UNIVERSITY OF OKLAHOMA
GRADUATE COLLEGE

ARTIFICIAL INTELLIGENCE AVERSION IN PUBLIC POLICY

A DISSERTATION
SUBMITTED TO THE GRADUATE FACULTY
in partial fulfillment of the requirements for the
Degree of
DOCTOR OF PHILOSOPHY

By
WILLIAM LEWERS LIVINGSTON
Norman, Oklahoma
2024

ARTIFICIAL INTELLIGENCE AVERSION IN PUBLIC POLICY

A DISSERTATION APPROVED FOR THE

DEPARTMENT OF POLITICAL SCIENCE

BY THE COMMITTEE CONSISTING OF

Dr. Joseph T. Ripberger, Chair

Dr. Hank C. Jenkins-Smith

Dr. Deven E. Carlson

Dr. Amy McGovern

Dedication

For my mother and father, Vanessa and Gilbert Livingston who have always encourage me to think and speak with honesty, curiosity, and truth. And also, for my loving husband Shane Webb who has supported me and cheered me on throughout this journey.

Acknowledgements

I am very lucky to have so many incredible people to thank for their role in getting this dissertation written. The academic support that I received as an undergraduate at Northern Arizona University sparked my love for the study of Political Science. Furthermore, when I attended Pepperdine University I learned invaluable lessons both on the importance of Public Policy as well as on how to thoroughly analyze it to best benefit society. I am eternally grateful to the staff and faculty at both institutions for their commitment and encouragement for me to further my education. There are two particular individuals I would like to recognize from that time, Dr. Ted McAllister and Dr. Robert Kaufman who both motivated me and challenged my ways of thinking, and ultimately inspired me to further pursue a PhD. I also would not have been able to achieve this goal without the support of the people at the University of Oklahoma who refined and sharpened my thinking as I pursued my PhD. I would like to especially thank my committee members for their support, encouragement, and insight as I wrote this dissertation. I would also like to acknowledge everyone at the Institute for Public Policy Research and Analysis whose care, advice, and incredible lunch-time conversation kept me motivated and moving forward in this dissertation. I would like to specifically thank Joe Ripberger for his incredible support throughout this process and for being a role model of the kind of researcher I can strive to be. I want to thank my incredible parents Vanessa and Gilbert Livingston who have insured I never went a day not knowing how loved and cared for I was. I would also like to thank my incredible extended family in Hawaii who inspire me to carry forward the aloha spirit in everything that I do. Finally, I want to thank my husband Shane Webb whose support, companionship, love, and humor have been invaluable in getting me here today. To all of you I sincerely say, Mahalo nui loa.

Table of Contents

List of Tables

List of Figures

Abstract

This dissertation advances our understanding of how individuals relate to the utilization of artificial intelligence within the public policy domain. This is done through theoretical development about the aversion individuals have towards artificial intelligence as well as through empirical examination of that artificial intelligence aversion and how it manifests. In this dissertation I identify that artificial intelligence exists as a unique concept to individuals in contrast to algorithms. I then develop an index to measure people's levels of aversion to artificial intelligence. I validate that index by assessing how well it does at predicting support for current and future uses of artificial intelligence.

Once my index is validated I then turn toward trying to understand what variables are contributing towards people's different levels of aversion. I first examine the role that perceptions about risk and subjectivity of the area the artificial intelligence is being used in has on people's aversion index scores. I then examine how demographics influence both perceptions as well as people's levels of aversion. In these examinations I find that: perceived risk and perceived subjectivity contribute in part to people's levels of aversion, with perceived risk having a larger effect; and that demographics play a key role in people's perceptions about the utilization of artificial intelligence. Demographics help to understand how personal levels of aversion to artificial intelligence differ and are identified as an important area of focus if policy makers want to reduce artificial intelligence aversion. This research paves the way for future examination into how aversion changes over time as artificial intelligence is increasingly utilized in people's lives.

Chapter 1. Artificial Intelligence and Public Policy

1.1 Introduction

The advent of artificial intelligence (AI) in the realm of public policy represents a pivotal shift in the landscape of governance and societal management. This dissertation explores the integration of algorithms and the emerging role of AI, reflecting humanity's enduring quest to transcend the limitations imposed by scarcity. Scarcity, a fundamental economic concept, is rooted in the imbalance between our desires and the ease of fulfilling them. The advent of AI and algorithms marks a significant transformation, promising to alleviate this imbalance by reducing the costs associated with both prediction and decision-making.

AI in public policy is not just about technological advancement; it is about how these technologies are received, interpreted, and ultimately utilized by the public. Understanding public response to AI's integration into policy-making is crucial. The technology's potential for societal shifts and efficiency gains is immense, but its acceptance and effectiveness hinge on public perception and trust. This dissertation delves into this intricate relationship, examining how public attitudes towards AI influence and are influenced by policy initiatives.

The scope of this research encompasses a comprehensive analysis of AI's role in public policy, with a specific focus on the AI Aversion Index (AIAI) developed in chapter two. This index serves as a novel tool for measuring public resistance or acceptance towards AI within policy-making contexts. Additionally, the dissertation investigates how individuals perceptions and various demographic variables contribute towards AI aversion, as explored in the chapters three and four respectively. These factors are pivotal in shaping public opinion and, consequently, the successful implementation of AI-driven policies.

In examining these aspects, the dissertation draws upon practical examples and theoretical frameworks. For instance, the application of the PATTERN algorithm in the U.S. criminal justice system, shows both the potential and pitfalls of AI in policy (Partnership on AI, 2020). PATTERN was an AI designed to automate the identification of people eligible for parole within the criminal justice system., and while it greatly enhanced the speed at which individuals were identified, it overestimated the likelihood for minority recidivism. While AI can significantly enhance efficiency and decision-making, it also raises concerns about biases and errors inherent in algorithmic predictions. The overestimation of recidivism rates among minorities by the PATTERN algorithm illustrates the critical need for a nuanced understanding of AI's role in public policy.

Ultimately, this dissertation aims to bridge the gap between the technological promise of AI and the realities of its implementation in public policy. By exploring the multifaceted responses of the public toward AI, it seeks to contribute to a more informed and effective integration of these technologies into the fabric of society. The goal is not only to understand AI's current impact but also to anticipate and shape its future role in public governance.

1.2 Previous Literature

1.2.1 AI and Algorithmic Aversion in Public Policy

The emergence of artificial intelligence in public policy marks a significant paradigm shift, transcending traditional algorithmic decision-making processes. This dissertation, at its core, seeks to elucidate the nuances of AI aversion, a phenomenon distinct yet related to algorithmic aversion, within the public policy domain. The intricate relationship between AI, algorithms, and public policy necessitates a thorough exploration, as the success of AI implementation in governance hinges on public acceptance and trust.

The Distinction between AI and Algorithms in Public Perception

At the outset, it is imperative to establish a clear distinction between AI and algorithms in the public consciousness. Past research on algorithmic aversion has been pivotal in understanding resistance to technology-based decision-making. However, AI aversion in public policy extends beyond these initial forays, involving more complex and nuanced public reactions (Davis, 1989; Venkatesh et al. 2003). Algorithms, fundamentally, are sets of rules leading to consistent outcomes, while AI encompasses learning capabilities and autonomous decision-making, features that evoke different public responses. This dissertation employs affective imagery measurements to dissect the public's conceptual distinction between AI and algorithms, a crucial step in understanding AI aversion (Szalay and Deese 1978).

Evolution of Theoretical Frameworks in Technology Acceptance

The exploration of AI aversion in public policy is grounded in foundational theoretical frameworks that have shaped our understanding of technology acceptance. Central to this discourse is Davis's Technology Acceptance Model (TAM), which posits that perceived usefulness and ease of use are the primary determinants of technology adoption. While TAM offers insights into general technology acceptance, its applicability to the advanced, autonomous nature of AI systems is limited (Scharre, 2018). Similarly, the Unified Theory of Acceptance and Use of Technology (UTAUT) integrates key elements from various theories to predict user intentions and behaviors regarding information systems. UTAUT's constructs of performance expectancy, effort expectancy, social influence, and facilitating conditions have been validated across various technologic contexts. Yet, the model requires adaptation to accommodate the unique attributes of AI, particularly in public policy settings, where AI's autonomous decision-making capabilities evoke distinct cognitive and emotional responses (Dietvorst, Simmons, & Massey 2016; Jussupow, Benbasat, & Heinzl 2020).

Differentiating AI Aversion from Algorithmic Aversion

AI aversion, while sharing similarities with algorithmic aversion, demands a distinct analytical approach. Previous methodologies focusing on algorithmic aversion through vignettes and controlled experiments have illuminated aspects of aversion in specific contexts. However, these approaches fall short in capturing the broader, more complex landscape of AI aversion. The intricacies inherent in AI, particularly in public policy domains, necessitates a refined examination, extending beyond the scope of traditional algorithmic evaluation (Dietvorst, Simmons, & Massey 2016). This dissertation pivots from these foundational methodologies, aiming to establish a more comprehensive understanding of AI aversion. It acknowledges that trust in AI involves not just an assessment of objective algorithmic performance but also subjective interpretations of AI's role in decision-making processes.

1.2.2 Factors Influencing Public Perception and Aversion to AI

Understanding the factors that influence public perception and aversion to artificial intelligence is crucial for its successful integration into public policy. This segment of the dissertation delves into the primary determinants of AI aversion: perceived risk and subjectivity, drawing upon extensive literature on algorithmic aversion to illuminate potential parallels and distinctions in public responses to AI.

Perceived Risk and AI Aversion

Perceived risk has been shown to play a pivotal role in shaping aversion towards algorithms in the past. Applying this aspect to AI aversion is conceptually tied to the construct of performance expectancy within the UTAUT, as elaborated by Venkatesh et al. However, the complexity of AI systems introduces an amplified dimension of risk that extends beyond traditional performance expectancy (Venkatesh et al. 2003). Slovic's seminal work on risk perception provides a crucial theoretical underpinning for this investigation, suggesting that risk perception involves not just the likelihood of negative outcomes but also public trust and perceived loss of control (Slovic, 1987). In the context of AI, this translates to concerns over unpredictability and potential errors in decision-making processes, particularly in high-stakes environments where AI systems operate.

Research indicates that the multifaceted concept of risk, encompassing potential errors, biases, and far-reaching consequences of AI decisions, significantly influences public trust in AI. This influence is seen within various domains, where the stakes, perceived risks, and the ethical considerations can vary substantially, impacting individuals' trust and acceptance of AI technologies (Filiz 2023).

Subjectivity in AI Tasks and Public Trust

The perceived subjectivity or objectivity of tasks assigned to AI significantly influences public trust in these systems. Subjective tasks, which involve elements challenging to quantify or influenced by human judgment, emotions, or opinions, are typically less trusted when performed by AI. In contrast, objective tasks, which are data-driven and can be assessed more quantitatively, garner comparatively more trust when performed by AI (Castelo, Bos, & Lehmann, 2019; Yeomans et al., 2019). Research by Castelo, Bos, and Lehmann found that distrust in AI is higher for tasks perceived as more subjective, such as those requiring creativity or personal interpretation. Conversely, AI systems are more trusted for objective tasks like data analysis. This distinction highlights how the nature of the task impacts public perceptions of AI's reliability.

Yeomans et al.'s research further underscores this point, showing that even when AI is known to outperform humans in subjective tasks, there is still a preference penalty. This aversion persists in subjective domains, demonstrating the public's reluctance to rely on AI for tasks involving personal preferences or judgments, even when AI's performance has been identified as superior (Yeomans et al., 2019).

Comparative Analysis of AI and Algorithmic Aversion

This dissertation extends the exploration of AI aversion by comparing it with previous findings on algorithmic aversion. The identified narratives in research on algorithmic aversion, namely perceived risk and subjectivity, serve as a foundation for understanding aversion towards AI in public policy (Purves, Jenkins, & Strawser 2015). Chapters two and three assesses whether the factors driving algorithmic aversion carry over to AI aversion and which of these - risk or subjectivity - is a more significant driver of aversion in the public policy domain. This comparative analysis is vital for identifying areas in public policy where AI integration might face resistance or acceptance, enabling policymakers to strategize effectively for higher approval and adoption rates.

1.2.3 The Need for a Comprehensive Approach to Measuring AI Aversion

In the evolving landscape of artificial intelligence in public policy, understanding and measuring AI aversion requires a comprehensive approach that transcends traditional methodologies. This next section of the dissertation underscores the necessity of developing a novel measure of AI aversion that is both robust and nuanced, considering the multifaceted and domain-specific nature of this phenomenon. This novel measure will be an index for identifying people’s different levels of aversion and allowing for direct comparisons between individuals as well as reproductivity in future research.

Developing the Artificial Intelligence Aversion Index

The development of the AIAI marks a significant advancement in the study of AI aversion. Prior research on algorithmic aversion, while insightful, often focused on single domains, limiting the broader applicability of their findings (Dietvorst, Simmons, & Massey 2016). The AIAI is designed to overcome this limitation by incorporating a range of factors across various policy domains, thereby enhancing both the internal and external validity of AI aversion measurements. This index allows for the assessment of AI aversion in a diverse array of policy contexts, reflecting the diverse values and concerns that people hold regarding different policy areas.

Crafting an index such as the AIAI necessitates a meticulous approach that accounts for the complexity and diversity of public opinions on AI. The index is designed to gauge AI aversion in a way that considers not only the direct responses to AI technologies but also the subtler, more subjective aspects of public sentiment. This approach provides a more holistic understanding of AI aversion, capturing nuances that might have been overlooked in narrower evaluations of the phenomena.

Assessing Demographic Influences on AI Aversion

A comprehensive approach to measuring AI aversion also involves considering the demographic variables that influence public perception. Research has shown that factors such as gender, age, race, political affiliation, and education level significantly impact how individuals perceive and interact with AI technologies (Rogers, 2009; Leonardi, 2012). For instance, gender differences in technology perception and acceptance have been observed, with women generally expressing more concerns about the ethical and social implications of AI, potentially leading to higher levels of aversion (Morris & Venkatesh, 2000). Similarly, generational differences play a critical role, with younger generations typically more comfortable with AI integration compared to older generations who may exhibit higher levels of aversion due to unfamiliarity and concerns about the pace of technological change (Williams, Anderson, & Drennan, 2010).

Racial and cultural factors also play a significant role in AI perception. Studies have indicated that experiences and societal contexts shape how different racial groups perceive and interact with AI, with minority groups often expressing concerns about biases and fairness in AI systems (Finucane et al. 2000). Political ideology further influences perceptions, with conservatives and liberals displaying differing attitudes towards AI based on their ideological beliefs (Smith & Anderson, 2019). Moreover, education level has been found to correlate with AI aversion, with higher educational attainment generally associated with lower perceived risk and therefore lower aversion towards AI (Bucchi & Neresini, 2008).

Cultural Theory and Its Role in Understanding AI Aversion

Cultural Theory offers a unique lens through which to examine AI aversion (Wildavsky 1987). The theory categorizes individuals based on their relation to society (group) and their view of societal structure (grid), proposing four distinct worldviews: Hierarchical, Egalitarian, Fatalist, and Individualist. Each group holds different perceptions of technology and risk, influencing their attitudes towards AI. For example, Hierarchists might view AI as a tool for maintaining structure, leading to less aversion, while Egalitarians, concerned with equality, could fear AI exacerbating social disparities.

1.3 Overview of the Dissertation

1.3.1 Structure and Content

The structure of this dissertation is designed to provide a comprehensive exploration of the public's perception and aversion to artificial intelligence in public policy. It consists of three main chapter, each contributing uniquely to the overarching theme and objectives of the dissertation.

Chapter 2: Establishing AI Aversion and Its Measurement

Chapter two lays the foundation for the dissertation by establishing the concept of AI aversion in the public policy domain. It differentiates AI aversion from algorithmic aversion, underscoring the necessity to treat AI as a distinct entity in public perception. This chapter also includes the development of the AIAI, a pioneering tool for measuring public aversion to AI across various policy contexts. This index is crucial for understanding the diverse and nuanced responses to AI, offering a robust method for quantifying aversion levels between different individuals. This chapters also tests the robustness of the AIAI by measuring how well it does at predicting current and future support for AI uses in public policy. This chapter contributes to the dissertation by setting the stage for a deeper exploration of AI aversion, providing a measurable framework to assess and analyze public sentiments towards AI in policy-making.

Chapter 3: Unraveling the Intricacies of AI Aversion

The third chapter of this dissertation delves into the intricacies of AI aversion, focusing on understanding its underlying causes. It compares AI aversion to algorithmic aversion, examining whether the determinants of algorithmic aversion, such as perceived risk and subjectivity, apply to AI aversion. This exploration is crucial in identifying specific factors that contribute to and shape public resistance towards AI integration in various policy frameworks. By utilizing the AIAI and survey data, this chapter assesses how perceived risks and the subjective nature of policy domains influence public aversion to AI. The insights from this chapter are essential for policymakers and AI developers, as is provides a nuanced understanding of the factors driving public aversion, enabling for more informed strategies for integration of AI into public policy.

Chapter 4: Demographic Influences on AI Aversion

Chapter four expands the scope of the research to include demographic factors and their influence on AI aversion. It explores how characteristics such as gender, age, race, political affiliation, and education level impact public perceptions and aversion to AI. This granular analysis provides a comprehensive understanding of the diverse attitudes towards AI, highlighting the complex interplay between demographic factors and perceptions of risk and subjectivity. This chapter contributes significantly to the dissertation by offering a detailed understanding of how different segments of the population respond to AI in public policy. This is critical for tailoring policy initiatives and communication strategies to diverse audiences who will have to interact with these new technologies.

Chapter 5: Conclusions, Implications, and Directions for Future Research

This dissertation concludes with a chapter that looks at how the results of the previous chapters relate to one another and what they mean for the understanding of AI aversion. This chapter will also provide an updated model of the relationship between the different variables that are utilized throughout this dissertation. Finally, this chapter will conclude on considerations of what some of the next steps for studying AI aversion should look like, identifying different potential paths for future research.

1.3.2 Conceptual Framework, Model and Methodology

The conceptual framework of this dissertation revolves around a comprehensive understanding of public aversion to artificial intelligence in public policy. Central to this framework are two key components: the AIAI and the modeled relationship between these different variables. These elements collectively provide a structured approach to analyzing how various factors contribute to AI aversion.

The Artificial Intelligence Aversion Index

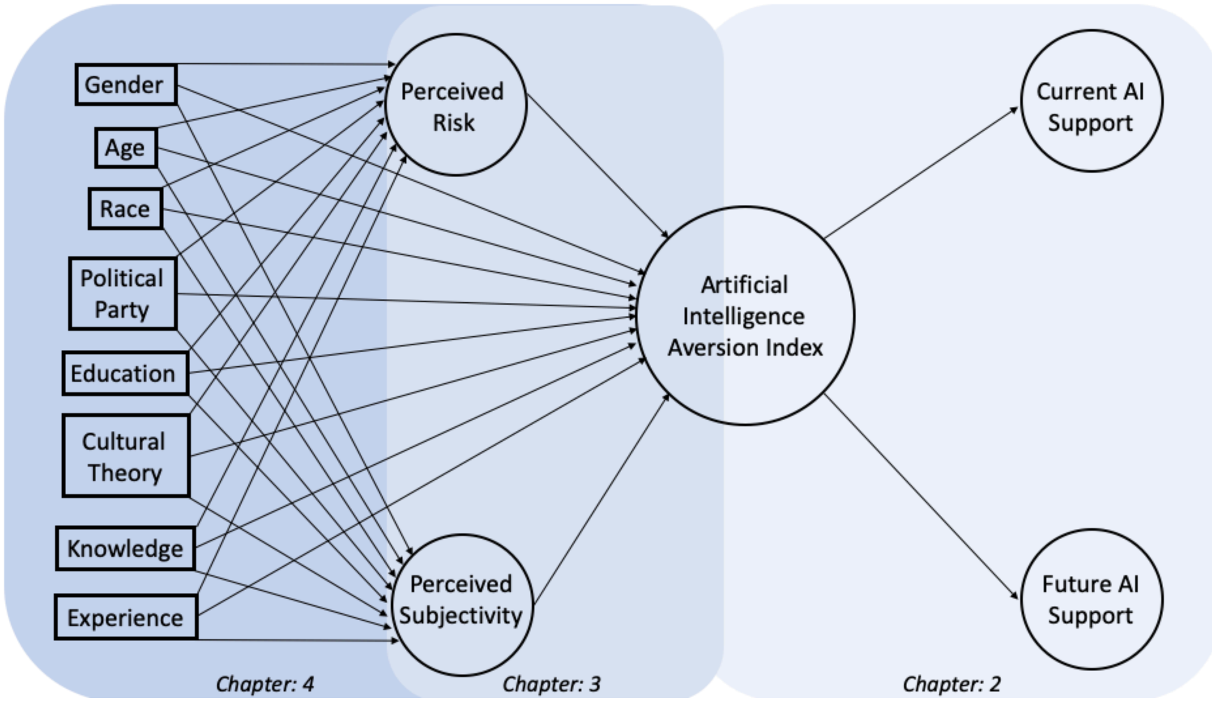
The AIAI, developed in chapter two, serves as a vital tool for quantifying public aversion to AI in the context of public policy. It represents a significant advancement in measuring AI aversion by encapsulating a wide range of factors that influence public sentiment. This index is designed to capture not only direct responses to AI technologies but also the subtler, more subjective aspects of public perception, such as trust, perceived risk, and the subjective nature of tasks assigned to AI.

Modeled relationship

The modeled relationship shown in Figure 1.1 is the theorized relationship between the variables that will be examined in the following chapters of this dissertation. Chapter two of this dissertation will look to evaluating the relationship that the AIAI has on individuals support for current AI uses in public policy as well as their support for future AI uses within public policy. Chapter three will look further upstream to see what if any effect perceived risk and perceived subjectivity have on peoples AIAI scores. This will begin to identify some of the driving forces underlying people's different levels of aversion. Chapter four will look at demographics and the role they play both in people's perceptions of risk and subjectivity with regards to AI as well as their overall effect on people's differential levels of AI aversion.

The conclusion section of this dissertation will return to this hypothesized model and refine it based on the results of these three chapters. This refined model will help to understand more clearly what role these different variables have on people's aversion towards artificial intelligence.

Figure 1.1: Modeled Relationship of Variables



Research Methodology Overview

The research methodology employed across these chapters is multi-faceted, combining quantitative and qualitative approaches to provide a comprehensive understanding of AI aversion. The primary methods include:

1. Affective Imagery Testing: Used in the second chapter, this method helps establish the public's conceptual distinction between AI, algorithms, and advanced technology. It involves measuring the emotional valence associated with self-generated images of AI, providing insights into public perceptions and attitudes.
2. Survey and Data Analysis: Extensively used in the third and forth chapters, survey methods and data analysis tools like factor analysis and linear regression are employed to assess the underlying causes of AI aversion. These methods help in quantifying the impact of perceived risks, subjectivity, and demographic factors on public aversion to AI.

3. Factor Analysis and linear regression: Factor analysis is used to ensure that metrics such as the AIAI are indeed measuring a single variable and allowing for the analysis of its impact on different variables such as current and future AI support. Linear regression is used to identify the effect of perceptions and different demographic variables on the measured levels of AI aversion.

1.3.3 Anticipated Contributions

This dissertation makes several significant contributions to the fields of political science and public policy, particularly in the context of AI. By exploring AI aversion within the public policy domain, this research not only enhances the understanding of public sentiment towards AI but also informs the development of more effective policy strategies for AI integration.

Enhancing Understanding of Public Sentiment Towards AI

A primary contribution of this dissertation is the deepened understanding of public sentiment towards AI. The development and application of the AIAI offers a nuanced measure of public aversion to AI. This tool goes beyond traditional methods, capturing a wide range of domains that influence public opinion on AI. The insights gained from this research provide a more comprehensive picture of public attitudes towards AI, a crucial factor for policymakers and AI developers.

Informing Policy Development and Implementation

The findings of this dissertation have direct implications for the development and implementation of AI-related policies. By identifying the key factors that drive public aversion to AI, such as perceived risk, subjectivity, and demographic influences, this research provides valuable insights for policymakers. Understanding these factors enables the design of policies that are more sensitive to public concerns, potentially enhancing public acceptance and support for AI initiatives.

Contributing to the Broader Discourse on AI and Technology Adoption

This dissertation contributes to the broader discourse on technology adoption and acceptance. The exploration of demographic influences on AI aversion, such as gender, age, race, political affiliation, and education level, enriches the understanding of how different segments of the population perceive and interact with AI technologies. This knowledge is vital for tailoring communication strategies and policy initiatives to diverse audiences, ensuring that AI benefits are equitably distributed and concerns are adequately addressed.

Implications for Future Research

The dissertation sets the stage for future research in several ways. Firstly, it establishes a methodological framework that can be applied to other domains where AI is being introduced. Secondly, the modeled relationship of aversion and policy support and AIAI provide a template for studying other emerging technologies and their public reception. Finally, the findings raise new questions about the role of cultural factors in technology acceptance, suggesting avenues for further exploration in the intersection of technology, culture, and policy-making.

Guiding Strategic Policy-Making and AI Implementation

The comprehensive analysis of AI aversion in public policy domains equips policymakers with the knowledge to make strategic decisions about AI implementation. Understanding the complex dynamics of public aversion to AI will help in designing more effective and publicly acceptable AI policies, thus fostering a more informed and receptive environment for technological advancements.

Chapter 2: Artificial Intelligence Aversion: Aversion Identification and Creating an Index of Aversion

2.1 Introduction

As explained in the introduction of this dissertation, this research builds upon past attempts to understand algorithmic aversion and applying the lessons learned to the similar phenomena of artificial intelligence (AI) aversion. This expansion is done through the vehicle of public policy. This work also goes further than past research by examining what some of the moderating principles are on aversion by incorporating past research into risk perceptions to identify what causes different people to respond differently to the experience of AI aversion.

For this research to work, I must first demonstrate that algorithms and artificial intelligence (AI) are distinct in the minds of members of the public. This is a necessary step to establish that the past research on algorithmic aversion is not sufficient in understanding AI aversion. Once the distinction is established I must then demonstrate that AI aversion exists in the public policy domain to be able to test the different moderating variables on that aversion. If there is no distinction between algorithmic aversion and AI aversion or there does not exist any amount of aversion when it comes to the domain of public policy, there would be no further benefit in trying to assess the effect of other variables on aversion. Finally, I must develop a metric for evaluating AI aversion that works, not only in this research, but can also be used in future research when examining this aversion in other fields outside of the policy domain.

This chapter will focus on that goal of establishing a clear distinction between AI and algorithms as well as developing a simple index to determine what an individual's level of AI aversion is. To achieve this, I construct a two-pronged measure for identifying whether or not this aversion exists in the domain of public policy. First, I establish that algorithms are distinct from AI as well as the broader concept of advanced technology through the use of affective imagery testing. Then I develop an index of AI aversion by subjecting survey respondents to a battery of vignettes designed to draw out considerations of AI adoption into public policy services. Once that index is developed I compare it to the results of another more comprehensive and explicit examination of AI aversion to both current and potential future uses of AI in public policy to establish the validity of the index at measuring the desired variable. I then proceed to examine the results of both the affective imagery and index measures. I conclude this chapter with a short discussion of what my findings are, which established a solid foundation for dissecting the intricate web of moderating variables that influence aversion. I finish by discussing the implications from this section on the following chapters of this dissertation.

