Scientists' Perspectives on the Social Impacts and Regulations of Artificial Intelligence: Values, Media Use, Professional Characteristics, and Openness to Inclusion of Social Science Input

By

Luye Bao

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Mass Communications)

at the

UNIVERSITY OF WISCONSIN-MADISON

2022

Date of final oral examination: 07/27/2022

The dissertation is approved by the following members of the Final Oral Committee:

Dominique Brossard, Professor and Chair, Life Sciences Communication
Daniel M. Bolt, Professor, Educational Psychology
Todd P. Newman, Assistant Professor, Life Sciences Communication
Dietram A. Scheufele, Vilas Distinguished Achievement Professor, Life Sciences Communication
Michael A. Xenos, Professor, Life Sciences Communication

Table of contents

Disclaimer

This dissertation is supported by the Wisconsin Alumni Research Foundation and the Morgridge Institute for Research. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of these organizations.

Acknowledgements

First, I am deeply indebted to my advisor, Dominique Brossard. Her invaluable scholarship, big-picture thinking, and emphasis on interdisciplinary research have shaped the type of work I aspire to do. Dominique has also provided immense encouragement and support to build my capacity and confidence as a scholar. I can't thank her enough for her guidance and patience through my dissertation design, survey implementation, and writing process.

My sincere gratitude also extends to my committee members. I wish to thank Dietram Scheufele for his critical input on my dissertation. His deep knowledge of theory and methods and craft in theory development is remarkably inspiring and influential to my approach to research. I also learned a great deal from him about teaching and interacting with students. I am grateful to Mike Xenos and Todd Newman. They have not only provided important feedback for my dissertation, but also have been very supportive and encouraging for research collaboration. I am grateful to Dan Bolt for his guidance on statistical methods and for his insightful perspectives on this dissertation.

This dissertation would not have been possible without the generous help from my survey participants, who altruistically devoted their time to completing the survey and sharing many constructive comments with me. I would also like to thank Steve Meyer and Yachao Qian for their tremendous help with sample collection for my survey.

I was lucky to have had opportunities to work with my wonderful peers in the Science, Media and Public (SCIMEP) research group. I would like to thank Mikhaila Calice, Shiyu Yang, Nicky Krause, Becca Beets, Lindsey Middleton, and Anqi Shao for their feedback on my survey. I wish to thank SCIMEP alumni, Emily Howell, Kate Rose, and Chris Wirz for their peer mentorship when I was a junior PhD student. The thought provoking conversations I had about research and life with my peers have made this doctoral journey more enjoyable. My special thanks go to Mikhaila, for her passion and support during our collaboration and for always sending me positive vibes.

During my time at LSC, I am grateful to Tera Wagner, who is always supportive and kind, from helping me with logistics at every step to accomplish this degree to sharing resources on and off-campus. I extend my thankfulness to LSC faculty members, including Kaiping Chen, Sedona Chinn, Nan Li, Bret Shaw, and Don Stanley. They have taught me skills in teaching, research, and the job market. I would also like to acknowledge the financial support and various opportunities for professional development that the department has provided for me.

Finally, my earnest thanks go to my parents for their support and encouragement.

Abstract

The wickedness of science and technology requires difficult compromises among social groups whose values and concerns are potentially in conflict. This has raised new challenges for scientists in terms of incorporating external expertise into their tasks as researchers, public communicators, and stakeholders in science policy decision-making. Prior studies have provided social psychological explanations for the potential differences between scientists and lay publics regarding how they view scientific issues and related regulations. Other studies have offered normative perspectives on responsible research and innovation that are grounded in understanding the social contexts of technology development and a wide representation of public input. However, prior research has fallen short in connecting the normative ideals and the descriptive components of scientists' views of the societal impacts of their work, related regulations, and scientific conduct. In addition, it is unclear how factors such as scientists' unique professional characteristics in comparison with other social groups, as well as factors such as their cognitive processing and media attention that are common to everyone else, weigh differently in affecting scientists' views of their work and related regulations.

This dissertation uses artificial intelligence (AI) as a case study. AI is one of the most recent examples of wicked science. This dissertation first synthesizes the importance of social science research in bridging the interface between scientists and lay publics. This argument is based on frameworks ranging from the broader impacts to responsible research and innovation. Before delving into factors that impact AI scientists' likelihood of incorporating social science input into AI research and development (Studies C & D), this dissertation first examines a) whether there exist attitudinal differences in risk and benefit perceptions of AI between AI scientists and lay publics (Study A) and b) whether AI scientists' exposure to social science

research on AI relates to a more socially desirable outcome for scientists, such as forming a more inclusive view of who should have a say in regulation development (Study B).

This dissertation shows heterogeneous and multifaceted segments of AI scientists in terms of how they view the societal impacts of AI. Although AI scientists and lay publics have similar levels of risk perceptions of AI, scientists consider AI more beneficial than their lay counterparts. In terms of the role of social sciences in bridging the science-society interface, disagreement with the hierarchical hypothesis of hard and soft sciences and reading social science research on AI are associated with more inclusive perspectives of who should have a say in AI regulation development, especially with an emphasis on including the voices of citizens, end-users, and civic groups. Moving forward to investigate AI scientists' likelihood of incorporating social science input into AI research and development, results reveal that AI scientists' self-estimation relies on their attitudes toward social sciences and interdisciplinary collaboration, whereas their estimation of most AI scientists depends on contextual factors, such as perceived risks and benefits of AI.

Societies will need to communicate and make decisions across diverse stakeholder groups to address the challenges of emerging science and technology. As scientists represent an influential stakeholder group, this dissertation responds to the theoretical inquiry of how dispositional, informational, and professional characteristics explain scientists' views of the societal impacts of their work, their roles in regulation development, and their approaches to conducting research. This dissertation also provides practical implications for ensuring responsible research and development of emerging technologies through incorporating social science input.

List of tables

List of figures

Chapter 1. Introduction and overview

In today's era of post-normal science, the high stakes and uncertainty of controversial science and technology issues highlight the value of involving expertise beyond technical aspects to address the ethical, legal, and social implications of science (Funtowicz & Ravetz, 1992). These characteristics of science and technology issues have led to changes regarding how scientists conduct research and engage in the governance of these issues. For instance, the US National Science Foundation (NSF) updated its proposal and award criteria by introducing the broader impacts criterion, which requires grant applications to consider "the potential to benefit society and contribute to the achievement of specific, desired societal outcomes" (National Science Foundation, 2020). The European Union has adopted the Responsible Research and Innovation (RRI) framework into its strategic plan, which obliges scientific work to follow "a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view to the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products (in order to allow a proper embedding of scientific and technological advances in our society)" (von Schomberg, 2011, p. 47). These new inquires on the relevance of science to society and making the scientific communities accountable are quite different from simply calling on public support for science by scientists decades ago (e.g., see Bush, 1945).

Science communication scholarship investigating the interface of science and society has indicated that societies will need to communicate and make decisions across all stakeholder groups to address the challenges of emerging and controversial science (for an overview, see Scheufele, 2022). It remains unclear how scientific experts view the societal impacts of their work and whether they acknowledge the importance of incorporating external expertise, from experts outside of their fields to lay publics to other stakeholders, into their research and regulatory practices. It is also unclear how factors such as scientists' unique professional characteristics in comparison with other social groups, as well as factors such as their cognitive processing and media attention that are common to everyone else, will affect scientific experts' perspectives on the societal impacts of their work.

To that end, this dissertation uses US-based scientific experts who have published academic articles in the fields related to artificial intelligence (AI) as a case study. My dissertation examines how professional, informational, and dispositional factors shape the multifaceted segmentation of AI scientific experts with regard to their views of AI and its regulation. Additionally, I explore how social sciences may bridge the interface between AI scientific experts and society and underlying mechanisms that incentivize the incorporation of social sciences into AI research. Before taking a closer look at the contents of this dissertation, I discuss background information of this dissertation by: 1) reviewing the wickedness of AI, 2) explaining the importance of incorporating expertise outside of one's fields (e.g., public input and social sciences) into AI scientific experts' research and regulatory practices, and 3) outlining factors that influence how scientific experts form attitudes toward scientific issues, related regulations, and research practices.

The wickedness of AI

AI is a fast developing area of science and technology. There are a wide range of definitions for AI. For example, a narrow definition of AI refers to "a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments" (National Artificial Intelligence Initiative

Act, 2020). On the other hand, for instance, a broad definition of AI views intelligences "as a process that converts unstructured information into useful and actionable knowledge. The scientific promise of artificial intelligence (AI) … is that we may be able to synthesize, automate and optimize that process, using technology as a tool to help us acquire rapid new knowledge in fields that would remain intractable for humans unaided" (Hassabis, 2017). Emerging research and applications of AI relate to various fields, such as medicine, economics, engineering, and military, among others (Stone et al., 2016). While AI research began in the 1950s, research in the fields related to AI has not proliferated until the last decade and has not reached the ambitious goals of developing artificial general intelligence or artificial superintelligence (Russell & Norvig, 2021). However, the current advanced AI systems have already caused profound societal impacts that call for societal debates about AI governance, such as using humanity to navigate the transition to advanced AI systems (e.g., Dafoe & Journal of International, 2018).

AI is an example of *wicked science*. Wicked science has various social impacts and raises complex questions that cannot be answered by scientists alone, but instead requires trade-offs between (often competing) values from a broader set of societal actors (Rittel & Webber, 1973). On the one hand, AI has deeply advanced human capabilities. For example, AI has been used for determining the 3D shapes of proteins that may transform biology (Callaway, 2020) or predicting potential nature disasters such as destructive earthquakes (Fuller & Metz, 2018). On the other hand, AI applications have generated complicated (un)intended consequences, including worsening societal inequalities in job recruitment, policing, and healthcare (Chakradhar, 2019; Heilweil, 2019; Selbst, 2017).

Like many other social contexts, the wickedness of AI indicates the possible coexistence of holding positive and negative evaluations independently and equivalently. The concurrence of predominantly positive and negative evaluations refers to ambivalent attitudes and an equivalent low level of positive and negative evaluations refers to indifferent attitudes (Kaplan, 1972; Priester & Petty, 2001). Empirical analyses across different science issues have shown that holding ambivalent views is prevalent among lay publics (e.g., Pew Research Center, 2022; Pidgeon et al., 2005). For instance, a 2021 Pew Survey found that nearly half of Americans (45%) felt equally excited and concerned about using AI in daily life, followed by 37% of them feeling more concerned than excited and 18% feeling more excited than concerned (Pew Research Center, 2022). A recent segmentation analysis on US citizens' risk and benefit perceptions of AI showed an asymmetric distribution. There was rarely a purely positive group that perceived high benefits and low risks of AI. However, more than a quarter of the sample held ambivalent attitudes toward AI (high risks/ high benefits), and a third of them held skeptical attitudes toward AI (high risks/ low benefits) (Bao et al., 2022). In the same analysis, the majority of respondents, including the ambivalent and the skeptical segments, has high agreement that the public should have a say in science and technology development, except two segments that perceive equivalently low or medium levels of benefits and risks and happened to include more minorities being disproportionately affected by AI (Bao et al., 2022).

From a normative perspective, complex policy decisions require a systematic process that delineates value dimensions of the pros and cons of the policy and collectively assesses each consequence (Edwards, 1979). For scientific experts, perceiving risks and benefits concurrently could potentially reflect the complexity of science and technology and increase the likelihood of being aware of the variety of values and interests that different social groups may have toward the same issue. The development of AI is facing social-technical problems with complicated ethical, legal, and social implications. These factors and the uncertainties regarding the quality of the science behind AI are unlikely to be effectively addressed by AI scientists alone (Hancock et al., 2019). AI scientists need to consult with various stakeholders and participate in societal debates on value trade-offs related to the development and use of AI (Buhmann & Fieseler, 2021). In fact, seeking public input about wicked science should occur sooner rather than later to ensure responsible research and innovation (Scheufele et al., 2021). A comprehensive understanding of the multifaceted segmentation of different stakeholders can ensure the inclusion of diverse opinions, knowledge, and judgment within a collective, which will be associated with more intelligent outcomes in societal debates and deliberation (Cappella et al., 2017). Hence, the first question of this dissertation is about how AI scientists position themselves within the risk-benefit typology, taking into account different professional, informational, and dispositional factors.

Scientists' roles in ensuring responsible AI development: From participating in regulation development to adopting professional norms in AI research and development

The (un)intended consequences of AI require that the development of AI be carefully regulated at the local, national, and international levels. Between 2016 to 2021, 14 countries have passed at least one AI-related bills into law, with the US having passed the most legislation, 13 bills in total (Zhang et al., 2022). Meanwhile, a surge of AI-related proposals has been introduced in US federal legislation, from 1 in 2015 to 130 in 2021 (Zhang et al., 2022). These legislation proposals range from regulations on facial recognition and biometric technology to the accountability of high-risk automated decision systems to AI training for the acquisition workforce. In 2021, the US federal government established a coordinating program, the National AI Initiative, to carry out missions like maintaining US's leading role in AI research and development, empowering US workforce with AI, developing trustworthy AI, and constantly seeking public input on federal AI policies (National Artificial Intelligence Initiative Act, 2020). As US market forces heavily drive AI development, industry-affiliated scientists have published the majority of the research on AI ethics (Zhang et al., 2022). Corporations have attempted to make explicit statements on their social responsibility of developing fair and trustworthy AI to address potential epistemological and reputational concerns (Buhmann & Fieseler, 2021). For instance, DeepMind implemented various ways to better understand AI's social impacts through public lecture, citizen engagement, and workshops with experts, advocates, and affected communities (DeepMind, n.d.). There have emerged intergovernmental efforts on developing AI regulations on various AI-related societal issues, such as a first-ever agreement on the ethics of AI among 193 countries led by UNESCO (United Nations, 2021). Additionally, there are ongoing conservations of potential bans or restrictions on lethal autonomous weapon systems in the UN disarmament committee (Kahn, 2021).

For democratic societies, "the primary concern for regulators … is not how to guard against capture by science but how to harness the collective expertise of the scientific community so as to advance the public interest" (Jasanoff, 1998, p. 250). Scientific experts have long been involved in science policymaking, such as the formation of independent science advisory committees, like EPA and FDA. Regulators tend to leverage scientists' expertise and credentials acknowledged by laypeople in the decision-making process (Jasanoff, 1998). The boundary-work between legitimate science and non-science has been constantly shifted, challenged, and subsequently defended by invested parties (Gieryn, 1983). The high stakes and uncertainty of controversial science and technology issues highlight the value of lay experience and cultural knowledge about the social and ethical implications of science, which urges the democratization of expertise (Funtowicz & Ravetz, 1992). Recently, in some extreme cases, scientists encounter challenges to their authority in knowledge production and science decision-making by populists (Mede & Schäfer, 2020).

It is also important to note that regulations alone are not panaceas for responding to the (un)intended consequences of AI in a timely and effective way, which shifts to the importance of reducing those (un)intended consequences in the AI research and development process. Recent controversy surrounding AI applications, such as predictive policing, has been a powerful reminder that applying AI systems to tackle social problems will not yield expected outcomes if scientists/developers do not fully consider the political, social, and cultural dimensions of the issue. For example, PredPol applies an earthquake prediction algorithm to predict crime based on the assumption that crimes operate in a similar manner to earthquake aftershocks, and that nearby crimes will be repeated (Bennett Moses & Chan, 2018). Accumulating evidence has shown that the flawed approach of predictive policing has not helped prevent violence, but has further reinforced racial bias and harmed civil liberty (Saunders et al., 2016; Selbst, 2017). In June 2020, Santa Cruz (CA) became the first US city to ban predictive policing (Asher-Schapiro, 2020). Though these legislative regulations are important, they are not able to reverse the damage. Debates over regulation development may also take a long period of time, thereby lagging scientific progress. For instance, the 2021 EU AI Act Proposal has received mixed reviews in which some view such regulation as too lenient and offering inadequate protection of fundamental human rights, whereas others believe that it will hamper scientific innovation (Lomas, 2021; McAfee, 2021). Therefore, effective professional norms of conducting research to ensure responsible AI research and development are essential and supplemental to political regulations, as learned from other scientific communities that have a longer history of research than AI, such as nuclear energy, climate change, and genome editing (Alexander et al., 2020; Freudenburg, 2008; Jasanoff & Hurlbut, 2018).

As AI development faces social-technical problems, the NSF has recently called for collaboration between researchers working in the fields of computer and information science and engineering (CISE) and social, behavioral, and economic sciences (SBE) to conduct research that improves the lives of people worldwide (NSF-CISE-SBE Virtual Roundtable, 2020). Addressing the social challenges that AI scientists face, such as understanding the social and historical context of their work, necessitates incorporating input from social sciences. By definition, social sciences refer to "the application of rigorous scientific method to the study of all human phenomena" (Bernard, 2012, p. 20798). For decades, social science scholars have been producing fundamental knowledge, methods, and tools to understand people and how they live (National Academies of Sciences, 2017). Social science studies can provide insight into the ethical, legal, and social implications (ELSI) of new scientific development that address different publics' needs and concerns. Social science studies can also focus on the ethics and responsibilities of scientists to ensure responsible innovation (e.g., Corley et al., 2016; von Schomberg, 2013). Although there is nascent cooperation across fields, computation scientists and social scientists differ in their normative beliefs about how to conduct and publish research (NSF-CISE-SBE Virtual Roundtable, 2020), which may lead to challenges for AI scientists to incorporate input from social sciences.

This dissertation will examine whether exposure to social science research on AI, such as those focusing on the societal impacts, ethics, and responsible development of AI, can be associated with a more inclusive view of who should have a say in AI regulation development among scientists. This dissertation will also examine underlying mechanisms that incentivize AI scientific experts' personal likelihood of incorporating social science input into AI research and development as well as their perception of these norms among most AI scientists. Before delving into more detail about my research, the following sections first review and compare theoretical frameworks that focus on guiding the research on and the use of science and technology from the perspective of social sciences, ranging from the broader impacts to responsible research and innovation. Additionally, I review how these principles and frameworks are reflected in the development of AI research and use.

Broader impacts

As discussed at the beginning of this Chapter, the NSF broader impacts criterion requires scientists to scrutinize the potential social groups that their work could empower or improve and actions or activities that make these impacts more likely to achieve while applying for grants (National Science Foundation, 2020). The idea of demonstrating the broader impacts of scientific work has gained increasing attention within the computing community. In a post published in the Association for Computing Machinery (ACM) Future of Computing Blog in 2018, 12 co-authored computer scientists from advanced university and industry AI research labs suggested including such rules in the peer review process of papers and grant proposals by asking authors to "add a "Broader Impacts" or a "Societal Impacts" section near the end of a paper, a la "Future Work" and "Limitations" (Hecht et al., 2018, p. 2). More importantly, these scientists argued for net benefits, which involve not only forecasting positive and negative impacts, but also initiating potential solutions to mitigate negative impacts through research or policy. The net positive impacts tap into the idea of cost benefit analysis. However, some of the negative impacts, such as replacement of human labor and extensive energy usage of computing, are far beyond questions that AI scientists can easily address, which will inevitably require expertise from other fields.

Since 2020, the Annual Conference on Neural Information Processing Systems (NeurIPS), one of the most prestigious conferences in machine learning, has started to ask submitters to consider the potential positive and negative societal impacts of their research, such as concerns about security and human rights (NeurIPS, 2020). Content analyses on the 2020 NeurIPS articles found that variations in statement length and broadness remained high, mentions of positive impacts were more often than the negative impacts, some authors did not clearly differentiate between technical impacts and societal impacts of their work (Ashurst et al., 2022; Nanayakkara et al., 2021). Whether these changes of including statements about broader impacts will be successful depends on the degree to which computer scientists acquire knowledge and skills to meaningfully anticipate the societal impacts and ethical issues of their work, or whether they collaborate with social scientists who study the lives of the affected publics and communities. Additionally, it is important to establish incentives and explain expectations and guidance for presenting the broader impacts (Ashurst et al., 2022), as well as to move forward from simply presenting these impacts after the research is done to actually engaging with various stakeholders that are likely to be affected by the use of these technology development (Nanayakkara et al., 2021).

Responsible Research and Innovation

While the broader impacts primarily focus on research outcomes, RRI underlines more fundamental changes in conducting research and developing innovation (Davis & Laas, 2014). Originating and widely used in European countries, RRI framework means “a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view to the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products (in order to allow a proper embedding of scientific and technological advances in our society)" (von Schomberg, 2011, p. 47). The definition of RRI is targeted to the normative anchor points from the Treaty on the European Union and emphasizes market products from a policymaking perspective (von Schomberg, 2011). Scholars have adopted RRI in a wide range of emerging science fields to discuss scientists' responsibilities and approaches to engaging with different societal actors (e.g., Gjefsen & Vie, 2021; Regan, 2021). However, others have argued that the RRI applies better to radically novel and unpredictable issues than to other issues like clinical trials, which prioritize observation of the predictable and the known (Mertens, 2018).

AI ethics metrics have been widely examined, and most of them are aligned with the RRI framework. The commonly referenced rules include "ensure safety, ensure accountability, ensure fairness, uphold human rights and values, respect privacy, reflect diversity/inclusion, promote collaboration, avoid concentration of power, provide transparency, acknowledge legal/policy implications, limit harmful uses of AI, and contemplate implications for employment" (Russell & Norvig, 2021, p. 987). Principles, such as privacy, fairness, robustness, and safety were most frequently mentioned by AI scientists in their broader impact statements (Ashurst et al., 2022). However, high-level principles may not be sufficient to guarantee ethical AI, given that the current field of AI "lacks common aims and fiduciary duties, professional history and norms, proven methods to translate principles into practice, and robust legal and professional accountability mechanisms (Mittelstadt, 2019)," as compared to other established fields, such as medical ethics. The application of technical AI ethics metrics is further experiencing the dilemma of breakthroughs of extra large-scale data, which strengthens the capacity of models but increases biases simultaneously (Zhang et al., 2022). To address these complicated challenges, scholars stressed the necessity of social-systems analysis that investigates all the possible effects of AI systems on all parties at all stages, including conception, design, deployment, and regulation (Crawford & Calo, 2016), which indicates concrete, case-by-case analyses in individual research and development on AI. For instance, applying AI to allocate healthcare resources needs a comprehensive understanding of existing disparity in insurance plan enrollments across different racial and socioeconomic groups (Obermeyer et al., 2019). Translating high-level principles into research practices may also be complicated by social contexts at the global scale, as the meanings of philosophical terms such as bias, discrimination, and liberty, differ by political systems and governance.

Challenges moving forward

If it is clear that AI needs to be thought through, monitored, and responsibly developed, the modalities to do so are less clear. Some principles need to be taken into account: First, integrating large principles for ethical and responsible development of AI, such as equity and inclusion, needs to be considered alongside cross-disciplinary expertise and public input. Second, international collaboration for AI development must tackle globally sociotechnical challenges, while also taking into consideration political and cultural differences. Third, the inclusion of public input should take place in the “upstream” process before the research has been done (Scheufele, 2011). Without upstream engagement, important societal debates about potential consequences of AI development could be delayed, and exploratory research that ends up having negative ramifications for society could remain unquestioned.

Notably, as an enabling technology, AI can be applied to a wide range of fields in which development and deployment potentially involves various parties. Therefore, it is difficult to directly implement experience from other fields, as done with the global observatory in the field of gene editing that provides a platform for accessing worldwide ethical and policy response to technology development and for tracking conceptual developments, tensions, and consensuses through international meetings and deliberations (Jasanoff & Hurlbut, 2018).

Factors that influence scientists' attitudes toward scientific issues, regulations, and scientific conduct

Previous sections have discussed how the wickedness of AI has led to the importance of understanding multifaceted segments of scientists based on their attitudes toward AI, acknowledging various stakeholders' roles in AI regulation development, and incorporating social science input into AI research and development among scientific experts. This section reviews factors that shape scientific experts' attitudes toward scientific issues, related regulations, and scientific conduct.

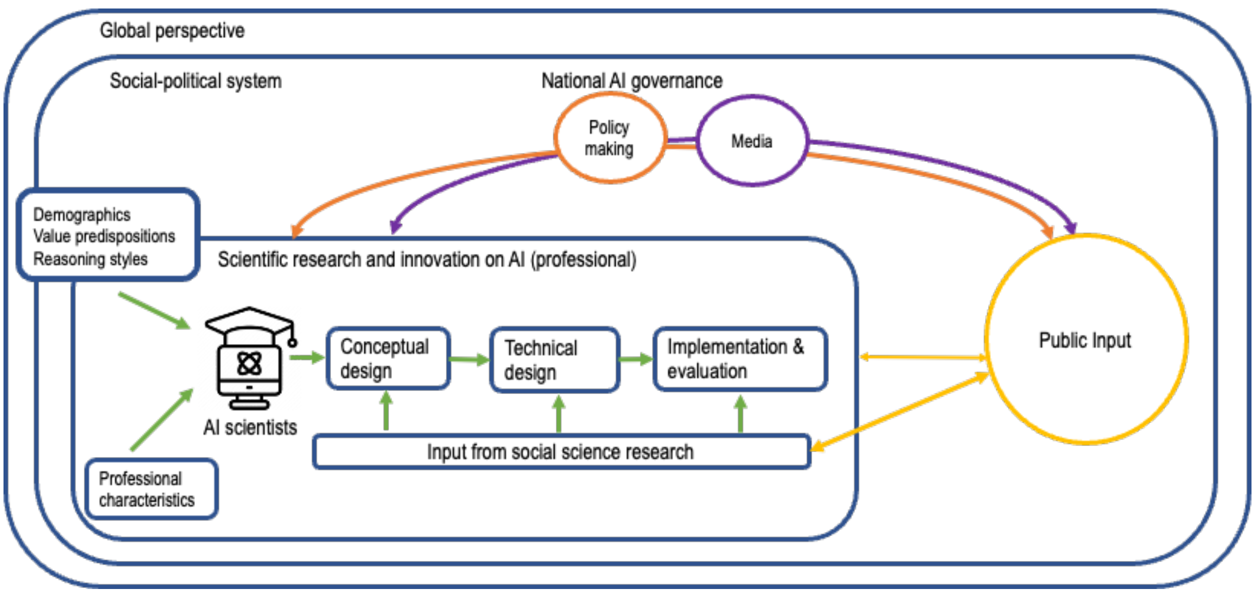
Prior literature on risk communication has distinguished between subjective perceptions of risks held by lay publics and scientific experts' perceptions of probabilities of certain technical outcomes (e.g., Fischhoff et al., 1983; Slovic, 1987). More recent science communication studies have found that scientific experts also use subjective value predispositions to form attitudes toward science and related regulations, such as trust in scientists, deference to scientific authority, and political ideology (e.g., Corley et al., 2009; Ho et al., 2011; Howell, Scheufele, et al., 2020; Su et al., 2016). Additionally, experts are susceptible to judgment biases because they might be "too quick to jump to conclusions, too slow to change their minds and too swayed by the trivia of the moment" (Kahneman, 2011, p. xix). These findings together reflect a complicated landscape regarding how scientific experts rely on various factors to process information and form attitudes.

To comprehensively investigate factors that shape scientific experts' views of AI, its regulation development, and the likelihood of incorporating social science input into AI research, this dissertation relies on a recent conceptual framework of three dimensions of science literacy—civic science literacy, digital media literacy, and cognitive science literacy (for a conceptual discussion, see Howell & Brossard, 2021). Accordingly, I develop three potential pathways, including value predispositions and reasoning styles (cognitive processing), media attention (information/media), and professional characteristics (science production), to explain scientific experts' views of AI, related regulations, and their scientific conduct. These three sets of factors encompass personal factors as well as the social dynamics of how society might influence science production and scientific experts' views of the societal impacts of their work.

Conclusion

Figure 1 presents an overview of the interface between AI scientists and society, as covered in this dissertation. While AI scientific experts are intellectual elites and cultural authorities who produce knowledge and develop applications of AI, they need input from lay publics and experts from other fields about the potential ethical, legal, and social implications of science development (e.g., Funtowicz & Ravetz, 1992; Wynne, 1992). AI experts are one of the key stakeholders when it comes to developing regulation, along with lay publics, civic groups, government branches, and tech companies. This dissertation examines how professional, informational, and dispositional factors may shape the multifaceted segmentation of AI scientific experts with regard to their views of AI and its regulation. Additionally, I explore how social sciences may bridge the interface between AI scientific experts and society. I also explore underlying mechanisms that foster the professional norms of incorporating social science input into AI research and development within the AI scientific communities.

Figure 1. Framework for the interface between AI scientists and society



Dissertation overview

In this dissertation, I conduct four related but independent studies that answer the research question of how input from social sciences may bridge the interface between AI scientists and lay publics. My dissertation is organized as follows.

In chapter 2, I discuss the survey methods used in this dissertation. The dissertation is based on a 2022 survey collected from scientific experts who have published articles in fields related to AI between 2010 and 2020. First, I present the procedure of participant recruitment through defining AI scientists, building a bibliometric definition of AI, collecting contact information, and sampling. Second, I discuss the procedure of fielding the online survey. Lastly, I present the descriptive statistics of the sample.

Chapter 3 presents study A, which identifies whether there are any attitudinal differences in views of the social impacts of AI between AI scientists and lay publics. What are the factors that account for the different views, such as demographics, value predispositions, and media use? The contributions of this study are two-fold. First, I compare the attitudinal differences using segmentation analyses guided by the risk-benefit typology. Second, I delve deep into the potential attitudinal differences within the scientific community to explore why some AI scientists have different views about AI outcomes compared to lay publics.

As discussed earlier in this introductory chapter, I propose the potential benefits of incorporating input from social sciences to connect AI scientists and lay publics. It is important to empirically examine whether such benefits exist. With these questions in mind, in Chapter 4, study B investigates the degree to which AI scientists have used social science research, ranging from reading articles to citing sources in their research. Furthermore, study B examines the impacts of incorporating input from social sciences on AI scientists' perceived roles of participating in AI regulation development. I analyze whether AI scientists think they alone can decide AI regulation development or whether they acknowledge the necessity to involve a diversity of social actors, such as policymakers, interest groups, and citizens. I construct different models of AI scientists' perceived roles in AI regulation development based on their views of who should have a say in AI regulation development. Additionally, I examine whether the use of social science research will make a difference in AI scientists' perceived roles in regulation development, among other factors.

Based on the proposed effectiveness of incorporating input from social sciences to bridge the interface between AI scientists and lay publics, this dissertation then explores how we can potentially incentivize AI scientists to incorporate social sciences' input. Prior research has demonstrated that scientific conduct is influenced by factors such as scientists' disciplines and value predispositions. It is unlikely sufficient to update AI scientists' views of how to conduct research by simply informing them of the necessity and importance of incorporating input from social sciences. In Chapter 5, study C analyzes different mechanisms that shape the likelihood of incorporating social sciences' input into AI research and development for individual AI scientists themselves and for most AI scientists. Study D embeds a survey experiment that takes social identity as one potential approach to compare the effectiveness of ingroup versus outgroup message cues about collaboration between AI scientists and social scientists. Participants were assigned to one of the four conditions (i.e., ingroup, outgroup, placebo, no stimuli) to read an introduction to a special issue. The study compares the effectiveness of ingroup versus outgroup cues on the likelihood of incorporating input from social sciences into different stages of AI research for individual AI scientists themselves and for most AI scientists.

