SETTING THE AGENDA FOR AI: ACTORS, ISSUES, AND INFLUENCE IN UNITED STATES ARTIFICIAL INTELLIGENCE POLICY

A Dissertation
Presented to
The Academic Faculty

By

Daniel S. Schiff

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Public Policy

Georgia Institute of Technology

August 2022

SETTING THE AGENDA FOR AI: ACTORS, ISSUES, AND INFLUENCE IN UNITED STATES ARTIFICIAL INTELLIGENCE POLICY

Thesis committee:

Dr. Juan Rogers
School of Public Policy
Georgia Institute of Technology

Dr. Jason Borenstein
School of Public Policy
Georgia Institute of Technology

Dr. Natália S. Bueno
Department of Political Science
Emory University

Dr. Scott C. Ganz
McDonough School of Business
Georgetown University

Dr. Ellen Zegura
School of Computer Science
Georgia Institute of Technology

Date approved: May 25, 2022

ACKNOWLEDGMENTS

It is a rare privilege to complete a doctoral dissertation on a subject one is passionate about, especially in such trying times. Getting to this point was only possible through the support and generosity of many individuals willing to take a chance on me.

Foremost, Kaylyn, I am grateful to have such a compassionate and brilliant life and research partner. You contributed above and beyond throughout critical stages of this dissertation, and served as my main source of meaning when we stopped work for the day.

To my chair Juan Rogers—as a true Renaissance man, you are a source of inspiration for the kind of interdisciplinary scholar and educator I wish to be. I was lucky to soak up just a fraction of your wisdom. Thank you for supporting my work and for holding up much of the School of Public Policy on the side.

To Jason Borenstein—one of my main reasons for coming to Georgia Tech was because of your research, but I didn't anticipate what a wonderful mentor and person you would be. Your encouragement, ethical leadership, collegiality, and support of me as a whole person has made all the difference these last few years.

To Ellen Zegura—you are an outstanding leader in every sense of the word, always stepping up and supporting me and others, and still finding the time focus on family and promote social good in every space you inhabit. I don't think there are many other Ellens out there, so I'm deeply grateful to count you as a mentor.

To Scott Ganz—you encouraged me to continually grow and shared insights I wouldn't have gotten anywhere else, leading to some of the most rewarding experiences of my doctoral journey. Beyond that, thanks for reminding me what's really important in life.

To Natália Bueno—you generously took me on as a student from another institution, and played a key role in shaping the kind of researcher I am. The principles by which you approach your teaching and research are ones I hope to emulate, and graduating from being your student to collaborator has been a highlight of my time in Atlanta.

This dissertation would not have been possible without the participation of folks at The Future Society including Nicolas Möes, Sam Curtis, and Yolanda Lannquist. Many thanks to Pam Greenberg at the National Conference of State Legislatures and Ashok Goel and Ron Arkin at Georgia Tech for their contributions as well. I am also grateful to discussants and chairs at APPAM, APSA, and other conferences, as well as to numerous scholars who generously provided their time and feedback on these projects.

There are many other individuals who have enriched my doctoral work. My deepest gratitude to policy, philosophy, and computing scholars like Justin Biddle, Michael Hoffmann, Gordon Kingsley, Aaron Levine, Julia Melkers, Ayanna Howard, Neha Kumar, Alan Marco, Emmanuelle Massetti, and Cassidy Sugimoto. It was a special privilege to be welcomed into the community of philosophers at Georgia Tech. My thanks as well to mentors during my years at Princeton University and the University of Pennsylvania who encouraged me along this path when AI policy and AI ethics were merely futuristic ideas.

I'm grateful to my brilliant collaborators, colleagues, and friends in the AI policy and ethics communities in the U.S. and abroad, and to my fellow travelers in other disciplinary spaces. Knowing that I work alongside such open-minded, supportive, and passionate people makes me hopeful that we can tackle these difficult issues together.

To my cohort mates and friends at Georgia Tech, Georgia State, and Emory, I'm fortunate to have grown, commiserated, and celebrated alongside such a talented and thoughtful group. I'm excited to see the wonderful contributions you all will continue to make.

To my friends who wondered what the heck I've been doing for the last five years, thanks for bearing with me and know that I'm proud of what you all are doing.

To my parents-in-law Dana and David, and siblings-in-law Elizabeth and Eric, I'm truly fortunate to have been adopted into such an amazing second family.

To my parents Patty and Art and sister Julia, I couldn't imagine a better way to spend these past five years than back at home in Atlanta with you all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

As research and adoption of artificial intelligence (AI) has significantly advanced in the early 21st century, determining how to govern AI has become a global priority. Key questions include how AI should be understood as a policy domain, which policy problems are most pressing, which solutions are most viable, and who should have a say in this process. This dissertation seeks to provide key insights into the early years of AI policy, focusing on the development of the emerging AI policy agenda in the United States. To do so, it examines and reveals which issues, actors, and influence efforts are playing a prominent role in the complex, ambiguous, and contested process of agenda-setting. The research performed draws on a variety of quantitative and qualitative methodologies, including document analysis, text-as-data and time series approaches, and experimental techniques. Data examined include text from U.S. federal AI policy documents, traditional and social media discourse from federal policymakers, media, and members of the public, and engagement data collected from state legislators who participated in a field experiment.

The results reveal that social and ethical dimensions of AI receive a heightened degree of attention in AI policy discourse. However, consideration of these issues remains partially superficial and subsumed into concern about AI's potential for economic innovation and role in geopolitical competition. Further findings demonstrate that policy entrepreneurs can use persuasive narratives to influence legislators about AI policy, and that these narratives are just as effective as technical information. Finally, despite pervasive calls for public participation in AI governance, the public does not appear to play a key role in directing attention to AI's social and ethical implications nor in shaping concrete policy solutions, such that the emerging AI agenda remains primarily expert-driven. The dissertation's findings and theoretical and methodological approaches offer key contributions to policy process scholarship and related fields of research, and provide a baseline on which to understand the evolution of the AI policy agenda and AI governance going forward.

CHAPTER 1

INTRODUCTION

This dissertation examines agenda-setting dynamics surrounding artificial intelligence (AI) policy in the United States (U.S.). AI policy–understood in the context of the early 21st century–is a relatively new endeavor. Major advances in research, development, and innovation surrounding AI in the past decade have already led to sweeping changes across many sectors of social and economic life. In turn, attention to AI's potential for innovation as well as associated social and ethical concerns is increasing.

As these issues are beginning to crystallize, a variety of stakeholders have proposed responses, including both informal and formal efforts to develop, manage, and regulate AI. Some proposals have included: increased funding for research, educational efforts to expand the labor supply of high-skill workers, ethical frameworks to mitigate harms of AI systems, and new domestic and international regulatory regimes to balance possible benefits and risks. While many actors will play a role in the collective effort to develop, manage, and regulate AI globally, including the private sector, this dissertation focuses on formal regulation and policymaking by federal and state government in the United States.

In short, there is a critical window now during which key decisions about AI's future impacts on society will be made, decisions which will shape how policies responding to AI are formulated, adopted, and ultimately implemented. To examine the dynamics that are manifesting during this window, I conduct a series of studies on agenda-setting of U.S. AI policy, drawing on policy process theory along with other related scholarship. The dissertation is structured into five chapters, including three research projects designed uniquely for this dissertation, and an Introduction and Conclusion. I begin below by introducing relevant definitions and concepts, reviewing the historical and modern context of AI, and describing the conceptual framework surrounding AI policy which underpins this project.

1.1 Foundational Concepts and Context

To understand the policy and agenda-setting dynamics that pertain to AI, it is first important to define, characterize, and contextualize AI. This section provides definitions, a brief history, and a discussion of the modern economic, social, and policy context under which agenda-setting is occurring for AI.

1.1.1 Defining AI

It is uncontroversial that AI refers to a multifaceted and complex set of systems, methods, and technologies, and that attempts to define it have been correspondingly diverse and contested. In part due to the slippery boundaries of what is considered AI that result from this definitional ambiguity, some find the use of the term 'AI' to be too broad altogether. Instead, they argue that it is more helpful in certain contexts to rely on alternative terms, such as autonomous or intelligent systems, algorithms, or machine learning.[1] Further complicating the issue is different treatment of AI depending on disciplinary approach. For example, a glossary created by IEEE (2020) identifies distinct definitions in ordinary language, computational disciplines, engineering, economics and social sciences, ethics and philosophy, and international law and policy.[2] Yet others have remarked that AI's definitional ambiguity has been productive in advancing the field (Agre, 1997) and making it accessible to many disciplines with a historical or current stake in AI, ranging from philosophy and linguistics to psychology and neuroscience.

Despite these complexities, some definitions have become more prominent than others. One such attempt to define AI comes from the technical literature. A foundational AI textbook by Russell and Norvig (1995) reviewed others textbooks' definitions and classified them as describing four types of AI systems: systems that attempt to 1) 'think' like humans,

[1] Machine learning refers to a subfield of AI in which algorithms dynamically adjust to new data, with limited human intervention, in order to improve the performance of a certain task or model (Alpaydin, 2016).

[2] Another conceptual feature of note is the paradoxical definitional shift known as the "AI effect," wherein once-challenging tasks that are finally achieved by AI are subsequently considered mainstream and then no longer attributed to AI (McCorduck, 2004).

2) think rationally (that is, in an idealized sense), 3) behave like humans, or 4) behave rationally. In Sweeney's (2003) review of around 1000 academic references in AI research, she found nearly 99% focused on rational or ideal-type definitions, suggesting some degree of consensus in the technical literature.[3]

Yet technical definitions may not be the most salient for policy contexts, as policy-relevant definitions may seek alternative objectives (Murdick et al., 2020). For example, in establishing its definition of AI for its landmark AI policy regulation, the European Union (EU) noted that an adequate definition for AI should be clear, narrow, and precise, but also "technology neutral" and flexible enough to be "future proof" in order to minimize policy and legal uncertainty (Martinez, 2019) as the techniques that constitute AI evolve over time (European Commission, 2021).[4] One such helpful definition aimed at policy and legal clarity—later adopted with modifications by the EU—was produced by the Organisation for Economic Co-operation and Development (OECD) through a multi-stakeholder process. The OECD (2019) defines an AI system as: "a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy." This operational definition recognizes a few elements common in many definitions, such as an AI system's ability to draw on or perceive data, make predictions or recommendations or otherwise act in some external environment, and operate with some level of autonomy and adaptability, potentially including a capacity for 'learning.'

It is also helpful to explicitly extend AI systems into their broader socio-technical context (Elish & boyd, 2018), conceiving of AI in terms of a multi-level taxonomy.[5] In this more granular sense, AI can be understood as 1) a set of mathematical, statistical, or com-

[3]Ongoing efforts from standards development organizations including the IEEE, ISO, and IEC towards developing definitional taxonomies and ontologies for AI are also relevant.

[4]Other policy-oriented goals related to defining AI adequately may include use of common language for communication, accountability, assessment of risks and harms, and measurement of progress.

[5]Also see Samoili et al. (2020) who develop a taxonomy of AI domains (e.g., reasoning, planning, learning) and subdomains (e.g., fuzzy logic, gradient descent, classification) based on a review of 55 definitions, or the OECD's (2022b) robust stakeholder-driven classification framework which identifies the key dimensions of AI as Data and Input, AI Model, Task and Output, Economic Context, and People and Planet.

putational techniques that are 2) applied to certain domains of computer science, and 3) implemented across social and economic systems. The statistical techniques include, for example, supervised machine learning methods such as regression and classification, unsupervised methods such as clustering and dimensionality reduction, symbolic logic- and knowledge-based approaches, and more: These range from relatively familiar (e.g., ordinary least squares regression) to relatively novel (e.g., generative adversarial networks or neuro-symbolic AI) (Fadlullah et al., 2017).[6] Domains of computer science include, for example, natural language understanding and generation, computer vision, and search and optimization (Elsevier, 2018). Finally, AI systems are then embedded into various social and economic sectors (Perset et al., 2020), covering everything from agriculture to transportation. This clarifies that they are not merely abstract software systems, but also part of a complex social and technical fabric that is indispensable to understanding the policy context.

A few other minor definitional points bear mention. One is that AI can be embodied in a physical system such as through industrial robotics or autonomous vehicles, or it may not be so embodied. A second point is the distinction between 'narrow' (or 'weak') AI, AI that that can be applied to only very specific pre-determined domains (such as playing chess), versus 'general' (or 'strong') AI, AI systems capable of operating with human-like intelligence across many domains.[7] While AI systems that exist today are all narrow systems, policy considerations for AI are not necessarily limited to narrow AI, as some actors have urged policymakers to be forward-looking in their deliberations (Baum, 2017; Hedlund, 2022; Naudé & Dimitri, 2020). Finally, in modern parlance, the terms and (concepts) 'AI' and 'automation' are at times distinguished and at times used interchangeably.