2.2 Acceptance and Aversion to Technology

Prior models in technology acceptance have laid out a crucial foundation in understanding how individuals adopt and interact with new technologies. Central to this discourse is Davis's Technology Acceptance Model (TAM), which posits that perceived usefulness and perceived ease of use are key determinants of technology adoption (Davis, 1989). Similarly, the Theory of Planned Behavior (TPB) by Ajzen (1991) provides a broader psychological framework, suggesting that individual behavior is driven by behavioral intentions, which in turn are influenced by attitudes, subjective norms, and perceived behavioral control. While these models have been instrumental in advancing our understanding of technology acceptance, they predominantly address general technologies and may not fully encapsulate the complexities associated with more advanced systems like AI. This gap in the literature signals a need for a more integrative approach that can account for the unique aspects of AI, an approach that begins to be addressed by the Unified Theory of Acceptance and Use of Technology (UTAUT) developed by Venkatesh et al. (2003).

UTAUT synthesizes key elements from existing theories to offer a comprehensive model that predicts user intentions to use an information system and subsequent usage behavior. The model's core constructs—performance expectancy, effort expectancy, social influence, and facilitating conditions—have been extensively validated in various technology contexts. This research seeks to extend the application of UTAUT by examining its relevance and adaptability to the unique domain of AI. Unlike traditional information systems, AI technologies present distinct characteristics, such as advanced autonomy and learning capabilities, that may elicit different cognitive and emotional responses from users. Thus, while UTAUT provides a robust starting point for understanding technology acceptance, this study diverges in its focus on AI-specific factors, particularly exploring how the unique attributes of AI might intensify or mitigate aversion. This extension of UTAUT into the realm of AI aversion not only enriches our understanding of public reactions to emerging technologies but also highlights the need for evolving existing models to accommodate the nuances of advanced digital innovations.

2.3 Distinguishing and Identifying Artificial Intelligence Aversion

While the UTAUT model provides a solid foundation for examining the antecedents of technology acceptance, it is imperative to refine these constructs when confronting the intricacies of AI systems. The intricacies inherent in AI, with its autonomous decision-making capabilities, demand an examination beyond the scope of traditional algorithmic evaluation. Previous research methodologies, notably those deploying vignettes and controlled experiments, have skillfully illuminated aspects of algorithmic aversion in specific contexts (Promberger & Baron, 2006; Önkal et al., 2009; Shaffer et al., 2012; Dietvorst, Simmons, & Massey, 2016), yet they fall short of capturing the broader, more complex picture of AI aversion. Recognizing this, my study seeks to pivot from these foundational methodologies towards a more holistic understanding of AI aversion, acknowledging that trust in AI entails not just an assessment of objective algorithmic performance but also subjective interpretations of AI's role in decision-making processes. This consideration ushers us into the next section, which critically evaluates past approaches to algorithmic aversion and carves out a distinctive niche for understanding AI aversion in a multi-domain landscape.

Strengths & Weaknesses of past methods

The utilization of vignettes in past research allows for the development of a general perspective on public values around specific instances of AI, usually focused on specific fields of research. While this allows for strong internal validity to establish algorithmic aversion in that particular domain, be it in medical treatment or financial investing (Dietvorst, Simmons, & Massey, 2016), its external validity is unconfirmed as it remains limited in its application to outside domains. Extrapolation of aversion outside of these domains is limited and difficult, making it crucial for future research to consider these external factors when examining AI aversion in broader contexts.

The external validity of these methods is important to consider because of the values associated with algorithmic usage in different domains. Studies have shown that the public focuses on the objectivity/subjectivity distinction in evaluating trust in algorithmic decision making (Castelo, Bos, & Lehmann 2019) while other studies have identified the risk associated with the decision-making domain as the determining factor (Scharre 2018). This dissertation will examine this distinction in detail in the following chapter. For this study, with the limited external validity of past research in this domain there is a need for the development of a more comprehensive system of establishing aversion across several domains to be able to measure both causes as well as mitigating variables effects on it.

Distinguishing Artificial Intelligence Aversion

In this dissertation I am interested in developing a way to identify if and how individuals distinguish AI from other advanced technology more broadly and algorithms more specifically. To this end, I have decided to utilize affective imagery measurements to identify how the public distinguished artificially intelligence from these other two domains. Affective imagery measures the images that arise when individuals are prompted by different topics as well as the emotional valence they identify with those self-generated images (Szalay and Deese 1978). The use of affective imagery allows for measuring the conceptual distinction between the subjects with regard to the terms. It also allows for statistical analysis of the difference in the emotional valence associated with the different terms.

While past studies have laid the groundwork for studying algorithmic aversion, examinations of the distinctions in public perception between algorithms and AI, as well as the underlying emotional responses, remain relatively unexplored (Dietvorst, Simmons, & Massey 2016; Jussupow, Benbasat, & Heinzl 2020). By delving into these nuances, this research seeks to provide a deeper understanding of the intricacies of AI aversion and how it may manifest differently than algorithmic aversion, contributing to the broader discourse on the public's relationship with emerging technologies, especially within the public policy context.

Identifying Artificial Intelligence Aversion

Once a clear differentiation between AI and other subjects has been established, the next crucial step is to devise a robust method for measuring AI aversion that maintains both high internal and external validity. I achieve this goal in the dissertation by crafting an index designed to gauge AI aversion across a diverse array of policy domains. Constructing an index permits the incorporation of variations in people's values concerning different policy areas, ultimately leading to the establishment of a comprehensive measure of overall AI aversion. This index possesses enhanced external validity, as it accounts for the diversity of values across different policy domains, while simultaneously upholding high internal validity by directly assessing AI aversion within multiple policy contexts (Glikson & Woolley 2020). Such an approach not only bolsters the reliability of the measurements but also offers a nuanced understanding of how AI aversion manifests across the spectrum of policy areas.

With the development of this comprehensive index, this research aims to address the limitations of past studies of algorithmic aversion that focused on single domains. It will provide a holistic perspective on AI aversion by examining its prevalence and characteristics across various policy areas. This approach recognizes that the factors contributing to aversion can vary substantially from one domain to another. For instance, trust in AI systems might be influenced differently in areas such as food assistance programs, criminal justice, and environmental regulation, as the stakes, perceived risks, and ethical considerations differ significantly. By creating an index that spans diverse policy domains, this research bridges the gap between the isolated findings of past studies and the complex reality of how AI aversion can be experienced in practice. This index will not only act as a valuable research tool but will also serve as a practical means to efficiently gauge AI aversion through a concise set of questions. This standardized measure of aversion will be instrumental in facilitating future research endeavors into the study of AI aversion. The index will also work as a useful metric for future evaluations of AI aversion that can be utilized by other researchers to further understand peoples differing levels of aversion to AI.

2.4 Data

The foundation of this study's analysis rests upon a comprehensive national survey encompassing 834 participants. This survey spanned a duration of 56 days, commencing on July 21 and concluding on September 15 2023. Respondents were drawn from an online panel of voluntary participants, maintained by my survey partner, Lucid. Lucid is a research technology platform which maintains and provides a nationally representative database of survey takers for both market and academic research. The survey's demographic makeup comprises adults aged 18 and above residing within the United States. A demographic breakdown of the respondents to the survey is produced below in Table 2.1, along with current US census estimates.

Table 2.1 Demographic Characteristics of Survey Sample

	US Census Estimates	Survey Sample
Gender		
Female	51%	59.7%
Male	49%	40.3%
Age		
18 to 29	20%	14%
30 to 49	33%	40%
50 to 69	32%	33%
70+	14%	14%
Ethnicity		
Non-Hispanic	83%	92%
Hispanic	17%	8%
Race		
White	77%	78%
African American	13%	13%
Asian	6%	4%
Other Race	3%	5%
NWS Region		
Northeast (CT, ME, MA, NH, NJ, NY, PA, RI, VT)	18%	18%
Midwest (IL, IN, IA, KS, MI, MN, MO, NE, ND, OH, SD, WI)	21%	20%
South (AL, AR, DE, DC, FL, GA, KY, LA, MD, MS, NC, OK, SC, TN, TX, VA, WV)	38%	41%
West (AK, AZ, CA, CO, HI, ID, MT, NV, NM, OR, UT, WA, WY)	23%	20%
Sample Size		**n = 834**

Rigorous data curation measures were implemented to ensure the reliability and validity of the survey responses. An essential component of this process was the incorporation of an attention check within the survey. This step was crucial to filter out participants who were not adequately engaged or attentive. As a result, responses from individuals who failed this attention check were excluded from the final dataset. Initially, the survey garnered a total of 882 responses. However, a further refinement of the data was conducted based on the completion time. Given that the mean completion time was a little over 14 minutes, and the median was just under 12, responses from participants who completed the survey in under 4 minutes were deemed as 'speeding' and subsequently removed. This adjustment resulted in a reduction of the sample size by 36. Moreover, an additional layer of data validation involved the verification of the respondents' geographic location. The survey required participants to be based within the United States, and therefore, responses from 12 individuals whose IP addresses indicated locations outside of the U.S. were also excluded. This was done following the methods in the Dennis, Goodson, & Pearson (2018) report to identify potential shortcomings in the data curation. These stringent data curation steps were critical in enhancing the accuracy and representativeness of the final sample, which stood at 834 participants, forming the basis for the study's analysis.

2.5 Methodology

Past research has shown that the emotional response people have towards a topic is linked to the perceptions that they have about that topic (Finucane et al. 2000; Slovic et al. 2007). To identify the emotional attachment that individuals have towards concepts such as algorithms and AI, this survey asked individuals to consider an important technological concept and then to record the first three images that come to mind when considering the term provided. These questions allow for affective imagery that is generated by the survey takers to be quantified and analyzed. For this survey respondents were randomly assigned to either consider *algorithms*, *artificial intelligence*, or *advanced technology*. Advanced technology was included to establish a baseline understanding of people's perceptions about the current development of futuristic technologies. After each image about the concepts was recorded by the respondents, they were then tasked with scoring the image on a five-point scale based on how positively or negatively they felt about it, shown in Table 2.2.

Table 2.2 Affective Imagery Experiment

Survey Question
This survey will focus on science and technology in public policy. We will start with some questions about the use of [artificial intelligence | algorithms | advanced technology] in public policy. Can you tell us the first three words or phrases that come to you when you think about [artificial intelligence | algorithms | advanced technology]?
When you think about this word or phrase, do you have positive or negative feelings?

In developing the AIAI, this study considers the UTAUT framework established by Venkatesh et al. (2003), which suggests that user acceptance is influenced by factors such as performance expectancy and effort expectancy. The AIAI aims to capture similar dimensions within the context of AI technologies. My survey endeavors to construct such an index by presenting participants with a series of distinct vignettes. These vignettes present scenarios in various policy domains wherein traditional government processes have been augmented with AI systems. Participants are then prompted to indicate the extent of their opposition or support for these novel systems. The vignettes strategically manipulate the services being altered to isolate the impact of introducing AI into policy domains. Subsequently, the responses are averaged and converted into an index, enabling me to quantify differences in responses across the various vignettes.

In addition to these vignettes, the survey also asked respondents to express their support or opposition to a series of specific instances where the government either currently implements AI in its service or could in the future begin to incorporate it into its services (See Appendix:

Table A2 & Table A3). To ensure the quality of the results from the survey I employed exploratory factor analysis (EFA) to validate the AIAI as well as the constructs representing public support for current and potential future uses of AI in policy-making. The EFA was conducted to ensure that the items used to measure each construct were consistent with the hypothesized factor structure, that there was a single dimension explaining peoples support for current AI uses, future AI uses, and that the AIAI was loading on a unidimensional axis. This statistical method provided a means to test whether the data fit a hypothesized measurement model based on established theories and previous empirical findings.

To assess the unidimensionality of the survey measures, parallel analysis was also conducted. This technique involved comparing the eigenvalues from the actual data with those generated from random data. By doing so, it was possible to determine the number of factors that should be retained for a reliable and valid factor structure. The parallel analysis served as a robust method to justify the factor retention decisions made in this study. The factor analysis was further extended to examine the predictive validity of the AIAI, a composite measure derived from five survey items related to different policy domains where AI is being used for decision-making.

Once the factor scores have been identified for the AIAI, current AI use, and future AI use I then use linear regression to evaluate how good the AIAI does at predicting current and future support towards AI. These methodological approaches allowed for a rigorous examination of the survey's measurement properties and the constructs' potential to predict relevant outcomes. The results from these analyses are presented in the subsequent section, providing empirical evidence to support the reliability and validity of the measures used in this study.

This comprehensive approach allows me to uncover the nuanced emotional dimensions that individuals attach to concepts such as algorithms and AI. This equips me with a versatile tool to understand the emotional dimensions that individuals attach to these and establish and evaluate an index for measuring AI aversion. This index can be utilized both by the following chapters in this dissertation, as well as in future research into AI aversion especially within the field of public policy.

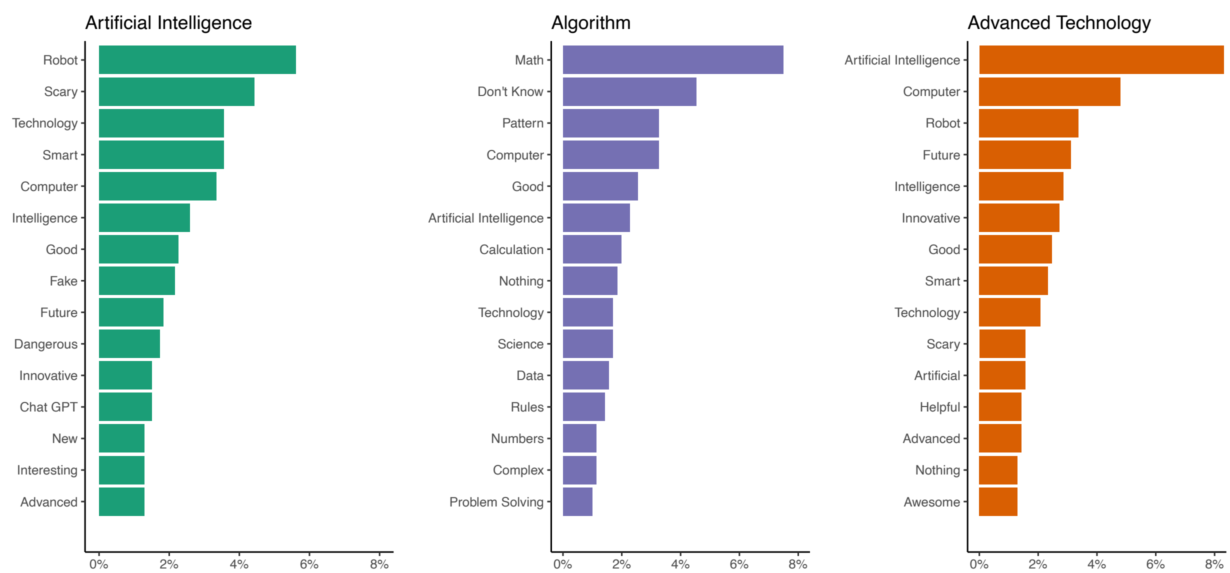
2.6 Results

Affective Imagery

Existing research has revealed that public perceptions concerning topics, such as AI, and their associated behaviors towards the topic can be deeply rooted in emotional responses to this innovative technology (Slovic et al. 2007; Finucane et al. 2000). These perceptions can be measured and categorized through the utilization of affective imagery association. Affective imagery can be understood as the spontaneous association's individuals form regarding AI or its implications, which in turn serve as a basis for evaluating their emotional reactions within the framework of the affect heuristic (Slovic et al. 2007). A two-step process of the free association method is used in the assessment of affective imagery. In the first step, respondents are prompted to express the words or phrases that spring to mind when they contemplate a specific object or stimulus (Szalay and Deese 1978). These articulated words or phrases essentially represent the "imagery" that respondents associate with that particular stimulus. Subsequently, respondents are asked to attribute a valence of positive, negative, or neutral to each word or phrase.

In the context of this study, I undertook an examination of respondents' affective imagery associated with AI aversion. I am cognizant of the fact that the choice of phrasing or terminology can substantially influence responses. To probe potential variations linked to different terms related to AI aversion, participants in my survey were randomly assigned to one of three prompts: "Artificial Intelligence," "Algorithm," or "Advanced Technology" (See Appendix: Table A1). Adhering to the method outlined by Szalay and Deese (1978), I requested respondents to list the first three words or phrases that come to mind upon encountering their assigned prompt. The responses displayed noteworthy diversity, as demonstrated by the compilation of the 15 most frequently cited associations for each prompt, presented in Figure 2.1 below.

Figure 2.1 Top 15 Most Frequent Images



The most common image returned for the prompt *Artificial Intelligence* was Robot, followed by Scary, Technology, Smart, and Computer. The most common image returned for the prompt *Advanced Technology* was AI, followed by Computer, Robot, Future, and Intelligence. Finally, the most common image returned for the prompt *Algorithm* was Math, Don't Know, Pattern, Computer, and Good. These results indicate that there is clearly a distinction between the most common images between these three different prompts. Amongst the top fifteen most common images the only ones shared between all three of these prompts are Technology, Computer, and Good.

After each image solicitation exercise respondents were asked "When you think about this word or phrase, do you have positive or negative feelings?" Respondents were given a five-point scale to answer the question with 1 being "Extremely negative" and 5 being "Extremely positive." Figure 2.2 shows the distribution of these scores as well as the mean of the self-reported valence scores for each. The mean score for *Artificial Intelligence* was 3.16, as compared to the mean scores for *Algorithm* and *Advanced Technology* which were 3.43 and 3.54 respectively.

Figure 2.2 Affect Distribution for Images

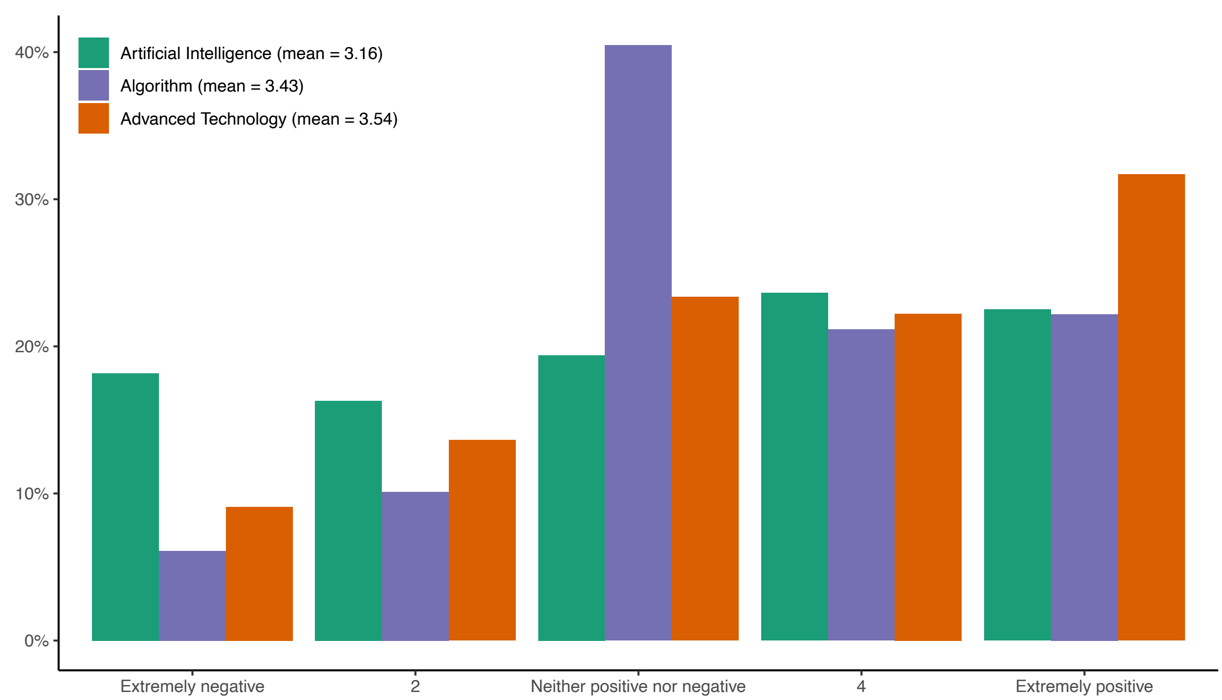
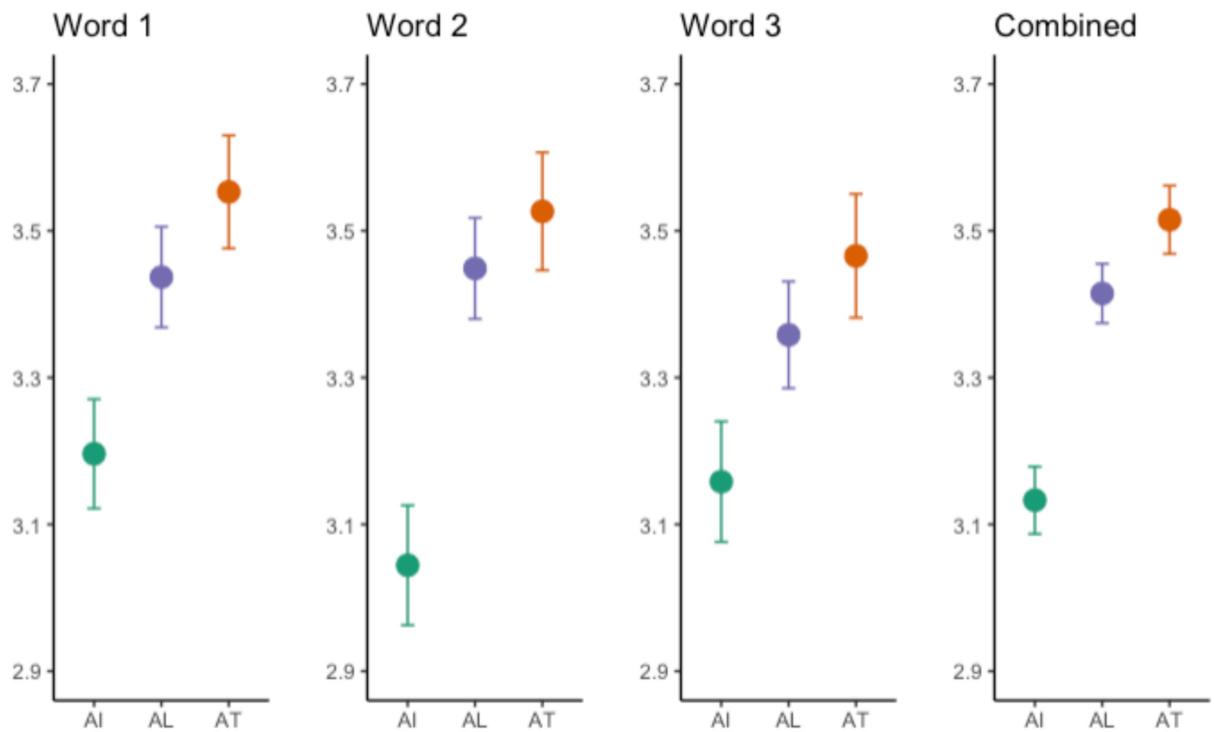


Figure 2.3 shows the mean valence score broken down by each individual word, as well as the combined mean valence score for each word. Across all three images generated by respondents *Advanced Technology* always received the highest mean valence score, followed

closely by *Algorithm*. In each instance the mean valence score for *Artificial Intelligence* was statistically lower than the other two terms.

Figure 2.3 Affect Distribution Breakdown by Word



These findings indicated that while respondents had a more positive valence score for the images associated with each prompt, algorithms and advanced technology have a much more positive valence than artificial intelligence. This analysis involved conducting a series of t-tests to compare the valence scores between the three prompts: artificial intelligence, algorithm, and advanced technology. The results of these tests are quite revealing, see Table 2.3. When comparing artificial intelligence and algorithms, a substantial and statistically significant difference in means is observed. The p-value for this comparison was very small ($p < 0.001$), providing strong evidence against the null hypothesis that there is no difference between the means. Furthermore, the confidence interval for this comparison ranged from -0.391 to -0.153, reinforcing the finding that artificial intelligence is associated with a significantly lower mean valence score when contrasted with algorithms.

Table 2.3 T-Test Comparisons of Perceptions Towards AI, Algorithms, and Advanced Technology

Comparison	t-Value	Degrees of Freedom	P-Value	95% Confidence Interval	Mean of X	Mean of Y
Artificial Intelligence vs. Algorithms	-4.472	1738	<0.00001	[-0.391, -0.153]	3.16	3.43
Artificial Intelligence vs. Advanced Technology	-5.820	1733.6	<0.00001	[-0.505, -0.251]	3.16	3.54
Algorithms vs. Advanced Technology	-1.723	1537.6	0.851	[-0.227, 0.015]	3.43	3.54

Similarly, when comparing valence scores for artificial intelligence and advanced technology, the results once again demonstrate a substantial and statistically significant difference in means. The p-value for this comparison was also very small ($p<0.001$), further corroborating the strong evidence against the null hypothesis. The confidence interval for this comparison ranged from -0.505 to -0.251, emphasizing that artificial intelligence has a significantly lower mean valence score compared to advanced technology.

In contrast, the comparison between algorithm and advanced technology revealed no significant difference in means. The p-value for this comparison was relatively large (0.085), and the 95 percent confidence interval ranged from -0.227 to 0.015, suggesting that there is no strong evidence of a difference in mean valence score between algorithms and advanced technology.

These results underscore a notable distinction in how respondents perceive artificial intelligence in comparison to both algorithms and advanced technology, with the former eliciting a more negative emotional response than the other two. However, it's important to recognize that algorithms and advanced technology do not exhibit such a pronounced contrast in their valence scores, as indicated by the lack of a significant difference in means between the two. These results clearly show that AI exist in a unique category in people’s minds when compared to how they think about algorithms and advanced technology.