The final chapter summarizes the research findings of this dissertation and discusses theoretical, methodological, and practical implications, as well as future research pathways to better connect AI scientists and lay publics to ensure effective science communication and responsible research and innovation.

Chapter 2. Methods

This dissertation used data from an online survey among AI scientists who are affiliated with institutions in the United States. The group of interest for my dissertation consists of AI scientists who have published papers related to AI.

Search string and sample collection

To compile a comprehensive pool of AI scientists, I used a bibliometric search method. This approach has been used in the contexts of other emerging science and technology issues, such as nanoscience and synthetic biology (e.g., Ho et al., 2011; Porter et al., 2008; Rose et al., 2018; Shapira et al., 2017). This approach identifies authors who have published research in AI from the Web of Science database. I chose the publication period from 2010 to 2020, because the number of AI research has a sharp increase since the last decade with the development of deep learning (Russell & Norvig, 2021) and news coverage of AI have rapidly increased since late 2009 (Fast & Horvitz, 2017). The eligible authors for this dissertation include those who are affiliated with U.S. institutions and who are correspondence authors.

Given the interdisciplinary nature of AI research, a bibliometric definition of AI faces the challenge of being either too broad or too narrow. To achieve a good balance of precision and recall, I combined deductive and inductive approaches to collect search terms and added conditional requirements to develop search strategies. The procedure is shown in Table 1.

Table 1. The procedure of search string development

Steps	Tasks and notes
1. Define AI and develop main categories	The definition of AI follows a widely used textbook, Artificial Intelligence: A modern approach. The following six disciplines consist of most of AI (Russell & Norvig, 2021): • natural language processing • knowledge representation • automated reasoning • machine learning • computer vision • robotics
2. Collect key search terms to develop the bibliometric definition of AI	• <u>Deductive:</u> I collected concepts introduced in the Russell & Norvig's textbook. • <u>Inductive:</u> I added additional keywords from an existing search string, which inductively extracted frequently used keywords from AI journal publications (Liu et al., 2021).
3. Add a conditional requirement to improve the precision	The conditional requirement includes: • major AI sub-categories (e.g., machine learning, natural language processing), • most widely used AI programming (e.g., deep learning, reinforcement learning) • keywords used in the most popular AI journals and conferences (e.g., machine intelligence).
4. Precision check	I manually checked the precision of each keyword by selecting a random sample of 25 retrieved articles and examined whether those articles were related to AI.

Table 2 shows the Boolean search strings in detail. A staff at the UW-Madison library assisted in extracting raw data from the Web of Science Dataset (WOS). The correspondence author information collected from WOS dataset contained email addresses. A researcher at UW-Madison Life Sciences Communication Department used Python to generate a list of US-based correspondence authors and eliminated duplicates, which includes a total of 35,833 authors with nonidentical emails.

Table 2. Boolean search strings

	Main categories	Key terms (Search: TS=Title/Abstract/Keywords)	Conditional requirement	Results (2010-20)
1	Artificial intelligence (inclusive)	TS = ("learning algorithm$" OR "artificial intelligen*" OR "computational intelligen*" OR "machine intelligen*" OR "human-level AI" OR "artificial general intelligence" OR "artificial superintelligence")	ALL = ("learning algorithm$" OR "artificial intelligen*" OR "computational intelligen*" OR "machine intelligen*" OR "machine* learning" OR "deep learning" OR "reinforcement learning" OR "knowledge representation" OR "natural language processing" OR "automated reasoning") OR (ALL = ("computer vision" NOT "computer vision syndrome"))	101,148
2	Natural language processing	TS = ("natural language processing" OR "natural language understanding" OR "natural language inference" OR "computational linguistics" OR "machine translation" OR "BERT" OR "language model*" OR "speech recognition" OR "recogniz* speech" OR "text-to-speech" OR "information extraction" OR "extract* information" OR "information retrieval" OR "retriev* information" OR "question answering")		30,570
3	Knowledge representation	TS = ("knowledge representation" OR "knowledge-based system$" OR "descript* logic$" OR "semantic network$" OR "ontological engineering" OR "default logic")		7,918
4	Automated reasoning	TS = ("automated reasoning" OR "forward chaining" OR "backward chaining" OR "logic programming" OR "first-order logic")		2,208
5	Machine learning	TS = ("machine learning" OR "deep learning" OR "reinforcement learning" OR "autonomic computing" OR "unsupervised learning" OR "supervised learning" OR "semisupervised learning" OR "semi-supervised learning" OR "transfer learning" OR "representation learning" OR "expert system$" OR "neural net*" OR "recurrent net*" OR "deep net*" OR "learning representation" OR "intelligent agent*" OR "multiagent" OR "multi-agent" OR "data mining" OR "gradient boosting" OR "backpropagation" OR "back-propagation" OR "long short-term memory" OR "perceptron$" OR "autoencoder*" OR "Q learning" OR "feedforward net*" OR "feed-forward net*" OR "hopfield net*" OR "boltzmann machine*" OR "activation function$" OR "generative adversarial net*" OR "ensemble Learning" OR "multitask learning" OR "multi* task learning" OR "random forest*" OR "support* vector* machine*" OR "genetic programming" OR "online learning" OR "kernel* method*" OR "fuzzy logic" OR "active learning" OR "markov decision process*" OR "bayes* net*" OR "gaussian process*" OR "genetic algorithm$" OR "probability logic" OR "probabilistic logic" OR "hidden markov model*" OR "evolution* algorithm*" OR "heuristic search" OR "anytime algorithm$" OR "graph* model*" OR "spatial reasoning" OR "em algorithm$" OR "markov chain*")		316,126
6	Computer vision	(TS = ("computer vision" NOT "computer vision syndrome")) OR (TS = ("convolutional net*" OR "deepfake*" OR "image classification" OR "classify* image*" OR "image understanding" OR "understand* image*" OR "pattern recognition" OR "recogniz* pattern*" OR "emotion recognition" OR "recogniz* emotion*" OR "facial recognition" OR "recogniz* face*"))		57,120
7	Robotics	TS = ("robot*" OR "robotic*" OR "imitation learning" OR "motion planning" OR "autonomous" OR "automate*" OR "automation*")		62,448
	1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7			405,131

Note. Key terms may relate to more than one main category.

Survey implementation and sample statistics

This project received Institutional Review Board approval for all parts of the data collection and analyses. Participants gave informed consent after receiving explanations of the study and possible consequences. Excluding 6,178 bounced emails, a total of 29,655 AI scientific experts with nonidentical emails were contacted to participate in a 15-minute web survey on Qualtrics. No incentives were provided. The survey was conducted from March to April 2022 with one initial email invitation and three subsequent reminders to non-respondents. The sample consisted of 2,352 respondents who had completed at least eighty percent of the survey, with a response rate of 8% (AAPOR, 2016; RR6).

To ensure the quality of the sample and examine non-response biases, I compared the demographic and professional characteristics of respondents across all four waves of email contact (Dillman et al., 2014). The wave refers to the time when respondents submitted their survey relative to the round of email invitations or reminders. Using Tukey's test for post-hoc analysis and the chi-square test, I did not find any significant individual differences in age, gender, and title across the four waves of email contact. There were some differences in professional characteristics. The average academic age of wave 2 is younger than that of wave 1. Compared to wave 1, wave 4 has more scientists from engineering but fewer from social, behavioral, and economic sciences. The components of scientists from the other fields remain similarly across the four waves of email contact. Besides, during the four weeks of the survey implementation, news coverage of AI, such as the use of AI in the Russia-Ukraine war, could have driven higher attention to the societal impacts of AI and indirectly influenced our response rate. However, no statistically significant differences were found in terms of attention to political and science news among respondents across the four waves of email contact. Therefore, I conclude that this sample has reached AI scientists from a wide range of fields, without overt self-selection biases.

The respondents in the sample represent a range of positions, including university faculty (62.8%), industry scientists and engineers (16.5%), government scientists (7.1%), graduate students (8.2%), and others (5.4%). Respondents focus on diverse subfields of AI, such as machine learning (80.7%), computer vision (25.5%), natural language processing (24.4%), knowledge representation (16.2%), ethics and societal impacts of AI (16.0%), robotics (10.5%), automated reasoning (9.3%), and other AI-related subfields (3.4%). 5.1% of the respondents identified that their work is not related to AI. The respondents also come from various academic fields, including computer and information science (48.4%), engineering (27.7%), medical sciences (19.1%), social, behavioral, and economic sciences (15.2%), mathematical and physical sciences (14.2%), biological sciences (14.2%), geoscience (4.0%), agriculture and food sciences (3.2 %), arts and humanities (3.2%), environmental resources and education (2.8%), and education and human resources (2.3%). Because this dissertation focuses on AI scientific experts, those who are from the fields of arts and humanities or those who do not identify their work as related to AI are not included for analysis. The number of eligible participants for analysis is 2,199.

Regarding demographics, the mean age is 45.3 years (*SD* = 13.1) and the mean academic age is 14.8 years (*SD* = 12.8). The gender distribution is as follows: 79.0 percent males, 18.1% females, and 0.8% non-binary. The sample has a majority of white (59.9%) and Asian (30.0%) respondents, followed by Hispanic or Latino (5.1%), Black or African American (2.0%), American Indian or Alaskan Native (0.5%), Native Hawaiian or the Pacific Islander (0.3%), and others (4.5%).

The demographic characteristics of this sample were compared with publicly available data on US computer scientists (ZIPPIA, 2022) and the US population (see Table 3). None of these percentages deviate more than six percentage points from the US computer scientists. Compared with the US Census data, the demographics of AI scientific experts have more males and Asians but fewer Black or African American people and White people. Demographic differences between a sample of experts and US Census data are similar to those found in other expert-public comparisons (see Ho et al., 2011; Howell, Scheufele, et al., 2020).

Table 3. Comparisons of demographics across the dissertation sample, US computer scientist data, and current US population

	The present work	US computer scientists (ZIPPIA, 2022)	US Census (2021)
Gender			
Female	18.1%	21.2%	50.5%
Male	79.0%	78.8%	49.5%
Non-binary	0.8%		
Not listed or prefer not to say	2.1%		
Race			
White	59.9%	66.1%	75.8%
Asian	30.0%	25.0%	6.1%
Black or African American	2.0%	1.0%	13.6%
America Indian and Alaska Native	0.5%	0.9%	1.3%
Native Hawaiian and Pacific	0.3%		0.3%
Two or more races			2.9%
Hispanic or Latino	5.1%	5.2%	18.9%
White alone, not Hispanic or Latino			59.3%

Note. In the US Census survey, Hispanics may be of any race, so they are included in applicable race categories. The other race categories include persons reporting only one race. In the present work, respondents can choose more than one race category.

Chapter 3. AI scientists' and lay publics' views of AI and its social impacts: A comparison of segmentation analyses (Study A)

A variety of AI applications have profound and diverse lifestyle and policy implications, associated with value and interest conflicts across different social groups without clear-cut solutions. Because risk assessments of science and technology are politically and culturally constructed, scientific experts may not be able to represent lay publics to frame risks (Sarewitz, 2015). Therefore, the prerequisite for societal debates about AI and effective communication between scientific experts and lay publics is to map the landscape of the variety of perspectives that the two stakeholder groups hold on AI and what factors shape the potential attitudinal differences within and across the two stakeholders. Without this nuanced understanding of how scientific experts and lay publics think about and respond to potential risks and benefits associated with AI, the adoption of AI systems may face a mismatch between the demand and the supply sides and difficulties in reaching value trade-offs across social groups.

Despite the fact that US citizens hold the highest trust in university scientists in keeping society's best interest in mind during AI development compared to their trust in other social actors or institutions (Calice et al., 2022), cumulative evidence shows that AI scientific experts and lay publics have various views of responsible AI development. For example, US citizens consider safety, performance, and privacy most important, whereas AI practitioners find fairness most important (Jakesch et al., 2022). The two groups also demonstrate different levels of trust in the entities that keep the best interests of the public in AI development. Compared to the general public, AI/ML researchers show higher levels of trust in scientific associations and international institutions, but they do not trust the military as much as lay publics do (Zhang et al., 2020).

These studies provide important insights to guide responsible AI development and effective AI governance. Yet, to the best of my knowledge, there have been no empirical studies comparing nuanced views of AI within and across different segments of scientific experts and lay publics. The comparison of multifaceted segmentation between scientific experts and lay publics can directly respond to the increasing concern that AI development lacks a diverse representation of social groups, especially minority groups. In a 2021 Pew survey, US citizens expressed concern that the experiences and views of women, Black, Hispanic, and Asian adults are less considered than men and White groups when AI programs are designed (Pew Research Center, 2022). A comprehensive understanding of the multifaceted segmentation of different stakeholders can ensure the inclusion of diverse opinions, knowledge, and judgment within a collective, which will be associated with more intelligent outcomes in societal debates and deliberation (Cappella et al., 2017).

Using a 2022 AI scientific expert dataset, this study first replicates a latent class analysis (LCA) that segments US citizens on their risk and benefit perceptions of AI. Not only do risk and benefit assessments shape individuals' support for science and related policy (e.g., Akin et al., 2017; Su et al., 2016), but risk and benefit assessments are also key information associated with new technology that policymakers communicate with the public based on the best available scientific evidence (Holdren et al., 2011). Methodologically, the LCA procedure allows for segmentation based on more complex constellations of attitudes, such as the coexistences of high risk and benefit perceptions (i.e., ambivalence), an understudied but important concept in examining attitudes toward wicked science issues (Pidgeon et al., 2005; Wirz, 2021). This study also moves beyond presenting descriptive analysis by conducting multivariate analyses to examine how value predispositions, news attention, and professional characteristics distinguish AI scientific experts' attitude types in different segments.

Existing typology of risk and benefit perceptions

In many social contexts, including forming attitudes toward science and technology, positive and negative evaluations can independently and equivalently coexist (Kaplan, 1972; Priester & Petty, 2001). The concurrence of predominantly positive and negative evaluations refers to ambivalent attitudes and an equivalent low level of positive and negative evaluations refers to indifferent attitudes (Kaplan, 1972). The degree of ambivalent attitudes can be assessed either by comparing separate scales of positive and negative evaluations (e.g., Kaplan, 1972) or by a unidimensional subjective measure capturing the feelings of tension or conflict when evaluating one object (e.g., Priester & Petty, 2001). Risk communication scholars have incorporated the measure of ambivalence to develop a four-way typology of different combinations of benefit and risk perceptions: positive (high benefit/low risk), negative (low benefit/high risk), ambivalent (high benefit/high risk), and indifferent (low benefit/low risk) (Pidgeon et al., 2005; Poortinga & Pidgeon, 2006).

As discussed in Chapter 1, for AI scientific experts, perceiving risks and benefits of AI concurrently reveal the complexity of AI, which may increase the likelihood of understanding that different social groups may hold a variety of values and interests related to the same AI application. Little research has conducted segmentation analysis and examined what factors may shape scientists' attitudinal ambivalence. Prior research on comparing scientists' and publics' views of science provide clues that scientist samples may posit differently in the risk-benefit typology. Lay publics usually perceive higher risks and lower benefits than scientists about emerging science, such as nanotechnology and synthetic biology (e.g., Ho et al., 2011; Howell, Scheufele, et al., 2020), though the two groups may worry about different risky consequences associated with science (Scheufele et al., 2007). For instance, scientists may worry more about environmental and long-term health problems caused by nanotechnology than lay publics, whereas lay publics may be more worried about the loss of jobs and an arms race (Scheufele et al., 2007). Given these considerations, the question remains as to how diverse scientific experts' views of AI are as compared to those of lay publics. How does the segmentation of AI scientific experts and laypeople differ? I therefore raise the following two questions:

RQ1: How can AI scientific experts be segmented based on their risk and benefit perceptions of AI?

RQ2: How does the segmentation of AI scientific experts differ from that of lay publics in terms of different combinations of risk and benefit perceptions?

Factors that distinguish scientific experts' segments on risk and benefit perceptions of AI

One recurrent explanation for attitudinal difference between scientists and lay publics is the degree of science literacy or familiarity with scientific issues. In some prior studies, public familiarity with the issue was found associated with reduced risk perceptions, though cultural factors may still play a role in explaining the difference in risk perceptions between informed publics and academic experts (e.g., Capon et al., 2015). Overstating the difference in issue familiarity between experts and lay publics might oversimplify mechanisms that shape attitudes toward science. While scientific experts are intellectual elites who produce knowledge about science and technology, it is debatable whether they excel in knowing potential ethical, legal, and social implications of science development (e.g., Funtowicz & Ravetz, 1992; Wynne, 1992).

Moreover, experts are susceptible to judgment biases because they might be "too quick to jump to conclusions, too slow to change their minds and too swayed by the trivia of the moment" (Kahneman, 2011, p. xix).

As discussed in Chapter 1, this study relies on a recent conceptual framework of three dimensions of science literacy—civic science literacy, digital media literacy, and cognitive science literacy (for a conceptual discussion, see Howell & Brossard, 2021). Howell and colleagues suggest that science literacy spans the lifecycle of science production, media and online information production, and opinion formation (Howell & Brossard, 2021). The framework is also applicable to explain scientists' attitudes towards issues on which they have expertise by considering the impacts of the societal contexts that shape science. The next sections review three potential pathways, including value predispositions (cognitive processing), media attention (information), and professional characteristics (science production) that may shape the segmentation of different types of risk and benefit perceptions of AI among scientific experts.

The role of value predispositions

Science communication studies have underscored the important role of value predispositions, such as political ideology, religiosity, and deference to scientific authority in shaping public attitudes toward science (e.g., Akin et al., 2017; Brossard & Nisbet, 2007; Scheufele & Lewenstein, 2005). Value predispositions serve as perceptual filters that help individuals form opinions or make decisions without abundant cognitive efforts (Fiske & Taylor, 1991; Scheufele, 2006). The use of value predispositions to form attitudes seems inevitable for anyone, including scientific and political experts who are considered more knowledgeable than lay publics on issues related to their work (e.g., Corley et al., 2009; Ho et al., 2011; Lee et al., 2020; Su et al., 2016). For instance, a paired survey found that government officials had more accurate factual beliefs than the public across a series of politically controversial issues. However, the two groups showed comparable partisan divides, with Republicans knowing better on issues like health care spending and Democrats on issues like climate change (Lee et al., 2020).

Despite the common use of value predispositions in shaping science attitudes, scientists and lay publics may rely on different value predispositions. While lay publics rely more heavily on values such as political ideology and religiosity (e.g., Ho et al., 2011; Howell, Scheufele, et al., 2020), scientists are more likely to use heuristics such as deference to scientific authority and trust in scientists (e.g., Ho et al., 2011).

This study focuses on three value predispositions that may relate to different levels of risk and benefit perceptions of AI: political ideology, deference to scientific authority, and belief in the authority of science as a way of knowing. Given the politicization of many science issues, individuals tend to form attitudes that are aligned with their ideologically congruent sources (Yeo et al., 2015). The impact of political ideology on scientists' attitudes has been found to be mixed. Social conservatism is significantly associated with lower levels of benefit perceptions of synthetic biology among synthetic biologists (Howell, Scheufele, et al., 2020), whereases there is no such difference across the political spectrum regarding scientists' risk perceptions of synthetic biology (Howell, Scheufele, et al., 2020) and nano scientists' support for academic and commercial regulation of nanotechnology (Su et al., 2016). AI applications have raised politically charged promises and concerns regarding national security, policing, elections, and misinformation (e.g., Knight, 2021; Lohr, 2018; Metz, 2019), which may make pre-existing political ideology more salient as a heuristic when forming risk and benefit perceptions of AI.

Although it is difficult to predict directional effects of ideology on risk and benefit perceptions of AI with an amalgam of value-laden considerations, I anticipate that scientific experts with less extreme political ideology are more likely to reconcile competing values associated with different risk and benefit perceptions of AI. Thus, I propose the following hypothesis:

> **H1:** AI scientific experts with more extreme political ideology will be less likely to perceive higher risks and benefits of AI simultaneously than those with moderate political ideology.

Deference to scientific authority is a consistent and strong value predisposition associated with pro-technology views, such as greater benefit perceptions and support for science across lay publics and scientists (Akin et al., 2017; Brossard & Nisbet, 2007; Ho et al., 2011). Deference to scientific authority taps into the belief that the knowledge production processes by scientists are in a systematic, objective, and apolitical manner and in the best interests of the public (for a review, see Brossard & Nisbet, 2007). The perceived exceptionalism of science as an institution is reinforced through US primary and secondary education systems (Brossard & Nisbet, 2007).

Despite similarities between the two concepts, deference to scientific authority and belief in the authority of science as a way of knowing are conceptually different in at least two aspects. First, US citizens do not always reveal consistent attitudes toward the scientific community and scientific research. For instance, those who identified as stable conservatives have less confidence in the scientific community, while simultaneously having more positive attitudes toward scientific research compared to those who are no longer conservatives or recently become conservatives (Mann & Schleifer, 2019). Second, deference to scientific authority indicates a more authoritative way of scientific conduct that excludes the public from consultation or involvement in science decision-making (Brossard & Shanahan, 2003; Howell, Wirz, et al., 2020). The level of belief in the authority of science as a way of knowing does not have such anti-democratic views of science decision making (Howell, Wirz, et al., 2020). Regardless of how lay publics treat the two concepts differently, no research has explicitly examined both value predispositions using a scientist sample. I therefore propose the following hypotheses:

H2a: Deference to scientific authority is associated with higher benefit perceptions and lower risk perceptions of AI among scientific experts.

H2b: Belief in the authority of science as a way of knowing is associated with higher benefit perceptions and lower risk perceptions of AI among scientific experts.

The role of news attention

Media plays a major role in connecting scientists and lay publics, as most lay publics are exposed to recent scientific development through mediated realties of science rather than through direct experience or engagement activities (Scheufele, 2014). Media coverage of science does not simply translate scientific discovery. Instead, journalists serve as gatekeepers that use news values to filter out the influential agendas for the public (Shoemaker et al., 2009). In light of this epistemology of science journalism, media coverage of science reduces the technical complexity of scientific facts but expands meaning making for public consideration and discussion of the impacts of science on society that are beyond scientific facts (Brennen, 2018).

In turn, mediated representation of science provides scientific experts with potential availability heuristics to understand how social contexts shape their work, including the potential risks and benefits of their work on society. Attention to news is particularly important to AI scientific experts, as they are sometimes criticized by the lack of experience and knowledge of how marginalized social groups might be underrepresented and vulnerable to AI applications (e.g., Benjamin, 2019). In a blog post titled "It's time to do something: Mitigating the negative impacts of computing through a change to the peer review process," published on the Association for Computing Machinery website, the world's largest educational and scientific computing society, a number of computer scientists suggested the surveillance function of media on their work, saying that "if there has been significant media coverage of a negative impact that is relevant to a paper or proposal, that impact should be addressed by the authors of that paper or proposal" (Hecht et al., 2018, p. 2).

The influence of news attention on science attitudes depends on how science is portrayed in media. Longitudinal, large-scale analyses of US and Western news coverage of AI over the last three to five decades show that AI related coverage is more positive than negative or has no negative bias against AI (Fast & Horvitz, 2017; Garvey & Maskal, 2019). However, as AI-related sociopolitical events have gained increasing attention in overtly policy arenas such as a surge of legislative proposals to Congress, recent research has discovered more nuances of AI-related coverage, showing a mix of risk and benefit considerations and potential partisan differences in news reporting. A recent study found that US news coverage on facial recognition between January and July 2020 has focused on themes like privacy and surveillance, racial bias, technology's ability to provide solutions, and technology's problematic development (Shaikh & Moran, 2022). Considering the problematic use of facial recognition, right-leaning media mentioned more abuses by foreign governments, whereas left-leaning media focused more on ethical problems, such as misusing personal data (Shaikh & Moran, 2022).

The formation of attitudes toward to AI is likely a function of media attention to different types of news, including science and political news. For lay publics, a relatively balanced diet of political and science news is associated with more ambivalent attitudes toward AI (Bao et al., 2022). Paying substantially more attention to political news over science news is associated with more skeptical attitudes toward AI (Bao et al., 2022). I anticipate similar effects of media attention on opinion formation for scientific experts and lay publics. Therefore, I propose the following hypotheses:

H3a: Attention to political news is associated with higher risk perceptions and lower benefit perceptions of AI among scientific experts.

H3b: Attention to science news is associated with lower risk perceptions and higher benefit perceptions of AI among scientific experts.

H3c: Attention to both political and science news is associated with higher risk perceptions and higher benefit perceptions of AI among scientific experts.

The role of professional characteristics

Because scientists do not constitute a monolithic group (Yeo & Brossard, 2017), I focus on how professional characteristics, such as scientists' fields and collaboration networks, may influence AI scientists' perceived risks and benefits of AI.

Research across multiple emerging science issues has demonstrated that scientists from various sub-fields will have different perspectives and bring different values and evidence to study the impacts of the same science and technology issue, especially the significance for society (Bertoldo et al., 2016; Sarewitz, 2015). In a study that examined scientists' views of nanotechnology, scientists across sub-fields had a high consensus on the benefits provided by nanotechnologies, but they disagreed with the extent to which nanotechnologies represented risks (Bertoldo et al., 2016). Scientists in the fields of physics and chemistry perceived benefits of nanotechnologies as predominantly outweighing risks, whereas toxicologists, life scientists, and social scientists perceived comparable levels of both benefits and risks of nanotechnology (Bertoldo et al., 2016). Given the interdisciplinary nature of AI research and its diverse AI applications, I anticipate a similar difference in attitudes among AI scientific experts from different subfields. Therefore, I propose the following research question:

RQ3: How do AI scientific experts from different subfields vary in their views of AI risks and benefits?

Interdisciplinary research collaboration is a way for scientists to acquire and implement scientific and technical human capital, which refers to the sum of a scientist's scientific, technical, and social knowledge, skills and resources as well as their social and professional network ties (Bozeman et al., 2001). Despite the potential barriers to collaboration, such as grant opportunities, collaboration across different genders, career status, institutions and industries can provide opportunities to create knowledge that would not have been possible without the expanded networks and multifaceted skills and resources (Bozeman & Corley, 2004; Bozeman et al., 2001). As for interdisciplinary collaboration among scientific experts from different subfields, it is likely that the exchange of heterogenous views of AI, training background, and protocols of research on human subjects will enable scientific experts to form more complicated views of AI and overcome overconfidence of one's knowledge. Thus, I raise the following hypothesis:

H4: AI scientific experts with a more diverse collaboration network will be more likely to perceive simultaneously higher risks and benefits of AI.

This study also considers the impacts of demographic differences, such as age and gender, on scientists' judgment of risks and benefits of AI. Older scientists are likely to perceive lower levels of benefits of emerging science and technology, such as nanotechnology (Ho et al., 2011) and synthetic biology (Howell, Scheufele, et al., 2020). Female scientists perceive higher risks of nanotechnology than their male peers (Ho et al., 2011).

Methods

Details of the sample are discussed in Chapter 2.

Measures

Dependent variables

Perceived risks and benefits of AI. I used the same measure from a segmentation study on a 2020 US public sample (Bao et al., 2022). Respondents were asked to rate the likelihood of ten possible risks and benefits of AI using a 7-point scale (1 = not at all likely, 7 = certain). Table 4 presents the comparisons between the AI scientific expert and the US public sample. Notably, factor analyses suggested the exclusion of one benefit item (i.e., reduce bias in human decision-making) and one risk item (i.e., change what it means to be human) for further analyses of the AI scientific expert sample due to cross-loading concerns. The remaining eight items were used for LCA. The index of benefit perceptions (M = 5.0, SD = 1.0, *Cronbach's* α = .81) and risk perceptions (M = 4.8, SD = 1.1, *Cronbach's* α = .77) were used for separate OLS regression models to test the research questions and hypotheses.

Table 4. Measures and descriptive statistics of dependent variables: How likely AI will…?

		2022 AI scientific expert sample		2020 US public sample	
		Mean	SD	Mean	SD
Perceived benefits	strengthen the U.S. economy	5.2	1.2	3.7	1.4
	increase national security	4.6	1.3	4.0	1.4
	improve individuals' health	5.3	1.1	4.0	1.4
	reduce bias in human decision-making	4.0	1.4	3.6	1.5
	help fight terrorism threats	4.8	1.3	4.1	1.4
Perceived risks	worsen societal inequalities	4.5	1.4	4.3	1.5
	give some people too much power	5.2	1.4	5.1	1.5
	threaten personal liberties	4.6	1.4	4.6	1.5
	change what it means to be human	3.8	1.7	4.1	1.7
	displace workers by automating their jobs	5.0	1.4	5.1	1.5

Independent variables

Political ideology was measured with a 5-point liberal-conservative scale for social (M = 2.2, SD = 1.0) and for economic issues (M = 2.6, SD = 1.0). I did not combine the two items due to a medium correlation (r = .51, p < .001).

Deference to scientific authority was adopted from established measures (Brossard & Nisbet, 2007; Howell, Wirz, et al., 2020), asking respondents' agreements on three items using a 5-point scale: "scientists know best what is good for the public;" "scientists should do what they think is best, even if they have to persuade people that it is right;" and "scientists should be able to conduct their research without consulting the public." Despite the high validity and reliability of the measure across multiple public samples, the three items yielded low reliability in our scientist sample (*Cronbach's* α = .48). I chose the item "scientists know best what is good for the public" for the analysis (M = 2.5, SD = 0.9). Belief in the authority of science as a way of knowing was measured by two items adopted from an established index (Howell, Wirz, et al., 2020). Respondents were asked their agreement on two items using a 5-point scale: "science is the best way to understand the world" and "science is the best way that society has of producing reliable knowledge." I averaged them to make an index ($M = 4.1$, $SD = 0.8$, $r = .62$, $p < .001$).

Attention to political news asked the extent to which respondents pay attention to news stories about a) national government and politics and b) international affairs. I averaged the two items to make an index (1 = none, 5 = a lot; $M = 3.6$, $SD = 0.9$, $r = .66$, $p < .001$). Attention to science news was an average of attention to a) recent developments in science and technology, b) ethical implications of emerging science and technology, and c) regulations of emerging science and technology ($M = 3.6$, $SD = 0.7$, *Cronbach's* $\alpha = .71$).

