (2021) 'Cultural Theory's Contributions to Risk Analysis: A Thematic Review with Directions and Resources for Further Research', *Risk Analysis*, 41: 429–55.
Johnson, B. B., Swedlow, B., and Mayorga, M. W. (2020) 'Cultural Theory and Cultural Cognition Theory Survey Measures: Confirmatory Factoring and Predictive Validity of Factor Scores for Judged Risk', *Journal of Risk Research*, 23: 1467–90.
Johnson, C. and Tyson, A. 'People globally offer mixed views of the impact of artificial intelligence, job automation on society' (Pew Research Center, 15 Dec. 2020).
Kahan, D. M. (2012). 'Cultural Cognition as a Conception of the Cultural Theory of Risk'. In: S. Roeser, R. Hillerbrand, P. Sandin, et al. (eds) *Handbook of Risk Theory: Epistemology, Decision Theory, Ethics, and Social Implications of Risk*, pp. 725–59. Dordrecht: Springer.
Kahan, D. M., Braman, D., Gastil, J., et al. (2007) 'Culture and Identity-Protective Cognition: Explaining the White-Male Effect in Risk Perception', *Journal of Empirical Legal Studies*, 4: 465–505.
Kahan, D. M., Braman, D., Slovic, P., et al. (2009) 'Cultural Cognition of the Risks and Benefits of Nanotechnology', *Nature Nanotechnology*, 4: 87–90.
Kahan, D. M., Jenkins-Smith, H., and Braman, D. (2011) 'Cultural Cognition of Scientific Consensus', *Journal of Risk Research*, 14: 147–74.
Kline, R. B. (2016) *Principles and Practice of Structural Equation Modeling*, 4th edn. p. 534. New York: The Guilford Press.
Krier, J. E. and Gillette, C. P. (1985) 'The Un-Easy Case for Technological Optimism', *Michigan Law Review*, 84: 405–29.
Landeta, J. (2006) 'Current Validity of the Delphi Method in Social Sciences', *Technological Forecasting and Social Change*, 73: 467–82.
Li, C.-H. (2016) 'Confirmatory Factor Analysis with Ordinal Data: Comparing Robust Maximum Likelihood and Diagonally Weighted Least Squares', *Behavior Research Methods*, 48: 936–49.
Liu, P., Guo, Q., Ren, F., et al. (2019) 'Willingness to Pay for Self-Driving Vehicles: Influences of Demographic and Psychological Factors', *Transportation Research Part C: Emerging Technologies*, 100: 306–17.
Lord, C. G., Ross, L., and Lepper, M. R. (1979) 'Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence', *Journal of Personality and Social Psychology*, 37: 2098–109.
Lupia, A. (2013) 'Communicating Science in Politicized Environments', *Proceedings of the National Academy of Sciences*, 110: 14048–54.
MacCallum, R. C., Wegener, D. T., Uchino, B. N., et al. (1993) 'The Problem of Equivalent Models in Applications of Covariance Structure Analysis', *Psychological Bulletin*, 114: 185–99.
Macnaghten, P. and Chilvers, J. (2014) 'The Future of Science Governance: Publics, Policies, Practices', *Environment and Planning. C, Government & Policy*, 32: 530–48.

Marangunic, N. and Granić, A. (2015) 'Technology Acceptance Model: A Literature Review from 1986 to 2013', *Universal Access in the Information Society*, 14: 81–95.

Marchant, G. E., Tournas, L., and Gutierrez, C. I. (2020) 'Governing Emerging Technologies through Soft Law: Lessons for Artificial Intelligence', *Jurimetrics*, 61: 1–18.

Mavrommati, G., et al. (2021) 'A Methodological Framework for Understanding Shared Social Values in Deliberative Valuation', *Ecological Economics*, 190: 107185.

Meadows, D. H., Meadows, D. L., Randers, J., et al. (1972) *The Limits to Growth: A Report for the Club of Rome's Project on the Predicament of Mankind*. New York: Universe Books.

Mehrabi, N., Morstatter, F., Saxena, N., et al. (2021) 'A Survey on Bias and Fairness in Machine Learning', *ACM Computing Surveys*, 54: 1–35.

Morning Consult. (2017) '*National Tracking Poll*' 170401. <https://morningconsult.com/wp-content/uploads/2017/04/170401_crosstabs_Brands_v3_AG.pdf> accessed 13 Nov 2020.

——— (2018) '*Morning Consult National Tracking Poll*' 180534. <https://morningconsult.com/wp-content/uploads/2018/05/180534_BRANDS_crosstabs_.pdf> accessed 13 Nov 2020.

——— (2021) '*A Perspective of Americans' Attitudes toward Artificial Intelligence*' <https://ml.globenewswire.com/Resource/Download/a1e27f5b-7edb-49a6-a13e-835a36e7ad2a> accessed 21 Nov 2020.

Organisation for Economic Cooperation and Development. (2019) *Recommendation on the Council on Artificial Intelligence* OECD/LEGAL/0449.

Parker, K., Morin, R., and Horowitz, J. M, 'Looking to the future, public sees an America in decline on many fronts' (21 Mar. 2019).

Peters, E. M., Burraston, B., and Mertz, C. K. (2004) 'An Emotion-Based Model of Risk Perception and Stigma Susceptibility: Cognitive Appraisals of Emotion, Affective Reactivity, Worldviews, and Risk Perceptions in the Generation of Technological Stigma', *Risk Analysis*, 24: 1349–67.