Index Construction

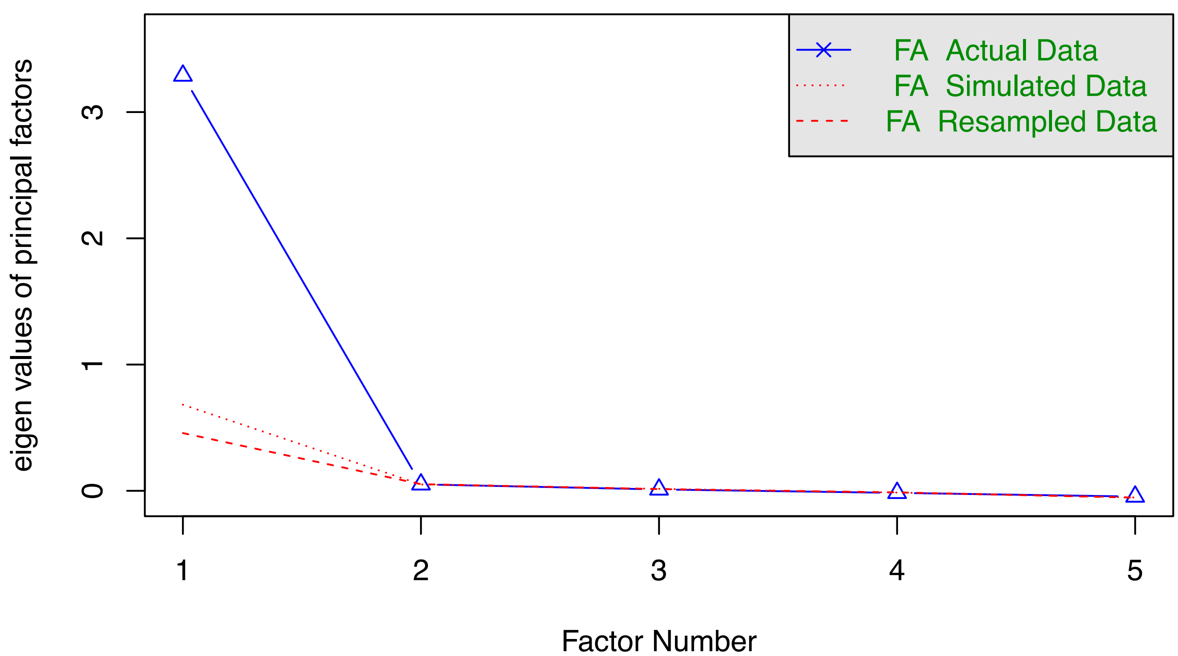
The AIAI is designed as a quantitative gauge of public sentiment toward AI across a range of policy sectors. By incorporating diverse areas such as criminal justice, disaster response, and economic regulation, the AIAI offers a nuanced view of the public’s stance on AI applications in key societal contexts (See Appendix: Table A4) . The AIAI was constructed from participant responses to vignettes that portrayed hypothetical scenarios involving AI across different policy domains. Responses were quantified on a scale from -10 to 10 including 0, reflecting varying degrees of aversion or acceptance toward AI. The index itself was calculated by averaging these scores, creating an aggregate measure that encapsulates the multifaceted nature of AI aversion.

Reliability Statistics

To assess the reliability of the AIAI, I utilized exploratory factor analysis (EFA) with the minimum residual (minres) method. This approach enabled me to examine whether the collected variables measure a single underlying construct. The EFA yielded high factor loadings across all variables, ranging from 0.75 (criminal justice) to 0.84 (environmental fines), indicating that they are all significantly correlated with the principal factor. This factor alone accounted for 66% of the variance, substantiating the index's focus on a singular construct of AI aversion.

The validation of these results was further reinforced by a parallel analysis scree plot in Figure 2.4. The parallel analysis revealed that the eigenvalues of the actual data decisively surpassed those of both the simulated and resampled data for the first factor. This graphical confirmation suggests that only one factor should be retained, corroborating the unidimensionality indicated by the EFA. The subsequent factors, with eigenvalues falling below or near the simulated data threshold, were deemed insignificant, thereby not warranting additional extraction. Goodness-of-fit measures were also encouraging, with a Tucker Lewis Index of 0.991 and a Root Mean Square Error of Approximation (RMSEA) of 0.052, both indicative of an excellent model fit.

Figure 2.4 Parallel Analysis Scree Plot for AIAI



I performed this same type of EFA and parallel analysis for my measure of support for current and future uses of AI (See Appendix: Table A2 & Table A3). The EFA for current uses of AI yielded high factor loadings across all variables, ranging from 0.74 (facial recognition software) to 0.85 (patent adjudication), indicating that they are all significantly correlated with the principal factor. This factor accounted for 63% of the variance in support for current AI uses. The EFA for future uses of AI also yielded high factor loadings across all variables, ranging from 0.74 (Minor Weather Forecasting) to 0.83 (City Planning & Urban Development), indicating that they are all significantly correlated with the principal factor. This factor also accounted for 63% of the variance in support for future AI uses in the policy domain.

I performed parallel analysis of the measures of current and future support for AI uses to ensure unidimensionality as well. Figure 2.5 shows the parallel analysis scree plot for current AI uses and reveals that the eigenvalues of the actual data decisively surpassed those of both the simulated and resampled data for the first factor. This confirms that only one factor should be retained from these survey questions, indicating unidimensionality. The goodness-of-fit measures for this model presented mixed results. The Tucker Lewis Index was very encouraging at 0.962, indicating a strong model fit. However, the RMSEA value of 0.09, while not excessively high, suggests a somewhat less optimal fit, falling into the range of a middling model approximation.

Figure 2.5 Parallel Analysis Scree Plot for Current AI Uses

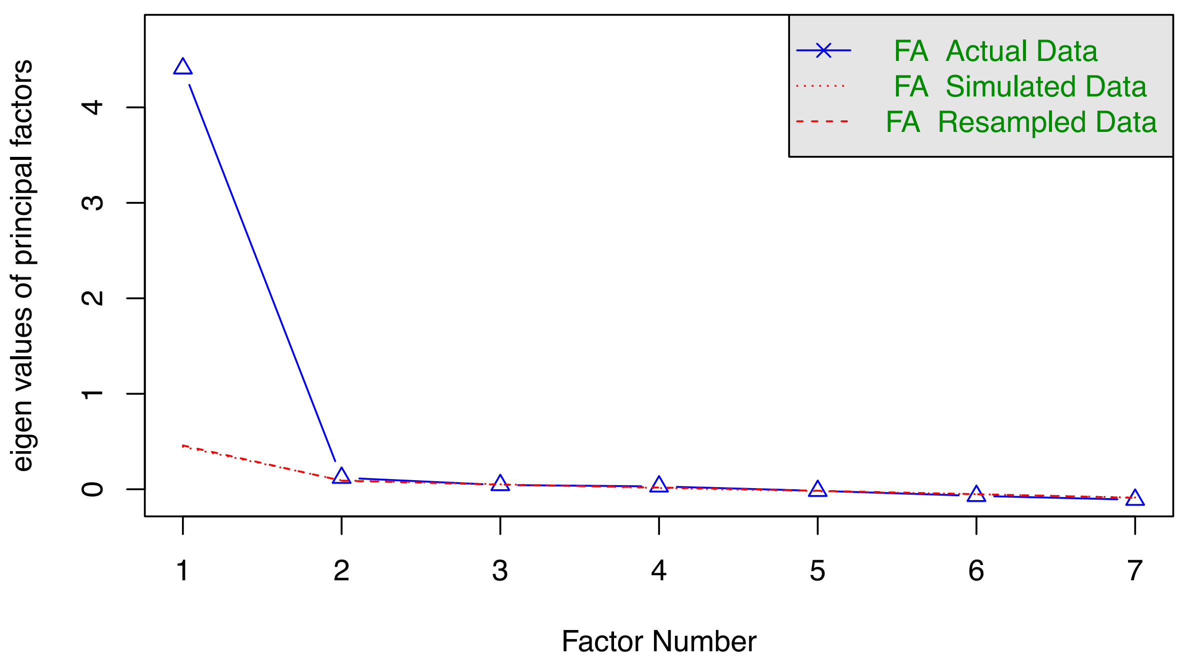
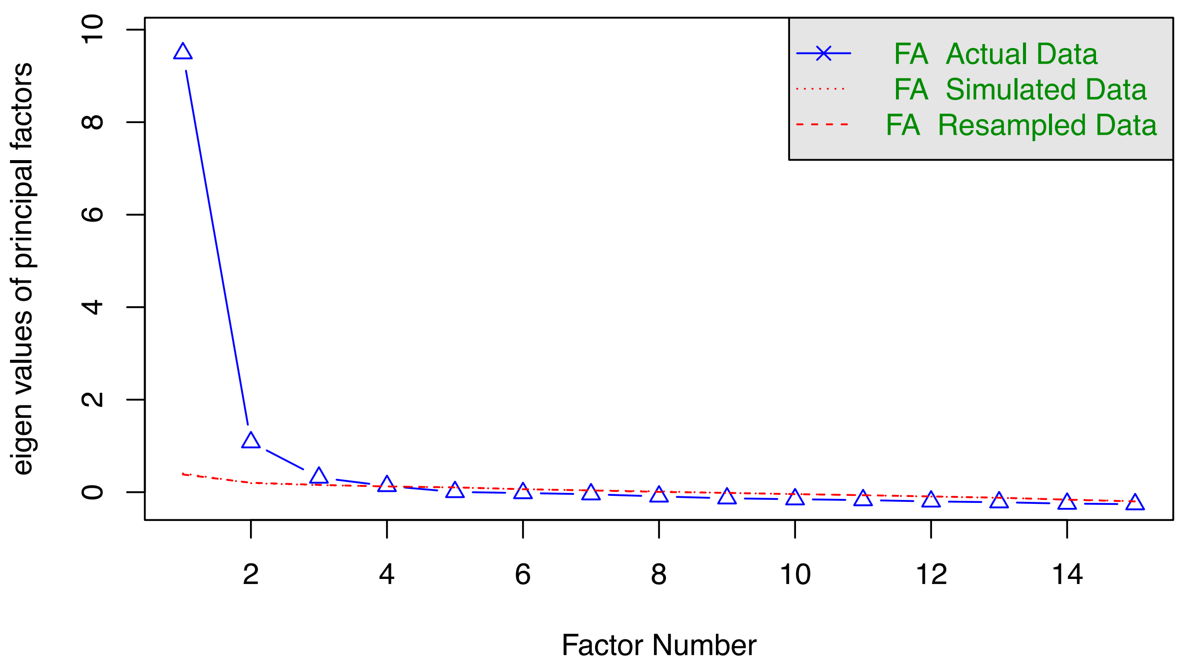


Figure 2.6 shows the parallel analysis for peoples support of future AI uses. The results once again show high eigen values on 1 factor being present in the data, however the eigenvalue of 2 factors was also slightly above the cutoff of 1 indicating that the survey results for future AI support could be loading onto two factors instead of just one.

Figure 2.6 Parallel Analysis Scree Plot for Future AI Uses



The assessment of the construct validity for future support of AI was conducted through both one-factor and two-factor models, as evidenced in Table 2.4. Despite the eigenvalues suggesting a unidimensional structure, the model fit indices approached, but did not exceed, the conventional thresholds for a robust fit. Specifically, the Tucker Lewis Index (TLI) fell slightly below the acceptable limit of 0.90, with values of 0.732 for the one-factor model and 0.88 for the two-factor model, suggesting a marginally inadequate fit (Hu & Bentler, 1999). Similarly, the Root Mean Square Error of Approximation (RMSEA) exceeded the preferred maximum of 0.10, with values of 0.197 for the one-factor model and 0.132 for the two-factor model, indicating a less than optimal model approximation (Browne & Cudeck, 1992).

Notwithstanding these deviations, the closeness of these indices to the respective cut-offs provides a reasonable justification for treating future AI support as a unidimensional construct in the forthcoming analyses. The potential second factor, although elucidating an additional 8% of variance, did not substantially refine the TLI or RMSEA to warrant its inclusion. Consequently, the analysis will proceed under the premise of unidimensionality for future AI support.

Table 2.4 EFA Comparison for Future AI Support

Model	Explained Var. 1st Factor	Explained Var. 2nd Factor	Tucker Lewis Index	RMSEA	Confidence Interval
One - Factor	63%	--	0.732	0.197	0.191 – 0.203
Two - Factor	64%	8%	0.88	0.132	0.125 – 0.139

The exploratory factor analysis and parallel analysis have provided a solid foundation for the construct validity of the AIAI, as well as the measures of support for current and future uses of AI. The AIAI and current AI support measures have demonstrated strong unidimensionality, indicating that they effectively capture a single underlying construct. Although the EFA for future AI support yielded fit indices that were slightly below the thresholds for optimal fit, the proximity of the TLI and RMSEA values to acceptable levels allows for the consideration of future AI support as a unidimensional construct as well.

With these findings in mind, the next phase of the analysis will employ linear regression to elucidate the predictive relationships between the AIAI and the factor scores for both current and future AI support. This approach is substantiated by the nearly sufficient unidimensionality of the future AI support measure, which justifies the use of its factor scores. By using factor scores for both constructs, the analysis will maintain consistency and methodological rigor.

The detailed outcomes of the linear regression models will be presented in the subsequent section, which will critically examine the role of the AIAI in shaping public attitudes toward AI implementation in various policy domains. This analysis will not only illuminate the current state of public opinion but also guide future investigations into the specific aspects of AI that engender support or aversion among the public.

Index Validation

The Artificial Intelligence Aversion Index (AIAI) and its relationship with public support for current and future uses of AI in policy were examined through linear regression models, see Table 2.5 and Figure 2.7. These models were essential to understanding the predictive power of the AIAI in gauging public sentiment towards AI applications within policy domains.

Table 2.5 Linear Model of AIAI and Current, Future, and Combined AI Support

	Dependent variable:		
	Current AI Support	Future AI Support	Combined AI Support
	(1)	(2)	(2)
Aversion Index	-0.760*** (0.022)	-0.876*** (0.022)	-0.782*** (0.019)
Constant	-0.000 (0.016)	-0.000 (0.016)	-0.000 (0.014)
Observations	834	834	834
R2	0.580	0.663	0.661
Adjusted R2	0.580	0.662	0.661
Residual Std. Error (df = 832)	0.469	0.453	0.406
F Statistic (df = 1; 832)	1,149.452***	1,634.565***	1,622.403***
Note	***p<0.1; **p<0.05, ***p<0.01**		

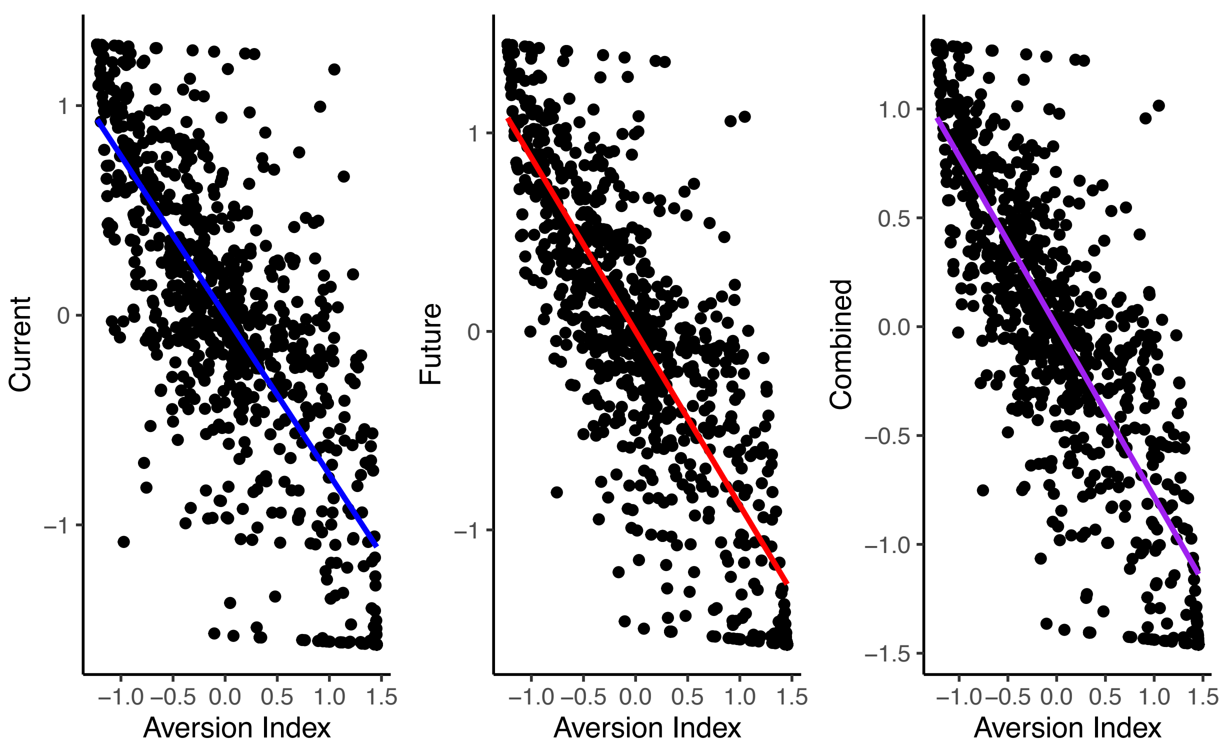
AIAI vs. Current AI Support

In the first model, the independent variable was the AIAI, and the dependent variable was the mean support for current AI uses. The results indicated a statistically significant negative relationship between current AI support and the AIAI (β = -0.760, $p < 0.001$). This implies that higher levels of aversion, as measured by the AIAI, are associated with lower support for current AI uses. The model accounted for approximately 58% of the variance in the Aversion Index (Adjusted $R^2 = 0.580$).

AIAI vs. Future AI Support

In the second model, support for future AI was used as the dependent variable. Similar to the first model, there was a significant negative association with the Aversion Index (β = -0.876, $p < 0.001$). This model explained a higher proportion of the variance (Adjusted $R^2 = 0.662$), suggesting that the AIAI is a more robust predictor of attitudes towards future AI uses.

Figure 2.7 Future AI Support EFA Comparison



[1]

[1] Robust regression analysis was also performed with this data to ensure accuracy. The results did not differ substantially from form the linear regression model shown here.

AIAI vs. Combined Current & Future AI Support

The third model integrated both current and future AI support. The combined measure of support for both also showed a significant negative relationship with the Aversion Index ($\beta = -0.782$, $p < 0.001$), with the model explaining about 63.6% of the variance (Adjusted $R^2 = 0.661$). This finding underscores the comprehensive predictive capability of the AIAI across different dimensions of AI applications in policy.

Interpretation

These results demonstrate the robustness of the AIAI as a predictor of public attitudes towards AI in policy. The consistent negative relationships across all models highlight that higher aversion levels are linked to lower support for both current and future AI applications. The increasing variance explained in the models incorporating future AI support indicates that public sentiments towards potential future uses of AI may be more sensitive to aversion levels than the current uses. This insight is pivotal for policymakers and researchers in understanding and addressing members of the publics concerns regarding AI implementation within various policy domains.

The models also reinforce the importance of distinguishing between current and future AI uses when assessing public attitudes. While there are commonalities in the factors influencing support for AI, the nuances between current and future applications require separate consideration. The stronger predictive power for future AI uses suggests that public apprehensions may be more pronounced when considering potential, yet unrealized, AI applications. This could be attributed to uncertainties in the potential biases of the AI system, the perceived risks associated with the future implications of AI in policy, or even the lack of experience with these yet unrealized usages of AI.

2.7 Conclusion

The research presented in this chapter has led to several significant findings and implications regarding public attitudes towards AI and its integration into public policy. Building upon previous research on algorithmic aversion, this work distinguishes and identifies AI aversion as a unique phenomenon. By employing affective imagery measures, this research reveals distinctions in the associations and valence that respondents attribute to AI compared to algorithms and advanced technology, showcasing the significant differences in public sentiment toward these concepts. The development and validation of the AIAI has been central to these insights gained from this research, allowing for a nuanced understanding of public sentiment towards AI.

The AIAI's validation was underpinned by exploratory factor analysis (EFA) and parallel analysis, which confirmed its unidimensionality and robustness in capturing AI aversion. This validation process underscored the index's reliability in representing a singular construct of AI aversion. The EFA results for measures of current and future AI uses, however, indicated a need for a more nuanced approach. While unidimensionality was assumed for future support of AI systems, it is important to consider what could have been causing the suboptimal fit for the EFA model. One potential cause could be an artifact of statistics in that there were more variables utilized in measuring future support (n=15) than in current support (n=7) or the index (n=5). Another potential cause could be the inherent uncertainty of imagining how future uses of AI could materialize with unknown risks and subjectivity to these future use cases. The following chapter will aim to shed some light on understanding the impact of risk and subjectivity perceptions in people’s aversion to AI.

Due to the assumption about unidimensionality of future AI support, I also conducted linear regression comparing the factors scores of the aversion index with mean scores for current and future support (See Appendix: Table A5). The findings were consistent with the results of the of the linear regression model assuming unidmensionality and using factor scores within Table 2.5.

Subsequent regression analyses involving the AIAI provided critical insights into the nature of peoples aversion. A significant negative relationship between the AIAI and support for both current and future AI uses was observed. This relationship was stronger for future AI uses, suggesting that public aversion is more pronounced when considering the potential implications of AI in policy.

The findings from this study pave the way for deeper exploration into the drivers of AI aversion. The next phase of research will focus on disentangling the roles of perceived subjectivity and perceived risk in shaping public attitudes towards AI. This investigation is critical, as understanding these underlying factors can inform strategies to address public concerns and enhance the acceptance of AI in policy.

This research has significantly advanced our understanding of public attitudes towards AI in policy. The AIAI serves as a valuable tool in quantifying public aversion and its impact on policy support. The decision to utilize mean scores for current and future AI support, based on model fit considerations, has provided a clearer picture of public sentiment.

As AI continues to evolve and find new applications in policy, understanding public attitudes will remain crucial. This study contributes to that understanding by highlighting the complexities of public sentiment towards AI and laying the groundwork for further research into the factors driving these attitudes. The next chapter will delve into whether perceived subjectivity or perceived risk is more influential in shaping public aversion to AI, offering deeper insights into the public's relationship with AI in the context of policy and governance.

Chapter 3: Unveiling the Dynamics of Artificial Intelligence Aversion: Risk and Subjectivity

3.1 Introduction

Having now confirmed the presence of AI aversion and also having crafted the AIAI for its systematic examination, this dissertation now pivots its focus towards unraveling the intricacies of the aversion individuals harbor toward the utilization of AI in diverse policy domains. In an effort to gain a deeper understanding of the dynamics inherent in this aversion, the present chapter endeavors to assess its underlying causes. Furthermore, I aim to draw insightful comparisons between the subtleties of AI aversion and the findings of previous research pertaining to algorithmic aversion. This shift in focus is driven by the overarching goal of shedding light on the specific factors that contribute to and shape the resistance observed in individuals towards the integration of AI within various policy frameworks.

The chapter expands upon into previous efforts to understand the origins of algorithmic aversion and contemplates whether they are equally applicable to the sphere of AI aversion. There are two narratives identified in the research as significant determinants in the realm of algorithmic aversion: the perceived risk associated with these technologies and the subjectivity of the domains in which these technologies are deployed in. Both of these have been identified as potential drivers for algorithmic aversion and therefore could also be drivers for AI aversion (Castelo, Bos, & Lehmann, 2019; Purves, Jenkins, & Strawser 2015).

This exploration utilizes an in-depth examination of survey results, factor analysis and linear regression, carefully designed to uncover the essential drivers of AI aversion. By scrutinizing public perceptions and sentiments, I aim to identify what plays a more substantial role in shaping aversion: the perceived risks inherent to AI adoption or the subjective nature of the policy domains in which AI is integrated. Both the AIAI created in the previous chapter and AI support measures in the survey data are utilized in the regression models to fully understand which driver is having a larger influence on AI aversion.

This chapter lays the groundwork for a comprehensive understanding of the driving force for AI aversion. By examining the compatibility of past research on algorithmic aversion with AI aversion and pinpointing the principal determinants of aversion, I set the stage for a deeper exploration of demographic variations in the subsequent chapter. The insights gained from this chapter will inform the strategies needed to harness the potential of AI while addressing the concerns of a diverse and dynamic society.

3.2 Past explanations for aversion in algorithms

AI aversion is the tendency people have to oppose the utilization of AI in decision making roles. While very similar to algorithmic aversion, research into people's aversion when the process that is being utilized is described as AI is minimal. In light of the limited research on AI aversion, this chapter adopts a novel strategy of drawing from the extensive literature on algorithmic aversion to elucidate the potential factors contributing to aversion towards AI systems. Prior research on algorithmic aversion primarily centers on two key explanations: the extent to which individuals perceive the task as subjective and the perceived risk associated with potential errors within the domain where the algorithm operates. These dimensions serve as fundamental underpinnings for investigating and understanding aversion to AI within public policy.

3.2.1 Perceived Subjectivity and Objectivity

One of the pivotal determinants of algorithmic aversion is the perceived subjectivity or objectivity of the tasks assigned to algorithms. Subjective tasks involve elements that are challenging to quantify or are influenced by human judgment, emotions, or opinions (Fisher 2022). In contrast, objective tasks are ones that are data-driven and can be assessed more quantitatively. Individuals tend to trust algorithms less for tasks that they consider to be subjective and instead trust algorithms more for tasks that they consider to be more objective (Castelo, Bos, & Lehmann, 2019; Yeomans et al., 2019).

Research by Castelo, Bos, and Lehmann (2019) found that distrust is higher when algorithms perform tasks that people consider to be more subjective. For instance, tasks such as composing a new song or writing a unique news articles are subject to personal interpretation, creativity, and values. These tasks are therefore less trusted by individuals to be appropriate domains for algorithms to be used within. Conversely participants were found to be more trusting when the algorithms were tasked with more objective goals such as the analyzing of data or in giving directions to a user. The perceived objectivity or subjectivity of a task impacts the extent to which individuals perceive algorithms to be reliable tools that they are comfortable to be using.

Yeomans et al. (2019) looked at how aversion exists within subjective domains even when participants know the algorithm is superior to the task than other humans. When asked to assess how funny another human thought a series of jokes were based on how funny they had ranked a different set of jokes, an algorithm consistently outperformed even friends and family members in accuracy at predicting individual preferences. However, when given the option of seeing a joke suggested by an algorithm instead of another human, participants were much more reluctant, regardless of how much more accurate the algorithm had been at assessing past subjective feelings about humor. This research shows that algorithms are suffering a penalty in preference when being utilized within subjective fields.

3.2.2 Risk Associated with Algorithms

In dissecting the multifaceted nature of public aversion to AI technologies, this study places a particular emphasis on the role of risk perception. This aspect is conceptually tethered to the construct of performance expectancy within the Unified Theory of Acceptance and Use of Technology (UTAUT) as posited by Venkatesh et al. (2003). However, the complexity of AI systems introduces an amplified dimension of risk, extending beyond the traditional scope of performance expectancy. This dimension is rooted in the public's concern over the unpredictability and potential for error in AI decision-making processes. Slovic's seminal work on the perception of risk (Slovic, 1987) provides a crucial theoretical underpinning for this investigation, suggesting that the perception of risk is not only about the likelihood of negative outcomes but also about the public's trust and their perceived loss of control. Aligning with Slovic's insights, this study probes into how these perceptions specifically influence aversion to AI algorithms, considering that AI systems often operate in high-stakes environments where the consequences of failure can be significant. The exploration of risk perception in the context of AI aversion is therefore an essential step in understanding the nuances of public sentiment toward these emerging technologies.

Risk has a strong influence in shaping aversion towards algorithms and AI systems within various domains. Risk, as a multifaceted concept, encompasses the potential for errors, biases, and the far-reaching consequences of algorithmic decisions. Understanding the role of risk is essential because it directly impacts individuals' trust and acceptance of these technologies. The algorithms that are being used are not devoid of flaws, biases, and errors due to the limitations in the data they use and the inherent uncertainty of human behavior. Risk in this context encompasses the potential for errors, biases, and the consequences of algorithmic decisions. Research indicates that risk has a substantial impact on individuals' trust in algorithms (Scharre, 2018; Purves & Davis, 2022; Filiz et al., 2023).