Research fields. I recoded respondents' research fields into four categories: a) computer and information science ($N = 1,111$; including those who also selected other fields); b) life sciences ($N = 382$) – agriculture and food; biological sciences; and medical sciences; c) physical sciences ($N = 506$) – engineering; geoscience; and math and physics; and d) social sciences ($N = 261$) – social, behavioral, and economic sciences; environmental resources and education; education and human resources (social science fields were given a priority over the life sciences and physical sciences). Each pair of the four categories were mutually exclusive except life scientists and physical scientists ($N_{\text{overlap}} = 63$).

Interdisciplinary collaboration. Respondents were asked how often they collaborated and/or coauthored with researchers from four areas, including life sciences, physical sciences, computer sciences, and social sciences, using a 5-point never–very often scale. Due to the skewness of distribution and low reliability across items, I recoded them into four binary variables (0 = never, 1 = ever) to make an additive index, ranging from 0 to 4 ($M = 2.9$, $SD = 1.1$).

Analysis

This study first used poLCA R package to classify respondents with similar levels of perceived risks and benefits of AI. To further differentiate characteristics of each class, I compared their demographics, value predispositions, news attention, views of scientific conduct, attitudes toward AI regulation, and concerns about AI-based discriminations with post hoc tests (see measurement in Table 5).

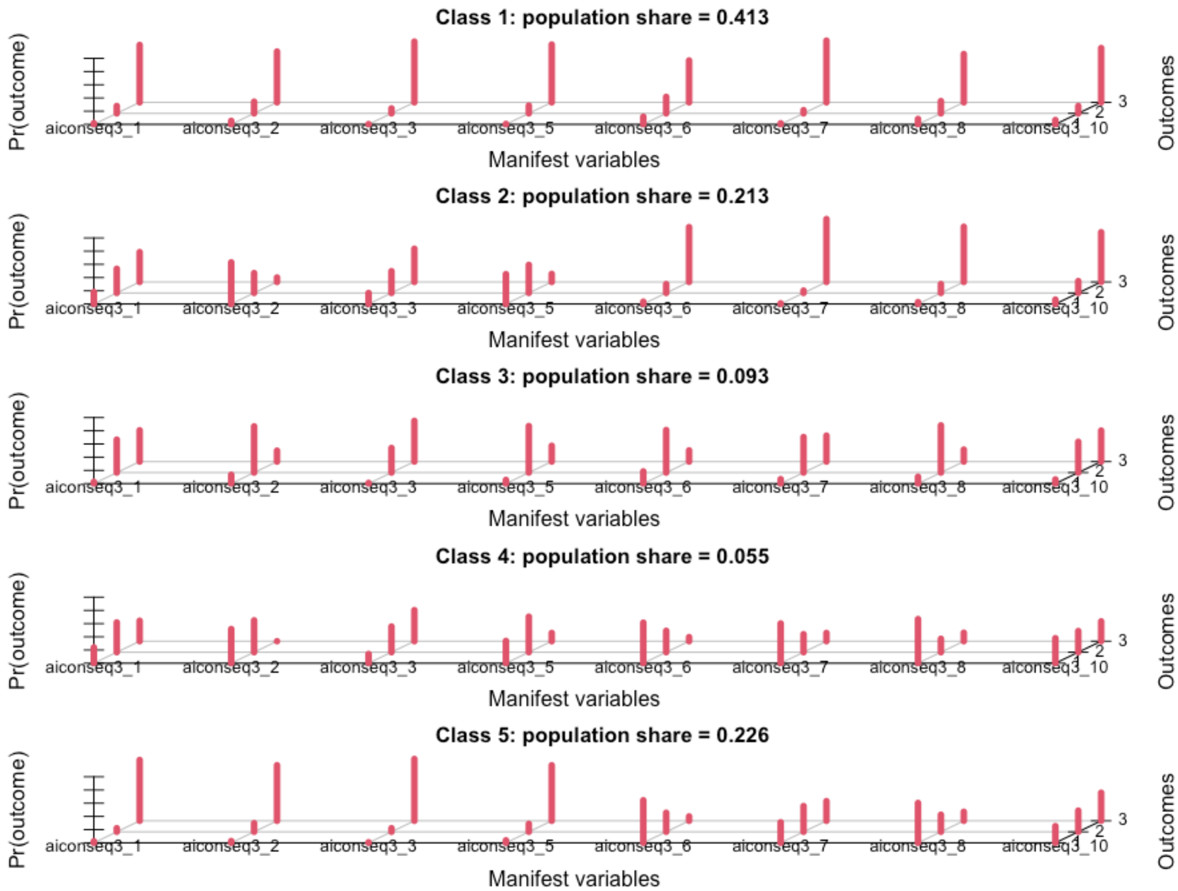
I then used a multinomial logistic regression model to examine how value predispositions, news attention, and professional characteristics distinguish different segments of AI scientific experts. All independent and control variables were entered in one model. I also applied OLS regression models to predict AI scientific experts' risk and benefit perceptions respectively to provide supplemental information. I focused on interpreting coefficients that are significant at 0.01 level to avoid the potential for Type I errors in a large sample (N = 2,199).

Results

Segmenting AI scientific experts on their risk and benefit perceptions

RQ1 asked what segments exist among AI scientific experts regarding their risk and benefit perceptions of AI. The five-class LCA model was chosen because it provides theoretically sound interpretations and it has the lowest BIC value, indicating a good model performance. Figure 2 presents the five classes of AI scientific experts: the ambivalent class (41%), perceiving high risks and high benefits; the supportive class (23%), perceiving high benefits and low risks; the skeptical class (21%), perceiving low benefits and high risks; the ambiguous-leaning attitude class (9%), perceiving medium levels of benefits and risks; the indifferent-leaning class (6%), perceiving low to medium levels of risks and benefits.

Figure 2. Estimated latent class conditional probabilities (AI scientific experts' risk and benefit perceptions of AI)



Note. Class 1: ambivalent; Class 2: skeptical; Class 3: ambiguous-leaning; Class 4: indifferent-leaning; Class 5: supportive. Respondents were asked to rate the likelihood of possible risks and benefits of AI in the following aspects: a. strengthen the U.S. economy (aiconseq3_1); b. increase national security (aiconseq3_2); c. improve individuals' health (aiconseq3_3); d. help fight terrorism threats (aiconseq3_5); e. worsen societal inequalities (aiconseq3_6); f. give some people too much power (aiconseq3_7); g. threaten personal liberties (aiconseq3_8); h. displace workers by automating their jobs (aiconseq3_10). 1 = "not at all likely," "very unlikely," and "unlikely," 2 = "somewhat likely," and 3 = "likely," "very likely," and "certain."

Table 5. Description of segments on AI scientific experts' characteristics and attitudes toward AI

	Ambivalent	Supportive	Indifferent leaning	Skeptical	Ambiguous leaning	Total	Scale
Age	45.65 [a]	46.49 [a]	45.50 [a]	43.64 [a]	44.60 [a]	45.28	numeric
Gender (male)	81% [a]	80% [a]	74% [a]	77% [a]	72% [a]	79%	categorical
Race							
White	59% [a, b]	54% [a]	60% [a, b]	69% [b]	55% [a, b]	60%	dummy
Asian	31% [a]	35% [a]	29% [a, b]	20% [b]	35% [a]	30%	dummy
Black	2% [a]	2% [a]	4% [a]	1% [a]	2% [a]	2%	dummy
Hispanic or Latino	4% [a]	5% [a]	9% [a]	5% [a]	7% [a]	5%	dummy
Media attention							
Attention to political news	3.65 [a, c]	3.38 [b]	3.38 [a, b]	3.72 [c]	3.55 [a, b, c]	3.58	1-5
Attention to science news	3.64 [a]	3.55 [a]	3.44 [a]	3.64 [a]	3.53 [a]	3.60	1-5
Value predispositions							
Economic ideology (high=conservatism)	2.65 [a]	2.84 [a]	2.66 [a]	2.26 [b]	2.69 [a]	2.61	1-5
Social ideology (high=conservatism)	2.18 [a]	2.43 [b]	2.37 [a, b]	1.78 [c]	2.26 [a, b]	2.16	1-5
Religious guidance	2.04 [a, b]	2.14 [a]	2.15 [a, b]	1.86 [b]	2.13 [a, b]	2.03	1-5
Science is the best way that society has of producing reliable knowledge.	4.36 [a]	4.36 [a]	4.11 [a, b]	4.16 [b]	4.18 [a, b]	4.29	1-5
Science is the best way to understand the world.	4.06 [a]	4.11 [a]	3.75 [a, b]	3.75 [b]	3.84 [a, b]	3.97	1-5
Scientists know best what is good for the public.	2.52 [a]	2.73 [b]	2.35 [a, c]	2.23 [c]	2.57 [a, b]	2.50	1-5
Scientists should do what they think is best, even if they have to persuade people that it is right.	3.60 [a]	3.60 [a, b]	3.36 [a, b]	3.40 [b]	3.37 [a, b]	3.53	1-5
Scientists should be able to conduct their research without consulting the public.	2.87 [a, b]	3.07 [a]	2.97 [a, b]	2.81 [b]	2.85 [a, b]	2.90	1-5
Scientists should pay attention to the wishes of the public, even if they think citizens are mistaken or do not understand their work.	3.59 [a]	3.40 [a]	3.32 [a]	3.62 [a]	3.54 [a]	3.54	1-5
It is appropriate for scientists to become actively involved in political debates about issues like AI.	4.11 [a]	3.96 [a, b]	3.74 [b]	4.15 [a]	4.02 [a, b]	4.06	1-5
Citizens should have a say in the development of these regulations in the United States (yes)	78% [a, c]	71% [a, b]	61% [b]	82% [c]	79% [a, b, c]	76%	dummy

	Ambivalent	Supportive	Indifferent leaning	Skeptical	Ambiguous leaning	Total	Scale
Attitudes toward scientific conduct							
Scientists are responsible for the way their discoveries are used by other people.	2.73 ^{a}	2.71 ^{a}	2.68 $^{a, b}$	3.01 ^{b}	2.80 $^{a, b}$	2.79	1-5
A discovery is in itself neither good nor bad, it is only the way the discovery is used that matters.	3.76 ^{a}	3.85 ^{a}	3.59 $^{a, b}$	3.37 ^{b}	3.61 $^{a, b}$	3.67	1-5
The authorities should require scientists to respect ethical standards.	4.25 ^{a}	4.08 ^{b}	4.04 $^{a, b}$	4.25 $^{a, b}$	4.09 $^{a, b}$	4.19	1-5
Scientists should be free to carry out the research they wish, provided they respect ethical standards.	4.15 ^{a}	4.16 ^{a}	3.95 $^{a, b}$	3.92 ^{b}	3.98 $^{a, b}$	4.07	1-5
Attitudes toward AI regulation							
Advancing AI quickly is more important than protecting society from the unknown risks of AI.	2.29 ^{a}	2.70 ^{b}	2.38 $^{a, b}$	1.83 ^{c}	2.35 ^{a}	2.29	1-5
Regulating AI may significantly slow down important scientific progress.	3.32 ^{a}	3.35 ^{a}	3.05 $^{a, b}$	2.95 ^{b}	3.21 $^{a, b}$	3.22	1-5
Existing regulations for AI research are sufficient.	2.70 ^{a}	2.96 ^{b}	2.80 $^{a, b}$	2.40 ^{c}	2.82 $^{a, b}$	2.71	1-5
Existing regulations for AI applications are sufficient	2.19 ^{a}	2.60 ^{b}	2.47 $^{a, b}$	1.87 ^{c}	2.38 $^{a, b}$	2.24	1-5
As a society, we are prepared for the potential effects of AI applications.	1.96 ^{a}	2.60 ^{b}	2.46 $^{b, c}$	1.61 ^{d}	2.22 ^{c}	2.07	1-5
There will be unintended consequences of AI applications.	4.43 ^{a}	3.89 ^{b}	3.88 $^{b, c}$	4.58 ^{d}	4.13 ^{c}	4.30	1-5
Inequality: Concern about AI worsening discrimination based on …							
gender	2.86 ^{a}	2.12 ^{b}	2.09 ^{b}	3.43 ^{c}	2.65 ^{a}	2.77	1-5
ethnicity or race	3.29 ^{a}	2.37 ^{b}	2.35 ^{b}	3.89 ^{c}	2.97 ^{a}	3.15	1-5
religion	2.63 ^{a}	1.96 ^{b}	1.90 ^{b}	3.03 ^{c}	2.37 $^{a, b}$	2.52	1-5
sexual orientation	2.73 ^{a}	1.95 ^{b}	1.98 ^{b}	3.28 ^{c}	2.55 ^{a}	2.63	1-5
income or social class	3.43 ^{a}	2.41 ^{b}	2.44 $^{b, d}$	3.98 ^{c}	2.95 ^{d}	3.24	1-5
health risk	3.19 ^{a}	2.24 ^{b}	2.22 ^{b}	3.82 ^{c}	2.82 ^{d}	3.05	1-5
age	2.95 ^{a}	2.16 ^{b}	2.18 $^{b, d}$	3.46 ^{c}	2.61 $^{a, d}$	2.83	1-5

Note. Means in the same row that have no superscript in common differ at $p < .01$ in the post hoc test (Scheffé).

Ambivalent attitudes (41%): The largest class had predominantly high levels of perceived risks and benefits of AI. This class held moderate value dispositions, with the level of ideology being close to the mean of the AI scientific expert sample. They are among the groups who are most deferent to scientific authority and belief in the authority of science as a way of knowing. AI scientific experts in this class paid equally great attention to political and science news. They shared a similar level of concern about the unintended consequences of AI applications as the skeptical class, though their concern that AI would worsen discrimination based on race, income, health risks, gender, age, religion, and sexual orientation was slightly lower than that of the skeptical class. The ambivalent class held a strong belief that the public should have a say in AI regulation development and acknowledged their own role in participating in political debates about AI.

Supportive attitudes (23%): This class perceived high benefits and low to mediocre risks of AI. The supportive class's benefit perceptions were comparable to those of the ambivalent class. On the contrary, the supportive class perceived low risks of AI worsening societal inequalities and threatening personal liberties, and they perceived mediocre levels of AI giving some people too much power and displacing workers by automating their jobs. The supportive class shared some similar characteristics with the ambivalent class: being highly deferent to scientific authority and the authority of science as a way of knowing. However, their relatively low levels of risk perceptions distinguish them from the ambivalent class with regards to their attitudes toward regulation and concern about AI-based discrimination. They were most positive about the sufficiency of existing regulations for AI research and applications. They held low levels of concern about AI-based discrimination against vulnerable social groups, parallel to the indifferent class. Additionally, among the five classes, the supportive class paid low levels of attention to political news.

Skeptical attitudes (21%): This class held high risks and low to mediocre benefits of AI, opposite to the supportive class. The skeptical class's risk perceptions were comparable to those of the ambivalent class. However, they perceived mediocre levels of benefits of AI strengthening the U.S. economy and improving individuals' health and low levels of benefits of AI increasing national security and helping fight terrorism threats. They were more liberal on economic and social issues than any other classes. They paid more attention to political news than the supportive class and the indifferent class. Their attention to political news is higher than their attention to science news. Despite the fact that the skeptical and the ambivalent classes shared a similarly strong agreement on the importance of the public's and scientists' active voices in debates on AI development, the skeptical class showed lower levels of deference to scientific authority and science knowledge than the ambivalent class. The skeptical class also claimed the least satisfaction in existing regulations for AI research and applications and the highest concern about potential AI-based discrimination against vulnerable groups.

The other two classes consisted of a small proportion of the sample, whose attitudinal strengths on AI were relatively weak. Ambiguous-leaning attitudes (9%): Most respondents in this class perceived mediocre levels ("somewhat likely") of some risks (i.e., worsening societal inequalities and threatening personal liberties) and benefits of AI (i.e., increasing national security and helping fight terrorism threats). Their perceived benefits of AI on positive economic and individual health consequences and perceived risks of AI on power disparities and job replacement were slightly higher than the other aspects of risk and benefit perceptions. Their concern about AI-based discrimination was comparable to the ambivalent class, but significantly lower than the skeptical class and higher than the supportive and the indifferent classes.

Indifferent-leaning attitudes (6%): The indifferent class perceived low levels of risks and benefits, though their benefit perceptions slightly outweighed their risk perceptions. They perceived most risks of AI unlikely to happen, except moderate levels of job replacement. They perceived moderate levels of benefits of AI, including strengthening U.S. economy and improving individuals' health. Their perception of other benefits, including increasing national security and helping fighting terrorism threats, was lower. The indifferent class paid the least attention to either political news or science news among the five classes. They not only showed low levels of deference to scientific authority and belief in the authority of science as a way of knowing, but also signaled the lowest agreement on the appropriateness for scientists to be actively involved in political debates about AI. They were least concerned about AI-based discrimination and unintended consequences.

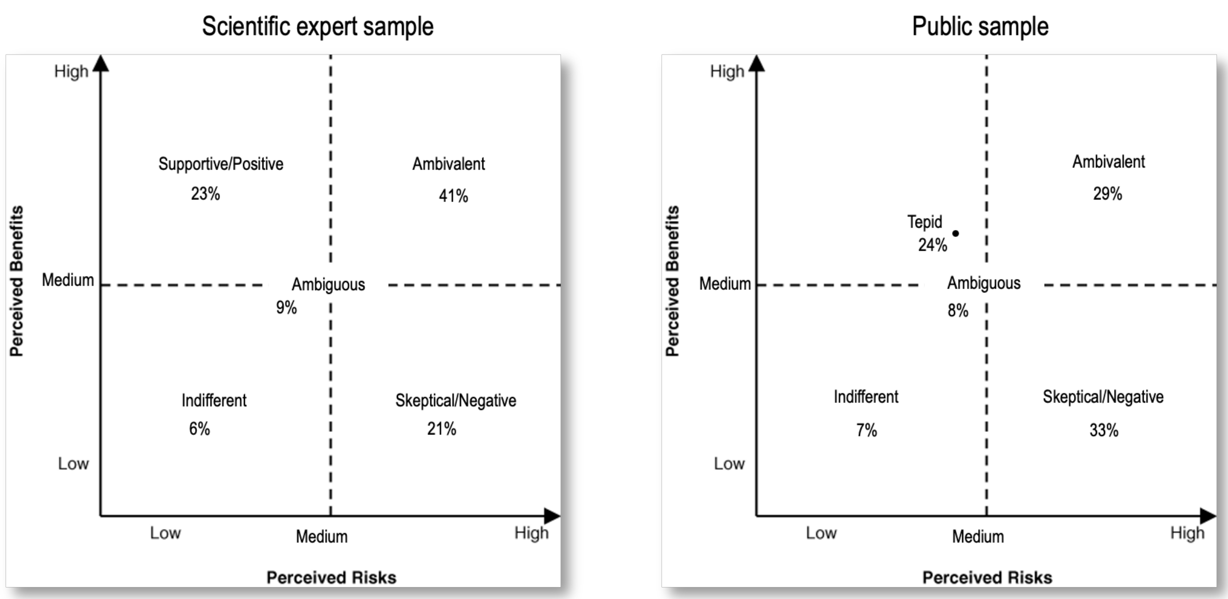
Comparing the segmentation results between AI scientific experts and US citizens

RQ2 asked how the segmentation of AI scientific experts differs from that of lay publics in terms of different combinations of risk and benefit perceptions. I compared the AI expert segmentation findings to the original study that used a nationally representative sample of 2,700 US adults aged 18 collected from February to March 2020 (see Bao et al., 2022 in detail). I discovered at least four differences regarding the type and size of segments.

First, the biggest difference is that the AI scientific expert sample had a supportive class who perceived high benefits and low risks of AI, whereas no such class appeared in the public sample. Second, the AI scientific expert sample had a larger proportion of the ambivalent class

(41%) than the public sample (29%). The size of the skeptical class in the AI scientific expert sample (21%) was smaller than the public sample (33%). Third, AI scientific experts in the skeptical class perceived slightly higher benefits of AI on economy and individual health than their laypeople counter parts in the skeptical class. Fourth, although the two samples consisted of a similarly small portion of the ambiguous and indifferent classes, risk and benefit perceptions were more spread out among AI scientific experts in those two classes than their lay counterparts.

Figure 3. Risk-benefit typology: Segmentation results of the scientific expert sample and the lay public sample



Factors that shape scientific experts' views of AI

Table 6 presents the results of the multinomial logistic regression model predicting different types related to risk and benefit perceptions of AI. Table 7 shows OLS regression models predicting risk and benefit perceptions of AI respectively. I discuss the results of the multinomial regression and the OLS regression models together following the order of our research hypotheses and questions.

Focusing first on value predispositions, I found that economic and social conservatism were associated with lower levels of risk perceptions ($\beta_{economic} = -.07$, $p < .01$; $\beta_{social} = -.11$, $p < .001$) and higher benefit perceptions of AI ($\beta_{economic} = .10$, p < .001; $\beta_{social} = .10$, $p < .001$). Likewise, as shown in the multinomial logistic regression model, those who were more liberal toward economic (*Exp(B)* = .75, $p < .001$) and social (*Exp(B)* = .72, $p < .001$) issues were more likely to have skeptical (low benefits/high risks) attitudes rather than having ambivalent attitudes toward AI. H1 was partially supported.

Deference to scientific authority was linked to lower levels of risk perceptions ($\beta = -.09$, $p < .001$) and greater benefit perceptions of AI ($\beta = .09$, $p < .001$). On the other hand, belief in the authority of science as a way of knowing was strongly associated with higher benefit perceptions of AI ($\beta = .22$, $p < .001$), but had no significant influences on risk perceptions of AI ($\beta = -.03$, *n.s.*). The multinomial logistic regression model showed similar results. Those who had lower levels of deference to scientific authority (*Exp(B)* = 0.75, $p < .001$) and beliefs in the authority of science (*Exp(B)* = 0.66, $p < .001$) were more likely to have skeptical attitudes than ambivalent attitudes. In contrast, those with high deference to scientific authority (*Exp(B)* = 1.27, $p < .001$) were more likely to have supportive rather than ambivalent attitudes. In short, H2a was supported and H2b was partially supported.

Table 6. Multinomial logistic regression model predicting different types of risk and benefit perceptions of AI

(Reference group: Ambivalent group)

	Supportive					Skeptical					Ambiguous					Indifferent				
	B	Std. Error	Wald	Exp (B)	Sig.	B	Std. Error	Wald	Exp (B)	Sig.	B	Std. Error	Wald	Exp (B)	Sig.	B	Std. Error	Wald	Exp (B)	Sig.
Male	-0.12	0.17	0.52	0.89	0.471	-0.07	0.16	0.20	0.93	0.659	-0.49	0.21	5.76	0.61	0.016	-0.40	0.25	2.48	0.67	0.115
Age	0.01	0.01	3.08	1.01	0.079	-0.01	0.01	3.82	0.99	0.051	0.00	0.01	0.11	1.00	0.746	0.00	0.01	0.20	1.00	0.655
Economic conservatism	0.10	0.08	1.82	1.11	0.177	-0.30	0.08	15.38	0.75	<.001	-0.01	0.10	0.00	0.99	0.951	-0.16	0.13	1.47	0.86	0.225
Social conservatism	0.14	0.08	3.37	1.15	0.066	-0.34	0.08	17.12	0.72	<.001	0.04	0.10	0.17	1.04	0.678	0.24	0.12	3.63	1.27	0.057
Deference to scientific authority	0.24	0.07	11.25	1.27	<.001	-0.29	0.07	15.68	0.75	<.001	0.12	0.10	1.43	1.13	0.232	-0.17	0.13	1.85	0.84	0.174
Belief in the authority of science	0.05	0.09	0.31	1.05	0.578	-0.42	0.08	26.32	0.66	<.001	-0.36	0.11	9.76	0.70	0.002	-0.30	0.14	4.88	0.74	0.027
Attention to political news	-0.28	0.08	13.09	0.75	<.001	0.04	0.08	0.28	1.04	0.598	-0.04	0.11	0.16	0.96	0.688	-0.21	0.13	2.55	0.81	0.110
Attention to science news	-0.02	0.10	0.03	0.98	0.859	0.08	0.10	0.76	1.09	0.384	-0.18	0.13	1.95	0.83	0.163	-0.20	0.16	1.55	0.82	0.213
Life sciences [a]	0.08	0.17	0.20	1.08	0.652	0.24	0.17	1.94	1.27	0.164	0.07	0.24	0.09	1.08	0.765	0.88	0.25	12.10	2.41	<.001
Physical sciences [a]	0.02	0.15	0.02	1.02	0.888	-0.08	0.16	0.23	0.93	0.634	0.05	0.21	0.06	1.05	0.812	0.07	0.26	0.07	1.07	0.786
Social sciences [a]	-0.44	0.23	3.60	0.64	0.058	0.04	0.19	0.04	1.04	0.837	0.30	0.26	1.35	1.35	0.246	0.14	0.35	0.16	1.15	0.689
Interdisciplinary collaboration	0.04	0.06	0.50	1.04	0.481	-0.12	0.06	4.31	0.89	0.038	0.02	0.08	0.08	1.02	0.779	-0.07	0.10	0.56	0.93	0.453

Note. a. Reference group: computer and information science.

Table 7. OLS regression models predicting risk and benefit perceptions of AI

	Risk perceptions				Benefit perceptions			
	Unstd Beta	Std. Error	Std Beta	Sig.	Unstd Beta	Std. Error	Std Beta	Sig.
Block 1: Risk/benefit perceptions								
Benefit perceptions	-0.04	0.02	-0.04	0.089				
Risk perceptions					-0.04	0.02	-0.04	0.089
Inc Adjusted R^2 (%)	0.8***				0.8***			
Block 2: Demographics								
Male	0.07	0.06	0.03	0.244	0.20	0.06	0.08	<.001
Age	-0.01	0.00	-0.09	<.001	0.00	0.00	0.03	0.159
Inc Adjusted R^2 (%)	0.6***				1.9***			
Block 3: Value dispositions								
Economic ideology	-0.07	0.03	-0.07	0.007	0.10	0.03	0.10	<.001
Social ideology	-0.12	0.03	-0.11	<.001	0.10	0.03	0.10	<.001
Deference	-0.10	0.03	-0.09	<.001	0.10	0.02	0.09	<.001
Belief in the authority of science	-0.03	0.03	-0.03	0.277	0.27	0.03	0.22	<.001
Inc Adjusted R^2 (%)	4.6***				8.6***			
Block 4: Media attention								
Attention to political news	0.19	0.03	0.16	<.001	-0.03	0.03	-0.03	0.213
Attention to science news	0.10	0.04	0.07	0.006	0.09	0.03	0.06	0.010
Inc Adjusted R^2 (%)	3.5***				0.3*			
Block 5: Professional factors								
Life sciences [a]	-0.16	0.06	-0.06	0.011	-0.14	0.06	-0.06	0.012
Physical sciences [a]	-0.06	0.06	-0.03	0.261	-0.02	0.05	-0.01	0.663
Social sciences [a]	0.04	0.07	0.01	0.589	-0.10	0.07	-0.03	0.147
Interdisciplinary collaboration	-0.06	0.02	-0.06	0.004	0.01	0.02	0.02	0.472
Inc Adjusted R^2 (%)	0.7***				0.2			
Total R^2 (%)	10.2***				11.8***			

Note. a. Reference group: computer and information science. **p* < 0.05; ***p* < 0.01; ****p* < 0.001

Regarding the impacts of media use, attention to political news (β = .16, p < .001) was associated with greater risk perceptions, but had no impacts on benefit perceptions (β = -.03, n.s.). Attention to science news was associated with higher risk perceptions (β = .07, p < .01) and benefit perceptions (β = .06, p = .010). Those who paid less attention to political news (*Exp(B)* = 0.75, p < .01) were more likely to have supportive attitudes than ambivalent attitudes. Therefore, H3a-c were partially supported.

Moving next to the disciplinary differences (RQ3), I found that life scientists perceived lower levels of AI risks and benefits compared to computer scientists. Accordingly, life scientists were more likely to be in the indifferent-leaning class than the ambivalence class, compared to computer scientists. I did not find any significant relationships between the diversity of interdisciplinary collaboration and perceived benefits of AI. Contrary to our hypothesis, interdisciplinary collaboration was associated with lower risk perceptions. H4 was not supported.

Age and gender also impacted risk and benefit perceptions of AI. Older scientific experts perceived lower levels of risks of AI (β = -.09, p < .001). Male scientific experts perceived more benefits of AI than non-male peers (β = .08, p < .001) and were more likely to form ambivalent than ambiguous attitudes (*Exp(B)* = .61, p = .016) compared to non-male peers.

Discussion

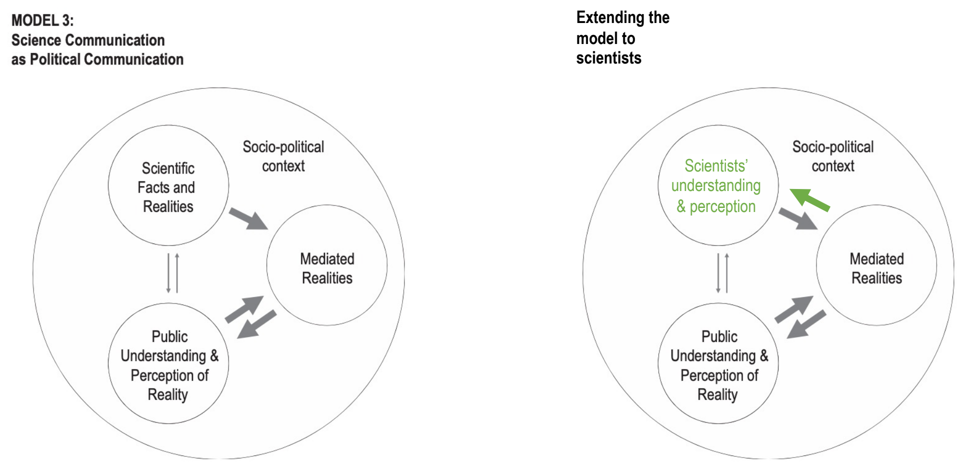
This study provides a systematic comparison of risk and benefit perceptions of AI between AI scientific experts and lay publics in the US. I find that AI scientific experts overall perceive substantially higher levels of benefits of AI than lay publics, whereases the two groups perceive similar levels of risks of AI. The LCA results indicate that AI scientific experts and lay publics consist of heterogenous subgroups within their own sample, such as holding ambivalent attitudes (high risks/high benefits). I further examine factors that influence AI scientific experts' perceived risks and benefits of AI.

Before discussing theoretical and practical implications in detail, I first address three main limitations of this study. First, the AI scientific expert dataset was collected about two years later than the public sample. Given the rapid progress of AI and fierce discussions of AI in public discourses, such as the use of facial recognition systems in detecting protesters and various AI tools used in the recent Russia-Ukraine war, current public views of AI may have shifted from when the data was gathered in 2020. Future studies will benefit from collecting data among different stakeholder samples simultaneously to improve the validity of conclusions. Second, the list of risky and beneficial outcomes of AI I used for LCA is by no means exhaustive. These items were selected from frequent media mentions. Future research could incorporate open-ended questions to enrich the data for more in-depth understanding of attitudes toward AI. Third, the comparison between the AI scientific expert sample and the lay sample relies on two separate analyses, which leads to difficulties in keeping the categorization labels consistent between the two analyses. Future studies might combine these two samples into one to conduct LCA. This approach will allow comparisons between the various portions of scientific experts and lay publics within each class.