[6]A comprehensive review or definition of the techniques that constitute AI historically is beyond the scope of this project, but for helpful summaries, see Metz (2021) and Wooldridge (2021). It is important to note that while symbolic methods, most prominently expert-based systems, were dominant until the 1980s (Brooks, 1999), machine learning techniques dominate the current paradigm of AI and feature heavily in modern public and policy discourse (Cockburn et al., 2018).

[7]More advanced forms of AI are sometimes described as artificial general intelligence (AGI), narrowly or radically transformative AI (TAI) or, even more distantly, artificial super intelligence (ASI) (Girasa, 2020; Gruetzemacher & Whittlestone, 2022).

While all of the classifications discussed above can be helpful in certain contexts, for the purposes of this dissertation it is especially important to approach AI in a way that is commensurate with policy discourse and often general public discourse. This is necessary because policy documents do not use the same definitions as do technical experts and researchers; they are more likely to use definitions that emphasize human thinking and behavior, or to fail to provide definitions at all (Krafft et al., 2020). Moreover, it is common for policymakers to (perhaps intentionally) treat AI as "a broad umbrella term" which loosely incorporates many of the characteristics and variations described above (Ulnicane et al., 2020, p. 2). This definitional ambiguity should thus be understood as a *feature* of AI policy discourse, one that engenders strategic and contentious attempts to frame AI in certain ways, amongst other effects. Accordingly, while there are certainly numerous component policy questions regarding specific types of AI systems and applications in various policy sectors, this thesis treats AI and AI policy at the high level of generality that most commonly marks the agenda-setting stage of AI policy and associated AI policy discourse.

1.1.2 History and Modern Context

The origin of the modern field of AI–including the coining of the term "artificial intelligence" by John McCarthy–is often taken as the 1956 Dartmouth Summer Research Project (Moor, 2006). This convening marked the beginning of a decade of investment, facilitated by the formation of the Advanced Research Projects Agency (ARPA) two years later, in 1958. Fifteen years after this history-making conference, the 1973 Lighthill Report commissioned by the British Scientific Research Council would cast considerable doubt on overpromises and a lack of progress in AI research, leading to the "First AI Winter" in the 1970s. This was followed by the Second AI Spring in the 1980s, a period in which expert systems were the dominant form of AI systems (Gonsalves, 2019). This wave of research was particularly facilitated by the 1983 Strategic Computing Initiative funded by ARPA's successor, the Defense Advanced Research Projects Agency (DARPA). Yet after progress slowed in the late 1980s, AI would undergo a Second Winter with significant funding cuts.

Progress in the 2000s and 2010s, including theoretical advances in neural networks and applications in image recognition and natural language processing, would set the stage for the 21st century paradigm of AI (Metz, 2021). Buttressed by increased processing power and an unprecedented amount of data made available from the Internet, AI researchers were able to use new and old techniques in a manner that quickly captured industry attention given a range of possible applications (Duan et al., 2019). Exemplifying this shift in focus, Google announced it was now an "AI-first" company instead of "mobile-first" in 2017, rebranding its Google Research division into Google AI (Lardinois, 2017). American technology giants like Microsoft, Facebook, Apple, and Amazon, and Chinese tech giants like Tencent, Alibaba, and Baidu expressed the same eagerness to embrace the new wave of innovation, publicly pronouncing new priorities around AI and massively increasing research and development efforts as well as competing to acquire smaller AI firms (Makridakis, 2017).

These efforts from the world's dominant technology companies are emblematic of the importance attributed to AI in the early 21st century. AI has been analogized to the Internet, electricity, and fire (Parker, 2018) and identified as a central technology to the so-called Fourth Industrial Revolution (Schwab, 2016). Estimates place AI's near-term global economic impact on the order of 10-15 trillion dollars annually (a total GDP increase of 26% by 2030), corresponding to more than 1% additional annual GDP growth (Bughin et al., 2018; Rao & Verweij, 2017). In turn, researchers and practitioners have discussed or posited impacts of AI across every sector of social and economic life: A partial list of key sectors includes agriculture (Jha et al., 2019), transportation (Bagloee et al., 2016), finance (Bredt, 2019), criminal justice (Završnik, 2019), education (Schiff, 2021), and healthcare (Yu et al., 2018). AI has seen similar growth in terms of research and innovation metrics, with multiplicative increases in the number of journal publications, patents, conference attendees, and a suite of new organizations focused on AI founded in the 2010s and 2020s

(Elsevier, 2018; D. Zhang et al., 2021).

A striking feature of this new Spring of AI research and development is the accompanying attention to social and ethical risks associated with AI-based innovation. New communities of researchers and practitioners have formed to highlight how AI raises concerns associated with accountability, bias, inequality, human rights, misinformation, manipulation, privacy, transparency, trust, and many more issues (Coeckelbergh, 2020; Fjeld et al., 2020; Schiff et al., 2021), both generally and related to specific social sectors and use cases. Some have warned of potentially devastating harms, such as large-scale labor displacement due to automation of work (Frey & Osborne, 2017), and even existential risks due to the possibility of self-improving AI systems that cease to be aligned with human values and goals (Baum, 2017; Bostrom, 2014). This attention to potential harms of AI has led to urgent work to devise technical and organizational practices, standards, and policies to mitigate risks and promote 'responsible' (or 'ethical' or 'trustworthy') AI research, development, and implementation (Golbin et al., 2020; Mökander et al., 2021; Stahl et al., 2021).

All of the elements above—relatively rapid uptake, substantial investment, multi-sector application, and concerns about AI's harms—constitute the modern context of AI and AI governance. Yet, especially in light of historical waves of progress and disillusionment, and a track record of faulty predictions, some have cautioned about the risks of continuing to hold either overly optimistic or pessimistic expectations (Armstrong et al., 2014; Oravec, 2019). According to this perspective, the transformations brought about AI may not be as radically good or bad as some imagine, or at least may not occur as rapidly as anticipated (Nazareno & Schiff, 2021). In that sense, AI in the 21st century may be understood more modestly, neither as wholly unique nor as driving a radical break with the past, but instead as taking its place alongside a history from which lessons can be drawn (Cave et al., 2018; Leung, 2020; Morley et al., 2019).

1.1.3 Policy Context

The elements above help to explain why policymaking for AI appears to constitute a relatively urgent priority. Indeed, according to the OECD's AI Policy Observatory, at least 600 AI policy initiatives were created by approximately 60 countries and territories between about 2016 and 2022 (OECD, 2022a). While even earlier policy initiatives surrounding digital technologies, automation, and robotics may also address AI to some extent, distinct efforts to specifically address AI policy by governments and intergovernmental bodies sharply increased around 2017 and 2018. Since then, and as of 2021, approximately three dozen countries have published formal AI strategies, with another couple dozen in the process of development, accompanied by increases in actual AI legislation (OECD, 2022a; D. Zhang et al., 2021).

While the majority of early-moving nations appear to be high-income countries and technological leaders (Schiff et al., 2020; The Future Society et al., 2020), attention to AI by policymakers is expanding globally. For example, UNESCO (2021) found that 21 countries in Africa consider development and use of AI as a priority in national development plans. Beyond national-level policymaking, similar efforts are also emerging bilaterally and multilaterally through entities such as the United Nations (UN), EU, G20, OECD, World Economic Form, Council of Europe, North Atlantic Treaty Organization (NATO), as well as newly formed organizations like the G7-led Global Partnership on AI (GPAI) (Cihon et al., 2020; Schmitt, 2021). Finally, these governmental and intergovernmental efforts are complemented by hundreds of frameworks, reports, and recommendations produced by civil society groups, industry bodies, academic collaborations, private firms, and various other groups interested in shaping AI policy (Jobin et al., 2019; D. Zhang et al., 2022).

These policy documents, particularly national documents, address similar issues to a significant extent (Schiff et al., 2021). They typically focus on how to facilitate and realize the transformative potential of AI while managing and mitigating associated social, ethical, economic, and legal risks. For example, documents often discuss mechanisms to advance AI research and development: These include direct funding, promotion of computer science education, creation of intermediary institutions to foster innovation, strategies to support startups and other enterprises, expansion of access to data and data infrastructure, and international cooperation. Notably, these policy objectives are hallmarks of traditional innovation policy (Edler & Fagerberg, 2017), focused on fostering technological advancement, economic growth, and resulting benefits that accrue. Governmental AI policy objectives also often include strategies for particular sectors recognized as growth areas or areas in need of development for a given country, such as healthcare, finance, or agriculture. Finally, one striking commonality across most national AI policy documents is an explicit emphasis on ethics, which involves articulating social and ethical risks associated with AI as well as principles and strategies that may be adopted to respond to those risks (Schiff et al., 2020; Ulnicane et al., 2021). This focus is not as typical in traditional innovation policy and, as this dissertation considers, may represent a newer paradigm in technology governance along the lines of what Diercks et al. (2019) call "transformative innovation policy."

In the United States, mainstream federal attention to AI began in 2016 and 2017 with three key documents: reports on Artificial Intelligence, Automation, and the Economy (2016) and Preparing for the Future of Artificial Intelligence (2016), and the National Artificial Intelligence Research and Development Strategic Plan (2016). Despite these relatively early efforts globally, the United States lagged behind countries like Canada, Japan, and the United Kingdom in advancing AI policy in subsequent years (Department of Public Administration, Italy, 2018; Washington Post Editorial Board, 2021), as other countries began to announce new initiatives, commissions, and preliminary policy strategies. The announcement of the 2019 American AI Initiative marked the beginning of the next wave of policy effort in the United States. While there had been little discussion of AI in legislation, hearings, Congressional Research Service reports, and so on, policy activity skyrocketed in 2018. Sustained attention resulted in dozens of hearings, numerous legislative proposals and Executive Orders, and the formation of new committees and organizational bodies within and across different branches of government, all aimed at AI policy (Chae, 2020; D. Zhang et al., 2022).

Prominent U.S. efforts include work by agencies such as the National Institute of Standards and Technology (NIST) (2021; 2022), the Federal Trade Commission (FTC) (2021), and the United States Patent and Trademark Office (USPTO) (2020); proposed legislation such as the Algorithmic Accountability Act (2019) and the Advancing American AI Act (2021); and the Office of Management and Budget's (OMB) (2020) Memorandum to Executive Agencies on Guidance for Regulation of AI Applications, amongst other key documents.[8] Notably, military-related activity has been arguably the fastest-moving effort, with more hearings in the Senate and House Committees on Armed Services addressing AI as compared to any other committee and the production of hundreds of pages of reports by the National Security Commission on AI (2021). Related developments include the announcement of the National AI Initiative and Office, National AI Advisory Committee, Select Committee on AI, Interagency AI Committee, and the elevation of the Joint Artificial Intelligence Center as part of the (2021) National Defense Authorization Act.

In light of all of these efforts, what do we know about the objectives of AI policymaking? According to the orienting missions statements, preambles, principles, and scope conditions commonly stated in key U.S. documents, primary goals include: to "sustain and enhance the scientific, technological, and economic leadership of the United States in AI," (White House, 2019, p. 3967), while also recognizing "safety, fairness, welfare, transparency, and other social goals" (Office of Management and Budget, 2020, p. 2). While these sentiments are reflective of commonly-espoused goals for AI policy in the United States and elsewhere, they are harder to achieve in practice, in part because they demand

[8] While presenting a comprehensive review of U.S. governmental efforts, even limited to the federal level, is difficult given the evolving landscape, the case study in the next chapter provides a more detailed discussion covering numerous such efforts. A large number of key documents can be found at the National AI Initiative's AI document repository currently hosted at AI.gov (National AI Initiative Office, Office of Science and Technology Policy, 2022).

careful tuning and balancing of risks and benefits throughout numerous governance instruments. For example, the European Commission, which has been more proactive on AI policy, notes that its objectives for AI policy include having "a balanced and proportionate" regulatory approach that is also "robust and flexible" and "comprehensive and future-proof," but nevertheless operates "without unduly constraining or hindering technological development" (European Commission, 2021, p. 3). This is no easy charge.