Scharre (2018) focuses on the incorporation of algorithms and AI systems within military applications and autonomous military functions. Military forces increasingly rely on algorithms, ranging from the guiding of heat seeking missiles, battlefield communications, and even automation of military action (Purves, Jenkins, & Strawser 2015). The Department of Defense has developed a three-level hierarchy for the utilization of automation within military action. The tiers are semiautonomous operations where the human is the deciding factor in a machine's actions, supervised autonomous operations where the decisions are made by the machine but the human can directly intervene if they desire to, and finally there are fully autonomous operations when humans cannot intervene in a timely fashion to alter the machines decisions (National Institute of Standards and Technology, 2008). The risk associated with these three levels of autonomy grows such that semi-autonomous systems are regulated to a lower degree than those of fully autonomous operations (Department of Defense, 2023).

Risk associated with AI exists beyond the military as well. Within the criminal justice context, the risk of misjudgment by algorithms can have massive effects on the lives of individuals (Davis et al. 2022). This is particularly important with respect to racial or ethnic biases as has been documented by Purves and Davis (2022). Their findings reveal that people are highly averse to algorithms when they perceive the risk of biased decisions or when errors could lead to severe consequences, such as wrongful incarceration or denial of parole (Partnership on AI, 2020; Hamilton et al. 2022). The risk of harm or injustice resulting from the adoption of an algorithm in the decision process intensifies the aversion people have towards algorithms.

Filiz et al. (2023) established what they identified as the tragedy of algorithmic aversion, which they describe as “algorithm aversion appears most frequently in cases where it can cause the most damage (Filiz et al. 2023).” This is because as the consequences of a decision increase, algorithmic aversion becomes more likely and the probability of success suffers due to the reluctance to use algorithms. It illuminates how heightened risk perception, especially in high-stakes situations, significantly exacerbates public aversion to algorithmic decision-making. This insight aligns directly with Slovic's theory of risk perception, emphasizing that aversion is not merely a product of potential negative outcomes but also a reflection of the public's distrust and their perceived loss of control over AI systems. In domains where the repercussions of errors are grave, such as healthcare or financial decision-making, this study reveals that the public's aversion intensifies proportionately with the perceived risk. This finding extends the conceptual framework of performance expectancy within the Unified Theory of Acceptance and Use of Technology (UTAUT), illustrating that the complex nature of AI systems introduces a dimension of risk perception that goes beyond traditional technology acceptance models. It underscores the need for understanding how the nuanced perception of risk in various application domains shapes the public's receptiveness towards AI and algorithmic technologies.

3.2.3 Subjectivity or Risk

Both the perceived subjectivity and risk are identified in the literature as potential factors influencing people’s algorithmic aversion. When algorithms are entrusted with subjective, high-risk tasks, individuals may exhibit a heightened level of aversion compared to more objective and low-risk ones. There remains little research examining which of these two is the bigger driver of aversion, and the extent to which they are contributing to peoples levels of aversion towards algorithms.

This chapter will examine both the effects of perceived risk and subjectivity on AI aversion to answer two questions. First, if the lessons from the research into algorithmic aversion carry over into AI aversion. Second, to identify which is the bigger driver for aversion in public policy, perceived risk or subjectivity. This will allow for future efforts at the utilization of AI into public policy to be able to identify what areas people might be more accepting of the incorporation of AI and which will receive more resistance. By disentangling which is a bigger driver for aversion, policymakers can act strategically in the adoption AI into the services they provide, allowing for higher acceptance and approval.

3.3 Data

This dissertation chapter will also make use of the comprehensive national survey used for the previous chapter. The survey (n=834) included a battery of questions that recorded both demographic information as well as perceptions and preferences when it comes to risk, subjectivity, and support as it relates to AI. The survey included an experimental treatment wherein participants were randomly assigned into two groups, one to measure perceived risk for AI and one to measure perceived subjectivity of AI. Table 3.1 shows the demographics of both experimental groups.

Table 3.1 Demographic Characteristics of Survey Sample

	US Census Estimates	Survey Sample	Perceived Risk	Perceived Subjectivity
Gender				
Female	51%	59.7%	62.6%	56.8%
Male	49%	40.3%	37.4%	43.2%
Age				
18 to 29	20%	14%	13%	14%
30 to 49	33%	40%	43%	36%
50 to 69	32%	33%	33%	33%
70+	14%	14%	11%	17%
Ethnicity				
Non-Hispanic	83%	92%	91%	93%
Hispanic	17%	8%	8%	5%
Race				
White	77%	78%	77%	78%
African American	13%	13%	14%	13%
Asian	6%	4%	3%	3%
Other Race	3%	5%	5%	5%
NWS Region				
Northeast (CT, ME, MA, NH, NJ, NY, PA, RI, VT)	18%	18%	19%	18%
Midwest (IL, IN, IA, KS, MI, MN, MO, NE, ND, OH, SD, WI)	21%	20%	21%	20%
South (AL, AR, DE, DC, FL, GA, KY, LA, MD, MS, NC, OK, SC, TN, TX, VA, WV)	38%	41%	42%	41%
West (AK, AZ, CA, CO, HI, ID, MT, NV, NM, OR, UT, WA, WY)	23%	20%	18%	22%
Sample Size		**n = 834**	**n = 417**	**n = 417**

3.4 Methodology

For the analysis of risk and subjectivity as determinants on AI aversion this survey will utilize both an experimental treatment and a series of linear regression models to identify which of the two is a larger driver for AI aversion. For the experimental treatment survey participants were randomly assigned to one of two different groups. Both groups were asked to consider a series of current and hypothetical future uses of AI in different policy domains.

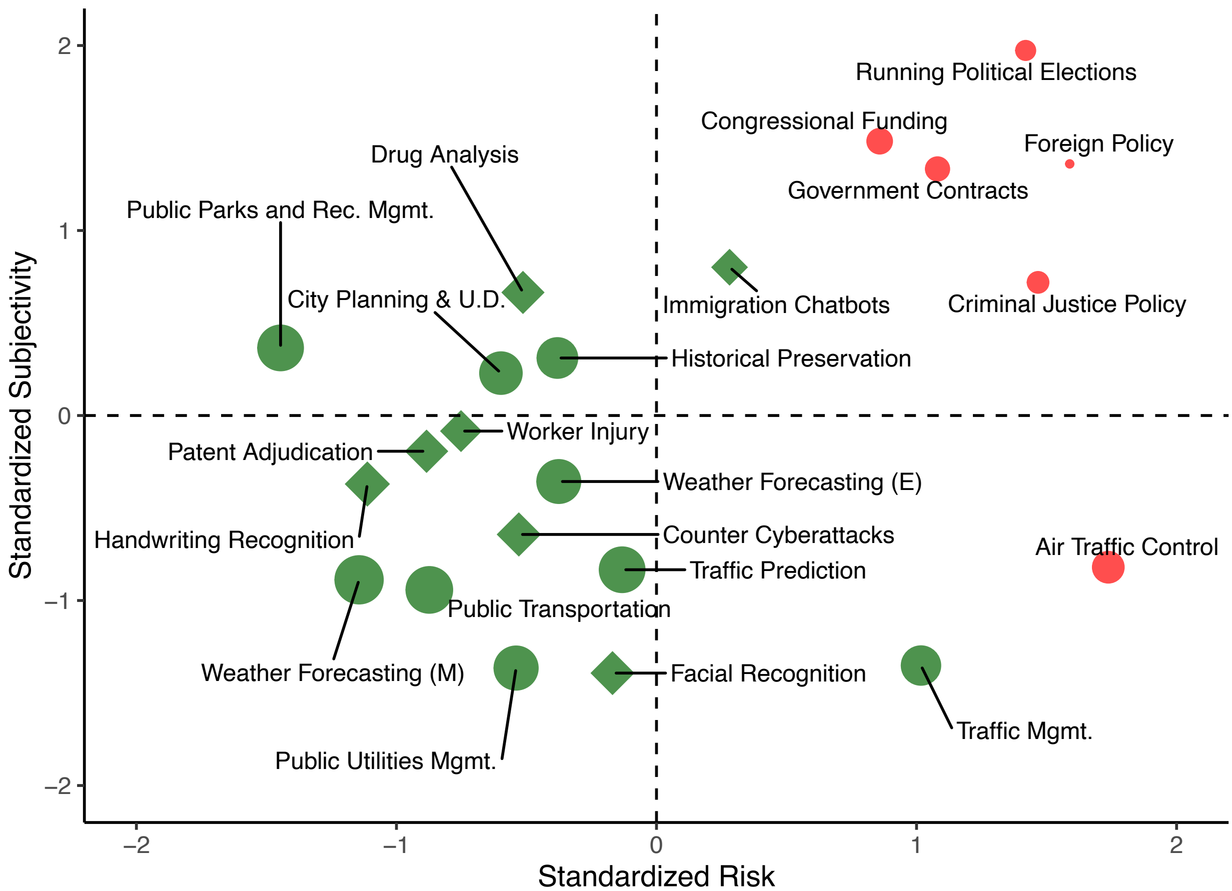
The first group was tasked with assessing their perceived risk for each of the different domains on a seven-point scale where 1 means “Not dangerous at all” and 7 means “Very dangerous”. The second group was with ranking the different domains on their perceived subjectivity where a 1 means “Objective” and a 7 means “Subjective.”

Both experiment groups were also asked to rank the same policy domains by how much they would support the AI being used. This will comprise the AI support variable used in the regression analysis. This analysis also utilizes the AIAI score generated for each of the respondents as described in the previous chapter of this dissertation. The AIAI is compiled from responses to a series of vignettes where the respondents are asked to consider diverse policy applications of the use of AI. Finally, the perceived risk and perceived subjectivity results from the experiment will be analyzed with exploratory factor analysis (EFA), just like in the previous chapter, to establish the unidimensionality of the results. Regression analysis will then be done to identify the relationship perceived risk and subjectivity have on peoples score on the AIAI.

3.5 Results

The results of the experiment are shown in Figure 3.1, as well as broken down further in the Appendix (Table A6, Figure A1, and Figure A2). Figure 3.1 shows the mapping of the standardized mean perceived risk and standardized mean perceived subjectivity of each of the different current and future policy domains as ranked by the survey participants. The size of each point shows the mean standardized support for each of the different domains, with ones that had more support than opposition shown in green and ones with more opposition than support shown in red. These standardized means are calculated as z-scores. The shape of each point also demarcates if the domain is one that is currently being used by the government (diamond) or could potentially be used in the future (circle). Interestingly the only domains with more opposition then support are ones with higher mean standardized perceived risk. However, there is clearly a relationship with subjectivity as well since *Traffic Management* is shown to have high mean standardized perceived risk yet still sees more support than opposition. The relationship between these two variables with support for AI will be explored in more detail in the following sections.

Figure 3.1 Standardized Experiment Results

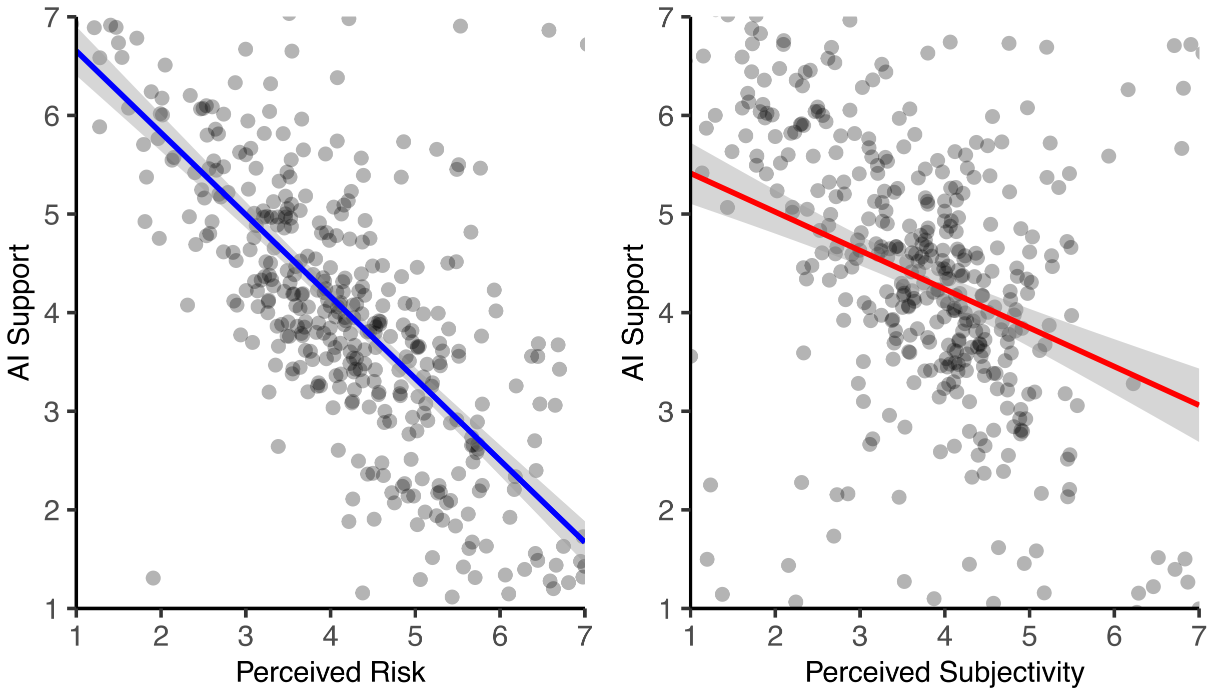


Note: The size of each point shows the mean standardized support for each domain, with higher support than opposition shown in green and higher opposition than support shown in red. These standardized means are calculated as z-scores. The shape of each point also demarcates if the domain is one that is currently being used by the government (diamond) or could potentially be used in the future (circle).

3.5.1 AI Support Regression Analysis

The first regression model looks at the relationship between support for AI with perceived risk and perceived subjectivity. This utilizes a direct comparison of AI support, perceived risk, and perceived subjectivity for the same series of real and hypothetical policy domains. The results of this regression model are displayed in Figure 3.2 and broken down in Table 3.3.

Figure 3.2 Perceived Risk & Subjectivity Effect on AI Support



[2]

Figure 3.2 highlights the difference in the relationship between support for AI and perceived risk and subjectivity. Support is measured between 1 and 7 with a 7 meaning completely support and a 1 meaning completely oppose. Risk and subjectivity are both also measure on a 1 to 7 scale where a 1 means not dangerous at all and a 7 means very dangerous with regards to risk and for subjectivity a 1 means objective and a 7 means subjective. Figure 3.2 shows that both perceived risk and subjectivity are negatively correlated with support for AI. Table 3.2 shows that for risk and support for AI the correlation is -0.754, while for subjectivity and support for AI the correlation is -0.341.

[2] Robust regression analysis was also performed with this data to ensure accuracy. The results did not differ substantially from form the linear regression model shown here.

Table 3.2 Correlation of AI Support with Risk and Subjectivity

	Correlation Coefficient	p-Value	Lower Confidence Interval	Upper Confidence Interval
Risk	-0.754	<0.001	-0.792	-0.701
Subjectivity	-0.341	<0.001	-0.423	-0.253

Table 3.3 shows the coefficient for risk was estimated to be approximately -0.830 with a standard error of 0.036 and a p-value well below 0.05. This indicates a statistically and substantively significant negative relationship between risk and levels of support for AI. This would mean that holding all else equal a 1 unit increase in the perceived risk of AI use in a policy domain equates to decrease in support for AI by 0.830 units.

Table 3.3 Comparing Mean AI Support with Mean Risk and Mean Subjectivity

	Dependent variable:	
	Artificial Intelligence Support	
	(1)	(2)
Risk	-0.830*** (0.036)	--
Subjectivity	--	-0.392*** (0.053)
Constant	7.484*** (0.160)	5.806*** (0.207)
Observations	417	417
R2	0.568	0.116
Adjusted R2	0.567	0.114
Residual Std. Error (df =417)	1.010	1.395
F Statistic	545.518***	54.444***
Note	***p<0.1; **p<0.05, ***p<0.01**	

Table 3.3 also shows that subjectivity is also both statistically and substantively significant with an estimated coefficient to be approximately -0.392 with a standard error of 0.053 and a p-value well below 0.05. For subjectivity this would mean that when holding all else equal a 1 unit increase in the perceived subjectivity of the domain AI is being used in translates to a decrease in support for AI in that domain by 0.392 units.

In examining the regression results in Table 3.3, it becomes evident that the perceived risk is playing a more prominent role in shaping support for AI than subjectivity is. The correlation between risk and the AI support is -0.754, reflecting a substantial negative relationship. This is reinforced by the negative coefficient of just over -0.8 showing that as perceived risk increases within policy domains, the support for AI decreases. In contrast, the correlation between subjectivity and the AI support, while still statistically significant, demonstrates a relatively smaller effect with a coefficient of -0.341. This also is reinforced by the smaller negative coefficient at just under -0.4, about half the coefficient for perceived risk. This would mean that while an increase in either perceived risk or perceived subjectivity, holding all else constant, would both cause decreases in the AI support, perceived risk is driving change at nearly twice the rate. This is also further reinforced by the considerably lower R-squared value for the regression model with subjectivity as the predictor (0.116) compared to the model with risk (0.568) which underscores the stronger explanatory power of perceived risk in predicting AI support.

3.5.2 Risk and Subjectivity Factor Analysis

Before I analyze the relationship between the AIAI and perceived risk and subjectivity I first need to establish the quality of my measure for risk and subjectivity. To do that, I once again use parallel analysis to identify the factor loadings of perceived risk (Figure 3.3) and perceived subjectivity (Figure 3.4). In both instances the eigen values on 1 factor are very high, indicating a very strong likelihood that a single factor is being measured by these questions. However, in each of the tables 2 factors is shown to be slightly above the cutoff of 1, indicating potential for 2 factors underlying the results. Table 3.4 and Table 3.5 show the results of Exploratory Factor Analysis (EFA) for both of these results, respectively.

Figure 3.3 Parallel Analysis Scree Plot for Perceived Risk

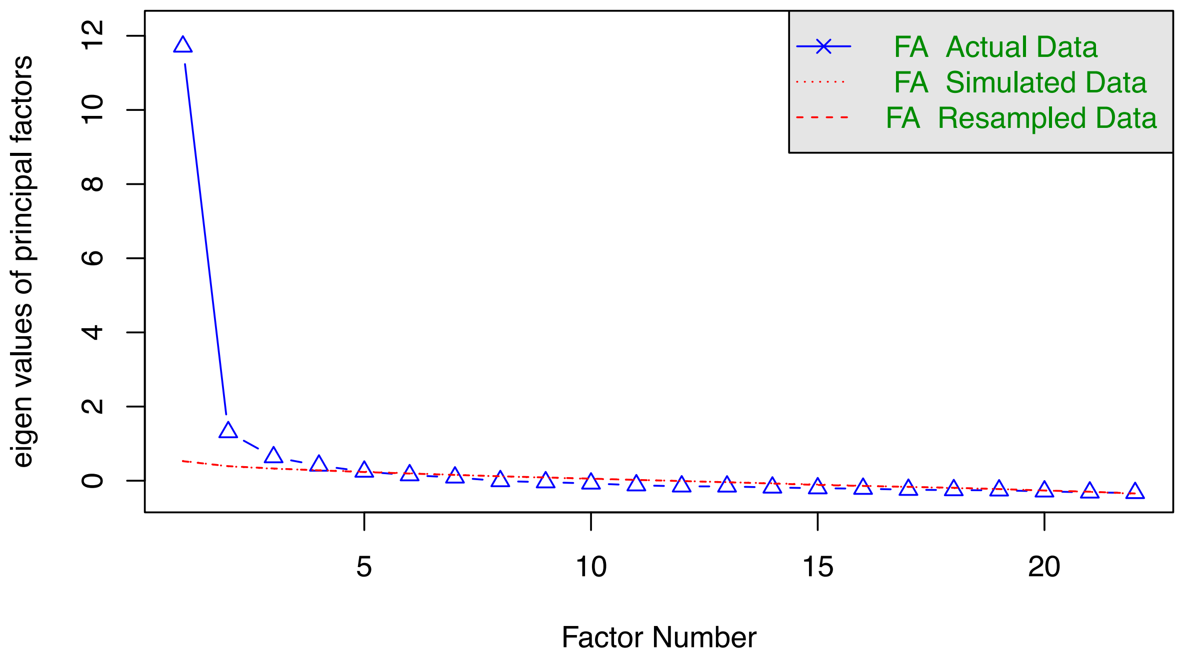


Table 3.4 EFA Comparison for Perceived Risk

Model	Explained Var. 1st Factor	Explained Var. 2nd Factor	Tucker Lewis Index	RMSEA	Confidence Interval
One - Factor	53%	--	0.719	0.147	0.142 – 0.153
Two - Factor	54%	7%	0.82	0.118	0.112 – 0.124

Perceived Risk

The EFA results for perceived risk, as detailed in Table 3.4, show that the model with one factor explained 53% of the variance, with a Tucker Lewis Index (TLI) of 0.719 and a Root Mean Square Error of Approximation (RMSEA) of 0.147. The inclusion of a second factor marginally increased the explained variance to 54% and improved the TLI to 0.82 and RMSEA to 0.118. These improvements, while notable, did not significantly alter the model's overall structure. Given the TLI's proximity to the commonly accepted threshold of 0.90 and the

RMSEA's approach towards 0.10, it is reasoned that the perceived risk measure is sufficiently unidimensional for the purpose of this study. Therefore, the one-factor model is adopted, attributing minor importance to the additional variance explained by the second factor.

Figure 3.4 Parallel Analysis Scree Plot for Perceived Subjectivity

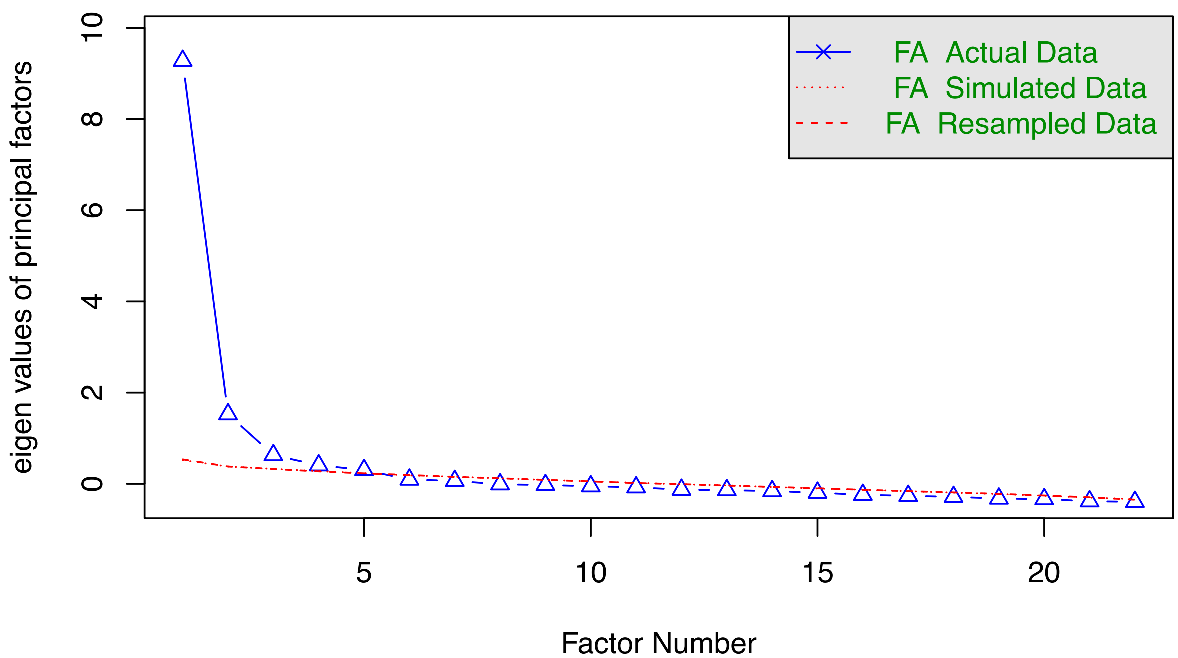


Table 3.5 EFA Comparison for Perceived Subjectivity

Model	Explained Var. 1st Factor	Explained Var. 2nd Factor	Tucker Lewis Index	RMSEA	Confidence Interval
One - Factor	42%	--	0.697	0.128	0.123 – 0.134
Two - Factor	43%	8%	0.836	0.094	0.088 – 0.101

Perceived Subjectivity

Similarly, the perceived subjectivity measure, as shown in Table 3.5, displayed a one-factor model accounting for 42% of the variance, with a TLI of 0.697 and an RMSEA of 0.128. Introducing a second factor marginally improved the variance explanation to 43% and enhanced the TLI to 0.836 and RMSEA to 0.094. Despite these improvements, the essential character of the model remained consistent with the one-factor solution. The TLI and RMSEA values, although not meeting the ideal thresholds, are sufficiently close to suggest a predominantly unidimensional construct. Hence, the one-factor model is chosen for its simplicity and adequacy in capturing the essence of perceived subjectivity.

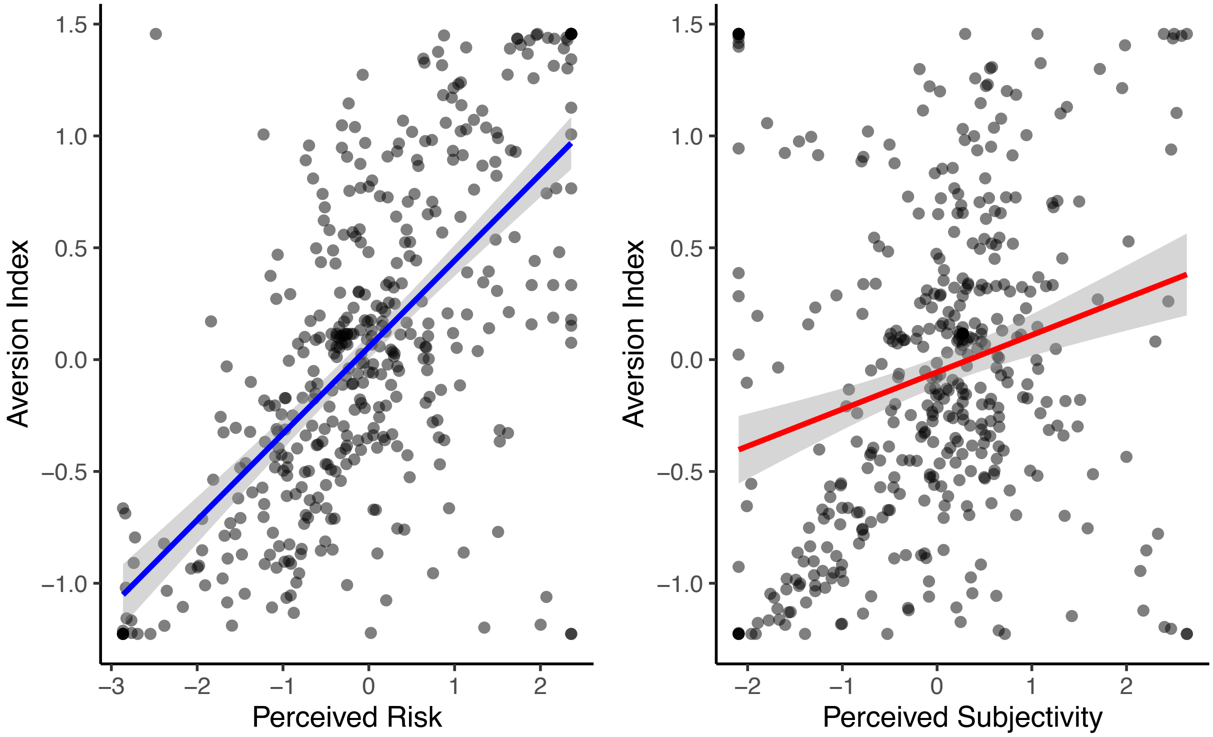
Implications for Analysis

In light of the EFA findings for both perceived risk and subjectivity, this research opts for the simpler, more parsimonious one-factor models for each construct. The choice is supported by the minimal influence of a second factor on the overall model fit and the near-adequacy of the fit indices. This decision allows for a focused analysis on the core elements of risk and subjectivity perceptions in the context of AI, using their factor scores to explore their relationships with the aversion index. The results highlight the nuanced but predominantly singular nature of public perceptions in these domains, a crucial consideration for comprehending and addressing AI-related apprehensions in policy and practice.