Despite these limitations, my study has three theoretical contributions. First, I extend the risk-benefit typology from lay populations to expert populations. Resonating with the discussions that science audiences are not monolithic (for reviews, see Metag & Schäfer, 2018; Scheufele, 2018), the extent of heterogeneity across subgroups is no less than that of the lay public sample. However, the components of the risk-benefit typology between the scientific expert and the public sample show different patterns. AI scientific experts consist of five segments: the ambivalent (high benefit/high risk: 41%), the supportive/positive (high benefit/low risk: 23%), the skeptical/negative (low benefit/high risk: 21%), the ambiguous-leaning (medium benefit/medium risk: 9%), and the indifferent-leaning (low benefit/low risk: 6%). US lay publics consist of the negative (33%), the ambivalent (29%), the tepid (24%), the ambiguous (8%), and the indifferent (7%). While US lay publics do not have a "purely supportive" group (high risks/low benefits), AI scientific experts have two subgroups with opposite attitudes toward AI—either supportive (high risks/low benefits) or skeptical (low risks/high benefits). Perhaps scientific experts with supportive attitudes predominantly perceive the promise of their work or are confident that they could address the potential risks. In contrast, lay publics are more likely to experience the negativity bias whereby negative evaluations are weighed more heavily than positive ones (Baumeister et al., 2001). For lay publics, the activation of benefit perceptions of AI will probably be accompanied by the activation of risk perceptions (Bao et al., 2022). Additionally, scientific experts' views of different risky and beneficial consequences are more nuanced than are those of their lay counterparts. For instance, scientific experts in the skeptical segment still perceive mediocre likelihood of AI strengthening the U.S. economy and improving individuals' health, and those scientific experts in the supportive segment perceive mediocre likelihood of AI displacing workers by automating their jobs. Lay publics do not show such nuances in their judgement of different risky or beneficial consequences of AI. This may indicate a "spillover effect" that cues from established issues inform judgments for new ones, which has also been found in public attitude formation of other science and technology issues (Akin et al., 2019; Ho et al., 2020).

Second, I shed light on factors that explain scientific experts' risk and benefit perceptions of AI based on a combination of segmentation analyses and regression analyses. Value predispositions have strong impacts on scientific experts' views of AI, which is consistent with prior research of scientists' views of other science and technology issues (e.g., Ho et al., 2011). Moreover, characteristics of different segments reveal that extreme value predispositions are associated with more one-sided attitudes and less ambivalent attitudes. For instance, the skeptical segment of AI scientific experts has the most liberal ideology and the lowest level of deference to scientific authority across five segments. However, paying a great amount of attention to both political and science news is likely to shape ambivalent attitudes toward AI that reconciles perceived risks and benefits of AI. Although AI scientific experts are more attentive to science news than their lay counterpart, a relatively balanced diet of political and science news attention is a key common characteristic shared by the ambivalent segment within the scientific expert sample and the lay public sample. This demonstrates the important role of media in the "science communication as political communication" model, referring to the notion that public understanding of science and perception of science reality are shaped by mediated realities (Scheufele, 2014). Such effects of mediated realities are also influential in shaping scientists' understanding and perception of scientific facts and realities (see Figure 4). Because scientists can potentially be better informed about the consequences of science and technology use through news coverage, especially for knowing more about social groups that may have different living experiences from scientists.

Figure 4. Extension of the science communication as political communication model



Scheufele DA (2014) Science communication as political communication. Proceedings of the National Academy of Sciences of the United States of America 111: 13585-13592.

Third, my study expands to examine the relationship between professional characteristics and scientific experts' risk and benefit perceptions of AI. As AI is interdisciplinary in nature and has a variety of applications used in different fields, I focus on the impacts of research fields and the diversity of interdisciplinary collaboration. AI scientific experts from the field of life sciences perceive lower levels of risks and benefits of AI as compared to scientific experts in the fields of computer and information sciences. Surprisingly, the diversity of interdisciplinary collaboration is associated with lower levels of risk perceptions, but has no effects on benefit perceptions or the segmentation of different types of attitudes toward AI. Perhaps the knowledge-based collaborations at the interdisciplinary level might increase AI scientific experts' confidence in addressing the potential risky concerns of AI with collective knowledge. Future research can analyze multidimensional collaborations, such as collaboration focusing on cultural perspectives and economic value (for an overview, see Bozeman et al., 2013).

To summarize, this study reveals the heterogeneity of AI scientific experts' attitudes toward AI, which is associated with their value predispositions and views of scientific conduct.

This study also shows potential attitudinal differences between scientific experts and lay publics, which calls for studies on the potential interventions that may increase scientific experts' consideration of public input. The next chapter will examine whether exposure to social science research on AI will make a difference in how scientists consider the role of themselves and other stakeholders, including citizens, in AI regulation development.

Chapter 4. How does exposure to social science research on AI shape AI scientists' perspectives on who should have a say in AI regulation development? (Study B)

As discussed in Chapter 1, a successful deliberative democracy depends on stakeholders acknowledging the legitimacy of other actors involved in the decision-making process (Bohman & Rehg, 1997). Therefore, before having debates on detailed regulations, it is essential to discuss who should have a say in regulation development. This chapter focuses on AI scientific experts' perspectives on who should have a say in AI regulation development, as scientific expert themselves are important stakeholders. Except for research investigating scientists' attitudes toward citizens' participation in regulatory development (e.g., Wirz, 2021), little research has examined the variety of social actors that scientists view as legitimate stakeholders in developing science and technology regulations.

The goal of this study is two-fold. First, this study aims to provide an empirical understanding of different models regarding whom scientists think should have a say in AI regulation development to make a connection to literature on linear versus stakeholder models. Second, given the desirable outcomes of a more inclusive view of stakeholders involved in societal debates on AI, this study investigates what factors shape a more inclusive view of who should have a say in AI regulation development among AI scientific experts. Specifically, I test factors including the extent to which AI scientists are exposed to social science input through reading and citing social science literature, value predispositions toward scientific conduct, and risk and benefit perceptions of AI.

Scientists' perceived roles in science decision-making: Linear vs. stakeholder models

Scientists may vary in perceiving their roles in the science policy making process. This study applies one classification dimension developed by Pielke (2007)—whether holding a linear or stakeholder model. The linear model indicates dominant positions of scientists in certain fields, referring to the notion that basic science knowledge is fundamental to applied research and then determines subsequent social benefits of science (Pielke, 2007). Guided by such beliefs, a scientific consensus is considered a prerequisite of political consensus, which endows basic science knowledge with high priority in science policy and decision making (Pielke, 2007).

The linear model has been historically popular in the US, because science has allowed the US to achieve an advantaged position in national security and competence since World War II. In 1945, Vannevar Bush delivered his famous report, *Science, The Endless Frontier,* calling for support for science from the government and citizens and the creation of the National Science Foundation to fund basic science research that will produce applied research and bring societal benefits (Bush, 1945). The linear model also suggests an unambiguous boundary-work between science and non-science, ensuring the intellectual authority of scientists. However, the roles of scientists holding a linear model are only useful when dealing with issues that have a broad consensus on the desired outcomes of certain science decision making across different social groups and low levels of uncertainty (Pielke, 2007). It also should be noted that recently, high-level leaders in the scientific community have argued against the linear model. For instance, Sethuraman Panchanathan, the new National Science Foundation Director, considers the relationship between basic research, applied research, and industry applications as symbiotic and synergistic rather than linear (Panchanathan, 2021).

In contrast to the linear model, a stakeholder model values the roles of non-expert users of science in the process of knowledge production and science decision making (Funtowicz & Ravetz, 1992; Pielke, 2007; Wynne, 2005). From the stakeholder perspective, scientists could play a role as either *issue advocates* that make policy recommendations aligned with their own political stance or *honest brokers of policy alternatives* that clarify or expand the scope of choice available to decision-makers and provide expertise on how these policy alternatives are consistent and inconsistent with scientific results (Pielke, 2007). For honest brokers, narrowing the scope of policy choice is based on political negotiation between conflicting interests and values, rather than scientists' political stances (Pielke, 2007). The similar idea of incorporating lay input through effective interactions between scientists and non-experts has also been indicated in other conceptual frameworks, such as research on two-way public engagement (e.g., Brossard & Lewenstein, 2010), responsible research and innovation (e.g., von Schomberg, 2013), and participatory technological assessment (e.g., Kaplan et al., 2021).

Despite the conceptualization of how scientists view their roles in decision making, ranging from a linear model to a stakeholder model, little research has empirically studied the prevalence of these models from the perspectives of scientists, including the context of AI. I therefore raise the following research question:

RQ1: In terms of determining who should have a say in AI regulation development, are AI scientists more aligned with a linear model (less inclusive) or a stakeholder model (more inclusive)?

Factors that shape scientists' perspectives of who should have a say in AI regulation

Why do some scientists form more inclusive views of who should have a say in AI regulation development than others? This section reviews factors that may shape their views of the legitimacy of various actors' involvement in AI regulation development, including scientists' value predispositions, views of regulation sufficiency, and professional characteristics regarding how they conduct their research.

The hierarchy hypothesis of the sciences: The hard and soft sciences

The linear model (discussed in the last section) shows an alignment with the hierarchy hypothesis of the sciences, as certain branches of sciences are perceived as more seminal than others. The hierarchy of the sciences was raised by Auguste Comte 200 years ago. Comte treated astronomy as the foundational stage of intellectual development, followed by physics, chemistry, biology, and sociology (cited fromCole, 1983). Scientific disciplines that closely relate to data and mathematics to validate natural laws are classified as "hard sciences," whereas "soft sciences" refer to disciplines that are arguably considered lacking methodological rigor and that their findings likely to yield common science, such as sociology and psychology (Cole, 1983; Seising & Sanz, 2012; Storer, 1967). Anecdotal evidence suggest that support for the conventional hierarchy of the sciences is particularly popular among a small group of STEM scientists who received their training long ago (Freudenburg, 2008). Some laypeople may also hold similar beliefs linking the major contributions of science to hard sciences, such as medicine, computer technology, space exploration, transportation, but not to social and behavioral sciences-based sciences (Bernard, 2012).

The hierarchy hypothesis of the sciences may cause biased understandings among scientists from different disciplines that make them hesitant to collaborate with each other (Freudenburg, 2008). For instance, the hierarchy hypothesis implies that few scientists know the difficult calculation of equations, but most scientists think themselves as having some knowledge of human behaviors ("In praise of soft science," 2005; Shermer, 2007). However, arguments against the necessity of ranking sciences state that the multifaceted and complex nature of human subjects and the research protocols dealing with human beings make social sciences the “hard sciences” (Freudenburg, 2008; Shermer, 2007).

Where does AI fall into the classification of the hard or soft sciences? There is unlikely a clear-cut, single answer, as AI research involves different paradigms and closely relates to both the so-called “hard” and “soft” sciences. The traditional paradigm, or so-called hard computing, uses explicit models to solve problems with exact and optimal solutions (Seising & Sanz, 2012). In contrast, soft computing uses more interdisciplinary approaches, such as fuzzy logic, which mimics the ability of human thinking with a higher tolerance of uncertainty and imprecision (Seising & Sanz, 2012). Considering these different paradigms, it is plausible that scientific experts who see a clear distinction of the hard and soft sciences may think human experience and the social, ethical, and legal implications of AI are less relevant or important in regulation development, compared to their fellow scientific experts who see blurred boundaries between hard and soft sciences. Given the link between having the linear model and the acceptance of the hierarchy of sciences, I propose the following hypothesis:

H1: The agreement that it is useful to classify hard and soft sciences is associated with a less inclusive perspective on who should have a say in AI regulation development.

Reading and citing social science literature on AI

The way scientists produce and communicate knowledge so as to relate to other actors in the network may indicate their own value systems and embed social power structure (Gieryn, 1983). For instance, among nano-scientists, citing publications on environmental, health, and safety issues of nanotechnology is more common for those with greater agreement that setting moral limits is the most important area that citizens should be informed of before casting their vote for funding nanotechnology research (Youtie et al., 2011). Scientists with more positive attitudes toward social sciences are less likely to have deficit thinking of lay publics that is proven ineffective for the public communication of science (Simis et al., 2016).

A growing number of social scientists have been involved in AI-related research, as society is wrestling with the impacts of various AI applications used in our daily lives. On the one hand, AI brings social and cultural transformations, and AI applications share functions of decision-making and meaning-making that were only possessed by humans in the past, such as providing caregiving, driving, and writing news articles (e.g., Crawford & Calo, 2016; Guzman & Lewis, 2019; Hancock et al., 2019). On the other hand, the development of AI, such as the selection of training data and parameters, is embedded in human and social values (Broussard et al., 2019). AI research and development is an active and dynamic process that depends on relations and connections with different actors and ideas.

Social science research on AI can provide AI research and development with “nontechnical” expertise in the following aspects: understanding social and political contexts of social-technical problems that AI systems tend to address (e.g., Crawford & Calo, 2016; Obermeyer et al., 2019); making sense of public interpretation of ethical and moral considerations related to AI systems (e.g., Ryan, 2020; Shin, 2022); accounting for biases in training data and mitigating reproductions of inequality through AI systems based on social theories (Zajko, 2021, 2022); interpreting data in ways grounded in norms and values of human and social constructs (boyd & Crawford, 2012; Natale & Guzman, 2022); as well as engaging with the affected publics at different stages from understanding their access to AI to learning consequences of AI deployment (e.g., Buhmann & Fieseler, 2022; Kaplan et al., 2021). As a result, exposure to social science research on AI is likely to increase scientific experts' acknowledgement of involving other societal groups in regulation development. This study explores the effects of reading and citing social sciences, considering both informal and formal ways. Thus, I propose the following hypothesis:

> **H2:** The use of social science research, including a) reading or b) citing social science literature on AI, is positively associated with a more inclusive perspective on who should have a say in AI regulation development.

Value predispositions: Deference to scientific authority

Moving beyond the effects of how AI scientific experts view and use scientific insights from different social disciplines, their views of who should have a say in AI regulation development may rely on their perceived exceptionalism of scientists as compared to other cultural authorities, known as the deference to scientific authority. The authoritarian views of democratic processes in science indicate that it is not the responsibilities of citizens to develop rules to regulate development of controversial science issues (Brossard & Nisbet, 2007; Brossard & Shanahan, 2003). Lay publics with high deference to scientific authority are less likely to acknowledge the necessity of being consulted by scientists or being involved in science decision-making (Howell, Wirz, et al., 2020). Furthermore, analyses of STEM scientists who exclude lay publics from policy decisions or see lay publics as having no right to express their opinions have found that these scientists had negative views of lay publics, portraying them as irrational or ignorant (Freudenburg, 2008). A recent study on synthetic biology scientists found a negative association between the degree of deference to scientific authority and the likelihood that those scientists think citizens should have a say in regulation development (e.g., Wirz, 2021). I anticipate a similar effect of deference to scientific authority on how scientists view other non-expert stakeholders' roles in regulation development. I therefore propose the following hypothesis:

> **H3:** The degree of deference to scientific authority is associated with a less inclusive perspective on who should have a say in AI regulation development.

Contextual factors: risk and benefit perceptions of AI and perceived regulation sufficiency

The intertwined link between science and society requires scientific research to be accessible to policy and decision-making. For instance, as far back as the 1970s, the Asilomar Conference on Recombinant DNA initiated regulation of biotechnology (Berg et al., 1975). Scientists use their risk perceptions of a scientific issue to form their views of regulation (Corley et al., 2013, 2016; Corley et al., 2009; Su et al., 2016). Perceived risks over benefits are associated with more support for academic and commercial regulation among scientists (Su et al., 2016). Scientific experts with greater risk perceptions or higher levels of ambivalence are more likely to favor citizen involvement in regulation development. However, those who perceive lower levels of risks favor a more closed system that excludes citizen involvement (Wirz, 2021). Hence, I propose the following hypotheses and question:

H4: Risk perceptions of AI is associated with a more inclusive perspective on who should have a say in AI regulation development.

RQ2: How do benefit perceptions of AI associate with different perspectives on who should have a say in AI regulation development?

H5: The perception of insufficient regulation of AI applications is associated with a more inclusive perspective on who should have a say in AI regulation development.

This study also considers the impacts of demographic differences, such as age and gender, on scientists' attitudes toward technology regulation development. Older scientists are more likely to oppose academic and commercial regulation of nanotechnology (Su et al., 2016). Male scientists are less likely to support regulation on nanotechnology than their female counterparts (Corley et al., 2009).

Methods

Details of the sample are discussed in Chapter 2.

Measures

Dependent variables

<u>Stakeholders involved in AI regulation development</u>. Respondents were asked to choose the groups that should have a say in the development of these regulations in the US, if there were new regulations for AI research and use. They were presented with 13 groups with a dichotomous (yes or no) measure. The groups include university scientists, industry scientists, congress, the U.S. court system, regulatory agencies that oversee AI applications (e.g., the Department of Transportation or the Department of Defense), the White House, law enforcement

agencies (e.g., the FBI), digital rights groups (e.g., the Electronic Frontier Foundation), citizens through public engagement mechanisms (e.g., citizen forums), large technological companies (e.g., Google and Microsoft), AI-related non-profit organizations (e.g., Partnership on AI), intergovernmental bodies or agreements, and end users. Detailed descriptive statistics are presented in the results section (See Table 10).

Independent variables and control variables were presented in Table 8.

Table 8. Measures and descriptive statistics of independent and control variables

Variable	**Measure**	**Scale**	**Statistics**
Perceived hierarchy of sciences	It is useful to classify science into hard and soft sciences.	1=strongly disagree, 5=strongly agree	M = 2.8, SD = 1.0
Reading social sciences	Please think about the AI-related journal articles, policy documents, and reports that you have read over the last three years. With that in mind, how often would you say you **read** about the following topics? a) societal implications of AI b) ethics and responsible development of AI	1=never, 5=very often	M = 3.2, SD = 1.0, r = .74
Citing social sciences	Please think about the AI-related journal articles, policy documents, and reports that you have written over the last three years. How often would you say you **cite sources** that discuss the following topics? a) societal implications of AI b) ethics and responsible development of AI	1=never, 5=very often	M = 2.5, SD = 1.2, r = .78
Deference to scientific authority	Scientists know best what is good for the public.	1=strongly disagree, 5=strongly agree	M = 2.5, SD = 0.9
Benefit perceptions	How likely AI will …? a) strengthen the U.S. economy b) increase national security c) improve individuals' health d) help fight terrorism threats	1=not at all likely, 7=certain	M = 5.0, SD = 1.0, *Cronbach's* α = .81

Risk perceptions	How likely AI will …? a) worsen societal inequalities b) give some people too much power c) threaten personal liberties d) displace workers by automating their jobs	1=not at all likely, 7=certain	*M* = 4.8, *SD* = 1.1, *Cronbach's* α = .77
Regulation sufficiency	Existing regulations for AI applications are sufficient.	1=strongly disagree, 5=strongly agree	*M* = 2.2, *SD* = 0.9
Age		Numeric	*M* = 45.3, *SD* = 13.1
Gender	Male (79.0%), female (18.1%), non-binary (0.8%)	Binary (male)	
Research field	a) computer and information sciences (50.5%) b) life sciences (17.4%) c) physical sciences (23.0%) d) social sciences (11.9%)	Dummy coding in the model	

Analysis

This study applied LCA to cluster respondents with similar perceptions of who should have a say in AI regulation development. The analysis took an advanced approach that integrated an SEM model to predict the LCA classification.

Results

I began with the correlation analysis between key independent variables. As shown in Table 9, most of the absolute values of correlations were below 0.20, except for the correlation between risk perceptions and perceived sufficiency of regulations for AI applications (r = -.33, p < .001) and that between reading and citing social science research on AI (r = .65, p < .001). Due to the high correlation between reading and citing social science research, I ran two sets of LCA models by entering either the reading or the citing variable separately to avoid misleading interpretations caused by multicollinearity problems.

Table 9. Correlations between independent variables (Study B)

	Perceived hierarchy of sciences	Reading social sciences	Citing social sciences	Deference to scientific authority	Age	Gender	Risk perceptions	Benefit perceptions	Regulation sufficiency	Research field-life sciences	Research field-physical sciences
Reading social sciences	-0.12***										
Citing social sciences	-0.09***	0.65***									
Deference to scientific authority	0.15***	-0.02	0.02								
Age	-0.02	0.06**	0.00	0.00							
Gender (male)	0.12***	-0.11***	-0.10***	0.01	0.04						
Risk perceptions	-0.04	0.09***	0.02	-0.11***	-0.08***	0.01					
Benefit perceptions	0.19***	-0.03	-0.01	0.17***	0.06*	0.14***	-0.08***				
Regulation sufficiency	0.15***	-0.19***	-0.11***	0.16***	0.05*	0.09***	-0.33***	0.16***			
Research field-life sciences	-0.03	-0.09***	-0.10***	0.03	0.13***	-0.02	-0.07**	-0.03	0.07***		
Research field-physical sciences	0.12***	-0.17***	-0.18***	0.02	-0.01	0.04	-0.06**	0.03	0.04	-0.07***	
Research field-social sciences	-0.17***	0.11***	0.13***	-0.05*	0.07**	-0.09***	0.05*	-0.07***	-0.03	-0.17***	-0.20***

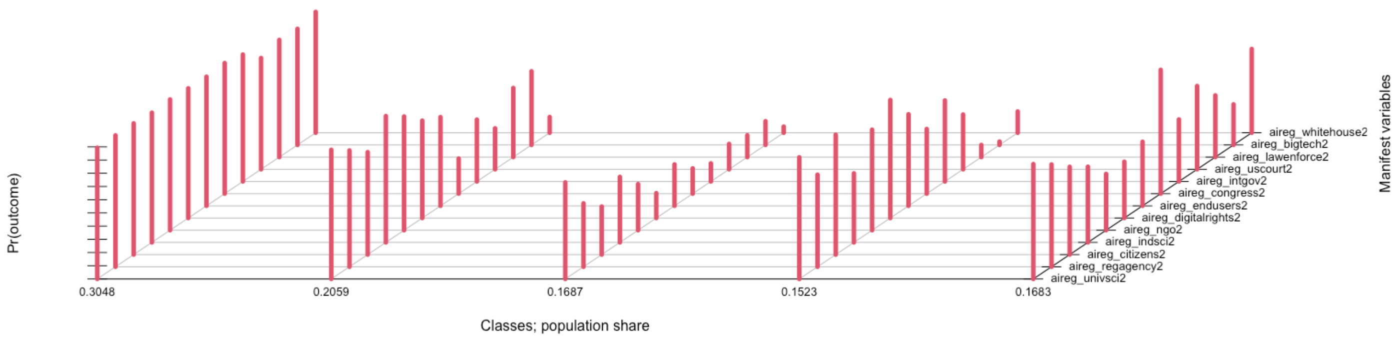
Note. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

RQ1 asked what different perspectives AI scientists have regarding who should have a say in AI regulation development. The five-class model was chosen (as shown in Figure 5 and Table 10), because the model had a theoretically grounded interpretation and a good model performance. Below, I discuss the features of different models regarding the inclusiveness of AI scientists' perspectives on who should have a say in AI regulation development.

Table 10. AI scientific experts' agreement on who should have a say in AI regulation development (sorted by each class)

	Inclusive (all stakeholders) (30%)	Industry-prioritized (21%)	Input agnostic (17%)	Experts & civic groups (15%)	Govern-ment prioritized (17%)	Overall sample
University scientists	99%	98%	73%	92%	87%	91%
Regulatory agencies that oversee AI applications (e.g., DoT or DoD)	99%	88%	48%	70%	77%	79%
Citizens through public engagement mechanisms (e.g., citizen forums)	99%	78%	36%	91%	66%	76%
Industry scientists	98%	96%	50%	52%	57%	75%
AI-related non-profit organizations (e.g., Partnership on AI)	99%	86%	35%	76%	43%	72%
End users	98%	67%	31%	69%	48%	68%
Digital rights groups (e.g., the Electronic Frontier Foundation)	98%	74%	19%	89%	43%	67%
Congress	99%	27%	19%	49%	94%	61%
Intergovernmental bodies or agreements	96%	47%	14%	61%	47%	57%
The U.S. court system	84%	31%	19%	41%	63%	52%
Law enforcement agencies (e.g., the FBI)	89%	52%	17%	9%	47%	49%
Large technological companies (e.g., Google and Microsoft)	88%	56%	18%	3%	31%	47%
The White House	92%	12%	5%	16%	63%	43%

Figure 5. Estimated latent class conditional probabilities (Who should have a say in AI regulation development)



Inclusive (all stakeholders) class (30%): Respondents in this class held the most inclusive view. They considered that all 13 listed actors should have a say, with only four actors receiving agreement below 95%, including the White House (92%), law enforcement agencies (89%), large tech companies (88%), and the U.S. court system (84%).

Industry-prioritized class (21%): The unique characteristic of this class was their relatively higher agreement on industry actors, such as industry scientists (96%) and large tech companies (56%), compared to other classes. This class was also open to the inclusion of university scientists (98%) and AI related non-profit organizations (86%), but was not open to the inclusion of government branches, such as the Congress (27%) and the White house (12%).

Input agnostic class (17%): This class held a negative attitude toward involving any social groups into AI regulation development. Although the majority of respondents in this class (73%) agreed that university scientists should have a say in AI regulation development, the percentage of agreement was still significantly lower than the other four classes.

Experts and civic groups (15%): This class predominantly supported the inclusion of university scientists (92%), citizens (91%), and digital rights groups (89%), whereas they were less open to industry related actors, such as industry scientists (52%) and big tech companies (3%). They were also not open to law enforcement agencies (9%).

Government-prioritized class (17%): This class had favorable attitudes toward all three branches of government regarding their participation in AI regulation development, including the Congress (94%), the White House (63%), the U.S. court system (63%), and regulatory agencies (77%). They were less open to include social actors such as AI related non-profit organizations (43%), digital rights groups (43%), and large tech companies (31%).

Next, I turned to H1-5 and RQ2 examining factors that determined the segmentation of scientists' views of who should have a say in AI regulation development. Table 11 shows the coefficients of variables that predicted the results of the latent class model. I chose the inclusive view class as the reference group to interpret results.

First, AI scientific experts who agreed with the hierarchy of sciences or the statement that it is useful to classify hard and soft sciences were more likely to be in the input agnostic class ($\beta = .40$, $p < .001$), the government-prioritized class ($\beta = .24$, $p = .01$), and the industry-prioritized class ($\beta = .22$, $p = .01$), as compared to the inclusive class. There were no differences between the experts and civic groups class and the inclusive class regardless of their views of perceived hierarchy of sciences. Therefore, H1 was mostly supported.

Table 11. Structural equation models predicting different types of who should have a say in AI regulation development – reading model (Reference group: The inclusive group)

	Industry-prioritized				Input agnostic				Experts & civic groups				Government-prioritized			
	Coeffi -cient	Std. error	t value	Pr(>|t|)	Coeffi -cient	Std. error	t value	Pr(>|t|)	Coeffi -cient	Std. error	t value	Pr(>|t|)	Coeffi -cient	Std. error	t value	Pr(>|t|)
(Intercept)	1.69	0.84	2.02	0.044	1.76	0.83	2.11	0.035	1.98	1.05	1.88	0.061	1.29	0.89	1.45	0.146
Age	-0.02	0.01	-2.51	0.012	-0.04	0.01	-5.09	<.001	-0.01	0.01	-1.51	0.132	-0.02	0.01	-2.48	0.013
Male	-0.82	0.20	-4.10	<.001	-0.56	0.23	-2.42	0.016	-0.42	0.24	-1.76	0.079	0.56	0.30	1.89	0.059
Reading social sciences	-0.22	0.09	-2.56	0.011	-0.25	0.09	-2.78	0.006	-0.07	0.10	-0.73	0.466	-0.24	0.10	-2.48	0.013
Perceived hierarchy of sciences	0.22	0.09	2.59	0.010	0.40	0.09	4.40	<.001	-0.11	0.11	-1.01	0.315	0.24	0.09	2.58	0.010
Deference to scientific authority	0.13	0.10	1.27	0.206	0.31	0.10	3.13	0.002	-0.22	0.11	-1.99	0.047	-0.10	0.10	-1.07	0.283
Risk perceptions	-0.33	0.09	-3.68	<.001	-0.58	0.09	-6.49	<.001	0.46	0.11	4.09	<.001	-0.27	0.09	-2.87	0.004
Benefit perceptions	0.10	0.10	0.99	0.323	-0.02	0.09	-0.18	0.856	-0.60	0.11	-5.70	<.001	-0.01	0.09	-0.16	0.876
Regulation sufficiency	0.14	0.11	1.30	0.195	0.61	0.11	5.70	<.001	-0.15	0.13	-1.15	0.251	0.03	0.11	0.27	0.787
Life sciences [a]	-0.44	0.23	-1.89	0.059	-0.33	0.24	-1.35	0.178	-0.08	0.27	-0.30	0.765	0.14	0.22	0.64	0.522
Physical sciences [a]	-0.09	0.20	-0.45	0.651	-0.04	0.21	-0.19	0.846	0.03	0.25	0.11	0.915	0.11	0.21	0.53	0.595
Social sciences [a]	-0.44	0.28	-1.61	0.107	-0.24	0.28	-0.83	0.408	-0.15	0.29	-0.51	0.608	-0.11	0.29	-0.40	0.692

Note. a. Reference group: computer and information science.