In turn then, several key and inter-related questions—not limited to those below—have permeated across AI policy discourse, reflecting both uncertainty over how to achieve these stated policy objectives, and tension between potentially competing goals and actors:

- **Timing**: How quickly should formal regulation of AI proceed? While some have argued that AI's rapid advancement is outpacing the ability of governments to respond to pressing problems (European Digital Rights (EDRi), 2021) and that legislation is an urgent priority (European Parliament, 2020), others have cautioned that governments need to understand AI's technical and other dimensions more or they risk suppressing innovation (Microsoft, 2018).

- **Centralization**: Should AI be governed by a centralized or decentralized regime? This question applies domestically as well as internationally, in that it is debated whether a few leading agencies or intergovernmental bodies should be tasked specifically with managing AI or whether responsibility should be distributed amongst individual countries or agencies (Cihon et al., 2020). For instance, the EU (2021) has articulated a common multi-sector or 'horizontal' regulatory regime vesting authority in a few institutions, while the United States seems to favor of an ad-hoc sectoral approach.

- **Self-regulation**. Is the private sector or government best suited to 'govern' (or 'manage') AI, given that the bulk of AI development occurs in private firms (Fischer et al., 2021)? Advocates of self-governance have noted that industry has greater technical expertise and has proposed numerous responsible AI principles or practices (Marchant, 2019). Yet skeptics warn that these proposals are insufficient (Baker, 2021) or not fully genuine, and strategically aimed at deterring sorely-needed formal regulation (Railsback & Rodrigues, 2020).

- **Coordination**: Should governance be a competitive or cooperative endeavor? An increasingly prominent frame shaping AI policy warns of great power competition and even an arms race, particularly between the United States and China (Lee, 2018). Yet there also numerous calls for, and increasing efforts surrounding, global cooperation in research, trade, data sharing, and alignment on ethical principles (Ulnicane et al., 2021).

- **Precaution**: Should regulators take a relatively 'pro-innovation' or 'precautionary' approach to AI? Some have argued for the latter, including banning certain applications of AI, such as facial recognition or social scoring, and requiring risk or impact assessments to minimize harms from AI systems (European Commission, 2021; European Digital Rights (EDRi), 2021). Yet others warn that these kinds of restrictions are burdensome and unnecessary, especially for smaller firms, and may deter valuable innovation (Barczentewicz & Mueller, 2021; Walker, 2020). For example, the (2020, p. 2) OMB guidance to U.S. agencies states that "Agencies must avoid a precautionary approach..." including "regulatory or non-regulatory actions that needlessly hamper AI innovation and growth."

In addition to further numerous details that will be worked out in the years ahead, including issues within specific sectors such as transportation or healthcare, these questions characterize some of the essential debates taking place as the AI policy agenda emerges.

1.2 Conceptualizing AI Policy

With an understanding of how AI is defined, its history and modern development, and the current policy context, it is now possible to conceptualize AI as an object of policy study.

1.2.1 AI as a Strategic, Emerging, and General Purpose Technology

What defining attributes of AI are especially relevant to how AI policy is understood and shaped? I draw on three conceptualizations of technology in order to help characterize AI: general purpose technology, strategic technology, and emerging technology. The resulting synthesis is reflected in the left-most panes in Figure 1.1.

First, as noted previously, AI is typically considered to be a *general purpose technology* (GPT). According to Bresnahan and Trajtenberg, who coined the term in 1995, GPTs function as enabling rather than final technologies, and are characterized by their pervasiveness, technological dynamism, and downstream complementarities for innovation. Related to this definition, Cockburn et al. (2018, p. 2) argue that a further transformative aspect of AI is that it may function as an "invention of a method of inventing" (Griliches, 1957) (or IMI), inducing even more returns to scale. While the exact course that a GPT takes over time may be unclear early on, as in the case of the internet (Ristuccia & Solomou, 2014), and the productivity gains of GPTs may be overstated or decline over time (Naughton, 2016), even the *potential* of major cross-cutting impacts of a GPT can drive substantial attention and investment (a self-fulfulling prophecy of sorts).

Second, Leung (2020, p. 5) further characterizes AI as a *strategic GPT*, or a GPT "which has the potential to deliver vast economic value and substantially affect national security, and is consequently of central political interest to states, firms, and researchers." It is strategic in that sense that it has both military *value* and can pose a security *risk* if not managed carefully, with risks ranging from accidents to malicious misuses to broader structural risks. As such, while the vast economic benefits would be likely to attract the attention of state actors as well as private industry, Leung notes (p. 36) that the security dimension of AI "*guarantees* the involvement of the state" [emphasis added]. Pervasive and sustained attention to military dimensions of AI in the United States by government, researchers, and security-oriented think tanks confirms as much (for example, see: Allen & Chan, 2017; Defense Science Board, 2016; Kratsios, 2019; Spiegeleire et al., 2017). Notably, the dual use nature of AI as reflected by its simultaneous value and risk profile is already characteristic of its status as a GPT; yet the stakes are significantly elevated because of AI's prominent economic and military relevance.

Third, Rotolo et al. (2015, p. 4) define an *emerging technology* as "a radically novel and relatively fast growing technology characterised by a certain degree of coherence persisting over time and with the potential to exert a considerable impact" across socioeconomic domains. Here, novelty refers to a radical discontinuity (Day & Schoemaker, 2000), either in the fundamentals of the technology or in new applications. Rapid growth may be measured in terms of factors like knowledge output, actors involved, funding, or new products and services. Coherence refers to the formation of a stable and distinct community of practice. Prominent impact, a characteristic of GPTs as well, refers to multi-level, cross-cutting impacts on organizations, institutions, technological regimes, and knowledge production itself. Yet, Rotolo et al. emphasize (p. 4) that the "most prominent impact, however, lies in the future and so in the emergence phase is still somewhat uncertain and ambiguous." As such, emerging technologies are those marked by radical novelty, fast growth, coherence, prominent impact, and uncertainty and ambiguity.

A fairly modest number of modern technologies might meet the definition of a strategic, emerging, and general purpose technology, including, for example, aerospace technology, automotive technology, biotechnology, computing technology, nanotechnology, and nuclear technology (Harris, 2016; Leung, 2020). Yet, each technology differs along the extent to which it meets the above conditions, and each brings with it a unique technical, social, and economic context that alters the nature of associated policymaking. This limits

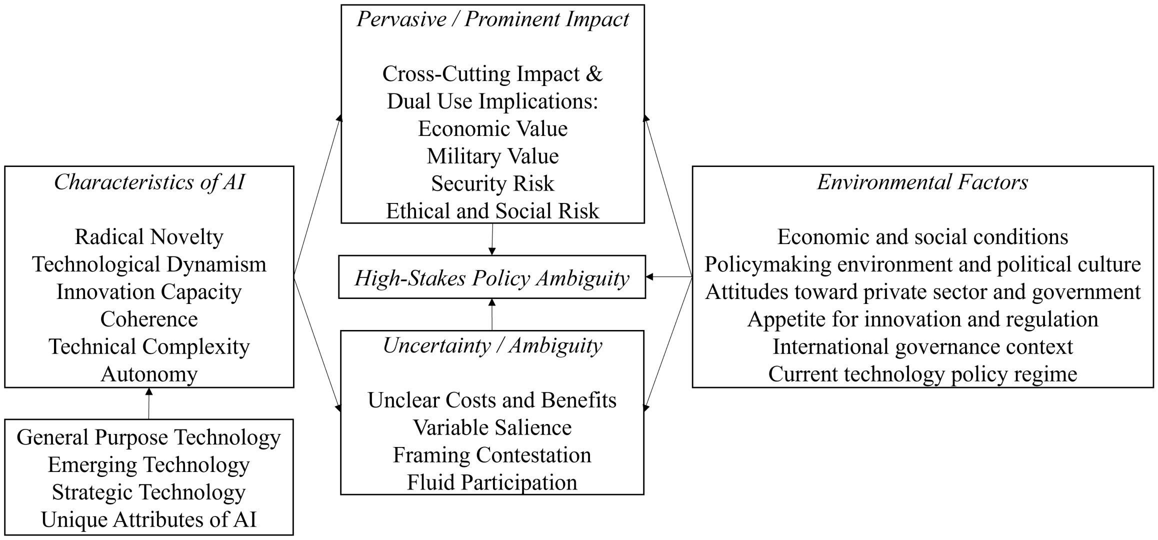

Figure 1.1: Conceptual framework: High-stakes policy ambiguity in AI policy

the extent to which learning about one case can guide insights about the others.

As such, two final characteristics associated with AI merit discussion. The first is autonomy, a common element of definitions of AI, referring to AI's capacity for semi-independent functioning. Autonomy merits special note because AI systems are argued or speculated to be capable of replacing humans in performing many tasks (Frey & Osborne, 2017); yet the bounds and timelines surrounding AI's future capabilities are unclear (Dreyfus, 1992; Müller & Bostrom, 2016). Autonomy thus compounds both technical and policy uncertainty. The second characteristic is technical complexity, which is not atypical for digital and information and communication technologies like AI. That is, while some topics of policy concern are relatively accessible to the general public, such as smoking policy and minimum wage regulation, AI represents a highly sophisticated domain traditionally considered to be dominated by highly-educated experts—an assumption this dissertation examines. AI's complexity is relevant as it further shapes how policy issues are understood, which participants are influential, and how challenging it is to construct effective and acceptable policy solutions (Carmines & Stimson, 1980; Gormley, 1986; Zito, 2001).

1.2.2 Prominent Impact and Uncertainty/Ambiguity

Overall, many of the above characteristics are discussed commonly in the context of AI, including in key policy documents (for example, see: European Commission, 2021; White House, 2019). Jointly, they establish the grounds for what I term *high-stakes policy ambiguity*, depicted in the middle panels of Figure 1.1. The high-stakes policy ambiguity associated with AI can be understood with respect to a few key elements, which I argue play a major role in shaping AI as an object of policymaking. These include AI's pervasive and prominent impact, the underlying conditions of uncertainty and ambiguity, and several influential environmental factors.

The first key element of high-stakes policy ambiguity is *pervasive and prominent impact*. As discussed above and in the literature, this refers to how a strategic, emerging, and general purpose technology like AI has cross-cutting impacts with associated dual use implications. As such, it offers both substantial economic and military value, while also posing security as well as ethical and social risks (Dwivedi et al., 2021; Kania, 2018).[9]

Second, characteristics like novelty, dynamism, complexity, and autonomy imply that AI will be marked by significant *uncertainty and ambiguity*. Much of the uncertainty stems from how costs and benefits resulting from AI may arise and be distributed (O'Shaughnessy et al., 2022). Some popular narratives feature clear 'losers,' such as workers displaced by automation (Berg et al., 2018) or racial minorities harmed by biased algorithms (Turner-Lee et al., 2019). Other narratives feature 'winners,' such as high-skill workers who remain employed (Frey & Osborne, 2017) or forward-thinking businesses savvy enough to adopt AI quickly (Makridakis, 2017). Still other discourse emphasizes both opportunities and risks for large and small businesses and governments, along with contests between actors that will determine the eventual winners and losers (Lee, 2018).

[9]Because a large body of work (for example, see: Wang & Siau, 2019) describes these pervasive and prominent impacts, including the implied dual use nature of AI applications, I will not reiterate these details again. However, it is worth emphasizing that this characterization of AI is extremely widespread and uncontroversial, and features prominently in policymaking considerations about AI.

These open-ended consequences are reflected in public opinion studies which demonstrate mixed and conflicting sentiment towards AI (Johnson & Tyson, 2020; Knight Foundation and Gallup, 2022), with differential attitudes depending on regional context and across economic, social, educational, and racial subgroups (Rainie et al., 2022). As an example of some associated nuances, while many workers expect automation to lead to widespread labor displacement, they do not expect it to threaten their own jobs (Smith & Anderson, 2017). Americans also tend to support AI innovation while simultaneously supporting careful management, though less so regulation (O'Shaughnessy et al., 2022), and attitudes vary somewhat based on the particular AI use case (U.S. Chamber of Commerce, 2022). Public understanding about AI is also relatively nascent. For example, B. Zhang and Dafoe (2019) find that the majority of the American public believes that virtual assistants, smart speakers, driverless cars, and autonomous drones use AI, but they wrongly assume that Google search, Netflix recommendations, and Facebook photo-tagging do *not* use AI. As another example, while 85% of a United Kingdom sample claim to have heard of AI (Cave et al., 2019), only 9% had heard of the term machine learning (Ipsos MORI, 2017). Overall, according to public opinion research performed by DeCario and Etzioni (2021), only about 16% of the American public is 'AI literate' as measured by scoring above 60% on a factual test about AI.