3.5.3 Artificial Intelligence Aversion Index

This next regression model will look at the relationship between the AIAI with perceived risk and perceived subjectivity. This will utilize a direct comparison of the factor score of the index generated from the previous chapter with the perceived risk and perceived subjectivity factor scores for the same series of real and hypothetical policy domains. It is important to point out that since the AIAI measures the level of aversion people have towards AI an increase in it is equivalent to a decrease in the level of support for AI that individual might have. This is why while the direction of the regression is inverted compared to the previous regression, the implications and interpretation is the same.

Figure 3.5 Perceived Risk & Subjectivity Effect on AI Aversion Index



[3]

Figure 3.5 displays the relationship between the factor scores for AIAI and the perceived risk and subjectivity of different adoptions of AI within policy domains. The index score is the same from the previous chapter of this dissertation, an aggregation of individuals level of support for a series of vignettes where AI is being utilized in different policy making domains. The correlation between risk and the AIAI is 0.647 while for subjectivity and the AIAI the correlation is 0.240, shown in Table 3.6. Both perceived risk and perceived subjectivity are positively correlated with the index score for people showing that as their perception of risk or subjectivity increases there score on the aversion index also increases, meaning they become more averse to the use of AI.

[3] Robust regression analysis was also performed with this data to ensure accuracy. The results did not differ substantially from form the linear regression model shown here.

Table 3.7 shows the coefficient for risk was estimated to be 0.387 with a standard error of 0.022 and a p-value well below 0.05. This indicates a statistically and substantively significant relationship between risk and the AIAI. This would mean that holding all else equal a 1 standard deviation increase in the perceived risk of AI an individual has equates to a 0.387 standard deviation increase in the AIAI score.

Table 3.6 Correlation of Aversion Index with Risk & Subjectivity

	Correlation Coefficient	p-Value	Lower Confidence Interval	Upper Confidence Interval
Risk	0.647	<0.001	0.588	0.700
Subjectivity	0.240	<0.001	0.147	0.328

Table 3.7 Factor Score Regression - Aversion Index, Perceived Risk & Subjectivity

	Dependent variable:	
	Artificial Intelligence Aversion Index:	
	(1)	(2)
Risk	0.387*** (0.022)	--
Subjectivity	--	0.166*** (0.033)
Constant	0.056** (0.027)	-0.056* (0.034)
Observations	417	417
R2	0.419	0.057
Adjusted R2	0.417	0.055
Residual Std. Error (df =417)	0.559	0.692
F Statistic	298.880***	25.307***
Note	***p<0.1; **p<0.05, ***p<0.01**	

Table 3.7 also shows that subjectivity is also both statistically and substantively significant with an estimated coefficient to be approximately 0.166 with a standard error of 0.033 and a p-value that is also well below 0.05. For subjectivity this would mean that when holding all else equal a 1 standard deviation increase in the perceived subjectivity of the domain AI is being used in translates to an increase in the aversion index score by 0.166 standard deviations.

In examining the correlation coefficients and regression results in Table 3.7, it becomes evident that by looking at the factor scores risk plays a more prominent role in shaping AI aversion than subjectivity. The correlation between risk and the AIAI stands at 0.647, reflecting a substantial relationship. This is reinforced by the coefficient of nearly 0.4 showing that as perceived risk increases within policy domains, the level of aversion to AI increase significantly. In contrast, the correlation between subjectivity and the aversion index, while still statistically significant, demonstrates a smaller effect with a correlation of 0.240 and a coefficient of 0.166. This would mean that while an increase in either perceived risk or perceived subjectivity, holding all else constant, would both cause decreases in the AIAI, perceived risk is driving change at a much faster rate.

The adjusted R-squared also highlights the different effect each of these variables has on aversion. The adjusted R-squared value for the model with risk as the predictor is 0.417, indicating that nearly 42% of the variance in AIAI is explained by the perceived risk alone. This substantial value highlights the significant role that perceived risk plays in shaping aversion to AI. In stark contrast, the adjusted R-squared value for the model with subjectivity as the predictor is much lower, standing at only 0.055. This implies that perceived subjectivity accounts for just about 5.5% of the variance in the AIAI.

These disparities in the correlation, coefficients, and the adjusted R-squared values illustrate that, while both perceived risk and subjectivity contribute to AI aversion, risk perception is a far more potent driver. The substantial difference in their explanatory powers underscores the predominant influence of perceived risk in public aversion towards AI. This finding suggests that concerns about the risks associated with AI, more than its perceived subjectivity, are key determinants in shaping public attitudes towards the adoption and utilization of AI technologies in various policy domains.

3.6 Conclusion

In this chapter of the dissertation, the exploration of Artificial Intelligence (AI) aversion has been expanded upon, focusing on the pivotal roles of perceived risk and subjectivity. Building upon the foundation laid by the creation of the Artificial Intelligence Aversion Index (AIAI) in the previous chapter, this segment has delved into the intricate dynamics of AI aversion in various policy domains. The aim was to identify the parallels and distinctions between AI aversion and the well-studied concept of algorithmic aversion, thereby enriching the understanding of public resistance to AI integration within diverse policy frameworks.

The primary conclusion drawn from this analysis is the dominant influence of perceived risk over AI aversion. Through rigorous evaluation of survey data, factor analysis, and linear regression, it became clear that perceived risk plays a more critical role in shaping aversion than the subjectivity of the policy domains where AI is implemented. This is substantiated by the significant negative correlation between perceived risk and support for AI (correlation coefficient of -0.754) and a notable positive correlation with the AIAI (correlation coefficient of 0.647), as presented in Table 3.3 and Table 3.7. On the other hand, subjectivity, though a contributing factor to AI aversion, exerts a comparatively weaker influence, with a correlation coefficient of -0.341 with AI support and 0.240 with the AIAI. This is also substantiated by the regression coefficients where both for support of AI as well as the AIAI perceived risk was more than twice the size of perceived subjectivity.

Further analysis via regression models reinforces these findings. The regression coefficient for perceived risk (AI support -0.830, AIAI 0.387) was more than twice the size of the regression coefficient for perceived subjectivity (AI support -0.392, AIAI 0.166) in both models. The model assessing risk as a predictor also accounts for a considerable proportion of the variance in AIAI (adjusted R-squared value of 0.417), highlighting the significant role of risk perception. In contrast, the model focusing on subjectivity explains a relatively minor portion of the variance (adjusted R-squared value of 0.055). These results underline the nuanced yet predominantly singular nature of public perceptions in these domains, with risk concerns overshadowing subjective perceptions of AI.

As this dissertation progresses to chapter 4, the focus will transition to examining demographic influences on AI aversion. This next phase will investigate how factors such as race, gender, age, and other demographic variables modulate attitudes towards AI. Leveraging the insights obtained from this chapter, the forthcoming analysis is poised to offer a comprehensive understanding of AI aversion across diverse societal segments. This endeavor is imperative for developing strategies that not only maximize AI's potential but also address the varying concerns of a heterogeneous society, fostering informed acceptance and supportive integration of AI into public policy and other sectors.

Chapter 4: Demographic Influences on Perceived Risk, Subjectivity, and Artificial Intelligence Aversion

4.1 Introduction

Chapters two and three of this dissertation focused on establishing a foundational understanding of AI aversion, exploring its various dimensions, and the underlying reasons that drive it such as perceived risk and perceived subjectivity. This was done through the development and subsequent testing of the AIAI. Having examined some of the influence of the different perceptions towards AI aversion it now becomes necessary to expand the research to other potential drivers towards differential AI aversion levels. Past research has shown that demographic differences contributes towards differential technology adoption (Venkatesh et al. 2003) as well as risk perception (Finucane, Slovic, Mertz, Flynn, & Satterfield 2000) so that will be the next subject of examination for this research.

In this chapter, the exploration focuses on a more granular analysis of how demographic characteristics such as gender, age, race, political affiliation, education level, and cultural theory categories (Wildavsky 1987) influence and potentially mitigate people's level of AI aversion. This chapter aims to dissect the complex interplay between these demographic factors and perceptions of risk and subjectivity in the context of AI, providing a deeper understanding of the diverse attitudes towards this transformative technology.

4.2 Theoretical Framework

The intersection of demographics and technology perception has been a subject of extensive research, highlighting how various factors influence the adoption and perception of new technologies. Venkatesh et al. (2003) in their Unified Theory of Acceptance and Use of Technology (UTAUT) emphasize the pivotal role of demographic variables such as age and gender in influencing technology adoption decisions. They argue that these factors significantly impact an individual's perception of technology's ease of use and usefulness. Further research underscores the importance of individual differences in predicting technology use, suggesting that demographic factors can offer predictive insights into technology adoption and aversion (Rogers 2003).

Socio-Technical Systems Engineering, as discussed by Baxter and Sommerville (2010), provides a framework for understanding the dynamic interaction between technology and society. This theory suggests that technology is not merely a product of its technical components but also of the social context in which that technology is embedded. Leonardi (2012) extends this perspective to highlight how different social groups interact with technology, shaping its development and use within that society.

Additionally, Rogers' (2003) Diffusion of Innovations Theory provides valuable insights into how demographic factors influence the adoption and diffusion of new technologies. This theory posits that socio-economic status, age, education, and other demographic variables play crucial roles in determining the speed and extent to which new technologies are adopted by different groups. Different socio-economic groups interact with new technologies at different rates and in different ways, allowing for idiosyncrasies in their perspectives on that technology to develop.

Research by Finucane et al. (2000) and Wildavsky (1987) also helps to frame the exploration of demographic influences on AI aversion. Finucane et al.'s study on gender, race, and perceived risk, provides crucial insights into how demographic factors like gender and race can shape perceptions of risk, an essential aspect of AI aversion. This research suggests that demographic characteristics profoundly influence how different groups perceive the risks associated with AI. Finucane et al. identified what has come to be known as the “white male effect" which is that white males' tolerance of risk is distinct and often higher than other gender and racial groups, especially when it comes to technological risks (2000).

Wildavsky's cultural theory offers a different perspective, proposing that people's preferences, including those related to technology, are shaped by institutional and cultural constructs. This framework is particularly relevant to understanding how cultural backgrounds might influence attitudes towards AI, suggesting that aversion or acceptance of AI could be deeply rooted in cultural belief systems and worldviews. Together, these articles provide a deep theoretical basis for examining the intricate relationship between demographics, cultural theory, and AI aversion.

4.2.1 Gender, Risk, and AI Aversion

Gender differences in AI aversion are a complex and multifaceted topic. Past studies have consistently shown variations in technology perception and acceptance between men and women. For instance, a study by Morris and Venkatesh (2000) revealed that women tend to exhibit higher levels of perceived ease of use and usefulness when it comes to new technology adoption. In the context of AI, these gender-based perceptions can significantly influence attitudes and levels of aversion.

Conversely, additional research has demonstrated that women are more likely to express concerns about the ethical and social implications of AI as well as its utility, which could potentially lead to higher levels of aversion (Ahmer, Altaf, Khan, Bhatti, & Naseem 2023). This is in line with the broader literature suggesting that women, in general, show greater sensitivity to risk of technology (Finucane et al., 2000), which may translate into heightened risk perception regarding AI technologies.

Given these findings, a proposed hypothesis is that gender significantly influences perceived risk about AI and a person's level of AI aversion. Specifically, it can be hypothesized that women, due to their higher sensitivity to technologic risk and their ethical concerns, are likely to perceive greater risks associated with AI and hence exhibit higher levels of AI aversion compared to men.

H_1: Women will score higher on the AIAI than Men.

4.2.2 Generation and AI Aversion

Generational differences play a crucial role in shaping attitudes towards AI, with each generation bringing its unique set of experiences and perspectives. Gen Z, having grown up in a digital-first world, is generally more comfortable and optimistic about the integration of technology like AI in their lives. In contrast, Millennials, while technologically adept, tend to be more cautious, often weighing the benefits against potential risks and ethical concerns.

Gen X, having witnessed the transition to a digital era, tends to display a more pragmatic approach to AI, balancing optimism with a healthy skepticism about technological advancements. Boomers, who did not grow up with technology as an integral part of their lives, often exhibit higher levels of aversion due to unfamiliarity and concerns about the rapid pace of technological change.

This has been substantiated in past research, Williams, Anderson, & Drennan (2010) explored how different generations interact with technology, finding that Gen Z and Millennials are generally more adept and comfortable with emerging technologies. On the other hand, Prensky (2001) noted that older generations, such as Boomers, often face a steeper learning curve with new technology, influencing their perceptions and adoption rates.

Given these insights, it can be hypothesized that generational differences significantly impact perceived risks and aversion towards AI. Younger generations, accustomed to rapid technological advancements, may exhibit lower levels of AI aversion, while older generations, who have not grown up with such technologies, may perceive higher risks and demonstrate greater aversion.

H_2: Gen Z and Millennials will score lower on the AIAI than Gen X and Boomers.

4.2.3 Race, Risk, and AI Aversion

The exploration of AI aversion across racial groups involves understanding the nuanced ways in which cultural and social factors influence technology perception. Studies have shown that experiences and societal contexts significantly shape the way that different racial groups perceive and interact with AI (Obermeyer, Powers, Vogeli, & Mullainathan 2019). For instance, Black and other minority groups often express concerns about biases and fairness in AI systems, reflecting broader societal issues of inequality and discrimination. These concerns can be exacerbated by high profile instances of discrimination involved in the utilization of AI such as was the case with PATTERN which disproportionately flagged Black inmates at being higher risk of recidivism and then recommended against moving them forward with parole (Partnership on AI, 2020).

White individuals, in contrast, may have different concerns or perspectives, influenced by their societal experiences and cultural context. These differences in perception are crucial in understanding the broader landscape of AI aversion.

The exploration of AI aversion across racial groups, considering cultural and social factors, gains further depth with the inclusion of Finucane et al. (2000) study. Their research, focusing on gender, race, and perceived risk, known as the 'white male effect,' reveals that perceptions of risk and technology can vary significantly across different racial groups. This study suggests that racial minorities, who often face systemic inequalities, might perceive higher risks associated with AI technologies, potentially leading to increased aversion. Combined with Obermeyer et al.'s findings, this underscores the importance of understanding diverse racial perspectives in AI perception and adoption.

Based on these insights, a proposed hypothesis is that one's race significantly influences perceived risks about AI and the level of AI aversion. Specifically, it can be hypothesized that racial minorities, due to concerns about systemic biases and fairness, may perceive higher risks associated with AI, potentially leading to greater levels of aversion compared to White individuals.

H_3: White individuals will score lower on the AIAI than Black individuals or other minorities.

4.2.4 Politics, Risk, and AI Aversion

Political ideology plays a significant role in shaping perceptions of AI. Research indicates that Democrats, Republicans, and Independents may have different attitudes towards technology based on their ideological beliefs. For instance, a study by Smith and Anderson (2019) found that conservatives (often Republicans) tend to be more skeptical of the impacts of science and technology compared to liberals (often Democrats). This is especially true in instances where perceptions of bias with the technology systems are higher, such as social media.

Independents, who might not align strictly with either ideology, could display varied perceptions, influenced by specific aspects of AI rather than an overarching political ideology. Given this, it can be hypothesized that political ideology significantly influences perceived risk about AI and the level of AI aversion. Conservatives may perceive higher risks and exhibit more aversion due to skepticism towards rapid technological changes, while liberals may be more accepting of AI advancements. This can in part be due to perceptions about political biases either in the developers of the technology or in the data used by the AI systems to develop and train.

H_4: Republicans will score higher on the AIAI than Democrats.

4.2.5 Education, Risk, and AI Aversion

Education level is a critical factor in shaping perceptions and attitudes towards technologic developments. Studies have shown that for individuals with higher levels of educational attainment, such as those with graduate degrees, they are generally more receptive to new technologies, including AI. These individuals tend to have a better understanding of the benefits and risks associated with this technology, leading to a more balanced view (Bucchi & Neresini, 2008). In contrast, those with lower educational levels might have had less exposure to newer technologies as well as lower levels of information about AI, potentially leading to higher levels of aversion due to uncertainty or misconceptions.

Therefore, it can be hypothesized that people with higher education levels are associated with lower perceived risk and aversion towards AI. This is because individuals with more education are likely to have a greater understanding of AI, leading to more informed and less fearful perceptions of these technologies as well as more experience with new and advanced technologies.

H_5: People with graduate degrees will score lower on the AIAI than people with only associates or bachelor's degrees.

H_6: People with graduate degrees will score lower on the AIAI than people with no college degree.

H_7: People with only an associates or bachelor's degree will score lower on the AIAI than people with no college degree.

4.2.6 Cultural Theory, Risk, and AI Aversion

Cultural Theory, as proposed by Wildavsky (1987), offers a unique framework for examining AI aversion. It identifies two metrics by which to evaluate how individuals relate to society around them, group and grid. Group refers to an emphasis towards community and cooperative systems, while people who are low in group prefer individualistic systems and self-reliance. Grid refers to how tightly structured people view society to be with people high in grid seeing society as a more hierarchical and rules-based system. Within these two metrics, cultural theory categorizes individuals into four groups: Hierarchical, Egalitarian, Fatalist, and Individualist, each with distinct worldviews affecting their perception of technology and risk. Hierarchists (high group, high grid), who value order and authority, might view AI as a tool for maintaining structure, leading to lower levels of aversion. Egalitarians (high group, low grid), concerned with equality, could fear AI exacerbating social disparities and therefore experience higher levels of aversion. Fatalists (low group, high grid), feeling powerless to influence outcomes, might be indifferent or resigned to AI risks. Individualists (low group, low grid), valuing autonomy and self-reliance, might embrace AI for personal gain as a tool that they themselves could utilize.

Based on these cultural perspectives, a hypothesis can be formulated: An individual's alignment with these cultural worldviews will significantly influence their perceived risk about AI and the level of AI aversion.

H_8: Hierarchalists and Individualists will score lower on the AIAI than Egalitarians and Fatalists.

4.2.7 The White Male Effect and AI Aversion

Finally, this chapter will look at the intersection of race and gender and how that intersection effects AI aversion. As highlighted previously multiple different studies have been conducted to show that an individual's race and gender can be predictive to how the relate to risk as well as the adoption of new technology. Finucane et al. (2000) focused on the intersection of these two demographic variables when they identified the white male effect or that "white males are less likely to rate a hazard as posing a 'high risk'." The hypothesis that can be generated for this examination will be a combination of the ones generated for race and gender: Both an individual's race and gender will significantly influence their perceived risk about AI and their level of AI aversion

H_9: White men will score lower on the AIAI than any other racial/gender group.

4.3 Data

This chapter will continue to make use of the primary survey of this dissertation to evaluate the relationship between different demographics and AI aversion. Once again this was a national survey of 882 total responses. The responses were collected in partnership with Lucid. Responses were curates such that those who were deemed to be speeding, finishing the survey in under 4 minutes when the mean length was 14, were removed. Those with IP addresses that were located outside of the United States were also removed, in accordance with the methods laid out by Dennis, Goodson, & Pearson (2018). The final total response rate was 834.

Table 4.1 shows the breakdown of survey respondents into the groups outlined in the different hypothesis from the earlier section. This table is slightly different from the ones in previous chapters because age is broken down by generation and not census age group and Asians are combined with the Other race category to allow for statistical analysis. Generations were used as the metric for age instead of the census format to enable the hypothesis testing outlined in the previous section. Political party, highest levels of education completed, and the different cultural theory groups are also displayed. Political party and highest level of education completed are collected by lucid prior to the application of the survey, while cultural theory group was determined by the survey which is explained in the following methods section.

Table 4.1 Demographic Characteristics of Survey Sample

	Survey Sample
Gender	
Female	59.7%
Male	40.3%
Age	
Gen Z (18 to 26)	9%
Millennial (27 to 42)	32%
Gen X (43 to 58)	28%
Boomers 2 (59 to 68)	16%
Boomers 1 (69 +)	15%
Race	
White	78%
African American	13%
Other	9%
Political Party	
Democratic	50%
Republican	32%
Independent	13%
Prefer not to say	6%
Education	
No college degree	45%
Associates/Bachelor's	37%
Graduate Degree	17%
Cultural Theory	
Hierarchical	18%
Egalitarian	22%
Fatalist	14%
Individualist	25%
Sample Size	**n = 834**

4.4 Methodology

A two-pronged approach was necessary to evaluate these hypotheses. First, I had to identify from the data I had available which cultural theory group respondents could be categorized into. Second, I utilize multiple linear regression in order to identify the impact of demographic differences on the three different dependent variables of my research: perceived risk, perceived subjectivity, and AI Aversion.

To determine the cultural theory groups survey respondents were asked a series of 4 questions describing an outlook on life and asked to evaluate how accurately they each aligned with them on a scale from 0 to 10 were 0 means not at all, and 10 means completely. For this section respondents were categorized only if one of the four categories had the highest value, otherwise they were excluded from the analysis. The questions and the frequency of the responses to them are shown in Table 4.2.

Having established which cultural theory group to place individuals in this research now turns to evaluating the influence of these different demographic variables on AI aversion. This was achieved through multiple linear regression models, aimed at quantifying the relationship between demographics and three key aspects: perceived risk, perceived subjectivity, and the AI Aversion Index.

Initially, the study focused on constructing a dataset comprising various demographic groups, including gender, age, race, political affiliation, education level, and cultural theory categories. Each of these demographics was hypothesized to have an impact AI perception. To assess these impacts, multiple linear regression models were employed. The models were structured to predict each of the three dependent variables - perceived risk, perceived subjectivity, and AI aversion - based on each of the demographic categories. Perceived risk and AI aversion were selected because they are directly hypothesized to vary with different demographic variables. Perceived subjectivity was also selected to be analyzed in relationship to these demographic variables because it was an important predictor for AI aversion as identified in chapter three of this dissertation. The regression analysis also included two additional control variables, knowledge about AI and experience with AI. These were included in the regression analysis in an attempt to minimize the effect of novelty of the artificial intelligence was having on people's levels of aversion.

Table 4.2 Cultural Theory Answer Frequency

Hierarchical	I am more comfortable when I know who is, and who is not, a part of my group, and loyalty to the group is important to me. I prefer to know who is in charge and to have clear rules and procedures; those who are in charge should punish those who break the rules. I like to have my responsibilities clearly defined, and I believe people should be rewarded based on the position they hold and their competence. Most of the time, I trust those with authority and expertise to do what is right for society.										
Mean	0	1	2	3	4	5	6	7	8	9	10
5.94	7%	2%	5%	7%	6%	16%	9%	14%	17%	6%	12%
Egalitarian	My most important contributions are made as a member of a group that promotes justice and equality. Within my group, everyone should play an equal role without differences in rank or authority. It is easy to lose track of what is important, so I have to keep a close eye on the actions of my group. It is not enough to provide equal opportunities; we also have to try to make outcomes more equal.										
Mean	0	1	2	3	4	5	6	7	8	9	10
5.81	8%	3%	4%	6%	9%	16%	9%	11%	14%	6%	14%
Fatalist	Life is unpredictable and I have very little control. I tend not to join groups, and I try not to get involved because I can't make much difference anyway. Most of the time other people determine my options in life. Getting along is largely a matter of doing the best I can with what comes my way, so I just try to take care of myself and the people closest to me. It's best to just go with the flow, because whatever will be will be.										
Mean	0	1	2	3	4	5	6	7	8	9	10
5.15	10%	5%	7%	10%	8%	13%	10%	13%	10%	6%	9%
Individualist	Groups are not all that important to me. I prefer to make my own way in life without having to follow other peoples' rules. Rewards in life should be based on initiative, skill, and hard work, even if that results in inequality. I respect people based on what they do, not the positions or titles they hold. I like relationships that are based on negotiated "give and take," rather than on status. Everyone benefits when individuals are allowed to compete.										
Mean	0	1	2	3	4	5	6	7	8	9	10
6.60	2%	1%	4%	5%	8%	14%	10%	15%	13%	9%	19%

The regression analysis involved examining the relationship between each of the different demographics and the dependent variables in each model. The final dependent variable, AI aversion, was analyzed on three separate models, AIAI (1) had the same independent variables as the other two models for perceptions. AIAI (2) controlled for perceived risk in addition to the demographic variables. AIAI (3) controlled for perceived subjectivity in addition to the demographic variables. Due to the experiment model of the survey it was not possible to create a model controlling for both of these perceptions because respondents were randomly assessed on only one of these perception measures. In example, for gender, five separate models were constructed: one for perceived risk, one for perceived subjectivity, one for the AIAI (1), one for AIAI controlling for perceived risk (2), and one for AIAI controlling for perceived subjectivity (3). The outputs, including regression coefficients and their respective significance levels, were then meticulously tabulated. Table 4.3 provides a clear, comparative view of these influences across different demographic segments.