Peyton, K., Huber, G. A., and Coppock, A. (2021) 'The Generalizability of Online Experiments Conducted During the COVID-19 Pandemic', *Journal of Experimental Political Science*, 1–16. <https://www.cambridge.org/core/journals/journal-of-experimental-political-science/article/generalizability-of-online-experiments-conducted-during-the-covid19-pandemic/977D0A898CD4EA803ABE474A49B719E0>.

Renn, O. and Benighaus, C. (2013) 'Perception of Technological Risk: Insights from Research and Lessons for Risk Communication and Management', *Journal of Risk Research*, 16: 293–313.

Ripberger, J. T., Song, G., Nowlin, M. C., et al. (2012) 'Reconsidering the Relationship between Cultural Theory, Political Ideology, and Political Knowledge', *Social Science Quarterly*, 93: 713–31.

Rosseel, Y. (2012) 'Lavaan: An R Package for Structural Equation Modeling', *Journal of Statistical Software*, 48.

Sambasivan, N., Arnesen, E., Hutchinson, B., et al. (2021) 'Re-Imagining Algorithmic Fairness in India and Beyond' in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency FAccT'21: 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 315–28. New York: ACM.

Schiff, D. S., Schiff, K. J., and Pierson, P. (2021) 'Assessing Public Value Failure in Government Adoption of Artificial Intelligence', *Public Administration*, 100: 1–21.

Sears, D. O. and Funk, C. L. (1991) 'The Role of Self-Interest in Social and Political Attitudes', *Advances in Experimental Social Psychology*, 24: 1–91.

Selwyn, N., Cordoba, B. G., Andrejevic, M., et al. (2020) *AI for Social Good? Australian Public Attitudes toward AI and Society*, Melbourne: Monash Data Futures Institute.

Sharma, P. (2010) 'Measuring Personal Cultural Orientations: Scale Development and Validation', *Journal of the Academy of Marketing Science*, 38: 787–806.

Sjöberg, L. and Drottz-Sjöberg, B.-M. (1993) *Attitudes to Nuclear Waste*, xii. Stockholm: Stockholm School of Economics, Center for Risk Research.

Sloane, M., Moss, E., Awomolo, O., et al. (2020) 'Participation Is Not a Design Fix for Machine Learning', *arXiv*: 2007.02423[cs].

Slovic, P., Flynn, J. H., and Layman, M. (1991) 'Perceived Risk, Trust, and the Politics of Nuclear Waste', *Science*, 254: 1603–7.

Smith, A., 'Public attitudes toward computer algorithms' (Pew Research Center, 16 Nov. 2018a).

Smith, A., Public attitudes toward technology companies, (Pew research Center, 28 Jun. 2018b).

Smith, A. and Anderson, M., 'Automation in everyday life' (Pew Research Center, 4 Oct. 2017).

Stirling, A. (2008) "'Opening Up" and "Closing Down": Power, Participation, and Pluralism in the Social Appraisal of Technology', *Science, Technology & Human Values*, 33: 262–94.

Stix, C. (2021) 'Actionable Principles for Artificial Intelligence Policy: Three Pathways', *Science and Engineering Ethics*, 27.

Swedlow, B., et al. (2020) 'Construct Validity of Cultural Theory Survey Measures', *Social Science Quarterly*, 101: 2332–83.

Tennant, C., Stares, S., and Howard, S. (2019) 'Public Discomfort at the Prospect of Autonomous Vehicles: Building on Previous Surveys to Measure Attitudes in 11 Countries', *Transportation Research. Part F, Traffic Psychology and Behaviour*, 64: 98–118.

Triantaphyllou, E. (2000) 'Multi-Criteria Decision Making Methods'. In: P. M. Pardalos and D. Hearn (eds) *Multi-Criteria Decision Making Methods: A Comparative Study*, pp. 5–21. Boston: Springer.

Ulnicane, I., Knight, W., Leach, T., et al. (2020) 'Framing Governance for a Contested Emerging Technology: Insights from AI Policy', *Policy and Society*, 40: 1–20.

United Kingdom Government. (2019) '*Artificial Intelligence: Public Awareness Survey*' <https://www.gov.uk/government/publications/artificial-intelligence-public-awareness-survey> accessed 4 Dec 2020.

United States Office of Management and Budget. (2020) 'Memorandum on Guidance for Regulation of Artificial Intelligence Applications M-21-06'.

Van der Linden, S. (2016) 'A Conceptual Critique of the Cultural Cognition Thesis', *Science Communication*, 38: 128–38.

Venkatesh, V. and Morris, M. G. (2000) 'Why Don't Men Ever Stop to Ask for Directions? Gender, Social Influence, and Their Role in Technology Acceptance and Usage Behavior', *MIS Quarterly*, 24: 115. JSTOR: 3250981.

Weber, E. and Hsee, C. (2000) 'Culture and Individual Judgment and Decision Making', *Applied Psychology*, 49: 32–61.

Wildavsky, A. and Dake, K. (1990) 'Theories of Risk Perception: Who Fears What and Why?', *Daedalus*, 119: 41–60.

Zhang, B., Anderljung, M., Kahn, L., et al. (2021) 'Ethics and Governance of Artificial Intelligence: Evidence from a Survey of Machine Learning Researchers', *Journal of Artificial Intelligence Research*, 71.

Zhang, B. and Dafoe, A. (2019) 'Artificial Intelligence: American Attitudes and Trends', Center for the Governance of AI, Future of Humanity Institute, University of Oxford.