While the specifics will no doubt shift over time, there are thus good reasons to think that awareness of and attitudes towards AI and its applications are highly variable. Applications such as autonomous vehicles may be perfectly salient (Smith & Anderson, 2017), whereas applications such as AI in banking may be relatively invisible (Ipsos MORI, 2017). Variable salience to the public as well as the cross-cutting nature of AI applications suggest that fluid participation (Cohen et al., 1972) will strongly characterize AI agenda-setting. Different actors, depending on their awareness, resources, and incentives, may engage with many, few, or no applications of AI. These complex boundaries entail that policymakers must consider one set of stakeholders regarding one AI sub-issue (e.g., autonomous vehicles) and another for a distinct issue (e.g., AI in healthcare, autonomous weapons, or facial recognition) (Peters, 2021). How these boundaries are negotiated implicitly or deliberately has important implications for policy influence efforts, agenda-setting, and ultimately AI governance.

Of particular relevance to this dissertation is that uncertainty over costs, benefits, and participation has downstream consequences for how AI is understood—or *framed*. Issue (or policy) frames refer to how policy-relevant problems, solutions, and issues generally are defined, and thus serve as "social constructs reflecting particular conceptions of reality" (Elder & Cobb, 1984, p. 115). The act of framing is arguably a necessary one for any policy topic, as how a topic should be understood is neither automatic nor obvious a priori (Gilardi et al., 2021). Functionally, issue frames highlight some aspects of an issue while minimizing others, creating boundaries and structuring the very categories associated with a topic (Baumgartner & Jones, 1993). When used effectively, frames can shape public and elite attitudes and decision-making by legitimizing certain policy concerns (or problems) and bringing particular policy alternatives (or solutions) to the decision agenda (Goddard & Krebs, 2015). Importantly, the act of framing is typically a strategic action and one that admits of disagreement, as competing actors may seek to expand or contain issues and build coalitions, in order to influence decision-makers (Baumgartner & Mahoney, 2008; Rein & Schön, 1996). As such, framing can affect the grounds of public and elite participation in policymaking itself, a topic I return to in the third chapter.

Yet, while framing has been shown to play a major role in policy topics ranging from immigration (Merolla et al., 2013) to public health (Menashe, 1998), some policy issues are arguably more susceptible to diverse framings than others. In particular, the context of high-stakes policy ambiguity implies that issue framing for AI will be unusually highly contested. Because of the cross-cutting and general purpose nature of AI policy, coupled with uncertainty and ambiguity, the range of potentially viable frames is extremely diverse (Cave et al., 2019; Cave et al., 2018; Imbrie et al., 2021). For example, while smoking as policy topic may be framed in terms of its public health implications or impacts on individual freedom, AI policy can be understood in terms of innovation, international competition, labor transformation, social and ethical risks, or existential harms, with variable implications for immigration, education, healthcare, transportation, and so on. In turn, because the impacts of AI are so significant and the winners and losers unclear, the stakes and motivations to exploit this diverse landscape of possible frames are high.

1.2.3 Environmental Factors

Finally, while a comprehensive review of social, technical, economic, and policy conditions shaping the policy context of AI is not possible, a few additional factors about this broader environment bear mention.

Economic and social conditions like the strength of the economy, inequality, market concentration, and social and political cohesion are likely to influence the kinds of AI policy issues or frames thought important, for example, by highlighting the economy, immigration, racial justice, or the environment as key concerns. Members of the public may have more or less positive attitudes towards technology or private sector firms responsible for most of AI development, as reflected in their opinion about big technology companies, for example. Trust in government and the general appetite for innovation versus regulation play a role here as well. Yet how these dynamics play out in decision-making also depends on the broader culture in policymaking and political circles, for example regarding political and electoral incentives to cooperate, compete, or frame issues in favorable ways.

The current technology policy regime, consisting of policies regarding data, privacy, the Internet, intellectual property, education, immigration, and so on, also critically sets the ground in which AI policymaking must occur. Finally, developments outside of the primary regional or national context also play an essential role. For example, the United States has signed onto the OECD AI principles and discussed cooperation with the EU, NATO, and UN. In this environment, U.S. policymakers may feel pressure to act in ways that minimize influence from the EU (Bradford, 2020) or strategic competitors like China.

This thesis cannot take full account of the large array of actors, influences, and conditions that will shape AI policymaking. Instead, the following chapters are circumscribed within the broader policy context described above and address a subset of relevant issues with a focus on U.S. AI policy dynamics. In particular, my investigations are informed by the high-stakes policy ambiguity that characterizes AI policy. As such, I pay special attention to specific actors—namely policy entrepreneurs, policymakers, the media, and the public—in light of their roles in agenda-setting—and examine key dynamics surrounding policy problems, solutions, issue frames, and influence strategies, amongst other topics.

1.3 Overview of Dissertation Project

1.3.1 Chapter Previews

Chapter 1 above represents the introduction of the topic. Here, I characterize AI, motivate the practical and scholarly importance of studying AI policy agenda-setting, and discuss the broader theoretical and methodological logic of the subsequent chapters. After briefly introducing the historical and modern context of AI and AI policy, including a review of review certain key policy questions, I present a simple conceptual framework to ground the dissertation's investigations into agenda-setting. I contend that AI, as a subject of policymaking, is a site of *high-stakes policy ambiguity* given its status as a strategic, general purpose, and emerging technology. In this sense, AI necessarily intersects with many policy problems, solutions, and actors with diverse and potentially competing interests. It is thus both highly impactful across a wide array of sectors and marked by uncertainty and ambiguity in terms of how policy issues are decided and who is empowered to shape the agenda.

Chapter 2 employs a case study of the U.S. federal AI policy agenda through qualitative and quantitative content analysis of 63 key strategic AI documents curated by the federal government and published between 2016 and 2020. I consider whether this emerging agenda better reflects a 'traditional' approach to innovation policy, focused on strategic economic and geopolitical goals with an expert orientation, or if it represents a shift towards a more 'transformative' paradigm, emphasizing social and ethical implications of technology policy and promoting greater inclusivity in the policy process. To develop these competing expectations in the context of the case study, I draw on a powerful tool in the policy scholar's arsenal for studying agenda-setting: the Multiple Streams Framework (MSF). In particular, the MSF focuses on three semi-independent streams of policy activity: 1) the problem stream, where issues and concerns about AI are defined, 2) the policy (or solution) stream, where policy proposals are formulated, and 3) the political stream, where features of the political environment such as elections and public opinion shape and constrain agenda-setting. The MSF thus directs attention to key conceptual elements of interest in each of the federal AI policy documents, helps to establish theoretical predictions associated with competing paradigms of innovation policy, and offers explanatory grounds for understanding *why* the early U.S. AI policy agenda has developed in its current fashion.

Based on case study and document analysis techniques including qualitative coding and pattern matching, I assess to what extent the emerging U.S. AI policy agenda indeed reflects a shift in how technology is governed, and explain why this shift has—or has not—occurred. To do so, I systematically examine the key U.S. AI policy documents and calculate the prevalence of certain kinds of focusing events, indicators, problem definitions, policy solutions, and issue frames, as well as evaluate the role of the public and experts in the agenda-setting process. To complement analysis of the primary dataset, I draw on additional sources such as reports from non-governmental sources and academic studies that address the AI agenda-setting process, such as documents that discuss indicators or problem definitions surrounding AI policy. This chapter therefore serves to characterize and distill the context and direction of early AI agenda-setting in the United States.

Chapter 3 continues to examine whether the U.S. AI agenda is reflective of an evolving approach to technology governance. In particular, it asks whether federal AI policy discourse from U.S. legislators is responsive to public participation and whether it emphasizes ethical and social dimensions of AI rather than only economic and geopolitical ones. To evaluate these questions, I draw on text analysis and time series analysis techniques using social media and traditional media data from members of the public, media, and policymakers. Contrary to expectations for complex, emerging technologies which are typically expert-dominated domains, I consider whether public attention to AI policy meaningfully predicts policymaker attention. Next, to examine the role of competing approaches to technology policy, I consider which issue frames put forward to define the problems and solutions surrounding AI are prevalent in public and policymaker discourse. I focus on three prominent issue frames, one of which emphasizes the potential of AI to enhance industry and national economic dynamism (the "innovation frame"), a second which emphasizes the social and ethical implications of AI (the "ethics frame"), and a third which emphasizes the international economic and military competitiveness dynamics surrounding AI (the "competition frame"). Further, and in line with the aspirations of a wide array of stakeholders calling for increased public participation in AI governance, I consider whether policymakers are especially likely to listen to the public regarding social and ethical implications of AI that may have a special stake for the public.

The primary data are social media messages on Twitter surrounding AI from members of the public and from members of the 115th and 116th Congresses, covering the time period from 2017 through 2019. I also draw on articles from the New York Times over the same period as an alternative proxy for public opinion or national mood, elements deemed important in the agenda-setting process according to the MSF. I employ text analysis methods to identify relevant messages and extract issue frames used over time, which I structure as time series data. Next, I use autoregressive integrated moving average (ARIMA) analysis to determine whether public (and media) issue attention to AI policy and to specific issue frames influences (or predicts) policymaker attention and use of those respective issue frames. I also consider a vector autoregression (VAR) approach to better handle endogeneity concerns given the likelihood of multiple actors influencing one another simultaneously, and apply fixed effects models to assess if there are heightened impacts on social media-engaged and AI-engaged policymakers. The study thus sheds light on how *sub-issue issue frames* are emerging, competing, and capturing policymaker attention and on whether the public has a meaningful role in this agenda-setting process, with implications for the future of AI governance.

Chapter 4 turns to policy entrepreneurs and the strategies they use to influence the AI agenda. It involves an informational field experiment of U.S. state legislators in the form of an audit or correspondence experiment, conducted in partnership with a leading non-partisan AI policy think tank, The Future Society. The design involves randomly assigning nearly all U.S. state legislators (approximately 7,350) to six variations of an email message about AI policy to assess their engagement with policy entrepreneurs and the relative effectiveness of various influence strategies. In particular, the experimental design contrasts two influence strategies used by policy entrepreneurs against one other and a control condition: the provision of expert information, thought to be important for complex technical domains, and the use of persuasive narratives, argued to be increasingly important in policymaking according to the Narrative Policy Framework. As a secondary focus and in continuation of the previous chapters, the experimental design also invokes issue framing, as state legislators are randomly assigned to receive messages emphasizing either an ethics or economic and technological leadership frame.

Policy entrepreneur influence is measured through policymaker engagement with the emails along a series of increasingly demanding tasks, including clicking on links to researcher-crafted informational resources and viewing information about and attending a webinar co-organized by the partner organization as part of the study. I use covariate-adjusted regression methods and a two-stage least squares instrumental variables approach to evaluate a range of pre-registered hypotheses surrounding the effectiveness of policy entrepreneurs, the dominance of the respective influence strategies or issue frames, how influence strategies' effectiveness may vary based on the issue frames used, and how policymaker characteristics such as prior experience with AI policy may moderate these influence dynamics. In line with open science principles, the study design, including the estimands and models, was pre-registered publicly in advance of analysis. The experimental approach applied through an authentic real-world study of AI policy influence buttresses this study's internal and external validity, and helps to bridge the scholarship on narratives with work on policy entrepreneurship strategy and agenda-setting more generally.

Finally, Chapter 5 concludes by synthesizing lessons learned from the preceding chapters regarding AI policy and agenda-setting for scholars of public policy and other disciplines (such as political science, communication and media studies, AI ethics, and innovation studies). In addition to contextualizing the study's findings, Chapter 5 also reviews theoretical and methodological contributions, limitations, and future research needs.

1.3.2 Logic of Dissertation Approach

This dissertation is theoretically centered on agenda-setting, drawing most heavily on the MSF in the initial case study. This framework has found broad popularity and arguably constitutes the most prominent framework for studying agenda-setting, in part due to its conceptual breadth, flexibility, and accessibility. Yet scholars have also observed corresponding weaknesses. Only a small minority (∼5%) of hundreds of MSF applications across more than 65 countries use quantitative methods, or clearly formulate (∼13%) much less test hypotheses (Jones et al., 2016). A sizable majority of MSF case studies focus on qualitative methods (approximately 88%), including document analysis, process tracing, and interviews (Jones et al., 2016). In contrast, a very small minority (∼5%) emphasize quantitative approaches such as regression (Robinson & Eller, 2010; Travis & Zahariadis, 2002) or agent-based modeling (Ganz, 2020; Rapaport et al., 2009). Many studies also treat the components of the MSF incompletely or superficially, leading to a lack of coordinated theory development (Cairney & Jones, 2016). Still others have observed the MSF may fail to fully acknowledge the ‘black box’ of government and may lack real-world relevance for policymakers (Anderson et al., 2020).