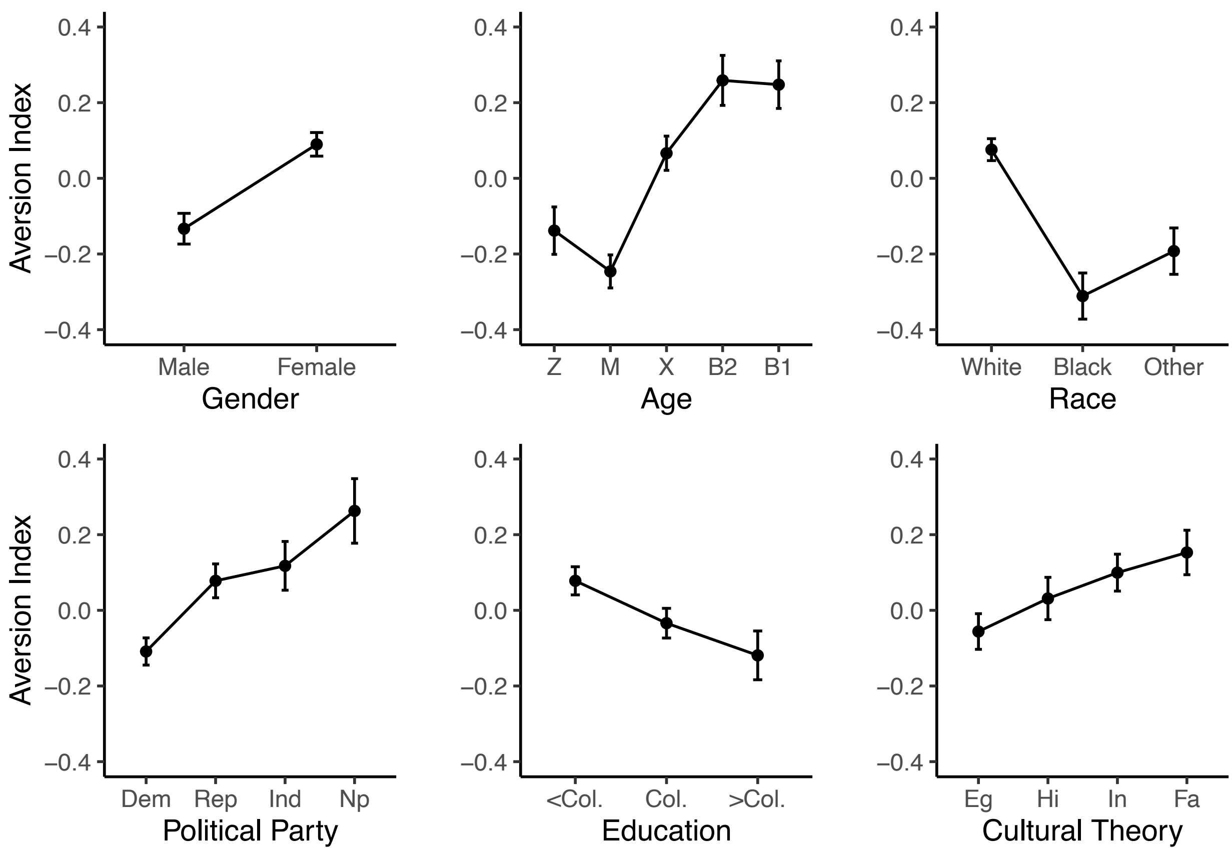
This same process was done for the analysis of gender and race together (**H_9**). Four different possible demographic combinations were created to enable this analysis: White Female, White Male, Black Female, and Black Male. The regression model for perceived risk, perceived subjectivity, and AIAI for these four categories are shown in Table 4.4. Other racial groups were not included in this analysis due to data limitations with the size of gender specific racial groups making proper analysis impossible. Once again, knowledge about AI and experience with AI were included in these regressions to minimize the effect that the novelty of artificial intelligence was having on people's aversion.

4.5 Results

The results of the multiple linear regressions for demographic predictions are shown in Figure 4.1 as well as in Table 4.3. The table shows the relationship between demographics and the three dependent variables: Perceived Risk, Perceived Subjectivity, and AI aversion. AI aversion is shown in three models, (1) without perceived risk or perceived subjectivity, (2) with perceived risk, and (3) with perceived subjectivity. Gender differences were notable, with females exhibiting a significantly higher scores on AI aversion only when risk perceptions were controlled for in the model. Interestingly, gender is the only demographic group which shows statistically significant findings for perceived subjectivity.

Age differences also yielded significant findings, with all groups older than Millennials showcasing significantly higher levels of perceived risk and higher scores on the AIAI. Gen Z the only group younger than Millennials was statistically indistinct from millennials across all models.

Figure 4.1 Demographics and AI Aversion



Note: **Age Figure**: Z represents Generation Z (age 18-26), M represents Millennials (age 27 – 42), X represents Generation X (age 43- 58), B2 represents Boomers 2 (age 59 – 68), and B1 represents Boomers 1 (age 69+). **Cultural Theory Figure**: Eg represents Egalitarians, Hi represents Hierechicalists, In represents Individualists, and Fa represents Fatalists.

Table 4.3 Multiple linear regression models predicting AI Aversion

		Perceived Risk	Perceived Subjectivity	AIAI (1)	AIAI (2)	AIAI (3)
Sex	Female (vs. Male)	0.052 (0.101)	-0.241** (0.108)	0.063 (0.052)	0.114* (0.061)	0.032 (0.074)
Age	Gen Z (vs. Millennial)	0.222 (0.167)	0.165 (0.196)	0.093 (0.091)	0.067 (0.101)	-0.032 (0.133)
	Gen X (vs. Millennial)	0.231* (0.126)	0.008 (0.146)	0.166** (0.067)	-0.002 (0.077)	0.182* (0.099)
	Boomers 2 (vs. Millennial)	0.622*** (0.152)	0.074 (0.171)	0.379*** (0.081)	0.241** (0.094)	0.230** (0.116)
	Boomers 1 (vs. Millennial)	0.336* (0.183)	0.068 (0.170)	0.192** (0.087)	0.022 (0.111)	0.224* (0.115)
Race	White (vs. Black)	0.114 (0.151)	-0.004 (0.166)	0.277*** (0.079)	0.236** (0.091)	0.252** (0.112)
	Other (vs. Black)	-0.026 (0.214)	-0.259 (0.230)	0.060 (0.111)	0.094 (0.129)	0.149 (0.156)
Political Party	Republican (vs. Dem)	0.113 (0.114)	0.159 (0.119)	0.059 (0.058)	0.073 (0.069)	-0.015 (0.081)
	Independent (vs. Dem)	0.267* (0.151)	0.050 (0.161)	0.118 (0.077)	0.167* (0.092)	-0.076 (0.109)
	No Party (vs. Dem)	0.439** (0.207)	-0.141 (0.230)	0.248** (0.108)	0.176 (0.126)	0.191 (0.156)
Education	<Col. (vs. >Col)	0.212 (0.141)	0.061 (0.149)	0.103 (0.072)	-0.017 (0.085)	0.116 (0.101)
	Col. (vs. >Col)	0.102 (0.135)	-0.000 (0.149)	-0.002 (0.071)	-0.087 (0.081)	0.035 (0.101)
Cultural Theory	Hierarchical (vs. Egalitarian)	0.080 (0.134)	0.169 (0.146)	0.131* (0.070)	0.157* (0.081)	0.052 (0.099)
	Individualist (vs. Egalitarian)	0.151 (0.122)	0.155 (0.131)	0.134** (0.063)	0.080 (0.074)	0.096 (0.089)
	Fatalist (vs. Egalitarian)	0.188 (0.146)	0.068 (0.155)	0.168** (0.075)	0.099 (0.088)	0.121 (0.104)
Control	Knowledge	-0.015 (0.020)	-0.062** (0.024)	-0.022** (0.011)	-0.016 (0.012)	-0.013 (0.016)
	Experience	-0.167** (0.069)	-0.160* (0.082)	-0.239*** (0.038)	-0.128*** (0.042)	-0.252*** (0.056)
Perception	Risk				0.429*** (0.036)	
	Subjectivity					0.158***(0.040)
	Constant	-0.250 (0.266)	0.561* (0.319)	-0.027 (0.146)	-0.022 (0.161)	0.011 (0.217)
	Observations	304	303	607	304	303
	R2	0.149	0.095	0.221	0.512	0.266
	Adjusted R2	0.098	0.041	0.199	0.481	0.220
	Residual Std. Error	0.810 (df = 286)	0.862 (df = 285)	0.600 (df = 589)	0.488 (df = 285)	0.583 (df = 284)
	F Statistic	2.935*** (df = 17; 286)	1.753*** (df = 17; 285)	9.835*** (df = 17; 589)	16.61*** (df = 18; 285)	5.729*** (18; 284)

Racial differences were distinct when looking at AI aversion, with White participants scoring higher on all indices, whereas the other group (a combination of all races other than white or black to enable statistical analysis) was not statistically distinct from Black individuals. Race was not a significant predictor in perceived risk or perceived subjectivity in the first two models. Political affiliation had mixed results with attitudes towards AI, with individuals who did not identify with any party having significant results across multiple models.

Educational level was very interesting in that across all five models it was not found to have a statistically significant effect on any of the dependent variables. Lastly, cultural theory analysis showed that Hierarchicalists, Individualists, and Fatalists had a statistically significant higher AI Aversion Index when compared to Egalitarians. This effect was reduced when risk perceptions were included in the model and vanished entirely when perceived subjectivity were included.

Figure 4.2 and Table 4.4 show the results of looking at the combined impacts of race and gender on people level of AI aversion. Echoing the earlier results on racial differences, both white females and white males scored higher on all three indices than either black males or black females. Though only the AIAI scores had statistically significant differences. Within racial groups the results of the earlier gender analysis are echoed, with white females having higher AI aversion than white males at the statistically significant level. Black females were just below the minimum cutoff of statistical significantly differences in AI aversion with black males ($p=0.102$). The gender differences were not enough to overcome the racial differences as white males scored higher on AI aversion than black females. Knowledge about AI and Experience with AI were once again included in this regression analysis to minimize the novelty effect that AI could have been having on results.

Figure 4.2 Gender and Race: AI Aversion

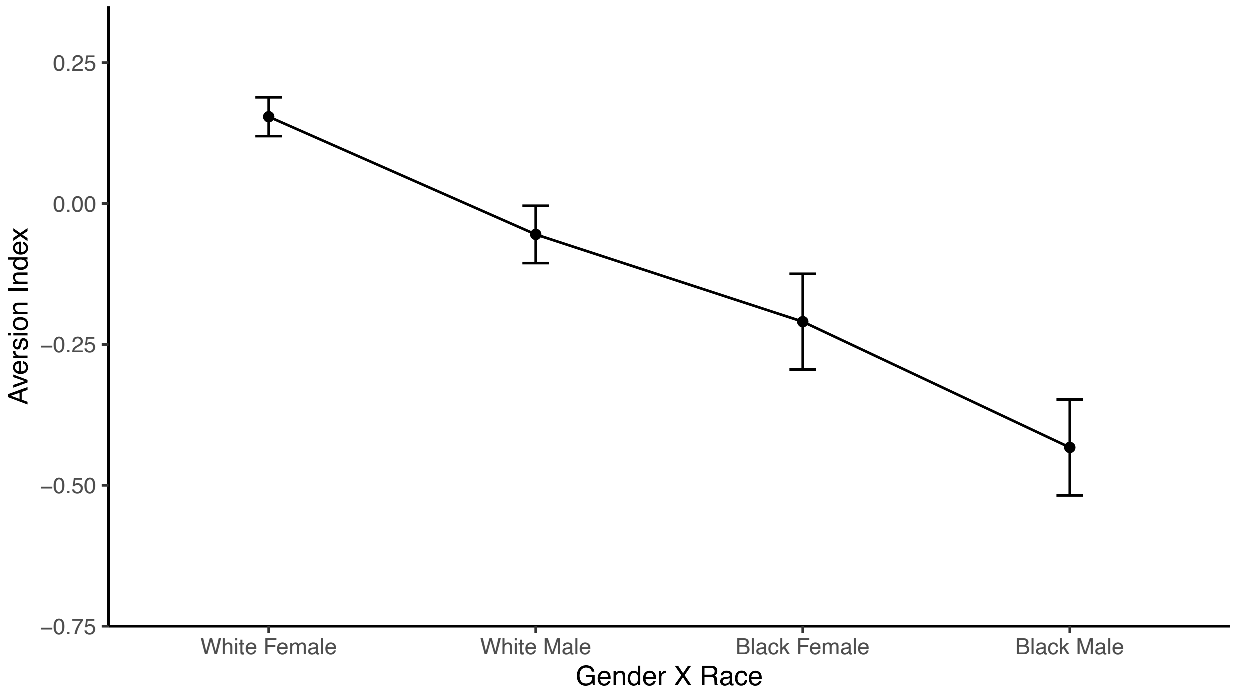


Table 4.4 Gender and Race Regression Analysis

	Perceived Risk	Perceived Subjectivity	AIAI
White Female (vs. Black Male)	0.326 (0.218)	-0.172 (0.229)	0.321*** (0.107)
White Male (vs. Black Male)	0.229 (0.221)	0.120 (0.225)	0.244** (0.107)
Black Female (vs. Black Male)	0.192 (0.268)	-0.008 (0.287)	0.109 (0.132)
Knowledge	-0.059*** (0.020)	-0.070*** (0.022)	-0.057*** (0.010)
Experience	-0.170** (0.073)	-0.183*** (0.081)	-0.294*** (0.037)
Constant	0.371 (0.254)	-0.132 (0.191)	0.580*** (0.126)
Observations	352	348	700
R2	0.079	0.080	0.228
Adjusted R2	0.066	0.067	0.222
Residual Std. Error	0.945 (df = 346)	0.972 (df = 342)	0.649 (df = 694)
F Statistic	5.957*** (df = 5; 346)	5.965*** (df = 5; 342)	40.979*** (df = 5; 694)

4.6 Analysis

The results of these multiple linear regressions shed some important light on the hypothesized relationships between demographics and AI aversion. In the gender analysis, the results substantiated the hypothesis that women demonstrate higher AI aversion than men, as indicated by the significant differences in the AI aversion index. The results show a significant difference in AI aversion index between the genders ((0.114*) when controlling for perceived risk.

The generational analysis revealed a distinct pattern that Gen Z and Millennials exhibit lower AI aversion compared to Gen X and both groups of Boomers, with Boomers showing notably higher aversion levels across all three models. This aligns with the hypothesis that younger generations would have lower aversion than older generations.

Contrary to expectations, the results of the regressions show that white individuals have higher AI aversion compared to black individuals (0.277***). These results endured in the models that controlled for perceived risk (0.236**) and perceived subjectivity (0.252**). This outcome contrasts hypothesis (**H_3**) that racial minorities, due to concerns about systemic biases and fairness, would perceive higher risks associated with AI, leading to greater levels of aversion. Therefore, the hypothesis should be rejected based on these results, suggesting a need to reevaluate the assumptions about how race influences AI aversion.

Political ideology's influence on AI aversion was mixed. The analysis showed that when controlling for the other demographic variables, Republicans and Democrats do not differ significantly from one another when it comes to AI aversion. This suggests that hypothesis (**H_4**) that political ideology influences AI aversion should be rejected. There was an interesting result in that those who did not identify with a party displayed higher AI aversion than both Republicans and Democrats, suggesting a potential area for future research to examine what may be causing that relationship.

The results of the effect of education on AI are very interesting, where in each of the three models it was not found to be statistically significant in its influence on AI aversion. This is particularly interesting when you look at Figure 4.1 which seems to indicate that as educational level increase people's levels of aversion decrease. There are several potential reasons why the effect of education may be disappearing. The first is that the effect of education is small and when so many other variables are used in the multiple regression analysis, the effect becomes indistinguishable from zero. Another potential cause is that the control variables of knowledge and experience with AI may have multicollinearity with education levels causing educations effect to disappear when these variables are included. Finally, this could be a result of the relatively small number of observations within the date that may be limiting the strength and precision of some of the results of this regression analysis.

The cultural theory results also challenged prior expectations. The regression shows Egalitarians had lower levels of AI aversion compared to Hierarchicalists, Individualists, and Fatalists. With Fatalists having the highest levels of aversion, followed by Hierarchicalists, Individualists, and finally Egalitarians. This would indicate that people with high grid (Fatalists and Hierarchicalists) had higher levels of aversion than those with low grid (Individualists and Egalitarians). This would also suggest that people with low group (Fatalists and Individualists) have higher levels of aversion than those with high group (Hierarchicalists and Egalitarians), though that the effect of group is smaller than the effect of grid.

Table 4.4 comparing gender and race in relation to perceived risk, perceived subjectivity, and the AIAI presents intriguing findings that both confirm and challenge the earlier hypothesis derived from Finucane et al. (2000) regarding the "white male effect." Contrary to the expectation that white males would score the lowest on the AIAI, indicating less perceived risk and aversion to AI, the results demonstrate that black males had significantly lower amounts of perceived risk and AI aversion than white males. White males did have lower levels of perceived risk and AI aversion than white females. The hypothesis that white men would be less averse to AI risk is only partially supported, indicating a more complex interplay of race and gender factors than the initial "white male effect" hypothesis suggests.

Black males, on the other hand, displayed a significant divergence from the other groups, with a strong negative perception of risk and the lowest AI aversion level. This finding is particularly striking as it indicates that the "white male effect" may not be as pervasive or as uniform across different domains as initially thought. Instead, these results point to the possibility of a "male effect" across different racial groups and domains. The data compels a reevaluation of the original hypothesis and suggests that future studies should explore the individual and combined effects of race and gender on technology perception more deeply.

For future research, these outcomes highlight the importance of intersectional analysis when examining attitudes towards technology and risk. The variations observed between the different demographic intersections suggest that attitudes toward AI are influenced by a combination of social, cultural, and possibly experiential factors that intersect in complex ways. Therefore, future studies should consider a broader range of demographic variables, possibly including socio-economic status, education, and cultural background, to unpack the intricate ways in which individuals perceive and respond to AI. Such nuanced understanding is crucial for policymakers, technologists, and educators who aim to foster a more inclusive approach to AI integration into society.

4.7 Conclusion

This chapter has delved into understanding the relationship between demographic factors and aversion to artificial intelligence (AI) in public policy. I have demonstrated through the use of regression analysis that demographic factors, including gender, age, and race, distinctly influence attitudes toward AI, with some deviations from traditional assumptions about risk perception. Contrary to the 'white male effect', which suggests a lower perception of risk among white males in technological domains, my findings indicate a more nuanced reality in how AI aversion manifests.

These insights have profound implications for understanding AI aversion. By challenging the white male effect, as well as some of the assumptions of the relationship between cultural theory and AI aversion this chapter underscores the complexity of perceptions around the adoption and use of AI within the area of public policy. This is especially true with regards to core demographic categories and how they evaluate the potential usage of AI. These results suggest that AI aversion is a multifaceted issue, influenced by a variety of socio-demographic factors, rather than being uniform across population segments.

The findings emphasize the importance of demographic-sensitive policies and AI implementation strategies. Policymakers and AI developers must consider these varying perceptions and attitudes towards AI across different demographic groups to ensure equitable and effective AI adoption in the public policy sphere.

This chapter significantly advances the dissertation's overarching theme of AI aversion in public policy. By dissecting the role of demographics, it provides a deeper, more granular understanding of the barriers to AI adoption, going beyond general aversions to uncover specific societal segments' concerns and perceptions.

Future research could explore the dynamic changes in AI aversion over time, considering the rapid evolution of AI technology and societal norms. Additionally, while this research was limited in its ability to look at multiple intersections of demographic categories, that area is primed for future in depth analysis. The examining the intersectionality of demographic factors and their combined impact on AI aversion could offer richer insights, further enriching the discourse on AI adoption in public policy. Finally future research into the roles of demographics on AI aversion should focus on large enough survey populations to enable more rigorous examination of the relationship between these different demographic groups.

This chapter has made a pivotal contribution to the discourse on AI aversion, especially in the context of public policy. By showcasing the limitations of past theories such as the white male effect in AI perception and emphasizing the importance of demographic nuances, it paves the way for more inclusive, effective, and tailored approaches to AI implementation in public policy sectors. The insights gained here not only enhance academic understanding but also offer practical guidance for policymakers and AI practitioners.

Chapter 5: Conclusions, Implications, and Directions for Future Research

This dissertation embarks on a nuanced exploration aimed at disentangling the complex relationship between public policy and AI, with a focus on the publics aversion to AI technologies. A primary goal is to delineate AI aversion as a phenomenon distinct from algorithmic aversion, addressing a significant gap in the current understanding of technology acceptance. By developing and rigorously validating the AIAI, this work provides a novel tool to quantify public sentiments towards AI, setting the stage for a deeper analysis of the societal readiness for AI integration in governance.

The dissertation further ventures into modeling the intricate relationships among a variety of variables that influence AI aversion. It scrutinizes demographic factors, perceived risks, and the subjective nature of policy domains to uncover how these elements collectively shape public attitudes towards AI technologies. Through this multifaceted approach, the research aims to offer a comprehensive understanding of the barriers to AI acceptance, proposing a framework that can guide policymakers, technologists, and researchers in developing strategies that bridge the gap between AI's potential and its societal acceptance. This endeavor not only enriches the academic discourse on AI aversion but also aims to pave the way for a more informed and effective integration of AI into the public sector, ensuring that technological advancements align with public values and expectations.

In chapter two, I introduce the Artificial Intelligence Aversion Index, a novel instrument developed to measure the public's aversion to AI within public policy settings. The chapter meticulously outlines the methodology behind the AIAI's creation, incorporating affective imagery and vignettes to gauge sentiment accurately. Validation processes, including exploratory factor analysis and regression models, underscore the index's reliability. Findings reveal a significant variance in AI aversion across different policy domains, highlighting the complexity of public sentiment towards AI integration.

Chapter three advances the discussion by examining the roles of perceived risk and subjectivity in shaping public AI aversion. Through a detailed analysis, the article establishes that perceived risks associated with AI technologies exert a more substantial influence on public aversion than the subjective interpretation of policy domains. This insight is pivotal, suggesting that mitigating perceived risks could be key to reducing AI aversion and enhancing public acceptance of AI in policymaking.

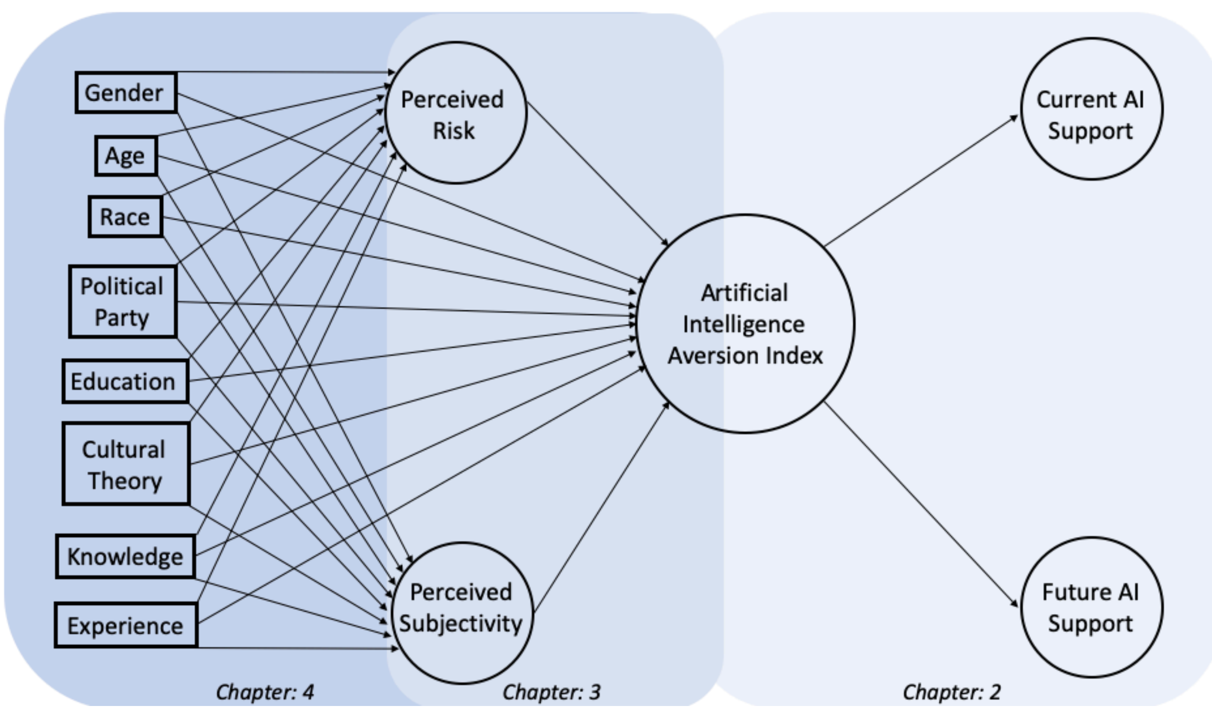
In chapter four, the focus shifts to the demographic dimensions of AI aversion. The research investigates how various demographic factors, such as age, gender, race, political affiliation, and education, influence attitudes towards AI. The findings illuminate notable differences in AI aversion across these groups, providing a nuanced understanding of the demographic underpinnings of AI sentiment. This exploration is crucial for tailoring policy and communication strategies to diverse public segments, aiming for a more inclusive approach to AI policy integration.

This dissertation systematically uncovers the layers of public aversion to AI within public policy frameworks. Chapter two lays the groundwork by introducing the AIAI pioneering a method to quantify public sentiment towards AI, thereby providing a foundation for understanding public resistance. Chapter three builds upon this base by examining the impact of perceived risks and subjectivity on AI aversion, suggesting that perceptions of risk significantly shape public attitudes more than the subjective nature of policy domains. Finally, Chapter four enriches this narrative by incorporating a demographic perspective, revealing how various demographic factors influence public sentiment towards AI, thus offering a comprehensive view of public aversion to AI in policymaking. Together, these chapters contribute distinct yet interconnected insights into the complex fabric of AI aversion, guiding the path toward more nuanced and effective policy interventions.

5.1 Updated Model

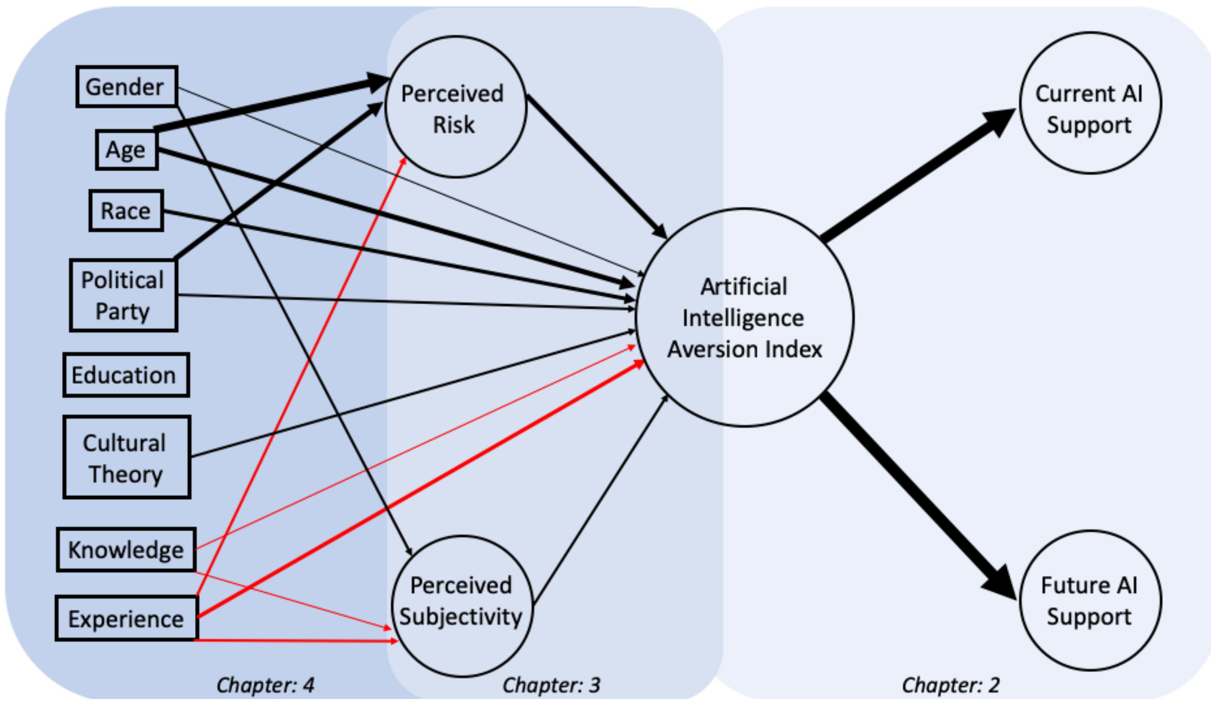
Having looked at the different ways in which these variables interrelate to one another we can again return to the model first introduced in the introduction. This model, Figure 5.1, identified the different possible variables and how they could influence one another. In it the different demographic variables, as well as knowledge and experience with AI are all shown influencing perceptions of risk and subjectivity as well as peoples AIAI scores. Perceptions of risk and subjectivity are also show to have an influence on peoples AIAI scores. Finally, the AIAI is shown influencing peoples current and future support for the use of AI.