Table 12. Structural equation models predicting different types of who should have a say in AI regulation development – citation model (Reference group: The inclusive group)

	Input agnostic				Government-prioritized				Experts & civic groups				Industry-prioritized			
	Coefficient	Std. error	t value	Pr(>|t|)	Coefficient	Std. error	t value	Pr(>|t|)	Coefficient	Std. error	t value	Pr(>|t|)	Coefficient	Std. error	t value	Pr(>|t|)
(Intercept)	2.23	1.07	2.08	0.038	0.57	0.83	0.68	0.495	1.11	0.87	1.27	0.203	1.17	0.80	1.45	0.147
Age	-0.01	0.01	-1.42	0.155	-0.04	0.01	-5.21	<.001	-0.02	0.01	-2.66	0.008	-0.02	0.01	-2.71	0.007
Male	-0.36	0.25	-1.46	0.143	-0.46	0.24	-1.94	0.053	0.62	0.30	2.07	0.039	-0.83	0.20	-4.16	<.001
Citing social sciences	-0.15	0.09	-1.74	0.082	0.03	0.08	0.36	0.721	-0.16	0.08	-1.95	0.052	-0.09	0.07	-1.25	0.212
Perceived hierarchy of sciences	-0.14	0.11	-1.29	0.196	0.41	0.09	4.56	<.001	0.24	0.09	2.57	0.010	0.23	0.09	2.61	0.009
Deference to scientific authority	-0.21	0.11	-1.84	0.066	0.30	0.10	3.00	0.003	-0.09	0.10	-0.90	0.370	0.12	0.10	1.23	0.219
Risk perceptions	0.46	0.12	3.99	<.001	-0.57	0.09	-6.36	<.001	-0.27	0.10	-2.84	0.005	-0.32	0.09	-3.49	<.001
Benefit perceptions	-0.62	0.11	-5.69	<.001	-0.01	0.09	-0.11	0.915	-0.04	0.09	-0.47	0.637	0.09	0.10	0.91	0.363
Regulation sufficiency	-0.19	0.14	-1.37	0.173	0.67	0.11	6.20	<.001	0.01	0.11	0.08	0.940	0.17	0.11	1.57	0.118
Life sciences [a]	-0.16	0.28	-0.57	0.569	-0.27	0.24	-1.10	0.273	0.17	0.22	0.79	0.431	-0.39	0.23	-1.69	0.091
Physical sciences [a]	-0.02	0.26	-0.08	0.939	0.07	0.21	0.33	0.741	0.06	0.21	0.28	0.783	-0.12	0.20	-0.57	0.571
Social sciences [a]	-0.15	0.29	-0.50	0.620	-0.30	0.29	-1.04	0.296	-0.13	0.29	-0.44	0.659	-0.43	0.27	-1.58	0.116

Note. a. Reference group: computer and information science.

Reading social sciences showed similar effects in distinguishing different views of who should have a say in AI regulation development. More frequent reading social sciences is likely to be in the class holding the most inclusive perspective than to be in the input agnostic class (β = -.25, p < .01), the government-prioritized class (β = -.24, p = .013), and the industry-prioritized class (β = -.22, p = .011). I concluded that H2a was mostly supported. Regarding citing social sciences, I failed to find any significant effects of citation behaviors on having a more inclusive view (see Table 12). Thus, H2b was not supported.

As expected, scientific experts with higher deference to scientific authority (β = .31, p < .01) were more likely to have the least inclusive view (i.e., input agnostic) than the most inclusive class. H3 was supported.

Risk perception was an important factor that segmented scientific experts' views of who should have a say. Scientific experts with higher levels of risk perceptions were more likely to have an inclusive view as compared to having views that prioritize government (β = -.27, p < .01), industry (β = -.33, p < .001), or that are input agnostic (β = -.58, p < .001). However, scientific experts with higher levels of risk perceptions were less likely to have an inclusive view as compared to those who value experts and civic groups (β = .46, p < .001). Therefore, H4 was partially supported. Regarding the effects of benefit perceptions on regulation views (RQ2), perceiving higher levels of benefits is less likely to be in the class supporting experts and civic groups (β = -.60, p < .001) than the inclusive class, showing opposite effects than risk perceptions on distinguishing these two classes. The perceived regulation sufficiency of AI applications (β = .61, p < .001) is associated with a less inclusive perspective of who should have a say in AI regulation development. H5 was supported.

Regarding demographic differences, older respondents were more likely to be in the inclusive class than the input agnostic (β = -.04, p < .001), the government-prioritized (β = -.02, p = .013), or the industry-prioritized (β = -.02, p = .012) classes. Male respondents were less likely to be in the industry-prioritized (β = -.82, p < .001) or the input agnostic (β = -.58, p = .016) classes, compared to the inclusive class. I did not find any significant effects of scientists' research fields on differentiating their views of who should have a say in AI regulation development.

Discussion

This study empirically analyzes different perspectives that AI scientific experts may have regarding who should have a say in AI regulation development. The inclusion of diverse stakeholders in regulation development is critical, as the post-normal science era necessitates the democratization of expertise (Funtowicz & Ravetz, 1992). Among all the potential stakeholders from the scientific community, industry, government, and interest groups, among others, AI scientific experts in this sample overwhelmingly consider themselves as key actors in developing AI regulation. They also have high agreement on citizens' involvement in AI regulation development. However, their views of what roles government and big tech companies play are more divisive.

Before I elaborate on the theoretical and practical implications, I first want to mention three limitations related to the measurements. First, regarding who should have a say in AI regulation development, I used a binary scale (yes or no). Compared to an interval scale, the dichotomous measure may fail to capture nuanced variations of different weights of each actor in regulation development. For instance, although most AI scientific experts agree with involving both scientists and citizens, some of them may see a more hierarchical relationship between scientists and citizens. Second, the list of actors is not exhaustive. Third, the measure of reading and citing social science literature related to AI is based on self-reported data. Future research can be improved by comparing self-reported data and actual citation behaviors documented in their publications.

Despite these limitations, this study provides insights on empirically analyzing different perspectives that scientists have on who should have a say in technology regulation development. The five segments show a tendency of a linear to inclusive spectrum, with the inclusive segment considering that all actors should have a say on the one end and the input agnostic segment on the other. Based on factors that shape the different attitude types, the segment that values experts and civic groups shares more common with the inclusive segment, whereas the government-prioritized and the industry-prioritized segments behave more similarly to the input agnostic segment. More specifically, compared to the inclusive segment, the three less inclusive segments have higher agreement on distinguishing hard sciences from soft sciences, read less research on AI's social impacts and ethical issues, and perceive lower levels of risks of AI. In contrast, the similarities between the segment who support the voice of experts and civic groups and the inclusive segment indicates the importance of including public values in science policy (Bozeman & Sarewitz, 2011). Due to the public funding for scientific and technology research and development, incorporating input from lay publics in decision-making is an important way to make science and technology development accountable for the public as well as encourage public engagement with science to achieve democratic ideals (Alperin et al., 2019).

My study also examines the impacts of exposure to social science research through formal citations and informal reading. I find a positive relationship between reading social science research on AI and the formation of a more inclusive view of who should have a say in AI regulation. This confirms the utility of incorporating social sciences into AI research and development to generate more socially desirable outcomes. Despite the high correlation between reading and citing social science research on AI, I do not find evidence of similar roles of citing social sciences on shaping different views of who should have a say in AI regulation development. Perhaps, reading social science research on AI, as an informal way, is more accessible and easier to comprehend for the AI scientific community. Reading may be effective enough in shaping scientific experts' attitudes toward science policy making. From a practical perspective, studies on ethics and societal impacts of AI should be popularized and published in non-expert formats in order to reach a wide range of audiences, such as via editorials and commentaries. Additionally, my study shows that perceived hierarchy of sciences is associated with a less inclusive view of who should have a say in AI regulation development among AI scientists. Scientists' agreement with hierarchical assumption of sciences resonates with the linear model that basic science determines applied science and the societal outcomes of science, which increases the likelihood of excluding citizens and other non-expert stakeholders in policy debates (Pielke, 2007).

Notably, this study is based on AI scientific experts' normative beliefs of who should have a say in AI regulation development. When it comes to science policy making, there are various challenges in designing infrastructures to ensure the diversity and inclusiveness of public participation (for a review, see Scheufele et al., 2021). Future research should study how scientific experts translate their normative beliefs into practical participation in interacting with different stakeholders in regulation development.

Chapter 5. AI scientists' likelihood of incorporating social science input into the AI development process: The role of values, science news attention, and attitudes toward social sciences (Study C & Study D)

As shown in study B, exposure to social science research can have positive impacts on forming more inclusive views of who should have a say in AI regulation development. With these positive effects of social science research in mind, this chapter further explores what factors shape scientific experts' likelihood to incorporate social science input into AI research and development.

As discussed in Chapter 1, it is urgent to foster effective professional norms for conducting responsible AI research and development that supplements political regulations. AI scientific communities are increasingly aware of the importance of scrutinizing the societal impacts of AI. A majority of AI/ML researchers think that AI safety research should be prioritized and that ML institutions should conduct pre-publication review to assess potential harms (Zhang et al., 2020). Scientific experts, on the other hand, are pessimistic about reaching a consensus about what an ethical AI would look like within the next decade (Pew Research Center, 2021).

To identify and address a blind spot of AI research, Crawford and colleagues have claimed the necessity to raise "a practical and broadly applicable social-systems analysis thinks through all the possible effects of AI systems on all parties. It also engages with social impacts at every stage — conception, design, deployment, and regulation (Crawford & Calo, 2016, p. 313)." Resonating with other frameworks, such as participatory and deliberative approaches and public engagement with wicked issues in science and technology, the inclusion of views of the affected publics should be involved in AI research and development. This inclusion should occur sooner rather than later to avoid neglecting possible problems that scientists may not be aware of and to ensure ethically responsible and inclusive innovation (Brey, 2017; Scheufele et al., 2021). Social science scholars and their work can potentially bridge lay publics and technical experts of AI, as the focus of social sciences is to understand people and how they live (National Academies of Sciences, 2017). To highlight the role of social scientists and their work is not to discount the value of direct involvement of lay publics in AI research. I argue that formal, interdisciplinary collaboration and informal ways of reading and citing interdisciplinary research can increase the scientific and technical human capital for each party, such as understanding the ethical and social issues of AI from various perspectives and disciplines (Bozeman et al., 2001; Ford, 2014). This is particularly important because changes such as writing broader impacts statements will not succeed unless computer scientists acquire knowledge and skills to meaningfully anticipate the societal impacts and ethical issues of their work (Nanayakkara et al., 2021). In practice, while there are some good models for interdisciplinary research institutions on AI, like the Stanford Institute for Human-Centered Artificial Intelligence, collaboration between computer scientists and social scientists remain to be widely fostered (NSF-CISE-SBE Virtual Roundtable, 2020).

Relying on concepts from expert political judgment, the theory of planned behavior, and science and risk communication, study C explores factors that shape AI scientific experts' tendency to incorporate social science input into their work as well as their estimation of most AI scientists. Particularly, I examine factors including value dispositions, science media attention, reasoning styles, attitudes toward social sciences and interdisciplinary collaboration, as well as risk and benefit perceptions of AI. This study will expand theoretical understanding of how scientists form social norms related to their research process and react to expertise outside of their fields, as well as provide practical implications on adopting these norms.

Study D embeds a survey experiment that takes social identity as one potential approach to compare the effectiveness of ingroup versus outgroup message cues about AI scientists collaborating with social scientists.

Study C

Adopting the norm: AI scientific experts' likelihood to incorporate social science input into AI research and development

This study examines scientific experts' self-reported intentions as well as their estimations of important others' intentions (i.e., their fellow AI scientists) in terms of incorporating social science input into AI research and development, which taps into the idea of their perceived peer norms. Social norms can be defined as shared conventions or rules that guide the behaviors of members of a group (Pepitone, 1976; Shaffer, 1983). Individuals perceive and interpret social norms based on observing other people's behaviors (Shaffer, 1983). Social norms also function as a "quasi-statistical sense" of how other people think, in this case, how most AI scientists respond to the use of social science in their work. Individuals may feel fear of isolation because they are deviated from the norms that the majority holds (Fung & Scheufele, 2014; Scheufele, 2007). It is likely that individual AI scientific experts tend to form parallel attitudes, perceptions, and behavioral intentions as their peers when it comes to incorporating social science input. This tendency could have been reinforced, as top-tier AI/ML conferences began to require broader impact statements evaluating the positive and negative impacts of each conference submission.

Despite the potential strong influences from peers on their own work protocol, AI scientific experts may rely on different mechanisms in evaluating their own likelihood and that of most AI scientists' tendency to incorporate social science insights into their own research. Drawing from the social information processing theory that individuals use contextual cues, such as relevant information from work and social environments to form meaning (Fulk et al., 1987; Salancik & Pfeffer, 1978), their estimation of others may rely more on social cues or social contexts, rather than their own value or attitudinal dispositions, such as reasoning styles. The next sections review factors that may influence AI scientific experts' likelihood of incorporating social science input into AI research for themselves and their estimation of most AI scientists' behavior intentions.

Reasoning styles

Forecasting the social impacts of AI is not easy because potential risks of AI are likely high, fast-changing, and unknown. Researchers have claimed that cognitive styles, or "how experts think," systematically explain why some experts consistently outperform others in forecasting future events (Tetlock, 2005). Tetlock's book on expert political judgment and subsequent studies have discovered one potential explanation for constantly better performance of prediction—fox-like reasoning versus hedgehog-like reasoning (Mellers et al., 2015; Tetlock, 2005). While hedgehogs know one thing very well, foxes have subtle knowledge of a wide range of issues. Experts with fox-like reasoning styles are more open-minded to outside expertise and are more willing to update their own beliefs, which increases their accuracy in predicting controversial political events than those with hedgehog-like reasoning styles (Mellers et al., 2015; Tetlock, 2005). Considering the context of understanding societal impacts of AI, it is likely that scientists with fox-like reasoning styles are more likely to form a more complex understanding of benefits and risks of AI. They could be more open to expertise from other disciplines, such as social science input. I therefore propose the following hypothesis:

> **H1:** The fox-like reasoning style is positively associated with higher likelihood of incorporating social science input into AI research for (a) themselves and for (b) most AI scientists.

Attention to science news

The epistemology of science journalism indicates that the job of science journalists is not only to reduce technical complexity of scientific facts, but also to expand meaning-making for public consideration and discussion of the impacts of science on society (Brennen, 2018). Mediated representation of science provides scientific experts with potential availability heuristics to understand how social contexts shape their work. As the majority of AI scientists are white, young, and East Asian people, AI scientists are sometimes criticized for the lack of diversity among them and the lack of training about the social, historical, and political contexts of their work (Ledford, 2019). The assumption about science media use is that media could potentially provide social debates on how AI may disproportionally influence social groups that have very different living experiences from those of AI scientists (Ledford, 2019; Shaikh & Moran, 2022). This pathway of science media use is related to the potential of adequate reasoning about the importance of incorporating social sciences. Furthermore, scientific experts are not only news consumers, but they can also be sources for news coverage on scientific discovery that are involved in the agenda and frame building of science news coverage. Media interactions subsequently increase scientists' publicity and reach to the general public. The reciprocal interaction between scientists and media may generate a new phenomenon, known as the medialization of science, making science profession more media-oriented (Weingart, 1998). Thus, I propose the following hypothesis:

> **H2:** Attention to science news is positively associated with higher likelihood of incorporating social science input into AI research for (a) themselves and for (b) most AI scientists.

Value predispositions

This study focuses on two value predispositions: deference to scientific authority and scientists' perceived responsibility to ensure safe use of science. As deference to scientific authority refers to ideas about scientific exceptionalism, the authority to make decisions on science in society is devoted to the scientific community rather than the general public (Brossard & Nisbet, 2007; Howell, Wirz, et al., 2020). Such an authoritative belief may prevent scientific experts from paying attention to public needs and wishes, accompanied with lower needs to incorporate social sciences to learn the potential societal impacts of their work. The perceived responsibility to ensure safe use of science, on the other hand, may yield opposite effects compared to deference to scientific authority. Given the diverse use of AI applications, such as those applications of large language models, being responsible for the way scientists' discoveries are used by other people means that AI scientists need to understand and forecast the social impacts of their work from the pre-deployment to post-market phases (Crawford & Calo, 2016; Nanayakkara et al., 2021). Understanding the ethical, regulatory, and societal concerns, as well as competing values and interests associated with AI applications, will require a greater amount of knowledge and skills that are grounded in social science research. Thus, I propose the following hypotheses on value predispositions:

H3: The degree of deference to scientific authority is negatively associated with higher likelihood of incorporating social science input into AI research for (a) themselves and for (b) most AI scientists.

H4: The degree of perceived responsibility to ensure safe use of science is positively associated with higher likelihood of incorporating social science input into AI research for (a) themselves and for (b) most AI scientists.

Risk and benefit perceptions

Scientists may use their risk perceptions of science issues to form views of science regulation and policy making (Corley et al., 2013, 2016; Corley et al., 2009; Su et al., 2016). Among scientists, risk perceptions are associated with greater support for technology regulation (Su et al., 2016) and inclusion of citizen engagement in regulation (Wirz, 2021). The association between increased risk perceptions and more support for technology regulation and the inclusion of public input in regulation development could be a proxy for how scientists consider the role of external expertise and regulation requirement for their own research process. Risk perceptions of AI may provide contextual cues that motivate scientific experts to devote more effort to observing and foreseeing the potential societal impacts of their own and others' work and develop interventions to reduce the potential negative impacts. However, the relationship between benefit perceptions of AI and the likelihood to incorporate external expertise in AI research and development is less clear than that of risk perceptions of AI. I therefore propose the following hypothesis and research question:

H5: Risk perceptions of AI are positively associated with higher likelihood of incorporating social science input into AI research for (a) themselves and for (b) most AI scientists.

RQ1: How do benefit perceptions of AI associate with the likelihood of incorporating social science input into AI research for (a) themselves and for (b) most AI scientists?

Attitudes toward social science research

Based on the theory of planned behavior (Ajzen, 1991), attitudes play a role in shaping behavioral intentions. Therefore, the likelihood of incorporating social science input into AI research and development relies heavily on the evaluation of social science research. As discussed in Study B, the hierarchy hypothesis of the sciences has established long-term bias against the so-called soft social sciences, which may prevent interdisciplinary collaboration (Freudenburg, 2008). Based on the hierarchy hypothesis of sciences, one major critique about social sciences is the lack of rigor in research methods ("In praise of soft science," 2005; Shermer, 2007), although social scientists have argued that what makes their work rigorous is the systematic way in which they collect and analyze data (McCloskey, 1990). Another bias is that many scientists may think that they know plenty about human nature, which discounts social scientists' research findings as common sense ("In praise of soft science," 2005).

Differences between computer sciences and social sciences have led to confusion that prevents potential collaboration (NSF-CISE-SBE Virtual Roundtable, 2020). It is apparent that each discipline speaks a different academic language and holds differing opinions regarding what questions are most important and what evidence should be provided. Therefore, researchers from different disciplines need to critically examine each other's conceptualizations and assumptions before they can work together on designing an interdisciplinary study. The process of conducting in-depth interdisciplinary research can be much more time-consuming than research conducted within a single discipline (Mynatt et al., 2020). Additionally, the two disciplines have different norms regarding what constitutes a valid contribution to their discipline. Social and behavioral scientists value more on theory-driven explanations, whereas computer scientists place greater emphasis on accurate predictions and/or working systems (NSF-CISE-SBE Virtual Roundtable, 2020). I therefore raise the following hypothesis:

H6: Positive attitudes toward social science research are positively associated with higher likelihood of incorporating social science input into AI research for (a) themselves and for (b) most AI scientists.

Attitudes toward interdisciplinary collaboration

Experiences of interdisciplinary collaboration in nuclear energy have shown that social scientists “have an opportunity and an obligation to ask not just what social science can contribute to STEM, but also, what working with STEM colleagues can contribute to the social sciences” (Freudenburg, 2008, pp. 279-280). To raise the awareness that social sciences are critical for understanding human behavior to solve social-technical problems, Nature has published an editorial, in praise of soft science, that asked hard scientists to “share methods that could help them (social scientists) address pressing societal problems” ("In praise of soft science," 2005, p. 1003). As big data, algorithms, and AI systems have created new shifts in how social scientists think about and conduct research, these technologies can not only provide social sciences with toolkits in collecting and analyzing data with an exceptional breadth, depth, and scale, but can also raise new questions about the applicability of social science theories in the algorithmically infused societies (boyd & Crawford, 2012; Lazer et al., 2009; Wagner et al., 2021). The mutual benefits may increase the likelihood of a reciprocal knowledge exchange and collaboration. Hence, I propose the following hypothesis:

> **H7:** Positive attitudes toward interdisciplinary collaboration between computer scientists and social scientists are positively associated with higher likelihood of incorporating social science input into AI research for (a) themselves and for (b) most AI scientists.

Methods

Details of the sample are discussed in Chapter 2. Table 13 listed the measures and descriptive statistics of dependent, independent, and control variables used in this study.

Table 13. Measures and descriptive statistics of variables

Variable	**Measure**	**Scale**	**Statistics**
Dependent variables			
Likelihood of incorporating social sciences for oneself	How likely do you think it is that **you** will incorporate insights from social sciences into the following aspects of your own work on AI? a) Conceptual design: Defining research objectives and proposing research questions b) Technical design: Developing model constructs, instruments, and attributes c) Execution and evaluation: Assessing the impacts of their research and applications	1=very unlikely, 5=very likely	M = 3.7, SD = 1.1, *Cronbach's* α = .87
Likelihood of incorporating social sciences among most AI scientists	How likely do you think it is that **most AI scientists** will incorporate insights from social sciences into the following aspects of their work on AI? a) Conceptual design b) Technical design c) Execution and evaluation		M = 3.1, SD = 1.0, *Cronbach's* α = .85
Independent variables			
Reasoning styles	In a famous essay, Isaiah Berlin classified thinkers as hedgehogs and foxes: The hedgehog knows one big thing and tries to explain as much as possible using that theory or framework. The fox knows many small things and is content to improvise	1=very much hedgehog-like, 5=very much fox-like	M = 3.4, SD = 1.0

	explanations on a case- by-case basis. When it comes to making predictions, would you describe yourself as more of a hedgehog than of a fox?		
Deference	Scientists know best what is good for the public.	1=strongly disagree, 5=strongly agree	*M* = 2.5, *SD* = 0.9
Scientist responsibility	Scientists are responsible for the way their discoveries are used by other people.		*M* = 2.8, *SD* = 1.1
Attention to science news	In general, how much attention do you pay to news stories about the following topics? a) recent developments in science and technology b) ethical implications of emerging science and technology c) regulations of emerging science and technology	1=never, 5=a lot	M = 3.6, *SD* = 0.7, *Cronbach's α* = .70
Risk perceptions	How likely AI will …? e) strengthen the U.S. economy f) increase national security g) improve individuals' health h) help fight terrorism threats	1=not at all likely, 7=certain	*M* = 5.0, *SD* = 1.0, *Cronbach's α* = .81
Benefit perceptions	How likely AI will …? e) worsen societal inequalities f) give some people too much power g) threaten personal liberties h) displace workers by automating their jobs		*M* = 4.8, *SD* = 1.1, *Cronbach's α* = .77
Attitudes toward social sciences	a) Most results in the social sciences tend to be common sense. b) Social science research is as rigorous as research in other fields of science.	1=strongly disagree, 5=strongly agree	a) *M* = 2.7, *SD* = 0.9 b) *M* = 3.3, *SD* = 1.1
Attitudes toward collaboration	Computer science research can greatly extend social theories by investigating social data that are co-shaped by algorithms and human behavior.		M = 3.9, SD = 0.8
Controls			
Age		Numeric	*M* = 45.3, *SD* = 13.1
Gender	Male (79.0%), female (18.1%), non-binary (0.8%)	Binary (male)	
AI subfield	Ethics and societal impacts of AI (16% Yes)	Binary	

Analysis

The survey used in the present study also embedded an experiment to examine how information sources may influence AI scientific experts' likelihood of incorporating social science input. Respondents were randomly assigned to a control or a treatment condition with stimuli from one of the three journal sources, including *Nature Machine Intelligence*, *Nature Human Behaviour*, and *Nature Biotechnology*. Though the experiment is not of the current study's primary interest, I controlled the confounding effects that the experimental conditions may cause by creating dummy variables of each experimental condition as covariates in the regression models.

I used hierarchical ordinary least squares regression (OLS) to test the research hypotheses and questions. To reduce the risk of Type I error caused by a large sample, I set the alpha level to establish significance level at 0.01. All independent variables were grouped in blocks and entered in the models based on the hypothesized order of causality (Cohen et al., 2003). The order of the blocks in the regression models were as follows:

- Block 0: experimental conditions;
- Block 1: demographics (gender and age);
- Block 2: career status (AI subfield);
- Block 3: reasoning styles;
- Block 4: value predispositions (deference to scientific authority and belief in scientists' responsibility);
- Block 5: attention to science news;
- Block 6: risk and benefit perceptions;
- Block 7: attitudes toward social sciences and interdisciplinary collaboration.

Results

I began with the correlation analysis between key independent variables. As shown in Table 14, all of the absolute values of correlations were below 0.27. I therefore concluded that there was no problematic multicollinearity in the analysis.

Table 14. Correlations between independent variables (Study C)

	Gender (male)	Age	AI Field	Reasoning styles	Deference	Scientist responsibility	Attention to science news	Benefit perceptions	Risk perceptions	Attitudes toward social sciences: results	Attitudes toward social sciences: methods
Age	0.04										
AI Field: Ethics and societal impacts of AI	-0.04	0.04									
Fox-like reasoning styles	-0.01	0.00	0.02								
Deference to scientific authority	0.01	0.00	-0.05**	-0.05*							
Scientist responsibility	-0.04	-0.05*	0.09***	0.01	0.06**						
Attention to science news	0.00	0.11***	0.24***	0.03	0.03	0.09***					
Benefit perceptions	0.14***	0.06*	-0.10***	-0.01	0.17***	-0.09***	0.07**				
Risk perceptions	0.01	-0.08***	0.13***	0.01	-0.11***	0.08***	0.11***	-0.08***			
Attitudes toward social sciences: Most results in the social sciences tend to be common sense.	0.06**	-0.05	-0.12***	-0.02	0.16***	0.03	-0.02	0.10***	-0.09***		
Attitudes toward social sciences: Social science research is as rigorous as research in other fields of science.	-0.11***	-0.05*	0.15***	0.02	0.01	0.14***	0.09***	-0.04	0.02	-0.27***	
Attitudes toward collaboration: Computer sciences research can greatly extend social theories by investigating social data that are co-shaped by algorithms and human behavior.	0.02	-0.04*	0.05*	-0.02	0.12***	0.01	0.12***	0.27***	-0.03	-0.01	0.19***

Note. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Table 15. OLS regression models predicting AI scientific experts' likelihood of incorporating social science input into AI research and development for themselves and for most AI scientists

	How likely do you think it is that you will incorporate insights from social sciences into your own work on AI?					**How likely do you think it is that most AI scientists will incorporate insights from social sciences into their work on AI?**				
	Unstandardized Beta	Std. Error	Standardized Beta	t	Sig.	Unstandardized Beta	Std. Error	Standardized Beta	t	Sig.
Block 0										
NMI	-0.05	0.06	-0.02	-0.82	0.413	-0.06	0.06	-0.03	-1.01	0.313
NHB	-0.04	0.06	-0.02	-0.60	0.549	-0.12	0.06	-0.05	-2.00	0.046
NB	-0.03	0.06	-0.01	-0.43	0.669	0.05	0.06	0.02	0.84	0.402
Inc Adjusted R^2 (%)	-0.1					0.4**				
Block 1										
Male	-0.21	0.06	-0.07	-3.56	<.001	0.02	0.06	0.01	0.44	0.663
Age	0.00	0.00	0.00	0.14	0.886	0.00	0.00	0.00	-0.04	0.972
Inc Adjusted R^2 (%)	1.0***					0.0				
Block 2										
AI subfield: Ethics and societal impacts of AI	0.54	0.07	0.18	8.19	<.001	-0.25	0.06	-0.09	-4.07	<.001
Inc Adjusted R^2 (%)	6.6***					0.9***				
Block 3										
Fox-like reasoning styles	-0.01	0.02	-0.01	-0.50	0.619	-0.05	0.02	-0.05	-2.16	0.031
Inc Adjusted R^2 (%)	0.0					0.2*				
Block 4										
Deference to scientific authority	-0.04	0.03	-0.03	-1.48	0.139	0.06	0.03	0.05	2.28	0.023
Belief in scientists' responsibility	0.11	0.02	0.11	5.10	<.001	0.09	0.02	0.10	4.56	<.001
Inc Adjusted R^2 (%)	1.9***					1.9***				

Block 5										
Attention to science news	0.18	0.03	0.11	5.26	<.001	0.09	0.03	0.07	2.83	0.005
Inc Adjusted R^2 (%)	1.7***					0.5**				
Block 6										
Benefit perceptions	0.00	0.03	0.00	-0.10	0.922	0.10	0.02	0.10	4.22	<.001
Risk perceptions	-0.02	0.02	-0.02	-0.84	0.400	-0.14	0.02	-0.15	-6.69	<.001
Inc Adjusted R^2 (%)	0.1					3.9***				
Block 7										
Attitudes toward social sciences: Most results in the social sciences tend to be common sense.	-0.03	0.03	-0.02	-1.03	0.303	0.07	0.03	0.06	2.60	0.009
Attitudes toward social sciences: Social science research is as rigorous as research in other fields of science.	0.25	0.02	0.25	11.10	<.001	0.04	0.02	0.04	1.79	0.073
Attitudes toward collaboration: Computer sciences research can greatly extend social theories by investigating social data that are co-shaped by algorithms and human behavior.	0.18	0.03	0.13	5.96	<.001	0.13	0.03	0.10	4.50	<.001
Inc Adjusted R^2 (%)	8.8***					1.4***				
Total R^2 (%)	20.0***					8.8***				

Note. *p < 0.05; **p < 0.01; ***p < 0.001.

Focusing first on the likelihood of incorporating social sciences for AI scientific experts themselves, the OLS regression model explained 20.0% of the variance (See Table 15). Attitudes toward social sciences (8.8%) explained the most of the variance, as followed by their research subfields (6.6%), value predispositions (1.9%), science media attention (1.7%), and demographics (1.0%). I failed to find any significant effects of reasoning styles (β = -.01, *n.s.*) on influencing the likelihood of incorporating social sciences for AI scientists themselves (H1a was not supported). Science media attention (β = .11, p < .001) was associated with greater likelihood of self-reported likelihood of incorporating social sciences into AI research, supporting H2a. There was no significant relationship between deference to scientific authority (β = -.03, *n.s.*) and individual AI scientists' likelihood of incorporating social science research (H3a was not supported). Those who had higher agreement that "scientists are responsible for the way their discoveries are used by other people" (β = .11, p < .001) were more likely to incorporate social sciences into their own research on AI, supporting H4a. I did not find any significant effects of risk perceptions (β = -.02, *n.s.*) and benefit perceptions (β = .00, *n.s.*) on AI scientific experts' likelihood of incorporating social sciences, which failed to support H5a and provided null findings for RQ1a. Positive attitudes toward social sciences research, i.e., agreeing that social science research is as rigorous as research in other fields of science (β = .25, p < .001), was linked to greater likelihood of AI scientific experts' incorporation of social science input into their work. Likewise, the likelihood of incorporating social science input into their work was higher when there was a higher agreement that computer sciences research can greatly extend social theories by investigating social data that are co-shaped by algorithms and human behavior (β = .13, p < .001). Notably, the other statement about social science research that most results in the social sciences tend to be common sense (β = -.02, *n.s.*) was no longer associated with the dependent variable after controls. Thus, H6a was partially supported and H7a was supported. Additionally, male respondents (β = -.07, p < .001) were less likely to incorporate social sciences research into their own work, whereases those who worked in the subfield of ethics and societal impacts of AI (β = .18, p < .001) had higher likelihood of incorporating social sciences research into their own work.