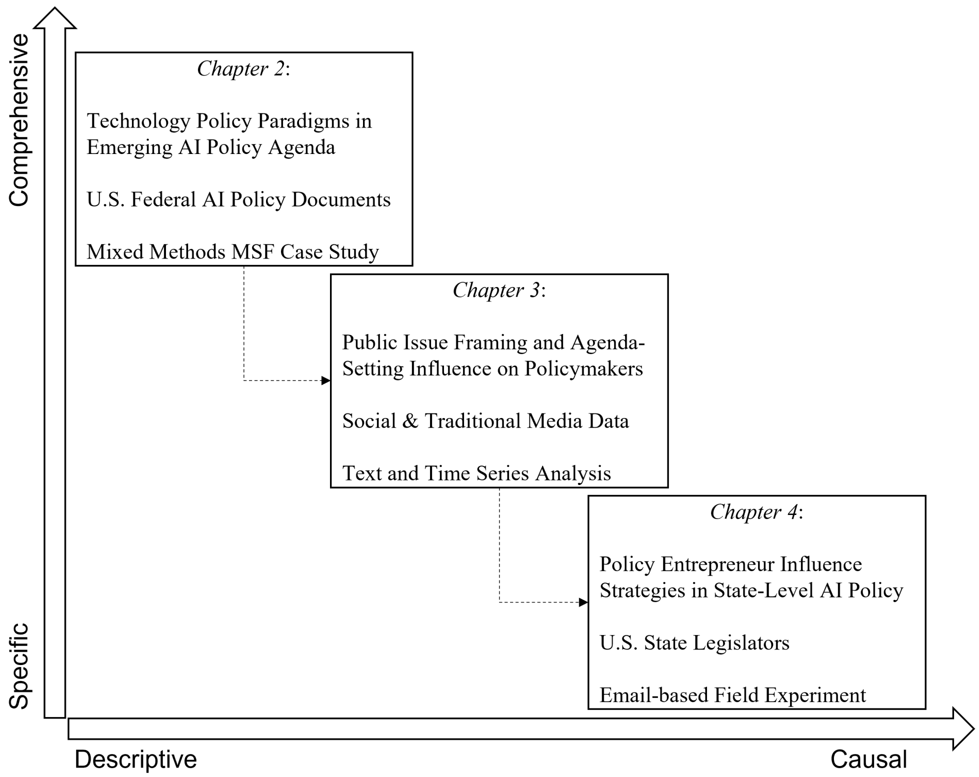

Figure 1.2: Structure of core dissertation chapters

This thesis responds to the limitations above in agenda-setting scholarship by proceeding gradually from a more exploratory and comprehensive focus to a more causal and narrow analysis in subsequent chapters, as pictured in Figure 1.2. That is, the initial MSF case study provides the best opportunity to retain consistency with the theoretical elements of agenda-setting scholarship with sufficient breadth. In turn, the subsequent chapters background this broader context and extract more specific research questions relevant to agenda-setting, allowing for quantitative modeling that can support hypothesis testing while drawing on insights from distinct bodies of theory. In combination, these inquiries help to unpack aspects of the policy process using primary and secondary data, qualitative and quantitative approaches, and theory from several scholarly disciplines, allowing for the integration of diverse forms of knowledge back into agenda-setting theory.

Finally, a novelty of this dissertation is its subject matter focus on AI policy. Policy scholars have observed that some areas of policymaking are becoming increasingly complex (Petridou & Mintrom, 2021) or 'wicked' while relatively little is understood about how policy processes operate in these technically complex policy domains. Currently, a large majority of agenda-setting studies focus on issues such as health and environment, while as few as 2% of studies explicitly address technology (Jones et al., 2016). Further, while AI policy research benefits in some ways from its interdisciplinary nature, it has also been correspondingly piecemeal and anecdotal, drawing on many disciplinary perspectives but too often lacking in theoretical definition and empirical rigor. In order to build knowledge and advance policymaking, there is a clear need to better understand how the dynamics of agenda-setting and policy process theory more broadly apply to or differ in the case of technologies such as AI. As such, this thesis draws on the conceptualization of AI as a strategic, emerging, and general purpose technology, and considers how important associated dimensions like AI's social and ethical implications and public participation are shaping AI governance. The findings from this research can therefore contribute to knowledge in a way that supports real-world relevance for policymakers and numerous other stakeholders interested in AI, while also advancing policy process theory and other scholarship related to AI and technology policy more generally.

1.4 References for Chapter 1

The Advancing American AI Act, S. 1353, 117th Cong. (2021). https://www.congress.gov/bill/117th-congress/senate-bill/1353

Agre, P. E. (1997). Toward a critical technical practice: Lessons learned in trying to reform AI. In G. Bowker, S. L. Star, L. Gasser, & W. Turner (Eds.), *Social science, technical systems, and cooperative work: Beyond the great divide*. Lawrence Erlbaum Associates, Inc.

Algorithmic Accountability Act of 2019, S. 1108, 116th Cong. (2019). https://www.congress.gov/bill/116th-congress/senate-bill/1108

Allen, G., & Chan, T. (2017). *Artificial intelligence and national security*. Belfer Center for Science and International Affairs. Boston, MA. https://www.belfercenter.org/sites/default/files/files/publication/AI%20NatSec%20-%20final.pdf

Alpaydin, E. (2016). *Machine learning: The new AI*. MIT Press.

Anderson, S. E., DeLeo, R. A., & Taylor, K. (2020). Policy entrepreneurs, legislators, and agenda setting: Information and influence. *Policy Studies Journal*, *48*(3), 587–611. https://doi.org/10.1111/psj.12331

Armstrong, S., Sotala, K., & hÉigeartaigh, S. S. Ó. (2014). The errors, insights and lessons of famous AI predictions – and what they mean for the future. *Journal of Experimental & Theoretical Artificial Intelligence*, *26*(3), 317–342. https://doi.org/10.1080/0952813X.2014.895105

Bagloee, S. A., Tavana, M., Asadi, M., & Oliver, T. (2016). Autonomous vehicles: Challenges, opportunities, and future implications for transportation policies. *Journal of Modern Transportation*, *24*(4), 284–303. https://doi.org/10.1007/s40534-016-0117-3

Baker, J. E. (2021). *Ethics and artificial intelligence: A policymaker's introduction*. Center for Security and Emerging Technology. Washington, D.C. https://doi.org/10.51593/20190022

Barczentewicz, M., & Mueller, B. (2021). *More than meets the AI: The hidden costs of a European software law*. Center for Data Innovation. Washington, D.C. https://www2.datainnovation.org/2021-more-than-meets-the-ai.pdf

Baum, S. (2017). *A survey of artificial general intelligence projects for ethics, risk, and policy* (SSRN Scholarly Paper No. ID 3070741). Social Science Research Network. Rochester, NY. https://doi.org/10.2139/ssrn.3070741

Baumgartner, F. R., & Jones, B. D. (1993). *Agendas and instability in American politics* (First edition). University of Chicago Press.

Baumgartner, F. R., & Mahoney, C. (2008). Forum section: The two faces of framing: Individual-level framing and collective issue definition in the European Union. *European Union Politics*, *9*(3), 435–449. https://doi.org/10.1177/1465116508093492

Berg, A., Buffie, E. F., & Zanna, L.-F. (2018). Should we fear the robot revolution? (The correct answer is yes). *Journal of Monetary Economics*, *97*, 117–148. https://doi.org/10.1016/j.jmoneco.2018.05.014

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies* (First edition). Oxford University Press.

Bradford, A. (2020). *The Brussels effect: How the European Union rules the world*. Oxford University Press.

Bredt, S. (2019). Artificial intelligence (AI) in the financial sector—potential and public strategies. *Frontiers in Artificial Intelligence*, *2*. https://doi.org/10.3389/frai.2019.00016

Bresnahan, T. F., & Trajtenberg, M. (1995). General purpose technologies 'engines of growth'? *Journal of Econometrics*, *65*(1), 83–108. https://doi.org/10.1016/0304-4076(94)01598-T

Brooks, R. A. (1999). *Cambrian intelligence: The early history of the new AI*. MIT Press.

Bughin, J., Seong, J., Manyika, J., Chui, M., & Joshi, R. (2018). *Modeling the global economic impact of AI*. McKinsey Global Institute. https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-frontier-modeling-the-impact-of-ai-on-the-world-economy

Cairney, P., & Jones, M. D. (2016). Kingdon's Multiple Streams Approach: What is the empirical impact of this universal theory? *Policy Studies Journal*, *44*(1), 37–58. https://doi.org/10.1111/psj.12111

Carmines, E. G., & Stimson, J. A. (1980). The two faces of issue voting. *The American Political Science Review*, *74*(1), 78–91. https://doi.org/10.2307/1955648

Cave, S., Coughlan, K., & Dihal, K. (2019). "Scary robots": Examining public responses to AI. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 331–337. https://doi.org/10.1145/3306618.3314232

Cave, S., Craig, C., Dihal, K., Dillon, S., Montgomery, J., Singler, B., & Taylor, L. (2018). *Portrayals and perceptions of AI and why they matter*. The Royal Society. https://doi.org/10.17863/CAM.34502

Chae, Y. (2020). U.S. AI regulation guide: Legislative overview and practical considerations. *The Journal of Robotics, Artificial Intelligence & Law*, *3*(1), 17–40.

Cihon, P., Maas, M. M., & Kemp, L. (2020). Fragmentation and the future: Investigating architectures for international AI governance. *Global Policy*, *11*(5), 545–556. https://doi.org/10.1111/1758-5899.12890

Cockburn, I. M., Henderson, R., & Stern, S. (2018). *The impact of artificial intelligence on innovation* (Working Paper No. 24449). National Bureau of Economic Research. https://doi.org/10.3386/w24449

Coeckelbergh, M. (2020). *AI ethics*. MIT Press.

Cohen, M. D., March, J. G., & Olsen, J. P. (1972). A garbage can model of organizational choice. *Administrative Science Quarterly*, *17*(1), 1–25. https://doi.org/10.2307/2392088

Day, G. S., & Schoemaker, P. J. H. (2000). Avoiding the pitfalls of emerging technologies. *California Management Review*, *42*(2), 8–33. https://doi.org/10.2307/41166030

DeCario, N., & Etzioni, O. (2021). America needs AI literacy now. https://pnw.ai/article/america-needs-ai-literacy-now/72515409

Defense Science Board. (2016). *Defense Science Board summer study on autonomy*. Defense Science Board, Department of Defense. Washington, D.C.

Department of Public Administration, Italy. (2018). *Artificial intelligence at the service of citizens*. Task Force on Artificial Intelligence of the Agency for Digital Italy and Department for Public Administration, Italy. Rome, Italy. https://ia.italia.it/assets/whitepaper.pdf

Diercks, G., Larsen, H., & Steward, F. (2019). Transformative innovation policy: Addressing variety in an emerging policy paradigm. *Research Policy*, *48*(4), 880–894. https://doi.org/10.1016/j.respol.2018.10.028

Dreyfus, H. L. (1992). *What computers still can't do: A critique of artificial reason*. MIT Press.