Figure 5.1: Modeled Relationship of Variables



The empirical evidence provided by the preceding chapters necessitates updating the model of how these variables relate to one another. This new model which incorporates the evidence from the past experiments is shown in Figure 5.2. This figure includes both the evidential effects of the variables on one another as well as the strength of their effect, which is reflected in the size of the arrows between the variables. The theorized relationships between perceptions of risk and subjectivity, the AIAI, and current and future support of AI are all retained in the new model. Interestingly, the demographic relationships are where the most updating occurred. None of the six demographic categories were found to influence all three of perceived risk, subjectivity, and the AIAI. Of the control variables, only experience of AI was found to have statistically significant influence on each of the three different variables, where knowledge of AI did not have a statistically significant influence on people's perceptions of risk holding all other variables constant. Perceived subjectivity saw the biggest drop off of demographic variable predictors, with only gender having a statistically significant difference on it. Perceived risk was only influenced by people's different generational groups and the race of the individuals. The AIAI was significantly influenced by all of the control variables and the demographic variables with the sole exception of education, which was not found to be statistically significantly influential on any of the dependent variables when holding all other independent variables constant.

Figure 5.2: Updated Modeled Relationship of Variables



5.2 Implications

Artificial intelligence has shown itself to be a vital tool for human flourishing and is here to stay. Since the release of public tools such as Chat GPT and Microsoft Edge to the public market, individuals have been able to experience first-hand the efficiency and power of these AI tools. Governments have also been quick to adopt and utilize AI services to better equip them to provide for their communities. While this research has focused on the aversion people have towards AI, it is important to be cognizant of the benefits that also come with AI. Increasingly people will be exposed both to benefits and costs with its adoption, and the findings of this dissertation will help to understand how their resultant aversion manifests and varies between different groups.

The findings from this dissertation underscore the critical need for policymakers, AI developers, and public administrators to proactively address public concerns and perceptions regarding AI in public policy. By understanding the factors that contribute to AI aversion, such as perceived risks and demographic differences, stakeholders can develop more targeted, inclusive, and effective strategies. Emphasizing transparent communication, engaging in public discourse, and incorporating public input into AI development and policy formulation are essential steps to bridging the gap between technological advancements and societal acceptance.

The research presented in this dissertation makes substantial contributions to the existing body of knowledge on public policy and AI integration. It challenges the traditional models of technology acceptance by unraveling the nuances of AI aversion and introducing the AIAI as a novel metric. The insights gained underscore the complexity of public sentiment and call for a revision of theoretical frameworks to encompass the affective and demographic dimensions of AI perception. This work sets a new precedent for future investigations into the societal implications of emerging technologies, proposing a more dynamic and inclusive approach to theoretical development in AI public policy research.

The research delineated in this dissertation offers pragmatic strategies for the incorporation of AI into public policy. It suggests that policymakers and AI developers could mitigate AI aversion by tailoring communication strategies to address specific concerns identified in the AIAI, such as focusing on addressing people’s perceptions of the risk associated with the use of AI in new policy domains, fostering a transparent development process to increase peoples levels of knowledge about the AI being used, and facilitating public engagement initiatives to help acclimate people to the AI by gaining direct experience with its use. By integrating these insights, stakeholders can create a more receptive environment, thereby enhancing the adoption and efficacy of AI solutions in public governance.

While this dissertation provides insightful contributions to the understanding of AI aversion in public policy, it acknowledges certain limitations that identify key areas for future research. The study's scope could be expanded by exploring additional demographic variables to understand AI aversion's intersectionality further. Data limitation prevented a more in-depth exploration of the role of race on people levels of aversion. Future research should focus specifically on the matter of race and how it influences people's level of aversion to AI. Longitudinal studies would also offer valuable insights into how AI aversion evolves over time, particularly as AI technologies become more prevalent and people become more aware of its potential uses and the potential risks associated with it. Additionally, applying the findings in practical policy experiments could test the effectiveness of strategies derived from this research, providing a feedback loop for continuous theoretical and practical refinement.

This dissertation highlights the significance of understanding and addressing AI aversion in the quest for harmonizing AI with public policy. It underscores a pivotal insight: the success of AI governance hinges not solely on technological robustness but equally on the nuanced acceptance by the public. AI could potentially revolutionize how we interact with one another as well as with the government through its policies, however if policymakers and AI adopters don't consider peoples levels of aversion and how best to address it going forward, we will struggle to reap the benefits that AI offers. As we stand on the cusp of a new era where AI's potential to revolutionize policy is immense, recognizing and mitigating aversion to AI is not just beneficial—it is imperative for the seamless, democratic, and effective adoption of AI in governance.

List of References

Ahmer, H., Altaf, S., Khan, H.M., Bhatti, I., & Naseem, S. (2023). Knowledge and Perception of Medical Students towards the use of Artificial Intelligence in Healthcare. JPMA. *The Journal of the Pakistan Medical Association*. 73.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179-211.

Baxter, G., & Sommerville, I., (2011) Socio-technical systems: From design methods to systems engineering, *Interacting with Computers*, 23 (1) 4–17,

Browne, M. W., & Cudeck, R. (1992). Alternative Ways of Assessing Model Fit. *Sociological Methods & Research, 21*(2), 230-258.

Bucchi, M., & Neresini, F. (2008). Science and public participation. In E. J. Hackett, O. Amsterdamska, M. Lynch, & J. Wajcman (Eds.), *The Handbook of Science and Technology Studies* (3rd ed., pp. 449-472). MIT Press.

Castelo N, Bos M.W., & Lehmann D.R. (2019). Task-Dependent Algorithm Aversion. *Journal of Marketing Research*, *56*(5), 809–825.

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3), 319-340.

Davis J., Purves, D., Gilbert J., & Strum S. (2022) Five ethical challenges facing data-driven policing. *AI Ethics* 2, 185–198. https://doi.org/10.1007/s43681-021-00105-9

Dennis, S. A., Goodson, B. M., & Pearson, C. (2020). Online Worker Fraud and Evolving Threats to the Integrity of MTurk Data: A Discussion of Virtual Private Servers and the Limitations of IP-Based Screening Procedures. *Behavioral Research in Accounting, 32* (1), 119 – 134.

Department of Defense. (2023). *DOD DIRECTIVE 3000.09: Autonomy in Weapon Systems*. Approved by Kathleen H. Hicks, Deputy Secretary of Defense.

Dietvorst, B.J., Simmons, J.P., & Masset, C. (2016) Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them. *Management Science* 64 (3), 1155-1170.

Eze, S.C., Awa, H.O., Chinedu-Eze, V.C.A., & Bello, A.O. (2021) Demographic determinants of mobile marketing technology adoption by small and medium enterprises (SMEs) in Ekiti State, Nigeria. *Humanities and Social Sciences Communications* 8, 82.

Filiz, I., Judek, J.R., Lorenz, M., & Spiwoks, M. (2023) The extent of algorithm aversion in decision making situations with varying gravity. *PLOS ONE 18*.

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of behavioral decision making*, 13(1), 1-17.

Fisher, E. (2022*). Algorithms and subjectivity: the subversion of critical knowledge*. Routledge.

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals,* 14(2), 627-660.

Hamilton Z., Duwe G., Kigerl A., Gwinn J, Langan N., & Dollar C. (2022) Tailoring to a Mandate: The Development and Validation of the Prisoner Assessment Tool Targeting Estimated Risk and Needs (PATTERN), *Justice Quarterly*, 39:6, 1129-115

Hu,L. T., & Bentler, P.M. (1999) Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Structural Equation Modeling, 6*, 1-55.

Jussupow, E., Benbasat, I., & Heinzl, A. (2020). Why are we averse towards algorithms? A comprehensive Literature Review on Algorithm Aversion. [Paper Presentation], ECIS Norway, https://aisel.aisnet.org/ecis2020_rp/168

Leonardi, P. M. (2012). Materiality, Sociomateriality, and Socio-Technical Systems: What Do These Terms Mean? How are They Related? Do We Need Them?. *Materiality and Organizing: Social Interaction in a Technological World.*

Morris, M. G., & Venkatesh, V. (2000). Age differences in technology adoption decisions: Implications for a changing workforce. *Personnel Psychology*, 53(2), 375-403.

National Institute of Standards and Technology. (2008). Autonomy Levels for Unmanned Systems (ALFUS) Framework: Volume I - Terminology (*NIST Special Publication 1011-I-2.0*, Version 2.0). Edited by Hui-Min Huang.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019) Dissecting racial bias in an an algorithm used to manage the health of populations. *Science* 366, 447-453.

Önkal D, Goodwin P, Thomson M, Gönül S, & Pollock A. (2009). The Relative Influence of Advice from Human Experts and Statistical Methods on Forecast Adjustments, *Behavioral Decision Making*, 22 (4), 390–409.

Partnership on AI. (2020) *Algorithmic Risk Assessment and COVID-19: Why PATTERN Should Not Be Used*. Issue Brief. Accessed online March, 1 2024: https://partnershiponai.org/wp-content/uploads/2021/07/ Why-PATTERN-Should-Not-Be-Used.pdf

Prensky, M. (2001). Digital natives, digital immigrants*. On the Horizon*, 9(5), 1-6.

Promberger M, & Baron J. (2006). Do Patients Trust Computers? *Journal of Behavioral Decision Making*, 19 (5), 455–68.

Purves D. & Davis J. (2022). *Criminal Justice Algorithms: Being race-neutral doesn't mean race-blind.* The Conversation.

Purves D., Jenkins R. & Strawser B.J. (2015) Autonomous Machines, Moral Judgment, and Acting for the Right Reasons. *Ethical Theory and Moral Practice* 18, 851–872. https://doi.org/10.1007/s10677-015-9563-y

Rogers, E. M. (2003). Diffusion of innovations (5th ed.). *Free Press*.

Scharre P. (2018). *Army of None: Autonomous Weapons and the Future of War*. New York, NY: W.W. Norton & Company.

Shaffer V.A., Probst C.A., Merkle E.C., Arkes H.R., & Medow M.A. (2013). Why Do Patients Derogate Physicians Who Use a Computer-Based Diagnostic Support System? *Medical Decision Making*, 33 (1), 108–118.

Smith, A., & Anderson, M. (2019). Public attitudes toward technology companies. *Pew Research Center.*

Slovic, P. (1987). Perception of Risk. *Science*, 236(4799), 280-285.

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). The affect heuristic. *European journal of operational research*, 177(3), 1333-1352.

Szalay L.B., & Deese, J. (1978). *Subjective Meaning and Culture: An Assessment through Word Associations*. Hillsdale N.J. L. Erlbaum Associates.

Venkatesh, V., Morris, M.G., Davis, G.B., & Davis, F.D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425-478.

Wildavsky, A. (1987). Choosing Preferences by Constructing Institutions: A Cultural Theory of Preference Formation, *The American Political Science Review*, 81 (1) 3-22.

Williams, A., Anderson, H., & Drennan, J. (2010). Is it the same for everyone? Exploring the generational divide in ICT workplace behaviors. *Journal of Managerial Psychology*, 25(1), 33-50.

Yeomans M, Shah A, Mullainathan S, & Kleinberg J. (2019). Making Sense of Recommendations, *Journal of Behavioral Decision Making*, (published online February 14), DOI:10.1002/bdm.2118.

Appendix

Table A1. Affective Imagery Survey Question

Survey Question	n	Mean	SD
Artificial Intelligence	316	3.13	1.41
Algorithms	254	3.42	1.12
Advanced Technology	264	3.52	1.31

Table A2. Current AI Support Survey Questions

Survey Question	Mean	SD
Facial Recognition Systems used by the Transportation Security Administration	4.40	1.99
Cataloging Worker Injury Narrative by the Bureau of Labor Statistics	4.30	1.79
Analysis of Adverse Drug Effects by the Food and Drug Administration	4.36	1.86
Adjudicating Patent and Trademark Applications by the US Patent and Trademark Office	4.37	1.81
Handwriting Recognition Tools used by US Postal Service	4.47	1.90
Chatbots used by the US Citizenship and Immigration Services	4.06	1.96
Tools to Counter Cyberattacks on Agency Systems by the Department of Homeland Security	4.42	1.95

Table A3. Future AI Support Survey Questions

Survey Question	Mean	SD
Air Traffic Control	3.83	2.07
Traffic Management	4.13	1.99
Traffic Prediction	4.47	1.93
Public Transportation Scheduling	4.50	1.86
Public Utilities Management	4.37	1.87
Public Parks and Recreations Management	4.45	1.82
City Planning & Urban Development	4.28	1.87
Running Political Elections	3.53	2.11
Determining Foreign Policy	3.43	2.03
Historical Preservation	4.19	1.92
Weather Forecasting (extreme weather/chance of tornado)	4.35	1.94
Weather Forecasting (minor weather/chance of rain)	4.59	1.89
Determining Criminal Justice Policy	3.56	2.05
Deciding what projects Congress should fund	3.65	2.03
Deciding whom to give Government Contracts too	3.62	2.03

Table A4. AI Aversion Index Survey Questions

Survey Question	Mean	SD
The criminal justice system is changed so that instead of judge determining the sentencing for guilty plaintiffs, an artificial intelligence will determine what the sentence should be. The artificial intelligence system will have access to a database of all past sentencing as well as information on recidivism or the amount of times past convicts have committed a crime again after being sentenced. The artificial intelligences goal would be to minimize the instances of future recidivism and would be able to update its sentencing based on the results of other decisions it had made.	-1.13	7.19
The national disaster warning system is changed so that instead of professionals determining when to issue a disaster warning, an artificial intelligence system will be making that decision. The artificial intelligence system will have access to a database of current conditions as well as a database on all past warning decisions. The artificial intelligences goal would be to minimize the instances of inaccurate disaster warnings and would be able to update its warning issuances based on the results of past decisions it had made.	1.47	6.62
The Supplemental Nutrition Assistance Program (SNAP) is changed so that instead of professionals determining whether someone qualifies for benefits, an artificial intelligence system will be making that decision. The artificial intelligence system will have access to a database of all past beneficiaries as well as information on fraudulent issuances of benefits. The artificial intelligences goal would be to minimize instances of fraud and to ensure people are receiving the proper benefits. The artificial intelligence system would be able to update its approval of the dispersion of benefits based on the results of past decisions it had made.	1.01	6.86
The amount of a tariff imposed on goods imported to the United States is changed so that instead of Congress members deciding tariffs on a case by case basis, an artificial intelligence system will be making the decisions. The artificial intelligence system will have access to a database of the current tariff policies of other countries as well as current market prices for good in the United States. The artificial intelligences goal would be to benefit producers of goods in the United States and would be able to update the tariffs based on the results of past decisions it had made.	1.38	6.39
The issuance of environmental fines is changed so that instead of government officials determining when to issue a fine for environmental damage to a company, an artificial intelligence system will be making that decision. The artificial intelligence system will have access to a database of all past fines as well as information on how companies changed future practices to reduce environmental harm. The artificial intelligences goal would be to minimize the instances of environmental harm and would be able to update its fines based on the results of past decisions it had made.	1.04	6.46

*Note: Scores were inverted when combined into the Aversion Index so that a higher Aversion Index score would equate to lower support for AI across different domains.

Table A5. Linear Model of AIAI and Mean Current, Future, & Combined Support

	Dependent variable:		
	Current AI Support	Future AI Support	Combined AI Support
	(1)	(2)	(2)
Aversion Index	-1.542*** (0.052)	-1.739*** (0.046)	-1.677*** (0.044)
Constant	4.340*** (0.038)	4.062*** (0.033)	4.150*** (0.032)
Observations	834	834	834
R2	0.512	0.629	0.636
Adjusted R2	0.511	0.629	0.636
Residual Std. Error (df = 832)	1.092	0.967	0.919
F Statistic (df = 1; 832)	871.849***	1,413.564***	1,455.602***
Note		***p<0.1; **p<0.05, ***p<0.01**	

Table A6: Experiment Results

Domain	**Current or Future**	**Risk**	**Subjectivity**	**Support**
Facial Recognition	C	4.20	3.43	4.40
Worker Injury	C	3.93	3.66	4.30
Drug Analysis	C	4.04	3.79	4.36
Patent Adjudication	C	3.87	3.64	4.37
Handwriting Recognition	C	3.77	3.61	4.47
Immigration Chatbots	C	4.40	3.82	4.06
Counter Cyberattacks	C	4.03	3.56	4.42
Air Traffic Control	F	5.06	3.53	3.83
Traffic Mgmt.	F	4.73	3.44	4.13
Traffic Prediction	F	4.21	3.53	4.45
Public Transportation	F	3.88	3.51	4.50
Public Utilities Mgmt.	F	4.03	3.43	4.36
Public Parks and Rec. Mgmt.	F	3.62	3.74	4.45
City Planning & U.D.	F	4.00	3.71	4.28
Running Political Elections	F	4.92	4.02	3.53
Foreign Policy	F	4.99	3.91	3.43
Historical Preservation	F	4.10	3.73	4.19
Weather Forecasting (E)	F	4.10	3.61	4.35
Weather Forecasting (M)	F	3.76	3.52	4.59
Criminal Justice Policy	F	4.94	3.80	3.56
Congressional Funding	F	4.66	3.94	3.65
Government Contracts	F	4.76	3.91	3.62

Figure A1: Mean Risk, Subjectivity, and AI Support Across Current Domains

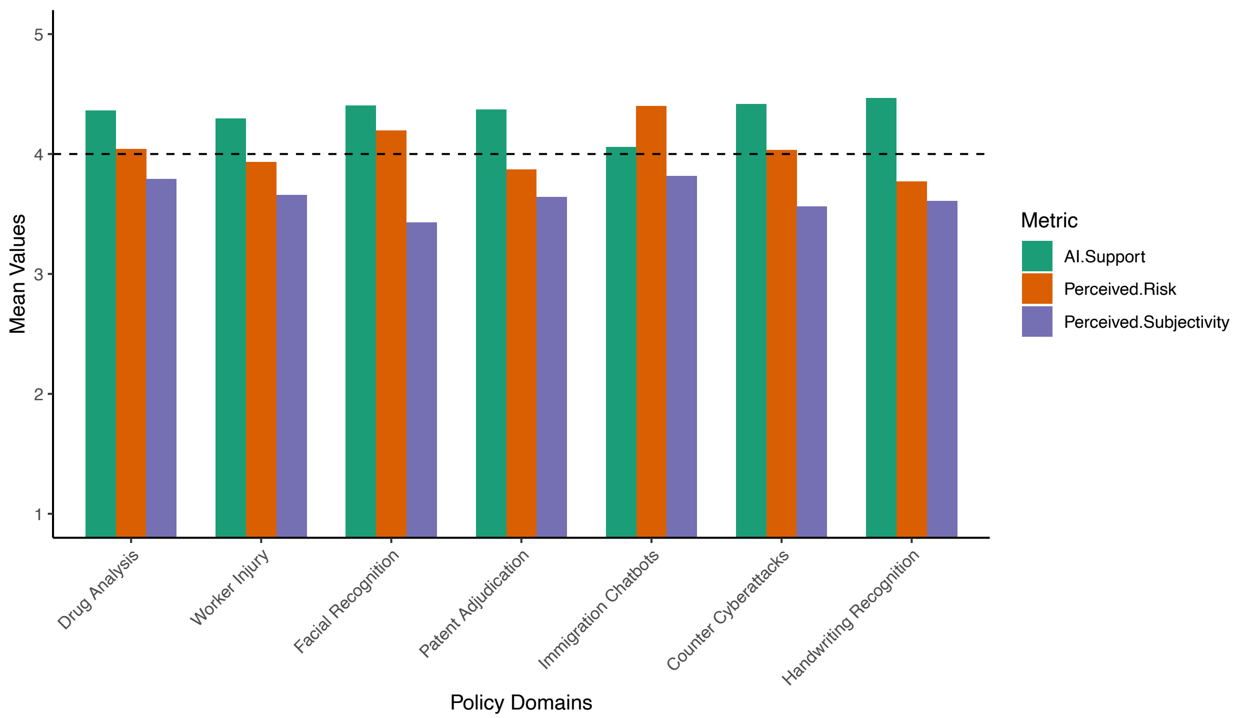
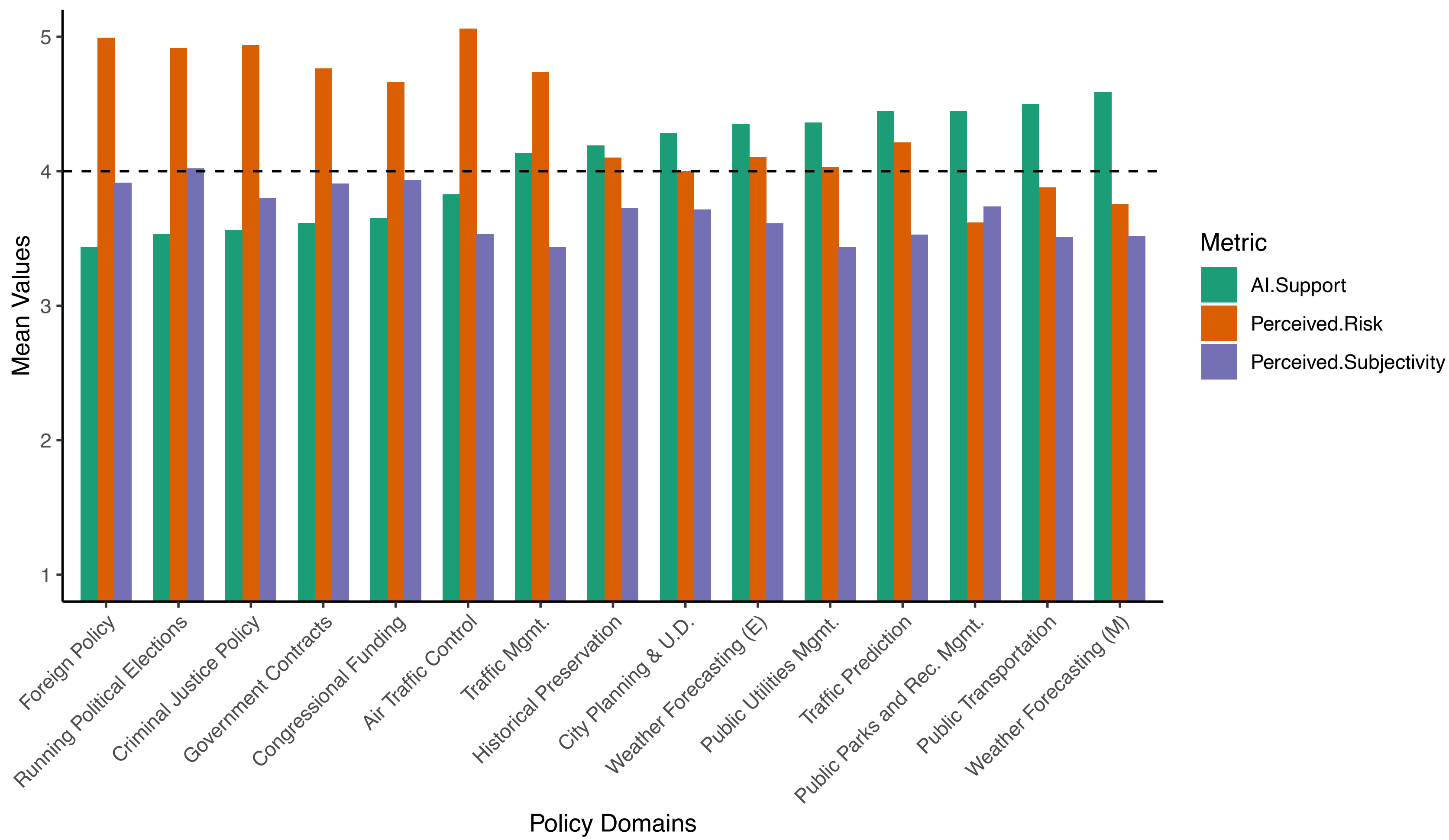


Figure A2: Mean Risk, Subjectivity, and AI Support Across Future Domains



Appendix: Survey Instrument

By answering the survey questions, I agree to participate in this research. Please print this page for your records.

consent: Do you consent to participate in this study?[4]

0 - No, I do not want to participate in this study
1 - Yes, I agree to participate in this study

------------------------End Web pg-----------------------

The rest of this survey will focus on science and technology in public policy. We will start with some questions about the use of [**rand_word:** ***algorithms* | *artificial intelligence* | *advanced technology***] in public policy. Can you tell us the first three words or phrases that come to you when you think about **rand_word**.

word_1: First word/phrase: [verbatim]

word_1_feel: When you think about this word or phrase, do you have positive or negative feelings?
1 - Very negative
2 - Negative
3 - Neither positive nor negative
4 - Positive
5 - Very positive

word_2: Second word/phrase: [verbatim]

word_2_feel: When you think about this word or phrase, do you have positive or negative feelings?
1 - Very negative
2 - Negative
3 - Neither positive nor negative
4 - Positive
5 - Very positive

word_3: Third word/phrase: [verbatim]

word_3_feel: When you think about this word or phrase, do you have positive or negative feelings?
1 - Very negative
2 - Negative

[4] Lucid the partner for this survey independently collects demographic data for their survey participants. This demographic data was the provided to the research in addition to survey responses.

3 - Neither positive nor negative
4 - Positive
5 - Very positive

----------------------End Web pg----------------------
Now we want you to think carefully about artificial intelligence. Please review this information before you continue.

For this survey we will be using the definition of artificial intelligence as a computerized system where data is taken in, processed, and a result is given. The system is capable of enhancing itself based on the data it is both receiving and producing. This allows it to change how it processes that data to give better results in the future.

ai_know: Before reading this information, had you heard about artificial intelligence?
0 - No
1 - Yes
2 - Not sure

----------------------End Web pg----------------------

[show if **ai_know** = 1]

knowledge: How would you rate your knowledge of artificial intelligence technology?
0 - Not at all knowledgeable
1
2
3
4
5
6
7
8
9
10 - Extremely knowledgeable

-----------------------End Web pg----------------------

For the next few questions we will ask you about the use of artificial intelligence in public policy.