I then focused on the OLS model predicting AI scientific experts' estimation of most AI scientists' likelihood of incorporating social sciences into AI development, which explained 8.8% of the variance (as seen in Table 15). Risk and benefit perceptions (3.9%) explained most of the variance, as followed by value predispositions (1.9%), attitudes toward social sciences (1.4%), and research subfields (0.9%). Contrary to the self-estimation model, those who worked in the subfield of ethics and societal impacts of AI (β = -.09, p < .001) considered most AI scientists less likely to incorporate social sciences research into AI research. Science media attention (β = .07, p < .01) was associated with greater likelihood that most AI scientists will incorporate social sciences into AI research, supporting H2b. Regarding value predispositions, belief in scientists' responsibility (β = .10, p < .001) were positively and significantly associated with higher likelihood that most AI scientists will incorporate social science input into their work, supporting H4b. Regarding RQ1b, benefit perceptions of AI (β = .10, p < .001) were associated with greater likelihood that most AI scientists will incorporate social science input into their work. Risk perceptions of AI (β = -.15, p < .001) yielded the opposite effects, suggesting that H5b was not supported. Like the self-estimation model, positive attitudes toward collaboration between computer and social scientists (β = .10, p < .001) was associated with higher likelihood of incorporating social science input into most AI scientists' work. H7b was supported. Agreement that most results in the social sciences tend to be common sense (β = .06,

$p < .01$) turned to be a significant predictor after controls, though the direction of effects was contradictory to the hypothesis. Agreement with the methodological rigor of social science research ($\beta = .04$, *n.s.*) was not associated with the dependent variable. H6b was thus not supported. Additionally, I did not find any significant effects of reasoning styles and deference to scientific authority on the dependent variable, failed to support H1b and H3b.

Discussion

This study analyzes what factors may impact scientific experts' likelihood of incorporating social science input into AI research and development, as incorporating social science input can help AI systems better tackle sociotechnical problems. Before I discuss the theoretical and practical implications in detail, I first want to flag two limitations of this study. First, the cross-sectional data cannot determine causality. The directional relationships between variables are theoretically grounded, but the variables may have reciprocal relationships. Consequently, I must interpret the results with caution, especially in order not to overestimate possible causalities. Second, many measures used in the model only contain one item, which might lead to random measurement errors. Although I adapted several multi-item measures from prior research, such as reasoning styles and attitudes toward social sciences, those measures in this data did not meet reliability tests. Future research can contribute to extracting latent dimensions for attitudes toward social sciences.

The findings show that AI scientific experts show higher tendency to incorporate social science input as compared to their estimation of most AI scientists doing so. The self-other discrepancy is largely driven by a small portion of scientific experts who study the social impacts and ethical issues of AI. It is not surprising that most AI scientists will not reach the degree of using social science input as these specialists do. The self-other comparison difference can also be influenced by the social desirability bias in a survey in which respondents rated themselves more favorably (Nederhof, 1985).

More importantly, I find that AI scientists will rely on different mechanisms to explain their own tendency and their estimation of most AI scientists' tendency to incorporate social science input into their research. For themselves, consistent with the theory of planned behavior (Ajzen, 1991), attitudes toward social sciences and interdisciplinary collaboration play the largest role, followed by their research focus on AI (i.e., whether their research focuses on the ethics and societal impacts of AI). When making an estimation of most AI scientists' likelihood to incorporate social science input into their work, the model overall explained less than 10 percent of the variance, smaller than half of that variance in the model predicting scientists' own likelihood. Among potential factors, risk and benefit perceptions show the strongest relationships with AI scientific experts' estimation of others' likelihood of incorporating social science input. This is in line with the social information processing theory positing that individuals use contextual cues to form meaning (Fulk et al., 1987; Salancik & Pfeffer, 1978). However, this study finds a counterintuitive effect that those perceiving high risks of AI consider that most AI scientists will less likely incorporate social science research, but that those who perceive greater benefits of AI consider that most AI scientists will likely incorporate social science input. Perhaps the negative association between risk perceptions of AI and perceived likelihood of most AI scientists' incorporation of social science input is an estimation of the current situation. Those who perceive more risks of AI are pessimistic about how most AI scientists may refine their research process. Future research could use focus groups and semi-interviews to delve deep into the potential reasons for why risk perceptions is associated with lower levels of perceived likelihood of most AI scientists' incorporation of social science input into AI development.

In terms of practical implications, to adopt norms of incorporating social science input into AI research and development at the large-scale scientific community level, it is important to deliver the social and contextual cues about potential risks and benefits of AI, especially the potential consequences that some AI scientists are not aware of. At an individual level, it is important to form positive evaluations of social science research regarding its values and rigorous research methods. Meanwhile, it is useful to emphasize how computer sciences may contribute to social science research, such as extending social theories by investigating social data that are co-shaped by algorithms and human behavior. Other promising incentives for incorporating social science input into AI research include highlighting scientists' responsibility to ensure safe research. To that end, collective efforts to incorporate social science input can further enforce social norms within the AI scientific community.

This study does not find evidence of the role of reasoning styles in shaping AI scientific experts' likelihood of incorporating expertise outside of their disciplines. It is likely that the dependent variable in this study, the likelihood of incorporating social science input into AI research and development, is less divisive or controversial, as compared to prior research that focuses on more complicated predictions of controversial, real-life political incidents.

In summary, AI research process needs frequent updates of the social impacts of AI among scientists. More collaboration is needed between AI scientists who design those applications and social scientists who understand the context of how AI applications may disproportionally influence certain social groups. The process of understanding the social impacts of AI and refining how to conduct AI research is an evolving process with constant belief updates into which social sciences can provide insight.

Study D

Study D examines interventions, if any, that may influence scientific experts' likelihood of incorporating social science input into AI research and development. In particular, based on social identity theory (Tajfel, 1974; Tajfel & Turner, 1986; Turner, 1999), this study investigates how ingroup versus outgroup cues embedded in journal sources and article languages may influence the extent to which individual scientists or most AI scientists are likely to incorporate social science input into AI research and development.

Social identity theory describes how individuals utilize their shared group identity to interpret new information in a way that is consistent with their group values (Tajfel, 1974; Tajfel & Turner, 1986; Turner, 1999). Social identity is an integral part of an individual's self-concept, obtained from the individual's membership in a group with emotional ties to the group (Tajfel, 1974). When a specific social identity is salient, individuals are likely to depersonalize themselves, make judgments that simulate group values, and form "us against them" distinctions (Greene, 2002; Hogg & Reid, 2006). Social identity, therefore, provides causal inferences of social changes in group settings (Tajfel & Turner, 1986). The differentiating effects of ingroup versus outgroup messages on attitudes and behaviors have been validated in various contexts, such as public perceptions of algorithmic news bias (e.g., Calice et al., 2021) and language use in describing fake news (e.g., Li & Su, 2020).

Although existing research on ingroup-outgroup differentiation is mostly done in the realm of partisan identity, this study applies the ingroup-outgroup differentiation to the context of interdisciplinary collaboration, for which scientific experts classify themselves based on areas of expertise. In this case, the group object is to form norms of incorporating social science input into AI research and development. When the message is delivered by important ingroup journals with ingroup identity language, I anticipate a higher likelihood of members who share the social identity to form attitudes and behavioral intentions that are aligned with group values. In contrast, outgroup messages that trigger group identity differences may have a countereffect in shaping attitudes and behavioral intentions with people who do not share the same social identity. Therefore, I propose the following hypotheses:

H1: Ingroup cues tend to have stronger and more positive effects than outgroup cues in predicting individual AI scientific experts' likelihood of incorporating social science input into AI research.

H2: Ingroup cues tend to have stronger and more positive effects than outgroup cues in predicting AI scientific experts' estimations of most AI scientists' likelihood of incorporating social science input into AI research.

Among AI scientists, people who have already focused on ethics and societal impacts of AI might have a higher likelihood of sharing identity with both computer scientists and social scientists. It is unclear how focusing on different subfields of AI research might moderate the effects of ingroup and outgroup cues on shaping the likelihood of incorporating social science input for themselves and for most AI scientists. Thus, I raise the following research question:

RQ1: How do different focuses of AI research moderate the effects of ingroup and outgroup cues on shaping the likelihood of incorporating social science (a) for individual scientists and (b) for their estimation of most AI scientists?

Methods

Respondents were randomly assigned to one of the four conditions. Three conditions attributed a short introductory article to three different publications: *Nature Machine Intelligence*

(NMI), Nature Human Behaviour (NHB), and *Nature Biotechnology (NB)* (see Appendix A for the stimuli). The fourth condition did not contain a stimulus. Stimuli for the NMI and NHB treatment conditions were excerpted and adapted from the same article "It's time to do something: Mitigating the negative impacts of computing through a change to the peer review process" (Hecht et al., 2018), published in the ACM Future of Computing Blog on March 29, 2018. I modified the titles and used ingroup versus outgroup language pronouns that made the NMI condition an ingroup cue and the NHB condition an outgroup cue for AI scientific experts. For instance, the title of the NMI stimulus (ingroup) was "We should collaborate with social scientists," whereas the title of the NHB stimulus (outgroup) was "AI scientists should collaborate with us." The NB stimulus was excerpted and adapted from an article titled "Regulating gene-edited crops" (Kuzma, 2018). The article was published in *Issues in Science and Technology* in Fall, 2018. The NB condition served as the role of a control condition, with stimuli covering gene-edited crops, a technology topic that was irrelevant to AI. All these stimuli used an identical format, adapted from the mobile interface of an editorial article from each journal. Measures used in study D were explained in study C.

Analysis

I first conducted mean comparisons to check the effectiveness of the experiment's randomization procedure. No significant differences across four conditions on key variables were found, such as age, gender, and research fields. I then used a two-way ANCOVA (Analysis of Covariance) to examine the main effects of the experimental conditions and the interacting effects between the experimental conditions and scientists' research focus on AI. Due to the successful randomization, I did not control for variables that measured individual-level demographic differences to examine the research hypotheses. The only covariate included was the extent of using social science in the pre-test. The variable was an averaged index of reading and citing research or reports on societal implications of AI and ethics and responsible development of AI, which was proposed to be associated with AI scientific experts' likelihood of incorporating social science input into research and practice for themselves and for most AI scientists in the future.

Results

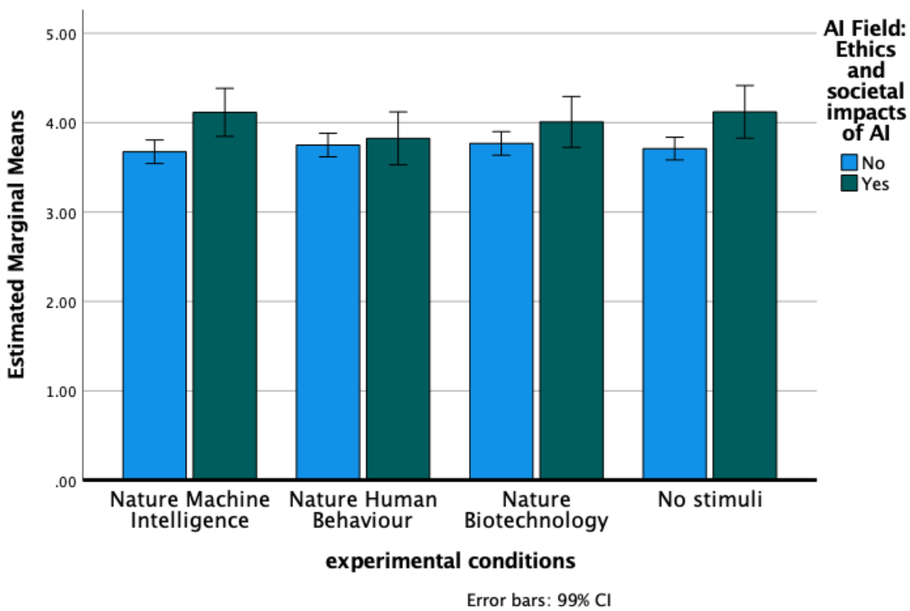
The two-way ANCOVA examined the main effects of the experimental conditions on the likelihood of incorporating social science input into AI research and development for individual AI scientists and for most AI scientists. The analysis also tested how the effects of experimental conditions may be moderated by whether an AI scientific expert's research focus was ethics and societal impacts of AI.

Table 16 shows the results of AI scientific experts' estimations of their own likelihood of incorporating social science input into AI research. I did not find any significant effects of the experimental conditions ($F(3, 2102) = 0.91$, $p = .435$, *n.s.*). Therefore, H1 was not supported. Regarding RQ1a, there were no interacting effects between the research focus on AI and the experimental conditions ($F(3, 2102) = 2.04$, $p = .107$, *n.s.*). However, the estimated marginal means suggested that scientific experts with a focus on ethics and societal impacts of AI in the NMI and the no stimuli conditions showed significantly higher likelihood of incorporating social science input into AI research than their counterparts who focused on technical aspects of AI (see Figure 6). No such differences were found in the NHB and the NB conditions.

Table 16. Two-way ANCOVA for predicting the likelihood of incorporating social science input into scientific experts' own work on AI

	Type III Sum of Squares	*df*	Mean Square	*F*	Sig.	Partial Eta Squared
Corrected Model	516.37	23	22.45	23.00	<.001	0.201
Intercept	13906.37	1	13906.37	14246.00	0.000	0.871
Experiment condition	2.67	3	0.89	0.91	0.435	0.001
AI subfield: ethics and societal impacts of AI	17.60	1	17.60	18.03	<.001	0.009
Pre-test social science use	336.85	16	21.05	21.57	<.001	0.141
Experiment condition * AI subfield	5.97	3	1.99	2.04	0.107	0.003
Error	2051.89	2102	0.98			
Total	32319.06	2126				
Corrected Total	2568.25	2125				

Figure 6. The likelihood of incorporating social science input into scientific experts' own work on AI based on experimental conditions and their research field



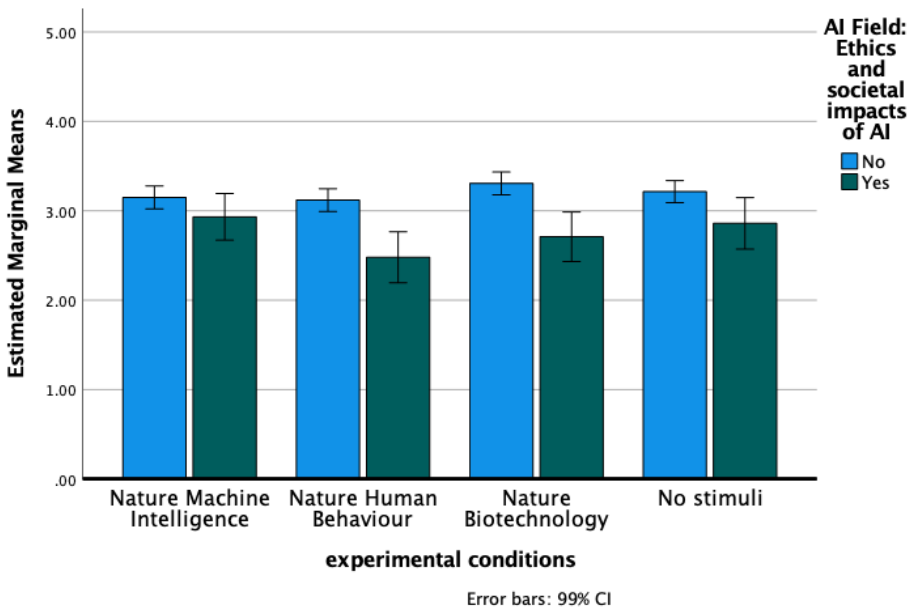
Note. Higher numbers on the (y-axis) indicate greater likelihood of incorporating social science input into respondents' own work on AI. Error bars indicate 99% confidence intervals.

Regarding AI scientific experts' estimations of most AI scientists (as seen in Table 17), there was a significant main effect of the experimental conditions on AI scientific experts' estimations of most AI scientists' likelihood of incorporating social science input into their work (F(3, 2126) = 3.99, p < .01), despite the small effect (η_p^2 = .006). Post hoc pairwise comparisons revealed that respondents in the NHB condition (M = 2.99) had a significantly lower level of estimation that most AI scientists will likely incorporate social science input into their work than the estimations in the two control groups (NB: *Mean Difference* = -.19, p < .01; No stimuli: *Mean Difference* = -.15, p = .049). Thus, H2 was partially supported. I discovered a marginal interacting effect between their research focus and the experimental conditions (RQ1b; F(3, 2126) = 3.15, p = .024). Based on the estimated marginal means (see Figure 7), AI scientific experts with a focus on ethics and societal impacts of AI in the NHB and the NB conditions perceived significantly lower likelihood that most AI scientists will incorporate social science input into AI research than their counterparts who focused on technical aspects of AI. No such differences were found in the NMI and the no stimuli conditions.

Table 17. Two-way ANCOVA for predicting the likelihood of incorporating social science input into most AI scientists' work on AI

	Type III Sum of Squares	*df*	Mean Square	*F*	Sig.	Partial Eta Squared
Corrected Model	95.02	23	4.13	4.43	<.001	0.046
Intercept	8328.57	1	8328.57	8933.85	0.000	0.808
Experiment condition	11.16	3	3.72	3.99	0.008	0.006
AI subfield: ethics and societal impacts of AI	42.81	1	42.81	45.92	<.001	0.021
Pre-test social science use	63.65	16	3.98	4.27	<.001	0.031
Experiment condition * AI subfield	8.81	3	2.94	3.15	0.024	0.004
Error	1981.96	2126	0.93			
Total	22731.25	2150				
Corrected Total	2076.98	2149				

Figure 7. The likelihood of incorporating social science input into most AI scientists' work on AI based on experimental conditions and respondents' research field



Note. Higher numbers on the (y-axis) indicate greater likelihood for incorporating social science input into most scientists' work on AI. Error bars indicate 99% confidence intervals.

Discussion

Based on the social identity theory, study D compares the effectiveness of ingroup versus outgroup cues on influencing the likelihood of incorporating input from social sciences into AI research for individual scientific experts and for their estimation of most AI scientists.

Overall, this experiment shows minimal effects of exposure to injunctive norms that are embedded with social identity cues. This study does not yield significant and strong effects on influencing individual AI scientific experts' own tendency to incorporate social science input into AI research and development. It is likely that a two-short paragraph stimulus is not strong enough to change how scientists view their research conduct, which is a stable behavioral intention. Meanwhile, the attention check shows that about a third of respondents in each condition did not correctly identify the journal source, which may further influence the effect size of the experiment.

The experiment provides more nuanced findings when it comes to AI scientific experts' estimations of most AI scientific experts' likelihood of incorporating social science into their work. The outgroup cue (*Nature Human Behaviour)* that calls for interdisciplinary collaboration between AI scientists and social scientists has a negative impact that makes AI scientific experts perceive lower levels of likelihood in terms of the collective use of social science input within the AI scientific communities. More importantly, this negative impact is more profound among AI scientific experts who study the ethics and societal impacts of AI. In other words, this particular social identity on professional field is more salient to those who study ethics and societal impacts of AI than those who study technical aspects of AI. Perhaps these AI scientific experts who focus on ethics and societal impacts of AI are concerned that outgroup messages would not be useful to change most AI scientists' views, as these AI scientists are likely more familiar with the challenges of studying ethics and societal impacts of AI. In order to call for interdisciplinary collaboration between computer scientists and social scientists, it might be useful to avoid outgroup identity language and to publish these calls in journals that are accessible to and valued by most AI scientists more often than journals with a focus on social science research. Future research might develop experiments that differentiate the effects of ingroup versus outgroup sources (AI journals/social science journals) and ingroup versus outgroup language usages (we/they) to provide more targeted recommendations.

Chapter 6. Conclusion

This chapter provides an overview of this dissertation by summarizing the findings of each study and discussing the theoretical, methodological, and practical implications.

AI has become an amalgam of scientific, political, regulatory, and ethical issues that require difficult compromises among social groups whose values and concerns are potentially in conflict. As AI scientists and developers alone cannot answer questions and solve problems raised by the use of AI, it is important to understand their views of the societal impacts of AI and their likelihood of incorporating external expertise into their research and regulatory practices to ensure responsible development of AI. Among all potential sources of external expertise, this dissertation focuses specifically on input from social sciences, as social scientists produce knowledge about the ethical, legal, and societal implications of emerging science and technology, as well as the ethics and responsibilities of scientists to ensure responsible innovation. Furthermore, this dissertation provides three sets of factors to understand AI scientific experts' views of the societal impacts, regulations, and scientific conduct related to AI. These factors include cognitive processing (value predispositions and reasoning styles), information (media attention), and science production (professional characteristics). These three pathways encompass factors such as scientists' unique professional characteristics in comparison with other social groups, as well as factors such as their cognitive processing and media attention that are common to everyone else.

Overview of findings

The broad purpose of my dissertation is to: 1) compare the potential attitudinal differences between AI scientific experts and lay publics; 2) examine the effects of AI scientific experts' exposure to social sciences on forming a more inclusive view of who should have a say in AI regulation development; and 3) investigate what factors shape the likelihood of incorporating social science input into AI research and development.

This study uses a 2022 sample collected from scientific experts who are correspondent authors of academic articles published in fields related to AI between 2010 to 2020 (N = 2,199). These articles were selected based on a bibliometric approach and retrieved from the WOS dataset. The sample consists of a wide range of scientific experts from the fields of computer and information sciences, engineering, life sciences, and social, behavioral, and economic sciences, among others.

Guided by the risk-benefit typology, study A uses segmentation analysis to compare perceived risks and benefits of AI between AI scientific experts and lay publics. Although AI scientific experts and lay publics have similar levels of risk perceptions of AI, scientific experts consider AI substantially more beneficial than their lay counterparts. The multifaceted segments of each sample indicate more nuanced attitudinal similarities and differences between the two stakeholder groups. AI scientific experts and lay publics both have large ambivalent segments (41% and 29%, respectively) that perceive equivalent levels of high risks and benefits of AI. The two segments pay a relatively balanced amount of attention to political and science news compared to other segments. Both samples have a segment with skeptical attitudes (low benefits/high risks), though the size of scientific experts (21%) is smaller than that of the lay publics (33%). These segments with skeptical attitudes, whether they are scientific experts or lay publics, hold lower levels of deference to scientific authority than the other segments in their respective sample. However, only the AI scientific expert sample has a supportive group that perceives high benefits and low risks. Study A reveals the heterogeneity of scientific experts' attitudes toward AI, which is associated with their value predispositions and views of scientific conduct. It also shows potential attitudinal differences between AI scientific experts and lay publics, which are influential to science decision making at the societal level.

With the normative belief that social science input may bridge the interface between AI scientific experts and lay publics, study B investigates whether such benefits exist by linking exposure to social science research on AI to scientistic experts' views of who should have a say in AI regulation development. Overall, AI scientific experts have high consensuses on the involvement of scientists, regulation agencies, and citizens, whereas their views of governmental branches and big tech companies are more divergent. Furthermore, I classify AI scientific experts into five segments based on their views of who should have a say in AI regulation development, including a) input agnostic, b) government prioritized, c) industry prioritized, d) valuing experts and civic groups, and e) valuing all stakeholders. AI scientific experts with more inclusive views of AI regulation development are associated with the following characteristics: having a lower agreement on the hierarchy of hard and soft sciences, reading research on societal impacts and ethical issues of AI more frequently, and having lower deference to scientific authority. Hence, study B reveals the value of exposure to social science research on AI in shaping a more inclusive model of who should have a say in AI regulation development.

Moving forward, studies C and D focus on forming professional norms of incorporating social science input into AI research and development within AI scientific communities. Study C examines what factors shape the tendency to incorporate social science input. AI scientific experts depend on different considerations in terms of estimating the likelihood that they and most AI scientists will incorporate social sciences into AI research and development. Their own likelihood of use is associated with their research focus, attitudes toward social sciences and interdisciplinary research between computer scientists and social scientists. Respondents' estimation of most AI scientists' likelihood of incorporating social science input into AI research and development relies more heavily on their risk and benefit perceptions of AI. Study D embeds a survey experiment that takes social identity as one potential approach to compare the effectiveness of ingroup versus outgroup message cues about AI scientists collaborating with social scientists. I do not find any significant findings of ingroup cues that call for collaboration from the AI scientific community. However, the countereffects of outgroup cues, especially on AI scientists who study the societal impacts and ethics of AI, indicate that a better way to shape the norms of incorporating social science input for most AI scientists is to avoid publishing articles that call for interdisciplinary research between computer scientists and social scientists in journals outside of AI disciplines and to avoid using outgroup identity languages, such as "they".

Limitations and future research

Before delving into theoretical, methodological, and practical implications in detail, it is worth noting that the findings of this dissertation must be considered against its several limitations as discussed below.

First, because AI development is highly embedded in social, political, and cultural contexts, it is vital to note that the findings should be interpreted with caution outside of the US. Respondents in my sample have affiliations with US-based institutions, though a small portion of them may have recently left the US or may work for institutions outside of the US simultaneously. Additionally, some of my research designs pertain to the US context, such as the list of potential social actors that should have a say in AI regulation development. Other political systems may have very different regulation systems (Dafoe & Journal of International, 2018) and different expectations of scientists' responsibilities and relationships to society (Gascoigne & Schiele, 2020). Future research can benefit from using cross-national samples, especially with a focus on non-English-language societies, to improve the generalizability of their research.

Second, this dissertation focuses more on the incorporation of social science input into AI research and development, rather than the other way around. It is important to note that this focus neither indicates that there is not such incorporation so far, nor discounts the importance of AI systems in advancing social science theories. My rationale is more of a response to current calls for understanding the social and political contexts of AI research at all stages, from design to deployment (e.g., Crawford & Calo, 2016; NSF-CISE-SBE Virtual Roundtable, 2020). Future research can focus on a more comprehensive view of the mutual benefits of interdisciplinary collaboration between computer scientists and social scientists.

Third, this study takes the first step to using self-reported data on gauging the effects of reading and citing social science literature related to AI. Self-reported data might not only inevitably have social desirability bias, but may also lack nuances compared to actual citation behaviors. Future research can be improved by connecting self-reported data from the survey and actual citation behaviors documented in respondents' publications. For instance, computational-assisted approaches, such as supervised machine learning, can help classify AI-related articles into different subfields. Future work can use network analysis to model citations outside of one's discipline, e.g., the number of social sciences articles that a computer scientist has cited and vice versa. This triangulation approach will further the understanding of the effects of reading and citing social science literature and connect to theories about how scientists conduct research.

Theoretical implications

First, this dissertation demonstrates the value of multifaceted segmentation in advancing theories to understand scientific experts' views of science and regulation. Prior research on risk communication has called for more nuanced approaches to understanding risk perceptions, such as using separate benefit and risk perceptions to compute an ambivalent bipolar continuum scale to capture attitude ambivalence (e.g., Wirz, 2021). Moving forward from a bipolar conceptualization, this dissertation indicates how the distributions of the AI scientific expert and the lay public samples in the risk-benefit typology are different. This dissertation also establishes the necessity of differentiating AI scientific experts with similarly low levels of ambivalence but opposite stances, namely the skeptical and the supportive segments.

More importantly, this dissertation provides input on whether assumptions about ambivalence and cognitive processing are applicable to the context of forming attitudes toward wicked sciences. For one thing, previous research suggests "asymmetries in attitude ambivalence," referring to the notion that ambivalence is higher when only negative evaluations are evoked than when only positive evaluations are evoked because negative evaluations usually emerge with at least weak levels of positive offsets (Cacioppo & Berntson, 1994; Cacioppo et al., 1997). In the context of attitudes toward AI, scientific experts have a purely supportive segment (high benefits/low risks), but lay publics do not. For lay publics, it is unlikely that only positive evaluations of emerging science and technology are activated, which may update the assumptions of asymmetries in attitude ambivalence. On the other hand, scientific experts have lower levels of asymmetries in attitude ambivalence because the activation of either risk perceptions or benefit perceptions is independent of the other one. Additionally, my findings may challenge the assumption that attitude ambivalence is associated with motivations to reduce one's cognitive dissonance and uncomfortable emotions (Clark et al., 2008; Newby-Clark et al., 2002). The ambivalent segments in both samples that are likely to reconcile risk and benefit perceptions of AI resulted from the great amount of news attention to science and political issues. In the context of wicked sciences, perceiving risks and benefits concurrently may not only reflect the complexity of science and technology, but also increase the likelihood of understanding the variety of values and interests that different social groups may have.

Second, this dissertation provides an empirical understanding of how scientific experts perceive their own roles in AI regulation development compared to the roles of other social actors. The differentiation between a linear model and an inclusive stakeholder model depends not only on scientists' perceived variety of actors that should have a say in AI regulation development, but also on their perceived hierarchy of different actors. The key pattern of an inclusive stakeholder model is the relatively equivalent importance of public voices and expert voices, which may increase the chances of mapping public values of the non-scientific, noneconomic goals of scientific endeavor onto those science policy-making processes (Bozeman & Sarewitz, 2005, 2011). On the contrary, any other models that prioritize the roles of scientists, governmental branches, or industry cooperation over citizens or civic groups are closer to the continuum end of a linear model, holding hierarchical views of different sciences and authoritative views of scientific conduct.

Third, I propose that social science input can bridge the interface between AI scientists and lay publics by synthesizing frameworks from the broader impacts to responsible research and innovation. The rationales for incorporating social science input into AI research and development are also grounded in empirical evidence shown in this dissertation: (a) there exist attitude differences between AI scientific experts and lay publics, with some scientific experts perceiving substantially higher benefits than their lay counterparts; and (b) the use of social science input on AI, such as reading research on the societal impacts and responsible development of AI, can foster a more inclusive stakeholder model that values public input in AI regulation development. To that end, this dissertation claims the importance of shaping social norms in the professional arena regarding incorporating social science input into AI research and development. I further differentiate the theoretical mechanisms that influence scientists' estimation of themselves and most AI scientists with regard to the likelihood of incorporating social science input into AI research and development. Self-estimations rely more on attitudes toward social science research and interdisciplinary collaboration, which are core components of the theory of planned behavior (Ajzen, 1991). In contrast, estimations of most AI scientists resonate more with the social information processing theory that scientists use contextual cues, such as risk and benefit perceptions of AI, to form their estimation (Fulk et al., 1987; Salancik & Pfeffer, 1978).