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of big data – evolution, challenges and research agenda. *International Journal of Information Management*, *48*, 63–71. https://doi.org/10.1016/j.ijinfomgt.2019.01.021

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., … Williams, M. D. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, *57*, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Edler, J., & Fagerberg, J. (2017). Innovation policy: What, why, and how. *Oxford Review of Economic Policy*, *33*(1), 2–23. https://doi.org/10.1093/oxrep/grx001

Elder, C. D., & Cobb, R. W. (1984). Agenda-building and the politics of aging. *Policy Studies Journal*, *13*(1), 115–127. https://doi.org/10.1111/j.1541-0072.1984.tb01704.x

Elish, M. C., & boyd, d. (2018). Situating methods in the magic of big data and AI. *Communication Monographs*, *85*(1), 57–80. https://doi.org/10.1080/03637751.2017.1375130

Elsevier. (2018). *Artificial intelligence: How knowledge is created, transferred, and used: Trends in China, Europe, and the United States*. Elsevier. https://www.elsevier.com/research-intelligence/resource-library/ai-report

European Commission. (2021). *Proposal for a regulation on a European approach for artificial intelligence* (COM(2021) 206 final). European Commission. Brussels, Belgium. https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence

European Digital Rights (EDRi). (2021). Open letter: Civil society call for the introduction of red lines in the upcoming European Commission proposal on artificial intelligence. https://edri.org/our-work/civil-society-call-for-ai-red-lines-in-the-european-unions-artificial-intelligence-proposal/

European Parliament. (2020). *Framework of ethical aspects of artificial intelligence, robotics and related technologies* (2020/2012(INL)). European Parliament. Brussels, Belgium. https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?lang=en&reference=2020/2012(INL)

Executive Office of the President, National Science and Technology Council (NSTC), Committee on Technology, United States. (2016). *Preparing for the future of ar-*

tificial intelligence. Executive Office of the President, National Science and Technology Council (NSTC), Committee on Technology, United States. Washington, D.C. https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

Executive Office of the President, United States. (2016). *Artificial intelligence, automation, and the economy*. Executive Office of the President, United States. Washington, D.C. https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF

Fadlullah, Z. M., Tang, F., Mao, B., Kato, N., Akashi, O., Inoue, T., & Mizutani, K. (2017). State-of-the-art deep learning: Evolving machine intelligence toward tomorrow's intelligent network traffic control systems. *IEEE Communications Surveys Tutorials*, *19*(4), 2432–2455. https://doi.org/10.1109/COMST.2017.2707140

Federal Trade Commission. (2021). Aiming for truth, fairness, and equity in your company's use of AI. https://www.ftc.gov/news-events/blogs/business-blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai

Fischer, S.-C., Leung, J., Anderljung, M., O'Keefe, C., Torges, S., Khan, S. M., Garfinkel, B., & Allan Dafoe. (2021). *AI policy levers: A review of the U.S. government's tools to shape AI research, development, and deployment*. Centre for the Governance of AI, Future of Humanity Institute, University of Oxford. Oxford, England. https://www.fhi.ox.ac.uk/wp-content/uploads/2021/03/AI-Policy-Levers-A-Review-of-the-U.S.-Governments-tools-to-shape-AI-research-development-and-deployment-%E2%80%93-Fischer-et-al.pdf

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI* (SSRN Scholarly Paper No. ID 3518482). Berkman Klein Center for Internet & Society. Rochester, NY. https://papers.ssrn.com/abstract=3518482

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, *114*, 254–280. https://doi.org/10.1016/j.techfore.2016.08.019

Ganz, S. (2020). Conflict, chaos, and the art of institutional design. *SocArXiv*. https://doi.org/10.31235/osf.io/2fzdt

Gilardi, F., Shipan, C. R., & Wüest, B. (2021). Policy diffusion: The issue-definition stage. *American Journal of Political Science*, *65*(1), 21–35. https://doi.org/10.1111/ajps.12521

Girasa, R. (2020). *Artificial intelligence as a disruptive technology: Economic transformation and government regulation*. Palgrave Macmillan.

Goddard, S. E., & Krebs, R. R. (2015). Rhetoric, legitimation, and grand strategy. *Security Studies*, *24*(1), 5–36. https://doi.org/10.1080/09636412.2014.1001198

Golbin, I., Rao, A. S., Hadjarian, A., & Krittman, D. (2020). Responsible AI: A primer for the legal community. *2020 IEEE International Conference on Big Data (Big Data)*, 2121–2126. https://doi.org/10.1109/BigData50022.2020.9377738

Gonsalves, T. (2019). The summers and winters of artificial intelligence. In M. Khosrow-Pour (Ed.), *Advanced methodologies and technologies in artificial intelligence, computer simulation, and human-computer interaction* (pp. 168–179). IGI Global. https://doi.org/10.4018/978-1-5225-7368-5.ch014

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

Griliches, Z. (1957). Hybrid corn: An exploration in the economics of technological change. *Econometrica*, *25*(4), 501–522. https://doi.org/10.2307/1905380

Gruetzemacher, R., & Whittlestone, J. (2022). The transformative potential of artificial intelligence. *Futures*, *135*, 102884. https://doi.org/10.1016/j.futures.2021.102884

Harris, E. D. (2016). *Governance of dual-use technologies: Theory and practice*. American Academy of Arts and Sciences.

Hedlund, M. (2022). Distribution of forward-looking responsibility in the EU process on AI regulation. *Frontiers in Human Dynamics*, *4*. https://www.frontiersin.org/article/10.3389/fhumd.2022.703510

Imbrie, A., Gelles, R., Dunham, J., & Aiken, C. (2021). *Contending frames: Evaluating rhetorical dynamics in AI*. Center for Security and Emerging Technology. Washington, D.C. https://doi.org/10.51593/20210010

Ipsos MORI. (2017). *Public views of machine learning: Findings from public research and engagement conducted on behalf of the Royal Society*. The Royal Society. London, UK. https://royalsociety.org/~/media/policy/projects/machine-learning/publications/public-views-of-machine-learning-ipsos-mori.pdf

Jha, K., Doshi, A., Patel, P., & Shah, M. (2019). A comprehensive review on automation in agriculture using artificial intelligence. *Artificial Intelligence in Agriculture*, *2*, 1–12. https://doi.org/10.1016/j.aiia.2019.05.004

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Johnson, C., & Tyson, A. (2020). Are AI and job automation good for society? Globally, views are mixed. https://www.pewresearch.org/fact-tank/2020/12/15/people-globally-offer-mixed-views-of-the-impact-of-artificial-intelligence-job-automation-on-society/

Jones, M. D., Peterson, H. L., Pierce, J. J., Herweg, N., Bernal, A., Raney, H. L., & Zahariadis, N. (2016). A river runs through it: A Multiple Streams meta-review. *Policy Studies Journal*, *44*(1), 13–36. https://doi.org/10.1111/psj.12115

Kania, E. (2018). *Technological entanglement: Cooperation, competition and the dual-use dilemma in artificial intelligence* (Policy Brief Report No.7/2018). International Cyber Policy Centre. Barton, Australia. https://apo.org.au/node/180151

Knight Foundation and Gallup. (2022). *Media and democracy: Unpacking America's complex views on the digital public square*. Knight Foundation and Gallup. https://knightfoundation.org/wp-content/uploads/2022/03/KMAD-2022-1.pdf

Krafft, P. M., Young, M., Katell, M., Huang, K., & Bugingo, G. (2020). Defining AI in policy versus practice. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 72–78. https://doi.org/10.1145/3375627.3375835

Kratsios, M. (2019). *How the United States can maintain its lead in the global AI race*. Information Technology and Innovation Foundation. Washington, D.C. https://itif.org/events/2019/09/10/how-united-states-can-maintain-its-lead-global-ai-race

Lardinois, F. (2017). In the AI wars, Microsoft now has the clearer vision. *TechCrunch*. https://social.techcrunch.com/2017/05/21/in-the-ai-wars-microsoft-now-has-the-clearer-vision/

Lee, K.-F. (2018). *AI superpowers: China, silicon valley, and the new world order*. Houghton Mifflin Harcourt.

Leung, J. (2020). *Who will govern artificial intelligence?* (Doctoral dissertation). Oxford University. Oxford, England. https://ora.ox.ac.uk/objects/uuid:ea3c7cb8-2464-45f1-a47c-c7b568f27665

Lighthill, J. (1973). *Artificial intelligence: A general survey*. British Scientific Research Council. Cambridge, United Kingdom. http://www.chilton-computing.org.uk/inf/literature/reports/lighthill_report/p001.htm

Makridakis, S. (2017). The forthcoming artificial intelligence (AI) revolution: Its impact on society and firms. *Futures*, *90*, 46–60. https://doi.org/10.1016/j.futures.2017.03.006

Marchant, G. (2019). *"Soft law" governance of artificial intelligence*. UCLA: The Program on Understanding Law, Science, and Evidence (PULSE). https://escholarship.org/uc/item/0jq252ks

Martinez, R. (2019). Artificial intelligence: Distinguishing between types & definitions. *Nevada Law Journal*, *19*(1015), 28.

Mattingly-Jordan, S., Donaldson, B., Gray, P., & Ingram, L. M. (2020). *Ethically aligned design: First edition glossary*. IEEE.

McCorduck, P. (2004). *Machines who think: A personal inquiry into the history and prospects of artificial intelligence* (Second edition). CRC Press.

Menashe, C. L. (1998). The power of a frame: An analysis of newspaper coverage of tobacco issues-United States, 1985-1996. *Journal of Health Communication*, *3*(4), 307–325. https://doi.org/10.1080/108107398127139

Merolla, J., Ramakrishnan, S. K., & Haynes, C. (2013). "Illegal," "Undocumented," or "Unauthorized": Equivalency frames, issue frames, and public opinion on immigration. *Perspectives on Politics*, *11*(3), 789–807. https://doi.org/10.1017/S1537592713002077

Metz, C. (2021). *Genius makers: The mavericks who brought AI to Google, Facebook, and the world*. Dutton.

Microsoft. (2018). *The future computed: Artificial intelligence and its role in society*. Microsoft. https://blogs.microsoft.com/blog/2018/01/17/future-computed-artificial-intelligence-role-society

Mökander, J., Morley, J., Taddeo, M., & Floridi, L. (2021). Ethics-based auditing of automated decision-making systems: Nature, scope, and limitations. *Science and Engineering Ethics*, *27*(4), 44. https://doi.org/10.1007/s11948-021-00319-4

Moor, J. (2006). The Dartmouth College Artificial Intelligence Conference: The next fifty years. *AI Magazine*, *27*(4), 87–87. https://doi.org/10.1609/aimag.v27i4.1911

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2019). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics*, *26*(4), 2141–2168. https://doi.org/10.1007/s11948-019-00165-5

Müller, V. C., & Bostrom, N. (2016). Future progress in artificial intelligence: A survey of expert opinion. *Fundamental issues of artificial intelligence* (pp. 555–572). Springer, Cham. https://doi.org/10.1007/978-3-319-26485-1_33

Murdick, D., Dunham, J., & Melot, J. (2020). *AI definitions affect policymaking*. Center for Security and Emerging Technology. Washington, D.C. https://doi.org/10.51593/20200004

National AI Initiative Office, Office of Science and Technology Policy. (2022). National Artificial Intelligence Initiative documents. https://www.ai.gov/documents/

National Institute of Standards and Technology. (2021). *Draft - taxonomy of AI risk*. National Institute of Standards and Technology. Washington, D.C. https://www.nist.gov/system/files/documents/2021/10/15/AI%20RMF_RFI%20Summary%20Report.pdf

National Science and Technology Council (NSTC), Networking and Information Technology Research and Development Subcommittee, United States. (2016). *The national artificial intelligence research and development strategic plan*. National Science and Technology Council (NSTC), Networking and Information Technology Research and Development Subcommittee, United States. Washington, D.C. https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/national_ai_rd_strategic_plan.pdf

National Security Commission on Artificial Intelligence. (2021). *Final report*. National Security Commission on Artificial intelligence. Washington, D.C. https://www.nscai.gov/wp-content/uploads/2021/01/NSCAI-Draft-Final-Report-1.19.21.pdf

Naudé, W., & Dimitri, N. (2020). The race for an artificial general intelligence: Implications for public policy. *AI & Society*, *35*(2), 367–379. https://doi.org/10.1007/s00146-019-00887-x

Naughton, J. (2016). The evolution of the internet: From military experiment to general purpose technology. *Journal of Cyber Policy*, *1*(1), 5–28. https://doi.org/10.1080/23738871.2016.1157619

Nazareno, L., & Schiff, D. S. (2021). The impact of automation and artificial intelligence on worker well-being. *Technology in Society*, *67*, 101679. https://doi.org/10.1016/j.techsoc.2021.101679

OECD. (2022a). OECD AI's live repository of over 260 AI strategies & policies. https://www.oecd.ai/dashboards

OECD. (2022b). *OECD framework for the classification of AI systems: A tool for effective AI policies* (No. 323). OECD. https://oecd.ai/en/p/classification

OECD. (2019). *Recommendation of the Council on Artificial Intelligence* (OECD/LEGAL/0449). OECD. Paris, France. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

Office of Management and Budget. (2020). *Memorandum for the heads of executive departments and agencies: Guidance for regulation of artificial intelligence applications*. Office of Management and Budget. Washington, D.C. https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf

Office of Science and Technology Policy. (2019). *The American AI Initiative*. Office of Science and Technology Policy. Washington, D.C. https://trumpwhitehouse.archives.gov/articles/accelerating-americas-leadership-in-artificial-intelligence/

Oravec, J. A. (2019). Artificial intelligence, automation, and social welfare: Some ethical and historical perspectives on technological overstatement and hyperbole. *Ethics and Social Welfare*, *13*(1), 18–32. https://doi.org/10.1080/17496535.2018.1512142

O'Shaughnessy, M., Schiff, D. S., Varshney, L. R., Rozell, C. J., & Davenport, M. A. (2022). What governs attitudes toward artificial intelligence adoption and governance? *Working paper*. https://doi.org/10.31219/osf.io/pkeb8

Parker, C. (2018). Artificial intelligence could be our saviour, according to the CEO of Google. https://www.weforum.org/agenda/2018/01/google-ceo-ai-will-be-bigger-than-electricity-or-fire/

Perset, K., Murdick, D., Clark, J., & Grobelnik, M. (2020). A first look at the OECD's framework for the classification of AI systems, designed to give policymakers clarity. https://www.oecd.ai/wonk/a-first-look-at-the-oecds-framework-for-the-classification-of-ai-systems-for-policymakers

Peters, B. G. (2021). *Advanced introduction to public policy*. Edward Elgar Publishing.