Public policy is what government does or does not do about a problem that comes before them for consideration and possible action.

social_1: How do you feel about the use of artificial intelligence in public policy?
1 - Very negative
2 - Negative

3 -	Neutral
4 -	Positive
5 -	Very positive

social_2: Are you excited or concerned about the use of artificial intelligence in public policy?
1 -	Very concerned
2 -	Concerned
3 -	Neutral
4 -	Excited
5 -	Very excited

-----------------------End Web pg----------------------

pc_1: Have you had any prior experiences with using artificial intelligence systems in decision making?
1 -	No, I have never had any experience with using artificial intelligence systems in decision making.
2 -	No, but I am interested in learning more about it.
3 -	Neutral
4 -	Yes, I have had some experience with using artificial intelligence systems in decision making.
5 -	Yes, I have had a lot of experience with using artificial intelligence systems in decision making.

pc_2: How knowledgeable do you feel you are about the use of artificial intelligence systems in public policy decision making?
1 -	Not at all knowledgeable
2 -	Not very knowledgeable
3 -	Neutral
4 -	Knowledgeable
5 -	Very knowledgeable

----------------------End Web pg----------------------

The usage of artificial intelligence has grown dramatically in many people's lives, often in ways they may be unaware of. Some common examples of artificial intelligence people use are personal assistants such as Siri and Alexa, navigation systems such as Google Maps and Waze, social media such as Facebook and Instagram, online shop-ping, and even online banking.

update: Having read that, would you like to update your previous answer about your past experiences with artificial intelligence?
0 - No
1 - Yes
2 - Not sure

----------------------End Web pg----------------------

[show if **update** = 1]

pc_update: Have you had any prior experiences with using artificial intelligence systems in decision making?
1 - No, I have never had any experience with using artificial intelligence systems in decision making.
2 - No, but I am interested in learning more about it.
3 - Neutral
4 - Yes, I have had some experience with using artificial intelligence systems in decision making.
5 - Yes, I have had a lot of experience with using artificial intelligence systems in decision making.

----------------------End Web pg----------------------

It is important to be aware of some of the common arguments advocates and opponents of the adoption of artificial intelligence make.

Advocates emphasize the **increased efficiency** of adoption of artificial intelligence as well as **increased consistency** as human bias is removed from the decision-making process.

Opponents emphasize the **lack of accountability** from the adoption of artificial intelligence as biases that exist in how the data is collected or included into the system can lead to **biased results**.

grisk_1: How worried are you about the lack of accountability from the adoption of artificial intelligence into pub-lic policy uses?
1 - Not worried at all
2 - Neutral
3 - Slightly worried
4 - Moderately worried
5 - Extremely worried

grisk_2: How worried are you about the potential for biased results from the adoption of artificial intelligence into public policy uses?
1 - Not worried at all
2 - Neutral
3 - Slightly worried
4 - Moderately worried
5 - Extremely worried

----------------------End Web pg----------------------

benefit_1: How beneficial do you think the increased efficiency of artificial intelligence use in public policy will be?

1 - Extremely beneficial
2 - Moderately beneficial
3 - Slightly beneficial
4 - Neutral
5 - Not beneficial at all

benefit_2: How beneficial do you think the increased consistency of results from artificial intelligence use in public policy will be?
1 - Extremely beneficial
2 - Moderately beneficial
3 - Slightly beneficial
4 - Neutral
5 - Not beneficial at all

------------------------End Web pg-----------------------

risk_benefit: If artificial intelligence systems were to be adopted in the use of public policy, which of these choices best reflects your views?
1 - The risks outweigh the benefits
2 - The benefits outweigh the risk
3 - The benefits and risks are equal
4 - Not sure

------------------------End Web pg-----------------------

trust_1: How much trust do you have in the government or other institutions to use artificial intelligence responsibly in public policy?
1 - Not at all
2 - Not very much
3 - Neutral
4 - Somewhat
5 - A lot

trust_2: How much trust do you have in the tech companies or other entities that develop artificial intelligence?
1 - Not at all
2 - Not very much
3 - Neutral
4 - Somewhat
5 - A lot

------------------------End Web pg-----------------------

cognitiveb_1: How confident are you that you understand how artificial intelligence works and its potential implications for public policy?
1 - Not at all confident

2 - Not very confident
3 - Neutral
4 - Somewhat confident
5 - Very confident

----------------------End Web pg----------------------

Sometime people answer these questions without first reading. Please select “Strongly agree” to ensure that you are paying attention.
1 – Strongly agree
2 - Agree
3 - Neutral
4 - Disagree
5 – Strongly disagree

----------------------End Web pg----------------------

For the next few questions you will read a few hypothetical scenarios where artificial intelligence has been incorporated into different policy areas. You will then be asked to rate the degree to which you would oppose or support the use of artificial intelligence in each scenario. [random order]

justice: The criminal justice system is changed so that instead of judge determining the sentencing for guilty plain-tiffs, an artificial intelligence will determine what the sentence should be. The artificial intelligence system will have access to a database of all past sentencing as well as information on recidivism or the amount of times past convicts have committed a crime again after being sentenced. The artificial intelligences goal would be to minimize the instances of future recidivism and would be able to update its sentencing based on the results of other decisions it had made.
-10 – Completely oppose
-9 -
-8 -
-7 -
-6 -
-5 -
-4 -
-3 -
-2 -
-1 -
0 – Neither oppose or support
1 -
2 -
3 -
4 -
5 -
6 -

7 -
8 -
9 -
10 – Completely support

disaster: The national disaster warning system is changed so that instead of professionals determining when to is-sue a disaster warning, an artificial intelligence system will be making that decision. The artificial intelligence system will have access to a database of current conditions as well as a database on all past warning decisions. The artificial intelligences goal would be to minimize the instances of inaccurate disaster warnings and would be able to update its warning issuances based on the results of past decisions it had made.
-10 – Completely oppose
-9 -
-8 -
-7 -
-6 -
-5 -
-4 -
-3 -
-2 -
-1 -
0 – Neither oppose or support
1 -
2 -
3 -
4 -
5 -
6 -
7 -
8 -
9 -
10 – Completely support

snap: The Supplemental Nutrition Assistance Program (SNAP) is changed so that instead of professionals deter-mining whether someone qualifies for benefits, an artificial intelligence system will be making that decision. The artificial intelligence system will have access to a database of all past beneficiaries as well as information on fraudulent issuances of benefits. The artificial intelligences goal would be to minimize instances of fraud and to ensure people are receiving the proper benefits. The artificial intelligence system would be able to update its approval of the dispersion of benefits based on the results of past decisions it had made.
-10 – Completely oppose
-9 -
-8 -
-7 -
-6 -
-5 -

-4 -
-3 -
-2 -
-1 -
0 – Neither oppose or support
1 -
2 -
3 -
4 -
5 -
6 -
7 -
8 -
9 -
10 – Completely support

tariff: The amount of a tariff imposed on goods imported to the United States is changed so that instead of Congress members deciding tariffs on a case by case basis, an artificial intelligence system will be making the decisions. The artificial intelligence system will have access to a database of the current tariff policies of other countries as well as current market prices for good in the United States. The artificial intelligences goal would be to benefit producers of goods in the United States and would be able to update the tariffs based on the results of past decisions it had made.
-10 – Completely oppose
-9 -
-8 -
-7 -
-6 -
-5 -
-4 -
-3 -
-2 -
-1 -
0 – Neither oppose or support
1 -
2 -
3 -
4 -
5 -
6 -
7 -
8 -
9 -
10 – Completely support

fines: The issuance of environmental fines is changed so that instead of government officials determining when to issue a fine for environmental damage to a company, an artificial intelligence system will be making that decision. The artificial intelligence system will have access to a database of all past fines as well as information on how companies changed future practices to reduce environmental harm. The artificial intelligences goal would be to minimize the instances of environmental harm and would be able to update its fines based on the results of past decisions it had made.
-10 – Completely oppose
-9 -
-8 -
-7 -
-6 -
-5 -
-4 -
-3 -
-2 -
-1 -
0 – Neither oppose or support
1 -
2 -
3 -
4 -
5 -
6 -
7 -
8 -
9 -
10 – Completely support

----------------------End Web pg----------------------

*****SPLIT A/B: testing perceived risk vs. perceived subjectivity*****

Track A (50%): perceived risk

Now we would like you to evaluate a series of instances where the government is already utilizing artificial intelligence in its federal administrative agencies.

Please evaluate them on how dangerous you feel it would be for a mistake to be made by the artificial intelligence being used.

Please respond to the following on a scale from one to seven, where one means very dangerous and seven means not dangerous at all.
[**random table for r_risk1—r_risk7**]

r_risk1: Facial Recognition Systems used by the Transportation Security Administration

1 – Very dangerous
2 -
3 -
4 -
5 -
6 -
7 – Not dangerous at all

r_risk2: Cataloging Worker Injury Narrative by the Bureau of Labor Statistics
1 – Very dangerous
2 -
3 -
4 -
5 -
6 -
7 – Not dangerous at all

r_risk3: Analysis of Adverse Drug Effects by the Food and Drug Administration
1 – Very dangerous
2 -
3 -
4 -
5 -
6 -
7 – Not dangerous at all

r_risk4: Adjudicating Patent and Trademark Applications by the US Patent and Trademark Office
1 – Very dangerous
2 -
3 -
4 -
5 -
6 -
7 – Not dangerous at all

r_risk5: Handwriting Recognition Tools used by US Postal Service
1 – Very dangerous
2 -
3 -
4 -
5 -
6 -
7 – Not dangerous at all

r_risk6: Chatbots used by the US Citizenship and Immigration Services

1 – Very dangerous
2 -
3 -
4 -
5 -
6 -
7 – Not dangerous at all

r_risk7: Tools to Counter Cyberattacks on Agency Systems by the Department of Homeland Security
1 – Very dangerous
2 -
3 -
4 -
5 -
6 -
7 – Not dangerous at all

----------------------End Web pg----------------------

Now we would like you to evaluate a hypothetical series of different policy domains where the government could one day incorporate artificial intelligence.

Please evaluate them based on how dangerous you feel it would be for a mistake to be made in the decision-making of each policy domain.

Please respond to the following on a scale from one to seven, where one means very dangerous and seven means not dangerous at all.
[**random table for risk1—risk15**]

risk1: Air Traffic Control
1 – Very dangerous
2 -
3 -
4 -
5 -
6 -
7 – Not dangerous at all

risk2: Traffic Management
1 – Very dangerous
2 -
3 -
4 -
5 -
6 -

7 – Not dangerous at all

risk3: Traffic Prediction
1 – Very dangerous
2 -
3 -
4 -
5 -
6 -
7 – Not dangerous at all

risk4: Public Transportation Scheduling
1 – Very dangerous
2 -
3 -
4 -
5 -
6 -
7 – Not dangerous at all

risk5: Public Utilities Management
1 – Very dangerous
2 -
3 -
4 -
5 -
6 -
7 – Not dangerous at all

risk6: Public Parks and Recreation Management
1 – Very dangerous
2 -
3 -
4 -
5 -
6 -
7 – Not dangerous at all

risk7: City Planning & Urban Development
1 – Very dangerous
2 -
3 -
4 -
5 -
6 -
7 – Not dangerous at all

risk8: Running Political Elections
1 – Very dangerous
2 -
3 -
4 -
5 -
6 -
7 – Not dangerous at all

risk9: Determining Foreign Policy
1 – Very dangerous
2 -
3 -
4 -
5 -
6 -
7 – Not dangerous at all

risk10: Historical Preservation
1 – Very dangerous
2 -
3 -
4 -
5 -
6 -
7 – Not dangerous at all

risk11: Weather Forecasting (extreme weather/chance of tornado)
1 – Very dangerous
2 -
3 -
4 -
5 -
6 -
7 – Not dangerous at all

risk12: Weather Forecasting (minor weather/chance of rain)
1 – Very dangerous
2 -
3 -
4 -
5 -
6 -
7 – Not dangerous at all

risk13: Determining Criminal Justice Policy
1 – Very dangerous
2 -
3 -
4 -
5 -
6 -
7 – Not dangerous at all

risk14: Deciding what projects Congress should fund
1 – Very dangerous
2 -
3 -
4 -
5 -
6 -
7 – Not dangerous at all

risk15: Deciding whom to give Government Contracts too
1 – Very dangerous
2 -
3 -
4 -
5 -
6 -
7 – Not dangerous at all

----------------------End Web pg----------------------

Track B (50%): perceived subjectivity

Now we would like you to evaluate a series of instances where the government is already utilizing artificial intelligence in its federal administrative agencies.

Please evaluate them on how subjective you feel the decision-making task is.

Please respond to the following on a scale from one to seven, where one means subjective (based on personal feelings, tastes, or opinions) and seven means objective or not subjective at all.

[**random table for r_subjective1—r_subjective7**]

r_subjective1: Facial Recognition Systems used by the Transportation Security Administration
1 – Subjective
2 -
3 -

4 -
5 -
6 -
7 – Objective

r_subjective2: Cataloging Worker Injury Narrative by the Bureau of Labor Statistics
1 – Subjective
2 -
3 -
4 -
5 -
6 -
7 – Objective

r_subjective3: Analysis of Adverse Drug Effects by the Food and Drug Administration
1 – Subjective
2 -
3 -
4 -
5 -
6 -
7 – Objective

r_subjective4: Adjudicating Patent and Trademark Applications by the US Patent and Trademark Office
1 – Subjective
2 -
3 -
4 -
5 -
6 -
7 – Objective

r_subjective5: Handwriting Recognition Tools used by US Postal Service
1 – Subjective
2 -
3 -
4 -
5 -
6 -
7 – Objective

r_subjective6: Chatbots used by the US Citizenship and Immigration Services
1 – Subjective
2 -
3 -

4 -
5 -
6 -
7 – Objective

r_subjective7: Tools to Counter Cyberattacks on Agency Systems by the Department of Homeland Security
1 – Subjective
2 -
3 -
4 -
5 -
6 -
7 – Objective

----------------------End Web pg----------------------

Now we would like you to evaluate a hypothetical series of different policy domains where the government could one day incorporate artificial intelligence.

Please evaluate them based on how subjective you feel the decision-making in the policy domain is.

Please respond to the following on a scale from one to seven, where one means subjective (based on personal feelings, tastes, or opinions) and seven means objective or not subjective at all.
[**random table for subjective1—subjective15**]

subjective1: Air Traffic Control
1 – Subjective
2 -
3 -
4 -
5 -
6 -
7 – Objective

subjective2: Traffic Management
1 – Subjective
2 -
3 -
4 -
5 -
6 -
7 – Objective

subjective3: Traffic Prediction

1 – Subjective
2 -
3 -
4 -
5 -
6 -
7 – Objective

subjective4: Public Transportation Scheduling
1 – Subjective
2 -
3 -
4 -
5 -
6 -
7 – Objective

subjective5: Public Utilities Management
1 – Subjective
2 -
3 -
4 -
5 -
6 -
7 – Objective

subjective6: Public Parks and Recreation Management
1 – Subjective
2 -
3 -
4 -
5 -
6 -
7 – Objective

subjective7: City Planning & Urban Development
1 – Subjective
2 -
3 -
4 -
5 -
6 -
7 – Objective

subjective8: Running Political Elections
1 – Subjective

2 -
3 -
4 -
5 -
6 -
7 – Objective

subjective9: Determining Foreign Policy
1 – Subjective
2 -
3 -
4 -
5 -
6 -
7 – Objective

subjective10: Historical Preservation
1 – Subjective
2 -
3 -
4 -
5 -
6 -
7 – Objective

subjective11: Weather Forecasting (extreme weather/chance of tornado)
1 – Subjective
2 -
3 -
4 -
5 -
6 -
7 – Objective

subjective12: Weather Forecasting (minor weather/chance of rain)
1 – Subjective
2 -
3 -
4 -
5 -
6 -
7 – Objective

subjective13: Determining Criminal Justice Policy
1 – Subjective
2 -

3 -
4 -
5 -
6 -
7 – Objective

subjective14: Deciding what projects Congress should fund
1 – Subjective
2 -
3 -
4 -
5 -
6 -
7 – Objective

subjective15: Deciding whom to give Government Contracts too
1 – Subjective
2 -
3 -
4 -
5 -
6 -
7 – Objective

*****End Split A/B*****

----------------------End Web pg----------------------

Now we would like you to reconsider the series of instances where the government is already utilizing artificial intelligence in its federal administrative agencies.

Please evaluate them on if you oppose or support the decision-making being done by artificial intelligence systems instead of humans.

Please respond to the following on a scale from one to seven, where one means Completely oppose and seven means Completely support.
[**random table for artificial_r1— artificial_r7**]

artificial_r1: Facial Recognition Systems used by the Transportation Security Administration
1 – Completely Oppose
2 -
3 -
4 -
5 -

6 -
7 – Completely Support

artificial_r2: Cataloging Worker Injury Narrative by the Bureau of Labor Statistics
1 – Completely Oppose
2 -
3 -
4 -
5 -
6 -
7 – Completely Support

artificial_r3: Analysis of Adverse Drug Effects by the Food and Drug Administration
1 – Completely Oppose
2 -
3 -
4 -
5 -
6 -
7 – Completely Support

artificial_r4: Adjudicating Patent and Trademark Applications by the US Patent and Trademark Office
1 – Completely Oppose
2 -
3 -
4 -
5 -
6 -
7 – Completely Support

artificial_r5: Handwriting Recognition Tools used by US Postal Service
1 – Completely Oppose
2 -
3 -
4 -
5 -
6 -
7 – Completely Support

artificial_r6: Chatbots used by the US Citizenship and Immigration Services
1 – Completely Oppose
2 -
3 -
4 -
5 -

6 -
7 – Completely Support

artificial_r7: Tools to Counter Cyberattacks on Agency Systems by the Department of Homeland Security
1 – Completely Oppose
2 -
3 -
4 -
5 -
6 -
7 – Completely Support

----------------------End Web pg----------------------

Now we would like you to reconsider the hypothetical series of different policy domains where the government could one day incorporate artificial intelligence.

Please evaluate them on if you oppose or support the decision-making being done by artificial intelligence systems instead of humans.

Please respond to the following on a scale from one to seven, where one means Completely oppose and seven means Completely support.
[**random table for artificial1—artificial22**]

artificial1: Air Traffic Control
1 – Completely oppose
2 -
3 -
4 -
5 -
6 -
7 – Completely support

artificial2: Coordinating Emergency Responses
1 – Completely oppose
2 -
3 -
4 -
5 -
6 -
7 – Completely support

artificial3: Traffic Management
1 – Completely oppose

2 -
3 -
4 -
5 -
6 -
7 – Completely support

artificial4: Traffic Prediction
1 – Completely oppose
2 -
3 -
4 -
5 -
6 -
7 – Completely support

artificial5: Public Transportation Scheduling
1 – Completely oppose
2 -
3 -
4 -
5 -
6 -
7 – Completely support

artificial6: Public Utilities Management
1 – Completely oppose
2 -
3 -
4 -
5 -
6 -
7 – Completely support

artificial7: Public Parks and Recreation Management
1 – Completely oppose
2 -
3 -
4 -
5 -
6 -
7 – Completely support

artificial8: City Planning & Urban Development
1 – Completely oppose
2 -

3 -
4 -
5 -
6 -
7 – Completely support

artificial9: Running Political Elections
1 – Completely oppose
2 -
3 -
4 -
5 -
6 -
7 – Completely support

artificial10: Determining Human Rights
1 – Completely oppose
2 -
3 -
4 -
5 -
6 -
7 – Completely support

artificial11: Determining Immigration Policy
1 – Completely oppose
2 -
3 -
4 -
5 -
6 -
7 – Completely support

artificial12: Determining Foreign Policy
1 – Completely oppose
2 -
3 -
4 -
5 -
6 -
7 – Completely support

artificial13: Determining National Security Policy
1 – Completely oppose
2 -
3 -

4 -
5 -
6 -
7 – Completely support

artificial14: Determining Arts & Culture Funding
1 – Completely oppose
2 -
3 -
4 -
5 -
6 -
7 – Completely support

artificial15: Historical Preservation
1 – Completely oppose
2 -
3 -
4 -
5 -
6 -
7 – Completely support

artificial16: Cyber Security
1 – Completely oppose
2 -
3 -
4 -
5 -
6 -
7 – Completely support

artificial17: Weather Forecasting (extreme weather/chance of tornado)
1 – Completely oppose
2 -
3 -
4 -
5 -
6 -
7 – Completely support

artificial18: Weather Forecasting (minor weather/chance of rain)
1 – Completely oppose
2 -
3 -
4 -

5 -
6 -
7 – Completely support

artificial19: Determining Criminal Justice Policy
1 – Completely oppose
2 -
3 -
4 -
5 -
6 -
7 – Completely support

artificial20: Determining National Security Policy
1 – Completely oppose
2 -
3 -
4 -
5 -
6 -
7 – Completely support

artificial21: Deciding what projects Congress should fund
1 – Completely oppose
2 -
3 -
4 -
5 -
6 -
7 – Completely support

artificial22: Deciding whom to give Government Contracts too
1 – Completely oppose
2 -
3 -
4 -
5 -
6 -
7 – Completely support

----------------------End Web pg----------------------

As technology advances and governments continue to try and best provide for their constituents, the adoption of artificial intelligence in public policy decision making may become inevitable. If that were to happen how much confidence would you have in the following different entities developing this artificial intelligence?

[**random table for gov_1—other_2**]

gov_1: U.S. military
1 - A great deal of confidence
2 - A fair amount of confidence
3 - Not too much confidence
4 - No confidence
5 - Don't Know

gov_2: U.S. Civilian government
1 - A great deal of confidence
2 - A fair amount of confidence
3 - Not too much confidence
4 - No confidence
5 - Don't Know

gov_3: NSA
1 - A great deal of confidence
2 - A fair amount of confidence
3 - Not too much confidence
4 - No confidence
5 - Don't Know

gov_4: FBI
1 - A great deal of confidence
2 - A fair amount of confidence
3 - Not too much confidence
4 - No confidence
5 - Don't Know

gov_5: CIA
1 - A great deal of confidence
2 - A fair amount of confidence
3 - Not too much confidence
4 - No confidence
5 - Don't Know

gov_6: NATO
1 - A great deal of confidence
2 - A fair amount of confidence
3 - Not too much confidence
4 - No confidence

5 - Don't Know

gov_7: Intergovernmental research organizations (e.g., CERN)
1 - A great deal of confidence
2 - A fair amount of confidence
3 - Not too much confidence
4 - No confidence
5 - Don't Know

tech_1: Tech companies
1 - A great deal of confidence
2 - A fair amount of confidence
3 - Not too much confidence
4 - No confidence
5 - Don't Know

tech_2: Microsoft
1 - A great deal of confidence
2 - A fair amount of confidence
3 - Not too much confidence
4 - No confidence
5 - Don't Know

tech_3: Google
1 - A great deal of confidence
2 - A fair amount of confidence
3 - Not too much confidence
4 - No confidence
5 - Don't Know

tech_4: Facebook
1 - A great deal of confidence
2 - A fair amount of confidence
3 - Not too much confidence
4 - No confidence
5 - Don't Know

tech_5: Apple
1 - A great deal of confidence
2 - A fair amount of confidence
3 - Not too much confidence
4 - No confidence
5 - Don't Know

tech_6: Amazon
1 - A great deal of confidence

2 - A fair amount of confidence
3 - Not too much confidence
4 - No confidence
5 - Don't Know

tech_7: Tesla
1 - A great deal of confidence
2 - A fair amount of confidence
3 - Not too much confidence
4 - No confidence
5 - Don't Know

person_1: Bill Gates
1 - A great deal of confidence
2 - A fair amount of confidence
3 - Not too much confidence
4 - No confidence
5 - Don't Know

person_2: Mark Zuckerberg
1 - A great deal of confidence
2 - A fair amount of confidence
3 - Not too much confidence
4 - No confidence
5 - Don't Know

person_3: Jeff Bezos
1 - A great deal of confidence
2 - A fair amount of confidence
3 - Not too much confidence
4 - No confidence
5 - Don't Know

person_4: Elon Musk
1 - A great deal of confidence
2 - A fair amount of confidence
3 - Not too much confidence
4 - No confidence
5 - Don't Know

other_1: University researchers
1 - A great deal of confidence
2 - A fair amount of confidence
3 - Not too much confidence
4 - No confidence
5 - Don't Know

other_2: Non-profit(e.g., OpenAI)
1 - A great deal of confidence
2 - A fair amount of confidence
3 - Not too much confidence
4 - No confidence
5 - Don't Know

----------------------End Web pg----------------------

Please rate the degree to which each of the following four groups of statements describes your outlook on
life, using a scale from zero to ten, where zero means not at all and ten means completely.
[**random order**]

h_rate: I am more comfortable when I know who is, and who is not, a part of my group, and loyalty to the group is important to me. I prefer to know who is in charge and to have clear rules and procedures; those who are in charge should punish those who break the rules. I like to have my responsibilities clearly defined, and I believe people should be rewarded based on the position they hold and their competence. Most of the time, I trust those with authority and expertise to do what is right for society.
0 - Not at all
1
2
3
4
5
6
7
8
9
10 - Completely

i_rate: Groups are not all that important to me. I prefer to make my own way in life without having to follow other peoples' rules. Rewards in life should be based on initiative, skill, and hard work, even if that results in inequality. I respect people based on what they do, not the positions or titles they hold. I like relationships that are based on negotiated "give and take," rather than on status. Everyone benefits when individuals are allowed to compete.
0 - Not at all
1
2
3
4
5
6
7

8
9
10 - Completely

e_rate: My most important contributions are made as a member of a group that promotes justice and equality. Within my group, everyone should play an equal role without differences in rank or authority. It is easy to lose track of what is important, so I have to keep a close eye on the actions of my group. It is not enough to provide equal opportunities; we also have to try to make outcomes more equal.
0 - Not at all
1
2
3
4
5
6
7
8
9
10 - Completely

f_rate: Life is unpredictable and I have very little control. I tend not to join groups, and I try not to get involved because I can't make much difference anyway. Most of the time other people determine my options in life. Getting along is largely a matter of doing the best I can with what comes my way, so I just try to take care of myself and the people closest to me. It's best to just go with the flow, because whatever will be will be.
0 - Not at all
1
2
3
4
5
6
7
8
9
10 - Completely

----------------------End Web pg----------------------

Some books and movies portray a future where technology provides products and services that make life better for people. Others portray a future where technology causes environmental and social problems that make life worse for people. How about you?

future_tech: Over the long term, do you think that technological changes will lead to a future where people's lives are better or to a future where people's lives are worse?
1 - A lot better
2 - Mostly better
3 - Neither better nor worse
4 - Mostly worse
5 - A lot worse

----------------------End Web pg----------------------

doright: On a scale from zero to ten, where zero means none of the time and ten means all the time, how much of the time do you trust the government in Washington to do what is right for the American people?
0 - None of the time
1
2
3
4
5
6
7
8
9
10 - All the time

----------------------End Web pg----------------------

comments: This survey is part of a project that focuses on ***artificial intelligence***. Is there anything you want to tell us about your views on ***artificial intelligence***? [verbatim]

----------------------End Web pg---------------------

We thank you for your time spent taking this survey.
Your response has been recorded.