Fourth, this dissertation provides input on factors that shape scientific experts' views of AI, related regulations, and their scientific conduct by comparing three sets of factors, including cognitive processing (value predispositions and reasoning styles), information (media attention), and science production (professional characteristics). Consistent with previous studies (e.g., Ho et al., 2011), value predispositions, especially the degree of deference to scientific authority, play an important role in shaping AI scientific experts' views of AI and its regulation. AI scientific experts with a more authoritative view of scientific authority perceive lower levels of risks of AI and have a less inclusive view of who should have a say in AI regulation development.

Media attention is influential in shaping scientific experts' views of AI and the likelihood of incorporating social science input into AI research and development. Paying a great, relatively balanced amount of attention to political and science news is associated with more ambivalent attitudes toward AI, which is a key common characteristic shared by the ambivalent segment within the scientific expert sample and the lay public sample. Scientific experts who pay more attention to science news have a greater likelihood of incorporating social science input into their work on AI than those who pay less attention to science news. This resonates with the important mediating role of media in the "science communication as political communication" model, referring to the idea that public understanding of science and perception of science reality is shaped by mediated realities over direct engagement between scientists and lay publics (Scheufele, 2014). The impact of science news attention on scientific experts' perceptions of how they conduct research suggests that the influence of mediated reality can be extended to scientists' understanding of science and their perception of science reality.

The influences of professional characteristics on scientific experts' views of AI, regulation, and scientific conduct are mixed. In general, scientific experts who perceive the hierarchy of different disciplines hold less inclusive attitudes toward external expertise. For instance, agreement with the usefulness of classifying hard sciences and soft sciences is associated with less inclusive views of who should have a say in AI regulation development. Positive attitudes toward collaboration between computer scientists and social scientists are associated with a higher likelihood of incorporating social sciences into their work on AI. However, professional characteristics are less influential than value predispositions and media attention in shaping scientific experts' views of the societal impacts of AI. The diversity of collaboration networks across disciplines is associated with lower levels of risk perceptions of AI but does not influence the level of benefit perceptions or the segmentation of different types of attitudes toward AI.

Methodological implications

First, this dissertation offers insights on conducting comparative studies between scientific experts and lay publics. Prior studies have concluded several challenges and strategies, such as the inclusion of a representative sample and the consideration of demographics in analysis and inference (e.g., Ho et al., 2011; Rowe & Wright, 2001). This dissertation reveals another potential challenge that some stable worldviews related to science and scientific conduct may be understood and performed differently by scientists and lay publics, especially when scientists themselves are key objects of those concepts. For instance, deference to scientific authority has been widely verified and examined in the public context (e.g., Brossard & Shanahan, 2003; Howell, Wirz, et al., 2020). When it comes to the scientist sample in this study, not only do items not scale together, but the concept also shows different relationships with external variables that are different from those of the public sample. Deference to scientific authority is highly correlated with liberal views among publics. The relationship between the two concepts is not significant in the scientist sample, despite the fact that scientists with skeptical views of AI are more liberal and less deferent to scientific authority than their peers in other segments. Perhaps some scientists may interpret deference to scientific authority as a sense of intellectual overconfidence. In short, this example indicates potential trade-offs between variable consistency and validity when comparing these two groups.

Second, study A and study B use segmentation analyses to understand the multifaceted, heterogeneous segments of AI scientific experts regarding their risks and benefits of AI as well as attitudes toward who should have a say in AI regulation development. Although results from the segmentation analysis are descriptive and exploratory, they serve as the foundations for creating targeted messages and choosing effective platforms to communicate with them (Slater, 1996). Furthermore, this dissertation combines multinomial logistic regression analyses comparing category membership based on segmentation analyses as well as OLS regression analyses that identify relationships between dependent and independent variables to examine research hypotheses and questions. This combination of various analyses is useful to examine relationships that are not simply linear. These analyses are also complementary to each other and provide a comprehensive approach for theory development.

Practical implications

Science communication scholarship that studies the interface of science and society has indicated that societies will need to communicate and make decisions across all stakeholder groups to encounter the challenges of emerging science, such as AI (Scheufele, 2022). This dissertation provides practical implications on how to better foster communication between scientists and other stakeholders, as well as how to form professional norms that ensure responsible AI development within scientific communities.

First, it is promising that the majority of AI scientific experts in this sample agree that citizens and end-users should have a say in AI regulation development, indicating opportunities for involving AI scientists to engage with lay publics to solicit diverse input from them. However, the variety of perspectives that scientists and lay publics hold on the risk and benefit perceptions of AI requires the inclusion of representatives from each segment when designing engagement activities. Because some AI applications have already disproportionally impacted vulnerable social groups who might not have a strong will to participate in societal debates about science and regulation development (Bao et al., 2022), it is important to identify appropriate groups of AI scientists, such as those with ambivalent attitudes, to highlight the value of public input for AI research that resonates with vulnerable social groups' value systems. Individual AI scientists should also compare their views of AI to other AI scientists and lay publics. The necessity for scientists to understand public opinion on AI is grounded in the fact that adopting AI applications will ultimately need public approval. Since risk and benefit perceptions of science are more cultural and political than technical, the potential attitudinal difference between scientists and lay publics may lead to failures in prioritizing certain research and use of AI that the public needs and wishes for.

Second, I provide recommendations for incentivizing the use of social science input in AI research that ensures the development of innovative AI applications while also considering the complex social, ethical, and legal implications of those developments. This approach to AI innovation will help scientists and other stakeholders direct the power of technology rather than let technology shape the meaning of democracy and citizenship (Jasanoff, 2016). To increase the likelihood of incorporating social science research into AI development for individual scientists, useful strategies include shaping positive attitudes toward social sciences, emphasizing the potential of computer sciences in expanding social science theories, and increasing scientists' perceived responsibility to ensure safe research. To increase the perceived social norms at the large-scale scientific community level, it is important to deliver social and contextual cues about the potential risks and benefits of AI.

References

AAPOR. (2016). *Standard definitions: Final dispositions of case codes and outcome rates for surveys* (9th ed.). American Association for Public Opinion Research.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179-211. https://doi.org/10.1016/0749-5978(91)90020-T

Akin, H., Rose, K. M., Scheufele, D. A., Simis-Wilkinson, M., Brossard, D., Xenos, M. A., & Corley, E. A. (2017). Mapping the landscape of public attitudes on synthetic biology. *Bioscience, 67*(3), 290-300. https://doi.org/10.1093/biosci/biw171

Akin, H., Yeo, S. K., Wirz, C. D., Scheufele, D. A., Brossard, D., Xenos, M. A., & Corley, E. A. (2019). Are attitudes toward labeling nano products linked to attitudes toward GMO? Exploring a potential 'spillover' effect for attitudes toward controversial technologies. *Journal of Responsible Innovation, 6*(1), 50-74. https://doi.org/10.1080/23299460.2018.1495026

Alexander, S., Atsbeha, E., Negatu, S., Kirksey, K., Brossard, D., Holzer, E., & Block, P. (2020). Development of an interdisciplinary, multi-method approach to seasonal climate forecast communication at the local scale. *Climatic Change, 162*(4), 2021-2042. https://doi.org/10.1007/s10584-020-02845-9

Alperin, J. P., Nieves, C. M., Schimanski, L. A., Fischman, G. E., Niles, M. T., & McKiernan, E. C. (2019). How significant are the public dimensions of faculty work in review, promotion and tenure documents? *Elife, 8*(e42254), 1-23, Article e42254. https://doi.org/10.7554/eLife.42254

Asher-Schapiro, A. (2020, June 24). *California city bans predictive policing in U.S. first*. Reuters. https://www.reuters.com/article/us-usa-police-tech-trfn/california-city-bans-predictive-policing-in-u-s-first-idUSKBN23V2XC

Ashurst, C., Hine, E., Sedille, P., & Carlier, A. (2022). *AI ethics statements: Analysis and lessons learnt from NeurIPS broader impact statements* 2022 ACM Conference on Fairness, Accountability, and Transparency, Seoul, Republic of Korea. https://doi.org/10.1145/3531146.3533780

Bao, L., Krause, N. M., Calice, M. N., Scheufele, D. A., Wirz, C. D., Brossard, D., Newman, T. P., & Xenos, M. A. (2022). Whose AI? How different publics think about AI and its social impacts. *Computers in Human Behavior, 130*, 107182. https://doi.org/10.1016/j.chb.2022.107182

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5*(4), 323-370. https://doi.org/10.1037/1089-2680.5.4.323

Benjamin, R. (2019). Assessing risk, automating racism. *Science, 366*(6464), 421-422. https://doi.org/10.1126/science.aaz3873

Bennett Moses, L., & Chan, J. (2018). Algorithmic prediction in policing: assumptions, evaluation, and accountability. *Policing and Society, 28*(7), 806-822. https://doi.org/10.1080/10439463.2016.1253695

Berg, P., Baltimore, D., Brenner, S., Roblin, R. O., & Singer, M. F. (1975). Asilomar conference on recombinant DNA-molecules. *Science, 188*(4192), 991-994. https://doi.org/10.1126/science.1056638

Bernard, H. R. (2012). The science in social science. *Proceedings of the National Academy of Sciences, 109*(51), 20796-20799. https://doi.org/10.1073/pnas.1218054109

Bertoldo, R., Mays, C., Poumadère, M., Schneider, N., & Svendsen, C. (2016). Great deeds or great risks? Scientists' social representations of nanotechnology. *Journal of Risk Research, 19*(6), 760-779. https://doi.org/10.1080/13669877.2015.1042503

Bohman, J., & Rehg, W. (1997). *Deliberative democracy: Essays on reason and politics*. MIT press.

boyd, d., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society, 15*(5), 662-679. https://doi.org/10.1080/1369118X.2012.678878

Bozeman, B., & Corley, E. (2004). Scientists' collaboration strategies: implications for scientific and technical human capital. *Research Policy, 33*(4), 599-616. https://doi.org/10.1016/j.respol.2004.01.008

Bozeman, B., Dietz, J. S., & Gaughan, M. (2001). Scientific and technical human capital: an alternative model for research evaluation. *International Journal of Technology Management, 22*(7-8), 716-740. https://doi.org/10.1504/ijtm.2001.002988

Bozeman, B., Fay, D., & Slade, C. P. (2013). Research collaboration in universities and academic entrepreneurship: the-state-of-the-art. *The Journal of Technology Transfer, 38*(1), 1-67. https://doi.org/10.1007/s10961-012-9281-8

Bozeman, B., & Sarewitz, D. (2005). Public values and public failure in US science policy. *Science and Public Policy, 32*(2), 119-136. https://doi.org/10.3152/147154305781779588

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

Brennen, J. S. (2018). Magnetologists on the beat: The epistemology of science journalism reconsidered. *Communication Theory, 28*(4), 424-443. https://doi.org/10.1093/ct/qty001

Brey, P. (2017). Ethics of emerging technology. In *The ethics of technology: Methods and approaches* (pp. 175-191).

Brossard, D., & Lewenstein, B. V. (2010). A critical appraisal of models of public understanding of science: Using practice to inform theory. In L. Kahlor & P. Stout (Eds.), *Communicating Science: New Agendas in Communication* (pp. 11-39). Routledge. https://www.taylorfrancis.com/books/e/9780203867631/chapters/10.4324/9780203867631-9

Brossard, D., & Nisbet, M. C. (2007). Deference to scientific authority among a low information public: Understanding US Opinion on agricultural biotechnology. *International Journal of Public Opinion Research, 19*(1), 24-52. https://doi.org/10.1093/ijpor/edl003

Brossard, D., & Shanahan, J. (2003). Do citizens want to have their say? Media, agricultural biotechnology, and authoritarian views of democratic processes in science. *Mass Communication and Society, 6*(3), 291-312. https://doi.org/10.1207/S15327825MCS0603_4

Broussard, M., Diakopoulos, N., Guzman, A. L., Abebe, R., Dupagne, M., & Chuan, C.-H. (2019). Artificial intelligence and journalism. *Journalism & Mass Communication Quarterly, 96*(3), 673-695. https://doi.org/10.1177/1077699019859901

Buhmann, A., & Fieseler, C. (2021). Towards a deliberative framework for responsible innovation in artificial intelligence. *Technology in Society, 64*, 101475. https://doi.org/10.1016/j.techsoc.2020.101475

Buhmann, A., & Fieseler, C. (2022). Deep learning meets deep democracy: Deliberative governance and responsible innovation in artificial intelligence. *Business Ethics Quarterly*, 1-34. https://doi.org/10.1017/beq.2021.42

Bush, V. (1945). *The endless frontier*. National Science Foundation.

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin, 115*(3), 401-423. https://doi.org/10.1037/0033-2909.115.3.401

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1997). Beyond bipolar conceptualizations and measures: The case of attitudes and evaluative space. *Personality and Social Psychology Review, 1*(1), 3-25. https://doi.org/10.1207/s15327957pspr0101_2

Calice, M. N., Bao, L., Freiling, I., Howell, E. L., Xenos, M. A., Yang, S., Brossard, D., Newman, T. P., & Scheufele, D. A. (2021). Polarized platforms? How partisanship shapes perceptions of "algorithmic news bias". *New Media & Society*. https://doi.org/10.1177/14614448211034159

Calice, M. N., Bao, L., Newman, T., Scheufele, D. A., Brossard, D., & Xenos, M. A. (2022). *U.S. public attitudes on artificial intelligence*. Department of Life Sciences Communication, University of Wisconsin-Madison. https://osf.io/k82d6

Callaway, E. (2020). 'It will change everything': Deepmind's AI makes gigantic leap in solving protein structures. *Nature, 588*(7837), 203-204. https://doi.org/10.1038/d41586-020-03348-4

Capon, A., Gillespie, J., Rolfe, M., & Smith, W. (2015). Perceptions of risk from nanotechnologies and trust in stakeholders: a cross sectional study of public, academic, government and business attitudes. *BMC Public Health, 15*(1), 424. https://doi.org/10.1186/s12889-015-1795-1

Cappella, J. N., Zhang, J., & Price, V. (2017). Collective intelligence: The wisdom and foolishness of deliberating groups. In K. Kenski & K. H. Jamieson (Eds.), *The Oxford Handbook of Political Communication*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199793471.013.014

Chakradhar, S. (2019, October 24). *Widely used algorithm for follow-up care in hospitals is racially biased, study finds*. STAT. https://www.statnews.com/2019/10/24/widely-used-algorithm-hospitals-racial-bias/

Clark, J. K., Wegener, D. T., & Fabrigar, L. R. (2008). Attitudinal ambivalence and message-based persuasion: Motivated processing of proattitudinal information and avoidance of counterattitudinal information. *Personality and Social Psychology Bulletin, 34*(4), 565-577. https://doi.org/10.1177/0146167207312527

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Lawrence Erlbaum Associates.

Cole, S. (1983). The hierarchy of the sciences. *American Journal of Sociology, 89*(1), 111-139. https://doi.org/10.1086/227835

Corley, E. A., Kim, Y., & Scheufele, D. A. (2013). The current status and future direction of nanotechnology regulations: A view from nano-scientists. *Review of Policy Research, 30*(5), 488-511. https://doi.org/10.1111/ropr.12036

Corley, E. A., Kim, Y., & Scheufele, D. A. (2016). Scientists' ethical obligations and social responsibility for nanotechnology research. *Science and Engineering Ethics, 22*(1), 111-132. https://doi.org/10.1007/s11948-015-9637-1

Corley, E. A., Scheufele, D. A., & Hu, Q. (2009). Of risks and regulations: How leading US nanoscientists form policy stances about nanotechnology. *Journal of Nanoparticle Research, 11*(7), 1573-1585. https://doi.org/10.1007/s11051-009-9671-5

Crawford, K., & Calo, R. (2016). There is a blind spot in AI research. *Nature, 538*(7625), 311-313. https://doi.org/10.1038/538311a

Dafoe, A., & Journal of International, A. (2018). Global politics and the governance of artificial intelligence. *Journal of International Affairs, 72*(1), 121-126. https://www.jstor.org/stable/26588347

Davis, M., & Laas, K. (2014). "Broader impacts" or "responsible research and innovation"? A comparison of two criteria for funding research in science and engineering. *Science and Engineering Ethics, 20*(4), 963-983. https://doi.org/10.1007/s11948-013-9480-1

DeepMind. (n.d.). *Exploring the real-world impacts of AI*. https://www.deepmind.com/about/ethics-and-society

Edwards, W. (1979). Multiattribute utility measurement: Evaluating desegregation plans in a highly political context. In R. Perloff (Ed.), *Evaluation Interventions: Pros and Cons* (pp. 13-54). Sage Publications, Inc.

Fast, E., & Horvitz, E. (2017). *Long-term trends in the public perception of artificial intelligence.* Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. AAAI Press. https://dl.acm.org/doi/10.5555/3298239.3298381

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1983). "The public" vs. "The experts": Perceived vs. Actual disagreements about risks of nuclear power. In V. T. Covello, W. G. Flamm, J. V. Rodricks, & R. G. Tardiff (Eds.), *The Analysis of Actual Versus Perceived Risks* (pp. 235-249). Springer US. https://doi.org/10.1007/978-1-4613-3760-7_16

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*. McGraw-Hill.

Ford, H. (2014). Big data and small: Collaborations between ethnographers and data scientists. *Big Data & Society, 1*(2), 2053951714544337. https://doi.org/10.1177/2053951714544337

Freudenburg, W. R. (2008). The social sciences and the fields of science, technology, engineering, and mathematics (STEM): Toward the building of improved, two-way bridges. In H. Hartman (Ed.), *Integrating the Sciences and Society: Challenges, Practices, and Potentials* (Vol. 16, pp. 279-300). Emerald Group Publishing Ltd. https://doi.org/10.1016/s0196-1152(08)16011-2

Fulk, J., Steinfield, C. W., Schmitz, J., & Power, J. G. (1987). A social information processing model of media use in organizations. *Communication Research, 14*(5), 529-552. https://doi.org/10.1177/009365087014005005

Fuller, T., & Metz, C. (2018, October 26). *A.I. is helping scientists predict when and where the next big earthquake will be*. The New York Times.

https://www.nytimes.com/2018/10/26/technology/earthquake-predictions-artificial-intelligence.html

Fung, T. K. F., & Scheufele, D. A. (2014). Social norms, spirals of silence and framing theory an argument for considering cross-cultural differences in media effects research. *Spiral of Silence: New Perspectives on Communication and Public Opinion*, 131-144.

Funtowicz, S. O., & Ravetz, J. R. (1992). Three types of risk assessment and the emergence of post-normal science. In S. Krimsky & D. Golding (Eds.), *Social Theories of Risk* (pp. 251-274). Praeger.

Garvey, C., & Maskal, C. (2019). Sentiment analysis of the news media on artificial intelligence does not support claims of negative bias against artificial intelligence. *OMICS: A Journal of Integrative Biology, 24*(5), 286-299. https://doi.org/10.1089/omi.2019.0078

Gascoigne, T., & Schiele, B. (2020). Introduction: A global trend, an emerging field, a multiplicity of understandings: Science communication in 39 countries. In T. Gascoigne, B. Schiele, J. Leach, M. Riedlinger, B. V. Lewenstein, L. Massarani, & P. Broks (Eds.), *Communicating science: A global perspective*. ANU Press.

Gieryn, T. F. (1983). Boundary-work and the demarcation of science from non-science - strains and interests in professional ideologies of scientists. *American Sociological Review, 48*(6), 781-795. https://doi.org/10.2307/2095325

Gjefsen, M. D., & Vie, K. J. (2021). Propping up interdisciplinarity: responsibility in university flagship research. *Journal of Responsible Innovation, 8*(1), 48-69. https://doi.org/10.1080/23299460.2021.1899520

Greene, S. (2002). The social-psychological measurement of partisanship. *Political Behavior, 24*(3), 171-197. http://www.jstor.org/stable/1558395

Guzman, A. L., & Lewis, S. C. (2019). Artificial intelligence and communication: A human–machine communication research agenda. *New Media & Society, 22*(1), 70-86. https://doi.org/10.1177/1461444819858691

Hancock, P. A., Nourbakhsh, I., & Stewart, J. (2019). On the future of transportation in an era of automated and autonomous vehicles. *Proceedings of the National Academy of Sciences, 116*(16), 7684. https://doi.org/10.1073/pnas.1805770115

Hassabis, D. (2017). *The mind in the machine: Demis Hassabis on artificial intelligence*. FT Magazine. https://www.ft.com/content/048f418c-2487-11e7-a34a-538b4cb30025

Hecht, B., Wilcox, L., Bigham, J. P., Schöning, J., Hoque, E., Ernst, J., Bisk, Y., De Russis, L., Yarosh, L., Anjum, B., Contractor, D., & Wu, C. (2018). *It's time to do something: Mitigating the negative impacts of computing through a change to the peer review* *process*. ACM Future of Computing Blog. https://acm-fca.org/2018/03/29/negativeimpacts/

Heilweil, R. (2019, December 19). *Artificial intelligence will help determine if you get your next job, but is it up to the task?* Vox. https://www.vox.com/recode/2019/12/12/20993665/artificial-intelligence-ai-job-screen

Ho, S. S., Goh, T. J., Chuah, A. S. F., Leung, Y. W., Bekalu, M. A., & Viswanath, K. (2020). Past debates, fresh impact on nano-enabled food: A multigroup comparison of presumed media influence model based on spillover effects of attitude toward genetically modified food. *Journal of Communication, 70*(4), 598-621. https://doi.org/10.1093/joc/jqaa019

Ho, S. S., Scheufele, D. A., & Corley, E. A. (2011). Value predispositions, mass media, and attitudes toward nanotechnology: The interplay of public and experts. *Science Communication, 33*(2), 167-200. https://doi.org/10.1177/1075547010380386

Hogg, M. A., & Reid, S. A. (2006). Social identity, self-categorization, and the communication of group norms. *Communication Theory, 16*(1), 7-30. https://doi.org/10.1111/j.1468-2885.2006.00003.x

Holdren, J. P., Sunstein, C. R., & Siddiqui, I. A. (2011). Memorandum: Principles for regulation and oversight of emerging technologies. Office of Science and Technology Policy. https://obamawhitehouse.archives.gov/sites/default/files/omb/inforeg/for-agencies/Principles-for-Regulation-and-Oversight-of-Emerging-Technologies-new.pdf

Howell, E. L., & Brossard, D. (2021). (Mis)informed about what? What it means to be a science-literate citizen in a digital world. *Proceedings of the National Academy of Sciences, 118*(15), e1912436117. https://doi.org/10.1073/pnas.1912436117

Howell, E. L., Scheufele, D. A., Brossard, D., Xenos, M. A., Kwon, S., Youtie, J., & Shapira, P. (2020). Scientists' and the publics' views of synthetic biology. In B. D. Trump, C. L. Cummings, J. Kuzma, & I. Linkov (Eds.), *Synthetic Biology 2020: Frontiers in Risk Analysis and Governance* (pp. 371-387). Springer International Publishing. https://doi.org/10.1007/978-3-030-27264-7_16

Howell, E. L., Wirz, C. D., Scheufele, D. A., Brossard, D., & Xenos, M. A. (2020). Deference and decision-making in science and society: How deference to scientific authority goes beyond confidence in science and scientists to become authoritarianism. *Public Understanding of Science, 29*(8), 800-818. https://doi.org/10.1177/0963662520962741

In praise of soft science. (2005). *Nature, 435*(7045), 1003. https://doi.org/10.1038/4351003a

Jakesch, M., Buçinca, Z., Amershi, S., & Olteanu, A. (2022). How different groups prioritize ethical values for responsible AI. *arXiv preprint arXiv:2205.07722*.

Jasanoff, S. (1998). *The fifth branch: Science advisers as policymakers*. Harvard University Press.

Jasanoff, S. (2016). *The ethics of invention: Technology and the human future*. WW Norton & Company.

Jasanoff, S., & Hurlbut, J. B. (2018). A global observatory for gene editing. *Nature, 555*(7697), 435-437. https://doi.org/10.1038/d41586-018-03270-w

Kahn, J. (2021, December 22). *The world just blew a 'historic opportunity' to stop killer robots—and that might be a good thing*. Fortune. https://news.un.org/en/story/2021/11/1106612

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kaplan, K. J. (1972). On the ambivalence-indifference problem in attitude theory and measurement: A suggested modification of the semantic differential technique. *Psychological Bulletin, 77*(5), 361-372. https://doi.org/10.1037/h0032590

Kaplan, L. R., Farooque, M., Sarewitz, D., & Tomblin, D. (2021). Designing participatory technology assessments: A reflexive method for advancing the public role in science policy decision-making. *Technological Forecasting and Social Change, 171*, 120974. https://doi.org/10.1016/j.techfore.2021.120974

Knight, W. (2021, May 24). *AI can write disinformation now—and dupe human readers*. https://www.wired.com/story/ai-write-disinformation-dupe-human-readers/

Kuzma, J. (2018). Regulating gene-edited crops. *Issues in Science and Technology, 35*(1), 80-85.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Van Alstyne, M. (2009). Social science. Computational social science. *Science (New York, N.Y.), 323*(5915), 721-723. https://doi.org/10.1126/science.1167742

Ledford, H. (2019). Millions of black people affected by racial bias in health-care algorithms. *Nature, 574*, 608-610. https://doi.org/10.1038/d41586-019-03228-6

Lee, N., Nyhan, B., Reifler, J., & Flynn, D. J. (2020). More accurate, but no less polarized: Comparing the factual beliefs of government officials and the public. *British Journal of Political Science*, 1-8. https://doi.org/10.1017/S000712342000037X

Li, J., & Su, M.-H. (2020). Real talk about fake news: Identity language and disconnected networks of the us public's "fake news" discourse on Twitter. *Social Media + Society, 6*(2), 2056305120916841. https://doi.org/10.1177/2056305120916841

Liu, N., Shapira, P., & Yue, X. (2021). Tracking developments in artificial intelligence research: constructing and applying a new search strategy. *Scientometrics, 126*(4), 3153-3192. https://doi.org/10.1007/s11192-021-03868-4

Lohr, S. (2018, February 9, 2018). *Facial recognition is accurate, if you're a white guy*. The New York Times. https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html

Lomas, N. (2021, November 30). *Europe's AI Act falls far short on protecting fundamental rights, civil society groups warn*. TechCrunch. https://techcrunch.com/2021/11/30/eu-ai-act-civil-society-recommendations/

Mann, M., & Schleifer, C. (2019). Love the science, hate the scientists: Conservative identity protects belief in science and undermines trust in scientists. *Social Forces*. https://doi.org/10.1093/sf/soz156

McAfee, A. (2021, November 30). *EU proposals to regulate AI are only going to hinder innovation*. Financial Times. https://techcrunch.com/2021/11/30/eu-ai-act-civil-society-recommendations/

McCloskey, D. N. (1990). Formalism in the social sciences, rhetorically speaking. *The American Sociologist, 21*(1), 3-19. https://doi.org/10.1007/BF02691776

Mede, N. G., & Schäfer, M. S. (2020). Science-related populism: Conceptualizing populist demands toward science. *Public Understanding of Science*, 0963662520924259. https://doi.org/10.1177/0963662520924259

Mellers, B., Stone, E., Atanasov, P., Rohrbaugh, N., Metz, S. E., Ungar, L., Bishop, M. M., Horowitz, M., Merkle, E., & Tetlock, P. (2015). The psychology of intelligence analysis: Drivers of prediction accuracy in world politics. *Journal of Experimental Psychology-Applied, 21*(1), 1-14. https://doi.org/10.1037/xap0000040

Mertens, M. (2018). Liminal innovation practices: questioning three common assumptions in responsible innovation. *Journal of Responsible Innovation, 5*(3), 280-298. https://doi.org/10.1080/23299460.2018.1495031

Metag, J., & Schäfer, M. S. (2018). Audience segments in environmental and science communication: Recent findings and future perspectives. *Environmental Communication, 12*(8), 995-1004. https://doi.org/10.1080/17524032.2018.1521542

Metz, C. (2019, November 11, 2019). *We teach A.I. systems everything, including our biases*. The New York Times. https://www.nytimes.com/2019/11/11/technology/artificial-intelligence-bias.html

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

Mynatt, E., Watts, D., Bliss, N., Nelson, A., Pearson, W., & Rutenbar, R. (2020). *Harnessing the computational and social sciences to solve critical societal problems*. National Science Foundation.

Nanayakkara, P., Hullman, J., & Diakopoulos, N. (2021). Unpacking the expressed consequences of AI research in broader impact statements. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 795–806). Association for Computing Machinery. https://doi.org/10.1145/3461702.3462608

Natale, S., & Guzman, A. L. (2022). Reclaiming the human in machine cultures: Introduction. *Media, Culture & Society*, 01634437221099614. https://doi.org/10.1177/01634437221099614

National Academies of Sciences, Engineering, and Medicine. (2017). *The value of social, behavioral, and economic sciences to national priorities: A report for the National Science Foundation*. National Academies Press.

National Artificial Intelligence Initiative Act of 2020, (2020). https://www.congress.gov/bill/116th-congress/house-bill/6216

National Science Foundation. (2020). *NSF proposal processing and review*. https://www.nsf.gov/pubs/policydocs/pappg20_1/pappg_3.jsp

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology, 15*(3), 263-280. https://doi.org/https://doi.org/10.1002/ejsp.2420150303

Newby-Clark, I. R., McGregor, I., & Zanna, M. P. (2002). Thinking and caring about cognitive inconsistency: When and for whom does attitudinal ambivalence feel uncomfortable? *Journal of Personality and Social Psychology, 82*(2), 157-166. https://doi.org/10.1037/0022-3514.82.2.157

NSF-CISE-SBE Virtual Roundtable. (2020, May 19). *2020 NSF CISE-SBE virtual roundtable on harnessing the computational and social sciences to solve critical societal problems*. https://cps-vo.org/group/CISE-SBE-VR20

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science, 366*(6464), 447-453. https://doi.org/doi:10.1126/science.aax2342

Panchanathan, S. (2021). "Talent and ideas are democratized in the sense that they are everywhere.”. *Issues in Science and Technology, 37*(3), 25-28.

Pepitone, A. (1976). Toward a normative and comparative biocultural social psychology. *Journal of Personality and Social Psychology, 34*(4), 641-653. https://doi.org/10.1037/0022-3514.34.4.641

Pew Research Center. (2021, June 16). *Experts doubt ethical AI design will be broadly adopted as the norm within the next decade*. https://www.pewresearch.org/internet/2021/06/16/experts-doubt-ethical-ai-design-will-be-broadly-adopted-as-the-norm-within-the-next-decade/

Pew Research Center. (2022, March 17). *AI and human enhancement: Americans' openness is tempered by a range of concerns*. https://www.pewresearch.org/internet/2022/03/17/ai-human-enhancement-appendix/

Pidgeon, N. F., Poortinga, W., Rowe, G., Horlick-Jones, T., Walls, J., & O'Riordan, T. (2005). Using surveys in public participation processes for risk decision making: The case of the 2003 British GM nation? Public debate. *Risk Analysis, 25*(2), 467-479. https://doi.org/10.1111/j.1539-6924.2005.00603.x

Pielke, R. (2007). *The honest broker: Making sense of science in policy and politics*. Cambridge University Press.