Petridou, E., & Mintrom, M. (2021). A research agenda for the study of policy entrepreneurs. *Policy Studies Journal*, *49*(4), 943–967. https://doi.org/10.1111/psj.12405

Rainie, L., Funk, C., Anderson, M., & Tyson, A. (2022). *AI and human enhancement: Americans' openness is tempered by a range of concerns*. Pew Research Center. Washington, D.C. https://www.pewresearch.org/internet/2022/03/17/how-americans-think-about-artificial-intelligence/

Rao, A. S., & Verweij, G. (2017). *Sizing the prize: What's the real value of AI for your business and how can you capitalise?* PricewaterhouseCoopers. https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf

Rapaport, O., Levi-Faur, D., & Miodownik, D. (2009). The puzzle of the diffusion of central-bank independence reforms: Insights from an agent-based simulation. *Policy Studies Journal*, *37*(4), 695–716. https://doi.org/10.1111/j.1541-0072.2009.00331.x

Rein, M., & Schön, D. (1996). Frame-critical policy analysis and frame-reflective policy practice. *Knowledge and Policy*, *9*(1), 85–104. https://doi.org/10.1007/BF02832235

Rességuier, A., & Rodrigues, R. (2020). AI ethics should not remain toothless! A call to bring back the teeth of ethics. *Big Data & Society*, *7*(2), 2053951720942541. https://doi.org/10.1177/2053951720942541

Ristuccia, C. A., & Solomou, S. (2014). Can general purpose technology theory explain economic growth? Electrical power as a case study. *European Review of Economic History*, *18*(3), 227–247. https://doi.org/10.1093/ereh/heu008

Robinson, S. E., & Eller, W. S. (2010). Participation in policy streams: Testing the separation of problems and solutions in subnational policy systems. *Policy Studies Journal*, *38*(2), 199–216. https://doi.org/10.1111/j.1541-0072.2010.00358.x

Rotolo, D., Hicks, D., & Martin, B. R. (2015). What is an emerging technology? *Research Policy*, *44*(10), 1827–1843. https://doi.org/10.1016/j.respol.2015.06.006

Russell, S., & Norvig, P. (1995). *Artificial intelligence: A modern approach* (First edition). Prentice Hall.

Samoili, S., López Cobo, M., Gómez, E., De Prato, G., Martínez-Plumed, F., Delipetrev, B., European Commission, & Joint Research Centre. (2020). *AI watch: Defining artificial intelligence: Towards an operational definition and taxonomy of artificial intelligence.* Publications Office of the European Union. Luxembourg. https://doi.org/10.2760/382730

Schiff, D. (2021). Out of the laboratory and into the classroom: The future of artificial intelligence in education. *AI & Society*, *36*(1), 331–348. https://doi.org/10.1007/s00146-020-01033-8

Schiff, D., Biddle, J., Borenstein, J., & Laas, K. (2020). What's next for AI ethics, policy, and governance? A global overview. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 153–158. https://doi.org/10.1145/3375627.3375804

Schiff, D., Borenstein, J., Laas, K., & Biddle, J. (2021). AI ethics in the public, private, and NGO sectors: A review of a global document collection. *IEEE Transactions on Technology and Society*, *2*(1), 31–42. https://doi.org/10.1109/TTS.2021.3052127

Schmitt, L. (2021). Mapping global AI governance: A nascent regime in a fragmented landscape. *AI and Ethics*. https://doi.org/10.1007/s43681-021-00083-y

Schwab, K. (2016). *The fourth industrial revolution* (First U.S. edition). Crown Business.

Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., & Hall, P. (2022). *Towards a standard for identifying and managing bias in artificial intelligence*. National Institute of Standards and Technology. Washington, D.C. https://doi.org/10.6028/NIST.SP.1270

Smith, A., & Anderson, M. (2017). *Americans and automation in everyday life*. Pew Research Center. http://www.pewinternet.org/2017/10/04/automation-in-everyday-life/

Spiegeleire, S. D., Maas, M., & Sweijs, T. (2017). *Artificial intelligence and the future of defense: Strategic implications for small- and medium-sized force providers*. The Hague Centre for Strategic Studies.

Stahl, B. C., Antoniou, J., Ryan, M., Macnish, K., & Jiya, T. (2021). Organisational responses to the ethical issues of artificial intelligence. *AI & Society*. https://doi.org/10.1007/s00146-021-01148-6

Sweeney, L. (2003). *That's AI?: A history and critique of the field* (CMU-CS-03-106). School of Computer Science, Carnegie Melon University. Pittsburgh, PA.

The Future Society, Global Partnership on AI, & International Centre of Expertise in Montreal for the Advancement of Artificial Intelligence. (2020). *Areas for future action in the responsible AI ecosystem*. The Future Society. Somerville, MA. https://thefuturesociety.org/wp-content/uploads/2020/12/Areas-for-Future-Action-in-the-Responsible-AI-Ecosystem.pdf

Travis, R., & Zahariadis, N. (2002). A multiple streams model of U.S. foreign aid policy. *Policy Studies Journal*, *30*(4), 495–514. https://doi.org/10.1111/j.1541-0072.2002.tb02160.x

Turner-Lee, N., Resnick, P., & Barton, G. (2019). *Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms*. Brookings Instittition. https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/

Ulnicane, I., Eke, D. O., Knight, W., Ogoh, G., & Stahl, B. C. (2021). Good governance as a response to discontents? Déjà vu, or lessons for AI from other emerging technologies. *Interdisciplinary Science Reviews*, *46*(1-2), 71–93. https://doi.org/10.1080/03080188.2020.1840220

Ulnicane, I., Knight, W., Leach, T., Stahl, B. C., & Wanjiku, W.-G. (2020). Framing governance for a contested emerging technology: Insights from AI policy. *Policy and Society*, *0*(0), 1–20. https://doi.org/10.1080/14494035.2020.1855800

UNESCO. (2021). *Artificial intelligence needs assessment survey in Africa*. UNESCO. Paris, France. https://unesdoc.unesco.org/ark:/48223/pf0000375322

United States Patent and Trademark Office. (2020). *Public views on artificial intelligence and intellectual property*. United States Patent and Trademark Office. Washington, D.C.

U.S. Chamber of Commerce. (2022). *AI and public opinion: A view from the American public*. U.S. Chamber of Commerce. Washington, D.C. https://americaninnovators.com/wp-content/uploads/2022/01/CTEC-US-Outlook-on-AI-Detailed-Analysis.pdf

Walker, K. (2020). Responding to the European Commission's AI white paper. https://blog.google/around-the-globe/google-europe/responding-to-the-european-commissions-ai-white-paper/

Wang, W., & Siau, K. (2019). Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: A review and research agenda. *Journal of Database Management*, *30*, 61–79. https://doi.org/10.4018/JDM.2019010104

Washington Post Editorial Board. (2021). The United States can't let other countries write AI policy for it. *Washington Post*. https://www.washingtonpost.com/opinions/global-opinions/the-united-states-cant-let-other-countries-write-ai-policy-for-it/2021/04/29/30c20142-a857-11eb-8d25-7b30e74923ea_story.html

White House. (2019). Executive Order on Maintaining American Leadership in Artificial Intelligence. https://www.federalregister.gov/documents/2019/02/14/2019-02544/maintaining-american-leadership-in-artificial-intelligence

William M. (Mac) Thornberry National Defense Authorization Act for Fiscal Year 2021, Pub. L. No. 116-283. (2021). https://www.congress.gov/116/plaws/publ283/PLAW-116publ283.pdf

Wooldridge, M. (2021). *A brief history of artificial intelligence: What it is, where we are, and where we are going* (B. Clark, Ed.). Flatiron Books.

Yu, K.-H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. *Nature Biomedical Engineering*, *2*(10), 719–731. https://doi.org/10.1038/s41551-018-0305-z

Završnik, A. (2019). Algorithmic justice: Algorithms and big data in criminal justice settings. *European Journal of Criminology*, 1477370819876762. https://doi.org/10.1177/1477370819876762

Zhang, B., & Dafoe, A. (2019). *Artificial intelligence: American attitudes and trends* (SSRN Scholarly Paper No. ID 3312874). Social Science Research Network. Rochester, NY. https://papers.ssrn.com/abstract=3312874

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

Zhang, D., Saurabh, M., Brynjolfsson, E., Etchemendy, J., Ganguli, D., Grosz, B., Lyons, T., Manyika, J., Niebles, J. C., Sellitto, M., Shoham, Y., Clark, J., & Perrault, R. (2021). *Artificial Intelligence Index report 2021*. AI Index Steering Committee, Human-Centered AI Institute, Stanford University. Stanford, CA. https://aiindex.stanford.edu/wp-content/uploads/2021/03/2021-AI-Index-Report_Master.pdf

Zito, A. R. (2001). Epistemic communities, collective entrepreneurship and european integration. *Journal of European Public Policy*, *8*(4), 585–603. https://doi.org/10.1080/13501760110064401

CHAPTER 2

LOOKING THROUGH A POLICY WINDOW WITH TINTED GLASSES: AGENDA-SETTING DYNAMICS IN U.S. AI POLICY

2.1 Introduction

Artificial intelligence (AI) policy is at a nascent stage. While dozens of countries have begun to put forward preliminary policy strategies, relatively few have adopted formal regulation (CIFAR, 2020; OECD, 2021). Indeed, the 2020s are expected to be a period of significant policymaking activity (Perry & Uuk, 2019; Zhang et al., 2022). As such, AI policy is arguably best understood to be undergoing agenda-setting, the often contentious process wherein policy actors compete to shape the key policy problems and solutions that will set the stage for future policymaking. Yet while policy process theory has been fruitfully applied to understand many policy domains, prominent agenda-setting frameworks like the Multiple Streams Approach or Framework (MSF) have rarely been applied to technology policy (Jones et al., 2016) and little is known about agenda-setting for AI policy in particular (Taeihagh, 2021), despite its sweeping social and economic implications.[1]

This chapter seeks to extend scholarly knowledge on the dynamics behind agenda-setting in AI policy, drawing on the MSF along with competing paradigms in innovation policy with disparate implications for technology governance. In particular, while traditional innovation policy often emphasizes strategic economic and geopolitical dimensions of technology underpinned by a strong expert orientation, newer paradigms emphasize societal objectives of innovation and call for broader public participation. How the AI policy agenda unfolds in this context helps to answer important questions such as the prospects for

[1] Most typically, research in science and technology using the MSF focuses on the environmental domain (Huber-Stearns et al., 2019). However, also see Goyal et al. (2021) for a recent broader review of the role of technology in the MSF and Justo-Hanani (2022) for a policy process oriented study of AI policy.

incorporating AI ethics into policymaking, which actors have power in the agenda-setting process, and how global AI governance is likely to take form.

To provide insight on these questions, this study performs qualitative and quantitative content analysis of 63 key strategic AI documents identified by the United States (U.S.) federal government and published between 2016 and 2020. Through extensive coding and analysis, it examines the policy problems, solutions, broader issue frames, and key focusing events and indicators reflected in U.S. AI policy discourse, as well as the role of experts and the general public in this process. Drawing on these data along with supporting documents and academic literature thus helps to unpack aspects of the policy process theorized as important in shaping the AI policy agenda. Based on this analysis and in light of competing visions and paradigms of technology governance, this study seeks to answer the following questions:

- Descriptively, it considers: *Does the U.S. federal AI policy agenda better reflect a traditional approach to innovation policy, emphasizing economic and geopolitical goals, or a 'transformative' paradigm, emphasizing social and ethical objectives and a public rather than expert orientation?*
- Explanatorily, it asks: *Why has U.S. AI policy taken this trajectory? What aspects of the policy process, discourse, and context surrounding AI have led to this result?*

Results indicate, on one hand, that advocates of AI ethics have had remarkable success in promoting attention to social and ethical dimensions of AI in policy discourse generally. Indeed, the degree of attention to ethics in technology governance may be unprecedented. Yet, such a transformation is partial at best and substantially tempered in practice. For instance, despite broad-based calls for attention to ethical implications at the mission statement level, the overwhelming majority of focusing events and indicators discussed in U.S. AI policy documents are traditional in nature, emphasizing economic benefits of AI and associated geopolitical concerns. Further, the majority of policy problems and solutions featured in these documents are also traditional in nature, including calls for increased research and development to realize AI's benefits, responses to military and security implications, expansion of access to datasets to support AI development, and improvement of the science, technology, engineering, and math (STEM) workforce through increased education. While some social and ethical problems do receive significant attention, such as those surrounding privacy, transparency, and trust, even these often take on a 'hybrid' mode where associated policy solutions are interpreted and justified in light of realizing traditional innovation goals, such as promoting trustworthy AI to increase consumer adoption. Meanwhile, other ethical concerns such as fairness, inequality, and human rights receive somewhat less attention. Finally, despite almost ubiquitous calls for public and diverse participation, the large majority of concrete recommendations specify industry or government experts, undermining the prospects for true public governance of AI.