Poortinga, W., & Pidgeon, N. F. (2006). Exploring the structure of attitudes toward genetically modified food. *Risk Analysis, 26*(6), 1707-1719. https://doi.org/10.1111/j.1539-6924.2006.00828.x

Porter, A. L., Youtie, J., Shapira, P., & Schoeneck, D. J. (2008). Refining search terms for nanotechnology. *Journal of Nanoparticle Research, 10*(5), 715-728. https://doi.org/10.1007/s11051-007-9266-y

Priester, J. R., & Petty, R. E. (2001). Extending the bases of subjective attitudinal ambivalence: Interpersonal and intrapersonal antecedents of evaluative tension. *Journal of Personality and Social Psychology, 80*(1), 19-34. https://doi.org/10.1037/0022-3514.80.1.19

Regan, Á. (2021). Exploring the readiness of publicly funded researchers to practice responsible research and innovation in digital agriculture. *Journal of Responsible Innovation, 8*(1), 28-47. https://doi.org/10.1080/23299460.2021.1904755

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences, 4*(2), 155-169. https://doi.org/10.1007/bf01405730

Rose, K. M., Howell, E. L., Scheufele, D. A., Brossard, D., & Xenos, M. A. (2018). The values of synthetic biology: Researcher views of their field and participation in public engagement. *Bioscience, 68*(10), 782-791. https://doi.org/10.1093/biosci/biy077

Rowe, G., & Wright, G. (2001). Expert opinions in forecasting: the role of the Delphi technique. In *Principles of forecasting* (pp. 125-144). Springer.

Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

Ryan, M. (2020). In AI we trust: Ethics, artificial intelligence, and reliability. *Science and Engineering Ethics, 26*(5), 2749-2767. https://doi.org/10.1007/s11948-020-00228-y

Salancik, G. R., & Pfeffer, J. (1978). A social information processing approach to job attitudes and task design. *Administrative Science Quarterly, 23*(2), 224-253. https://doi.org/10.2307/2392563

Sarewitz, D. (2015). CRISPR: Science can't solve it. *Nature, 522*(7557), 413-414. https://doi.org/10.1038/522413a

Saunders, J., Hunt, P., & Hollywood, J. S. (2016). Predictions put into practice: a quasi-experimental evaluation of Chicago's predictive policing pilot. *Journal of Experimental Criminology, 12*(3), 347-371. https://doi.org/10.1007/s11292-016-9272-0

Scheufele, D. A. (2006). Messages and heuristics: How audiences form attitudes about emerging technologies. In J. Turney (Ed.), *Engaging science: Thoughts, deeds, analysis and action* (pp. 20-25). The Wellcome Trust.

Scheufele, D. A. (2007). Opinion climates, spirals of silence and biotechnology: Public opinion as a heuristic for scientific decision-making. *Public, the Media and Agricultural Biotechnology*, 231-244. https://doi.org/10.1079/9781845932046.0000

Scheufele, D. A. (2011). *Modern citizenship or policy dead end? Evaluating the need for public participation in science policy making, and why public meetings may not be the answer*. Harvard University.

Scheufele, D. A. (2014). Science communication as political communication. *Proceedings of the National Academy of Sciences of the United States of America, 111*, 13585-13592. https://doi.org/10.1073/pnas.1317516111

Scheufele, D. A. (2018). Beyond the choir? The need to understand multiple publics for science. *Environmental Communication, 12*(8), 1123-1126. https://doi.org/10.1080/17524032.2018.1521543

Scheufele, D. A. (2022). Thirty years of science–society interfaces: What's next? *Public Understanding of Science, 31*(3), 297-304. https://doi.org/10.1177/09636625221075947

Scheufele, D. A., Corley, E. A., Dunwoody, S., Shih, T.-J., Hillback, E., & Guston, D. H. (2007). Scientists worry about some risks more than the public. *Nature Nanotechnology, 2*(12), 732-734. https://doi.org/10.1038/nnano.2007.392

Scheufele, D. A., Krause, N. M., Freiling, I., & Brossard, D. (2021). What we know about effective public engagement on CRISPR and beyond. *Proceedings of the National*

Academy of Sciences of the United States of America, 118(22), e2004835117. https://doi.org/10.1073/pnas.2004835117

Scheufele, D. A., & Lewenstein, B. V. (2005). The public and nanotechnology: How citizens make sense of emerging technologies. *Journal of Nanoparticle Research, 7*(6), 659-667. https://doi.org/10.1007/s11051-005-7526-2

Seising, R., & Sanz, V. (2012). From hard science and computing to soft science and computing – an introductory survey. In R. Seising & V. Sanz González (Eds.), *Soft Computing in Humanities and Social Sciences* (pp. 3-36). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-24672-2_1

Selbst, A. D. (2017). Disparate impact in big data policing. *Georgia Law Review, 52*(1), 109-195. https://www.georgialawreview.org/article/3373-disparate-impact-in-big-data-policing

Shaffer, L. S. (1983). Toward pepitone's vision of a normative social psychology: What is a social norm? *Journal of Mind and Behavior, 4*(2), 275-294.

Shaikh, S. J., & Moran, R. E. (2022). Recognize the bias? News media partisanship shapes the coverage of facial recognition technology in the United States. *New Media & Society*, 14614448221090916. https://doi.org/10.1177/14614448221090916

Shapira, P., Kwon, S., & Youtie, J. (2017). Tracking the emergence of synthetic biology. *Scientometrics, 112*(3), 1439-1469. https://doi.org/10.1007/s11192-017-2452-5

Shermer, M. (2007). The really hard science. *Scientific American, 297*(4), 44-46.

Shin, D. (2022). How do people judge the credibility of algorithmic sources? *AI & SOCIETY, 37*(1), 81-96. https://doi.org/10.1007/s00146-021-01158-4

Shoemaker, P. J., Vos, T. P., & Reese, S. D. (2009). Journalists as gatekeepers. In K. WahlJorgensen & T. Hanitzsch (Eds.), *Handbook of Journalism Studies* (pp. 73-87).

Simis, M. J., Madden, H., Cacciatore, M. A., & Yeo, S. K. (2016). The lure of rationality: Why does the deficit model persist in science communication? *Public Understanding of Science, 25*(4), 400-414. https://doi.org/10.1177/0963662516629749

Slater, M. D. (1996). Theory and method in health audience segmentation. *Journal of Health Communication, 1*(3), 267-284. https://doi.org/10.1080/108107396128059

Slovic, P. (1987). Perception of risk. *Science, 236*(4799), 280. https://doi.org/10.1126/science.3563507

Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., Hirschberg, J., Kalyanakrishnan, S., Kamar, E., Kraus, S., Leyton-Brown, K., Parkes, D., Press, W., Saxenian, A., Shah, J., Tambe, M., & Teller, A. (2016). *"Artificial intelligence and life in 2030." One hundred year study on artificial intelligence: Report of the 2015-2016 study panel*. http://ai100.stanford.edu/2016-report

Storer, N. W. (1967). The hard sciences and the soft: some sociological observations. *Bulletin of the Medical Library Association, 55*(1), 75-84. https://pubmed.ncbi.nlm.nih.gov/6016373

Su, L. Y. F., Cacciatore, M. A., Brossard, D., Corley, E. A., Scheufele, D. A., & Xenos, M. A. (2016). Attitudinal gaps: How experts and lay audiences form policy attitudes toward controversial science. *Science and Public Policy, 43*(2), 196-206. https://doi.org/10.1093/scipol/scv031

Tajfel, H. (1974). Social identity and intergroup behaviour. *Social Science Information, 13*(2), 65-93. https://doi.org/10.1177/053901847401300204

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behaviour. In S. Worchel & W. Austin (Eds.), *Psychology of inter-group relations* (2nd ed., pp. 7-24).

Tetlock, P. E. (2005). *Expert political judgment: How good is it? How can we know?* Princeton University Press.

Turner, J. C. (1999). Some current issues in research on social identity and self-categorization theories. In N. Ellemers, R. Spears, & B. Doosje (Eds.), *Social Identity: Context, Commitment, Content* (pp. 6-34). Blackwell.

United Nations. (2021, November 25). *193 countries adopt first-ever global agreement on the Ethics of Artificial Intelligence*. UN News. https://news.un.org/en/story/2021/11/1106612

von Schomberg, R. (2011). Prospects for technology assessment in a framework of responsible research and innovation. In M. Dusseldorp & R. Beecroft (Eds.), *Technikfolgen abschätzen lehren: Bildungspotenziale transdisziplinärer Methoden* (pp. 39-61). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-93468-6_2

von Schomberg, R. (2013). A vision of responsible research and innovation. In *Responsible Innovation* (pp. 51-74). https://doi.org/https://doi.org/10.1002/9781118551424.ch3

Wagner, C., Strohmaier, M., Olteanu, A., Kıcıman, E., Contractor, N., & Eliassi-Rad, T. (2021). Measuring algorithmically infused societies. *Nature*. https://doi.org/10.1038/s41586-021-03666-1

Weingart, P. (1998). Science and the media. *Research Policy, 27*(8), 869-879. https://doi.org/https://doi.org/10.1016/S0048-7333(98)00096-1

Wirz, C. D. (2021). *Risk perceptions for wicked issues: Toward more nuanced risk communication* [Doctoral dissertation, University of Wisconsin-Madison]. Madison, WI.

Wynne, B. (1992). Misunderstood misunderstanding: Social identities and public uptake of science. *Public Understanding of Science, 1*(3), 281-304.

Wynne, B. (2005). Risk as globalizing "democratic" discourse? Framing subjects and citizens. In M. Leach, I. Scoones, & B. Wynne (Eds.), *Science and citizens: Globalization and the challenge of engagement* (pp. 66-81). Zed Books London.

Yeo, S. K., & Brossard, D. (2017). The (changing) nature of scientist–media interactions: A cross-national analysis. In K. H. Jamieson, D. Kahan, & D. A. Scheufele (Eds.), *The Oxford Handbook of the Science of Science Communication* (pp. 261-272). Oxford University Press.

Yeo, S. K., Xenos, M. A., Brossard, D., & Scheufele, D. A. (2015). Selecting our own science: How communication contexts and individual traits shape information seeking. *Annals of the American Academy of Political and Social Science, 658*(1), 172-191. https://doi.org/10.1177/0002716214557782

Youtie, J., Carley, S., Shapira, P., Corley, E. A., & Scheufele, D. A. (2011). Perceptions and actions: relationships of views on risk with citation actions of nanotechnology scientists. *Research Evaluation, 20*(5), 377-388. https://doi.org/10.3152/095820211x13176484436014

Zajko, M. (2021). Conservative AI and social inequality: Conceptualizing alternatives to bias through social theory. *AI & SOCIETY, 36*(3), 1047-1056. https://doi.org/10.1007/s00146-021-01153-9

Zajko, M. (2022). Artificial intelligence, algorithms, and social inequality: Sociological contributions to contemporary debates. *Sociology Compass, 16*(3), e12962. https://doi.org/https://doi.org/10.1111/soc4.12962

Zhang, B., Anderljung, M., Kahn, L., Dreksler, N., Horowitz, M. C., & Dafoe, A. (2020, October 1). *Ethics and governance of artificial intelligence evidence from a survey of machine learning researchers*.

Zhang, D., Maslej, N., Brynjolfsson, E., Etchemendy, J., Lyons, T., Manyika, J., Ngo, H., Niebles, J. C., Sellitto, M., Sakhaee, E., Shoham, Y., Clark, J., & Perrault, R. (2022). *The AI Index 2022 Annual Report*. AI Index Steering Committee, Stanford Institute for Human-Centered AI, Stanford University. https://aiindex.stanford.edu/wp-content/uploads/2022/03/2022-AI-Index-Report_Master.pdf

ZIPPIA. (2022). *Computer scientist demographics and statistics in the US*. ZIPPIA. https://www.zippia.com/computer-scientist-jobs/demographics/

Appendices

Appendix A: Study D experimental conditions

On the next screen, you will see an article from 2021 introducing a special issue in the journal ***Nature Machine Intelligence*** on the potential social impacts of AI.

Please read it carefully, afterwards we will ask you a few questions about this topic.

nature
machine intelligence

Search Login

Content About Publish

Editorial | Published: 16 November 2021

We should collaborate with social scientists

Nature Machine Intelligence **3**, 925 (2021) | Cite this article

We in the AI research community must do more to address the downsides of our innovations, including the erosion of privacy, threats to democracy, and automation's effect on employment. Papers in this special issue raise these concerns and call for further evaluation of the potential negative effects of our work.

We need to engage more closely with social scientists and their literature. They have expertise in understanding the social impacts of diverse types of innovations and can improve our ability to assess the social impacts of AI.

In which journal was the article you just read published?

- Nature Machine Intelligence
- Nature Human Behaviour
- Nature Biotechnology
- Nature Climate Change
- Don’t know

On the next screen, you will see an article from 2021 introducing a special issue in the journal ***Nature Human Behaviour*** on the potential social impacts of AI.

Please read it carefully, afterwards we will ask you a few questions about this topic.

nature
humanbehaviour

Search Login

Content ⌄ About ⌄ Publish ⌄

Editorial | Published: 18 November 2021

AI scientists should collaborate with us

Nature Human Behaviour **3**, 925 (2021) | Cite this article

The AI research community must do more to address the downsides of their innovations, including the erosion of privacy, threats to democracy, and automation's effect on employment. Papers in this special issue raise these concerns and call for further evaluation of the potential negative effects of their work.

The AI research community needs to engage more closely with social scientists and our literature. We have expertise in understanding the social impacts of diverse types of innovations and can improve AI scientists' ability to better assess the social impacts of AI.

In which journal was the article you just read published?

- Nature Machine Intelligence
- Nature Human Behaviour
- Nature Biotechnology
- Nature Climate Change
- Don't know

On the next screen, you will see an article from 2021 in the journal ***Nature Biotechnology*** introducing gene-edited crops.

Please read it carefully, afterwards we will ask you a few questions.

nature biotechnology

Search Login

Content About Publish

Editorial | Published: 27 October 2021

Second-Generation Gene-Edited Crops

Nature Biotechnology **3**, 925 (2021) | Cite this article

Whereas human gene editing continues to garner public attention for its future promises and risks, agricultural applications are meanwhile rapidly emerging and slated to enter the market in the next few years.

In the United States, numerous field tests are under way, and companies and academic developers are switching from first-generation transgenic biotechnology approaches to gene editing. Dozens of gene-edited varieties have been produced, with hundreds more in research and development, including vegetable and specialty crops such as non-browning mushrooms, as well as commodity crops, such as herbicide-tolerant soybeans and corn.

In which journal was the article you just read published?

- Nature Machine Intelligence
- Nature Human Behaviour
- Nature Biotechnology
- Nature Climate Change
- Don't know

Appendix B: Survey questionnaire

Title of the Study: Scientists' perspectives on artificial intelligence and its potential social impacts

Principal Investigator: Dominique Brossard (phone: 608-263-0373) (email: dbrossard@wisc.edu)

Researcher: Luye Bao (phone: 608-262-1464) (email: lbao6@wisc.edu)

Description of the Research

You are invited to participate in a research study about scientists' perspectives on AI. You are part of a small number of scientists who have been asked to participate because you have published in an area related to AI.

The purpose of the research is to better understand how scientists understand AI and its potential social impacts. This research is conducted on the internet using a web survey. Survey data will be combined with publicly available data from Web of Science.

What Will My Participation Involve?

If you decide to participate in this research, you will be asked to complete an online questionnaire. Your participation will last approximately 15 minutes. You will be asked to complete only one survey.

Are There Any Risks to Me?

We don't anticipate any risks to you from participation in this study.

Are There Any Benefits to Me?

We don't expect any direct benefits to you from participation in this study.

How Will My Confidentiality be Protected?

While there will probably be publications as a result of this study, your name will not be used. We will not directly quote any comments you make in the survey. Only group characteristics will be published.

Whom should I contact if I have questions?

You may ask any questions about the research at any time. If you have questions about the research after you finish today you should contact the Principal Investigator Dominique Brossard at dbrossard@wisc.edu, or at 608-263-0373. You may also contact the researcher Luye Bao at lbao6@wisc.edu or 608-262-1464.

If you are not satisfied with response of research team, have more questions, or want to talk with someone about your rights as a research participant, you should contact the Education Research and Social & Behavioral Science IRB Office at 608--263--4312.

Your participation is completely voluntary. You have the right to withdraw from the study at any time.

Selecting 'Yes' below indicates that you have read this consent form, had an opportunity to ask any questions about your participation in this research, and voluntarily consent to participate.

- Yes, I agree to participate in this study and have read the consent form above.
- No, I do not want to participate in this study.

For the purpose of this survey, we would like to clarify what we mean when we mention the following concepts in the survey.

Scientists: scientists are from fields that include computer sciences, life sciences, physical sciences, and social sciences.

Computer sciences: the sciences concerned with the scientific study of computers and computational systems.

Life sciences: the sciences concerned with the scientific study of living organisms.

Physical sciences: the sciences concerned with the scientific study inanimate natural objects.

Social sciences: the sciences concerned with scientific study of human society and social relationships.

When these terms appear throughout the survey, they will be highlighted in blue with a mouse over definition for your reference.

---

1. To begin, please confirm which of the following titles best describe your current position. As we mentioned earlier, for the purpose of this survey, scientists are from fields that include computer sciences and engineering, life sciences, physical sciences, and social sciences.
    - o University scientist
    - o Industry scientist
    - o Government scientist
    - o Graduate student
    - o Other, please specify ______

2. What field are you in? Please check all that apply.
    - ☐ Agricultural and food sciences
    - ☐ Arts and humanities
    - ☐ Biological sciences
    - ☐ Computer and information science
    - ☐ Education and human resources
    - ☐ Engineering
    - ☐ Environmental research and education
    - ☐ Geosciences
    - ☐ Mathematical and physical sciences
    - ☐ Medical sciences
    - ☐ Social, behavioral, and economic sciences
    - ☐ Other, please specify ___________

3. What area of artificial intelligence (AI) do you work in? Please check all that apply.
    - ☐ Machine learning
    - ☐ Natural language processing

- ☐ Computer vision
- ☐ Automated reasoning
- ☐ Knowledge representation
- ☐ Robotics
- ☐ Ethics and societal impacts of AI
- ☐ Other AI-related subfields, please specify_____
- ☐ My work is not related to AI.

First, we are interested in understanding the topics you pay attention to in the news.

4. In general, how much **attention** do you pay to news stories about the following topics?

None	Very little	Some	Quite a bit	A lot
1	2	3	4	5

A. Science and technology
B. Ethical implications of emerging technologies
C. Regulations of emerging technologies
D. International affairs
E. National government and politics

Next, we would like to ask you some questions about your research and publication practice.

5. Please think about the AI-related journal articles, policy documents, and reports that you have *read* over the last three years.

With that in mind, how often would you say you ***read*** about the following topics?

Never	Rarely	Sometimes	Often	Very often	Not relevant to my work
1	2	3	4	5	

A. Societal implications of AI
B. Ethics and responsible development of AI
C. Commercialization and market transfer of AI
D. Technical aspects of AI

6. Please think about the AI-related journal articles, policy documents, and reports that you have *written* over the last three years.

With that in mind, how often would you say you ***cite sources*** that discuss the following topics?

Never	Rarely	Sometimes	Often	Very often	Not relevant to my work
1	2	3	4	5	

A. Societal implications of AI
B. Ethics and responsible development of AI
C. Commercialization and market transfer of AI
D. Technical aspects of AI

7. Please think about the AI-related journal articles, policy documents, and reports that you have *written* over the last three years.

With that in mind, how often would you say you ***collaborate and/or coauthor*** with researchers in the following areas?

Never	Rarely	Sometimes	Often	Very often	Not relevant to my work
1	2	3	4	5	

A. Life sciences
B. Physical sciences
C. Computer sciences
D. Social sciences

Now thinking about your communication practices if any.

8. In the past three years, have you done any of the following to communicate about AI? Check all that apply.

☐ Participated in interviews with journalists
☐ Interacted with government bodies or officials
☐ Collaborated with industry or professional stakeholders on research
☐ Presented research to the public
☐ Included members of the public directly in the research process
☐ Other, please specify_____
☐ None of the above

[Follow-up questions]

- Thinking about your **interviews with journalists,** how often would you say you **mention sources** that discuss the following topics?
- Thinking about your **interactions with government bodies or officials**, how often would you say you **mention sources** that discuss the following topics?
- Thinking about your **collaboration with industry or professional stakeholders,** how often would you say you **mention sources** that discuss the following topics?
- Thinking about when you **present research to the public,** how often would you say you **mention sources** that discuss the following topics?
- Thinking about when you **include members of the public directly in the research process,** how often would you say you **mention sources** that discuss the following topics?

Never	Rarely	Sometimes	Often	Very often	Not relevant to my work
1	2	3	4	5	

A. Societal implications of AI
B. Ethics and responsible development of AI
C. Commercialization and market transfer of AI
D. Technical aspects of AI

---

9. Below are some statements people have made about computer sciences and social sciences. How much do you agree or disagree with the following statements?

Strongly Disagree	Disagree	Neither disagree nor agree	Agree	Strongly Agree
1	2	3	4	5

A. Most results in the social sciences tend to be common sense.
B. Social science research is as rigorous as research in other fields of science
C. Computer science research does not focus enough on abstract and theoretical questions.
D. Computer sciences research can greatly extend social theories by investigating social data that are co-shaped by algorithms and human behavior.

10. Now thinking about the relationship between scientists (people like you) and the public, how much do you agree or disagree with the following statements?

Strongly Disagree	Disagree	Neither disagree nor agree	Agree	Strongly Agree
1	2	3	4	5

A. Scientists know best what is good for the public.
B. Scientists should do what they think is best, even if they have to persuade people that it is right.
C. Scientists should be able to conduct their research without consulting the public.
D. Science is the best way that society has of producing reliable knowledge.
E. Science is the best way to understand the world.
F. Scientists should pay attention to the wishes of the public, even if they think citizens are mistaken or do not understand their work.
G. Public opinion is more important than the scientists' opinions when making decisions about the ethical implications of scientific research.
H. It is appropriate for scientists to become actively involved in political debates about issues like AI.

11. Please indicate the extent to which you disagree or agree with each of the following statements about science, scientists, and their scientific conduct.

Strongly Disagree	Disagree	Neither disagree nor agree	Agree	Strongly Agree
1	2	3	4	5

A. Scientists are responsible for the way their discoveries are used by other people.
B. A discovery is in itself neither good nor bad, it is only the way the discovery is used that matters.
C. The authorities should formally require scientists to respect ethical standards.
D. Scientists should be free to carry out the research they wish, provided they respect ethical standards.
E. It is useful to classify science into hard and soft sciences.

Now, we would like to learn more about your thoughts on AI.

12. How likely do you think it is that AI will. . .?

Not at all likely	Very unlikely	Unlikely	Somewhat likely	Likely	Very likely	Certain
1	2	3	4	5	6	7

A. strengthen the U.S. economy
B. increase national security
C. improve individuals' health
D. reduce bias in human decision-making
E. help fight terrorism threats
F. worsen societal inequalities

G. give some people too much power
H. threaten personal liberties
I. change what it means to be human
J. displace workers by automating their jobs

13. How concerned are you about AI worsening discrimination against people based on. . .?

Not at all concerned	Slightly concerned	Moderately concerned	Concerned	Very concerned
1	2	3	4	5

A. gender
B. ethnicity or race
C. religion
D. sexual orientation
E. income or social class
F. health risk
G. age

Next, we are going to ask you some questions about AI-related regulations.

14. How much do you agree or disagree with the following statements?

Strongly Disagree	Disagree	Neither disagree nor agree	Agree	Strongly Agree
1	2	3	4	5

A. Advancing AI quickly is more important than protecting society from the unknown risks of AI.
B. Regulating AI may significantly slow down important scientific and innovation progress.
C. Existing regulations for AI **research** are sufficient.
D. Existing regulations for AI **applications** are sufficient.
E. As a society, we are prepared for the potential effects of AI applications.
F. There will be unintended consequences of AI applications

15. How much do you agree or disagree that the U.S. federal government should regulate the following AI applications?

The U.S. federal government should regulate …

Strongly Disagree	Disagree	Neither disagree nor agree	Agree	Strongly Agree
1	2	3	4	5

A. chatbots
B. deepfakes
C. social scoring
D. surveillance technology
E. smart personal assistants
F. autonomous vehicles
G. lethal autonomous weapons
H. personalized medicine
I. data-driven policing
J. brain organoid technology

16. If there are new regulations for AI research and use, which of the following groups should have a say in the development of these regulations in the United States? Please check all that apply.

- ☐ University scientists
- ☐ Industry scientists
- ☐ Congress
- ☐ The U.S. court system
- ☐ Regulatory agencies that oversee AI applications (e.g., the Department of Transportation or the Department of Defense)
- ☐ The White House
- ☐ Law enforcement agencies (e.g., FBI)
- ☐ Digital rights groups (e.g., the Electronic Frontier Foundation)
- ☐ Citizens through public engagement mechanisms, such as citizen forums
- ☐ End users
- ☐ Large technological companies, such as Google and Microsoft
- ☐ AI-related non-profit organizations, such as Partnership on AI
- ☐ Intergovernmental bodies or agreements
- ☐ Other, please specify_________________

[Survey experiment vignettes (See Appendix A)]

We have a few more questions about human-level AI.

17. For the purpose of the survey, human-level AI is defined as a machine being able to learn to do *anything* a human can do.

Do you think it is possible to predict when human-level AI will be achieved?
- o Yes, it is possible to predict
- o No, it is not possible to predict
- o No, human-level AI will never be achieved.
- o Don’t know

18. If you had to guess, by which year will human-level AI be achieved? Please move the slider to indicate the year.

	2025 2060 2095 2130 2165 2200
Year	
	☐ Beyond 2200

19. How confident are you about your prediction of when human-level AI will be achieved?

Not at all confident	Slightly confident	Moderately confident	Confident	Extremely confident
1	2	3	4	5

20. How likely do you think it is that **you** will incorporate insights from social sciences into the following aspects of your own work on AI?

Very unlikely	Unlikely	Neither unlikely nor likely	Likely	Very likely	This is not an aspect of my work
1	2	3	4	5	

A. Conceptual design: Defining research objectives and proposing research questions
B. Technical design: Developing model constructs, instruments, and attributes
C. Execution and evaluation: Assessing the impacts of your research and applications

21. How likely do you think it is that **most AI scientists** will incorporate insights from social sciences into the following aspects of their work on AI?

Very unlikely	Unlikely	Neither unlikely nor likely	Likely	Very likely	This is not an aspect of my work
1	2	3	4	5	

A. Conceptual design: Defining research objectives and proposing research questions
B. Technical design: Developing model constructs, instruments, and attributes
C. Execution and evaluation: Assessing the impacts of your research and applications

We have just a few questions left! In the last section, please tell us more about yourself, such as your demographics, professional biography, values, and reasoning style.

22. What is your gender?
    - o Male
    - o Female
    - o Non-binary
    - o Not listed, please tell us_____

23. In what year were you born?

24. Check all of the following categories that describe your race.
    - ☐ American Indian or Alaskan Native
    - ☐ Asian
    - ☐ Black or African American
    - ☐ Hispanic or Latino
    - ☐ Native Hawaiian or Other Pacific Islander
    - ☐ White
    - ☐ Other, please specify___________

25. What is your highest level of education?
    - o 4-year degree (or equivalent)
    - o Master's (or equivalent)
    - o Professional degree
    - o Doctoral degree (e.g., Ph.D.)
    - o Medical degree (e.g., M.D., D.D.S., or D.V.M.)

26. In what year did you complete your highest degree?

27. If you are currently supported by grants or contracts, whether as principal investigator (PI), co-PI or affiliated researcher, please indicate the source of this support. Please check all that apply.

    - ☐ NSF
    - ☐ NIH
    - ☐ DOE
    - ☐ Private sector, please specify ________
    - ☐ Other, please specify ________
    - ☐ I am not a PI, co-PI, or affiliated researcher.

28. How important is religion in your life?
    - o Not at all important
    - o Not too important
    - o Somewhat important
    - o Very important

29. The terms “liberal” and “conservative” may mean different things to people, depending on the kind of issue one is considering. Many people’s views do not fit perfectly into one of the categories below, so please indicate which one you think could best align with your views.

Very liberal	Liberal	Moderate	Conservative	Very conservative
1	2	3	4	5

A. In terms of **economic issues**, would you say you are …
B. In terms of **social issues**, would you say you are …

30. The last two questions are about your reasoning style. How much do you agree or disagree with the following statement about your reasoning style?

Strongly Disagree	Disagree	Neither disagree nor agree	Agree	Strongly Agree
1	2	3	4	5

A. Even after making up my mind, I am always eager to consider a different opinion.
B. I dislike questions that can be answered in many ways.
C. When considering most conflicts, I can usually see how both sides could be right.

31. In a famous essay, Isaiah Berlin classified thinkers as hedgehogs and foxes: The hedgehog knows one big thing and tries to explain as much as possible using that theory or framework. The fox knows many small things and is content to improvise explanations on a case- by-case basis. When it comes to making predictions, would you describe yourself as more of a hedgehog than of a fox?

Very much hedgehog-like	Somewhat hedgehog-like	Neither hedgehog-like nor fox-like	Somewhat fox-like	Very much fox-like
1	2	3	4	5

32. Finally, there are discussions and disagreements on the definition of AI. How do you define AI in your own words?

Please tell us if you have any comments about this study. If you have no comment, please leave it blank or type “no comment” in the space provided.

[Condition: NMN & NHB]
Before you conclude your participation, it is important for you to know that the article you read was excerpted and adapted from “It’s time to do something: Mitigating the negative impacts of computing through a change to the peer review process,” published in the ACM Future of Computing Blog on March 29, 2018 (link).

In the present study, respondents to the survey were assigned to read either an article about AI or an article about gene-edited crops that was attributed to different sources depending on the experimental condition. The goal of the experiment was to find out how views regarding AI and the conduct of AI research differ depending on the information source and content.

If you are interested in the results of this study, please provide your email address:

[Condition: NB]
Before you conclude your participation, it is important for you to know that the article you read was excerpted and adapted from “Regulating gene-edited crops” published in Issues in Science and Technology in Fall, 2018 (link).

In the present study, respondents to the survey were assigned to read either an article about AI or an article about gene-edited crops that was attributed to different sources depending on the experimental condition. The goal of the experiment was to find out how views regarding AI and the conduct of AI research differ depending on the information source and content.

If you are interested in the results of this study, please provide your email address:

[Condition: No stimuli]
If you are interested in the results of this study, please provide your email address:

We appreciate your time and consideration in completing the survey; thank you for participating in the study. It is only through the help of scientists like you that we can understand the potential social impacts of AI.