The study offers potential explanations for these findings, drawing on the MSF and innovation policy literature. First, some ethical problems are thought to be addressable through technical fixes, and thus more concrete proposals surrounding topics like privacy benefit from the ostensible 'technically feasibility' of these solutions. Meanwhile, broader calls for societal transformation that might challenge innovation-oriented goals, such as addressing inequality, may run afoul of 'value acceptability' in the U.S. context. Third, high-level rhetorical interest in social and ethical implications of AI, especially in documents with a government-wide scope, may fail to translate into action as government agencies in specific sectors interpret these questions through traditional norms and policy instruments.

In light of the importance of the United States in global technology governance as well as for AI research and development, these findings are important to international stakeholders interested in AI policy, as well as scholars of policy process theory and innovation policy interested in the extent to which new technology policy domains are characterized by more participatory, societal objectives and processes. Ultimately, despite some striking success for stakeholders invested in AI's social and ethical implications, the translation of ethics into policy appears currently limited. Instead of reflecting a wholesale evolution from traditional innovation policy to more transformative notions of technology governance, the

U.S. AI policy agenda instead reflects a layering or subsumption of the former into the latter. Nevertheless, policy entrepreneurs who are better able to translate their social and ethical policy concerns into concrete solutions may find a willing audience. While the policy agenda is taking shape, the policy window is not yet closed.

2.2 Theoretical Background

2.2.1 Agenda-Setting and the Multiple Streams Framework

The MSF is the first of two theoretical lenses used to study the policy context and agenda-setting process surrounding AI. Originally drawing on the concept of "organized anarchies," (Cohen et al., 1972), but applied to institutional rather than organizational settings, the MSF's basic assumptions involve ambiguity, unclear technology, problematic preferences, time constraints, and fluid participation (Cairney & Jones, 2016; Herweg et al., 2017). In short, the agenda-setting process is understood to be complex, characterized by ambiguity in issues, processes, and participation, leading to similar complexity in formulating policy preferences under time and attentional constraints. As such, competition over ideas plays a key role in how policy agendas are formulated (Greer, 2016), and constitutes a focus of this study.

The MSF is centered around three partially independent 'streams' of policy-relevant activity. The problem stream considers how conditions in society come to be considered as problems that demand policy action. Policy entrepreneurs and problem brokers (Knaggård, 2015) draw on various focusing events, quantitative indicators, and feedback from existing policy programs to identify and define (or indeed, construct) the relevant problems. Meanwhile, relatively loose networks of often less visible actors work to develop policy alternatives or solutions in the policy (or solutions) stream. Preliminary ideas, part of a "primeval soup" (Kingdon, 1984), are then refined through a "softening" process by policy communities with relevant expertise (Herweg, 2016) according to criteria such as technical and financial feasibility and value acceptability. Finally, in the political stream, national mood or public opinion, interest group activity, and the composition of the current government (Herweg et al., 2015) shape receptivity to the proposed policy problems and solutions. Here, policy entrepreneurs play a key role in coupling the streams together to place a package on the policy agenda (Mintrom & Norman, 2009; Roberts & King, 1991).[2]

This study highlights a subset of theoretical elements deemed important by the MSF, and uses these to guide the empirical approach in light of the policy context and data analyzed. Regarding the problem stream, it emphasizes which problems are identified as essential, highlighting indicators and focusing events as AI policy is arguably too nascent for policy feedback to be substantially in play. Regarding the policy or solution stream, it considers which policy solutions are proposed, how solutions are matched to problems, and whether features like value acceptability and technical feasibility seem to affect the coupling process. Relatedly, the study examines how the definition and coupling of problems and solutions effectively cultivates particular issue frames, for example, surrounding AI's potential for innovation or its social and ethical risks. Finally, the study pays less attention to the politics stream, as it essentially takes place during a single presidential administration (2016-2020) with relatively stable and substantially bipartisan ideological preferences. However, it does evaluate whether the agenda-setting process is expert-dominated or involves substantial public engagement and attention to public opinion (or national mood).

Figure 2.1 provides an overview of elements from the MSF and their role in agenda-setting dynamics for AI policy in light of the subsequent discussion on competing policy paradigms. Key elements theorized or determined to be important are emphasized.

[2] A traditional focus of the MSF is explaining the opening of a policy window, which can occur in either the problem or politics stream, for example after a significant focusing event or election, respectively. Yet this study is especially interested in how AI policy actors define or frame policy issues, problems, and solutions to shape the policy agenda, and is less interested in policy windows themselves. This is because the existence of a policy window in AI is uncontroversial; ample evidence of increased AI policy discourse and activity starting around 2016 confirms this (Perrault et al., 2019).

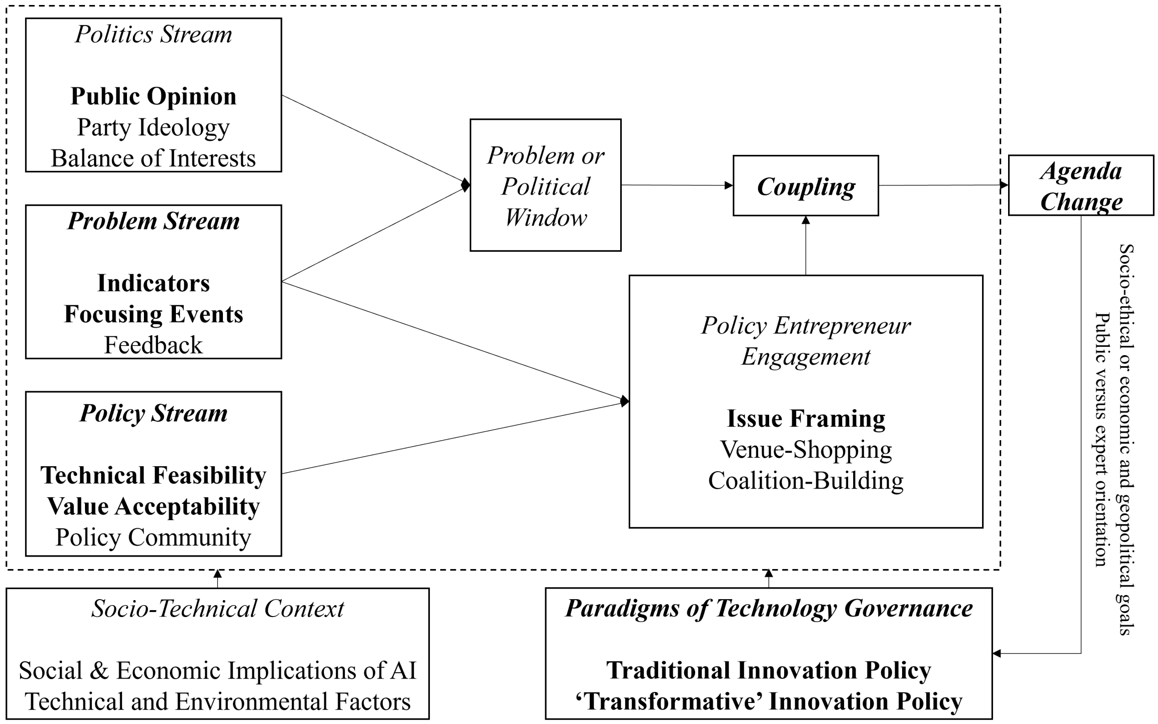

Figure 2.1: Conceptual framework: Agenda-setting dynamics in AI policy

2.2.2 Competing Paradigms of Technology Governance

The second conceptual pillar for the case study is the contrast between competing paradigms of innovation (or technology) policy. Here, a paradigm is understood to encompass the intersubjectively held normative goals behind policymaking, conceptions of underlying policy problems and appropriate instruments used to achieve these goals, and the underlying terminology and ideas used to frame this process (Daigneault, 2014; Hall, 1993). Indeed, policy paradigms have a storied history in the realm of technology and innovation policy (Morlacchi & Martin, 2009). Particularly relevant to this study is the extent to which policy paradigm shifts involve radical breaks in the Kuhnian sense, or more evolutionary transitions within the bounds of 'normal' policymaking (Princen & 't Hart, 2014) as part of the sociological and ideational process of change. To assess this process in the context of AI policy, this study seeks to provide "direct evidence of policy actors' ideas and beliefs" (Daigneault, 2014, p. 463) and provide measurable evidence of the "power and salience of the underlying ideas” (Baumgartner, 2014, p. 478) through quantifying attention to policy problems, solutions, and underlying issue frames.

In particular, the first paradigm this study considers is termed the ‘traditional’ approach to innovation policy, in which an economic, firm-centered, and expert-driven logic dominates agenda-setting and policy design (Edler & Fagerberg, 2017). For such an approach, strategies aimed at boosting the workforce, fostering economic growth and innovation, and accelerating adoption of emerging technologies are paramount. Departing somewhat from traditional industrial policy, modern innovation policy is especially focused on high-tech industries and technological advances that may lead to growth acceleration (Soete, 2007; Vu et al., 2020). Associated with this logic is an issue frame which emphasizes the economic growth potential of AI in terms of increased productivity, efficiency, industrial profit, national GDP, entrepreneurial activity, and so on (Imbrie et al., 2021; Ulnicane et al., 2022).

Typical strategies aligned with traditional innovation policy thus often center on advancing STEM through supply-side reforms, particularly expanding the size and skill of the workforce through education, federal research and development funding, and export and investment controls (Atkinson & Mayo, 2010; Fischer et al., 2021). Other innovation instruments include empowering regions, clusters, or hubs of innovation, supporting small- and medium-sized enterprises (SMEs), and aligning academic, industry, and public sector activity (Smits & Kuhlmann, 2004). Reference to the innovation frame is pervasive throughout policy discourse including in national AI policy strategies and key U.S. policy documents (Schiff, 2021).

A second and competing paradigm of technology policy deviates from a more economic, expert-driven, and firm-centered logic, emphasizing instead societal objectives and public participation in policy. Such an approach is in part a response to failures of traditional science and technology policy and innovation systems policy to solve major societal challenges such as inequality and climate change (Ulnicane, 2016). This perspective can be understood as associated with movements such as transformative innovation policy (Grillitsch et al., 2018; Schot & Steinmueller, 2018), mission-oriented innovation policy (Mazzucato, 2018) and Responsible Innovation (de Saille, 2015), jointly representing a "more recent shift in technology policy towards societal objectives" (Ulnicane et al., 2021, p. 80). Proponents of this paradigm are arguably more prone to advocate for stricter regulation and precaution in policymaking in recognition of the social and ethical risks of AI (Owen et al., 2013; Stirling, 2016). This approach thus acknowledges the possibility that innovation can lead to ethical and societal harms, such that a ubiquitous pro-innovation bias may be problematic. Relatedly, it tends to avoid a narrow focus on supply-side strategies aimed at economic growth, and embraces broader societal objectives across multiple sectors, potentially with a mission-oriented and global scale (Kuhlmann & Rip, 2018). For simplicity, this study term this the 'transformative' paradigm, but does not treat it as synonymous with transformative innovation policy (Diercks et al., 2019).

This approach to technology governance also implies and typically explicitly recommends a more inclusive and participatory process in shaping the policy agenda through incorporation of a broader array of actors beyond firms, academia, and government, such as civil society and the public (Warnke et al., 2016). Already, there have been repeated and widespread calls for diverse and public participation in AI policy from governments, academics, civil society, and industry (Schiff et al., 2021; Stix, 2021; Ulnicane et al., 2020; Vesnic-Alujevic et al., 2020).[3] Indeed, there is increased attention to the role of the public in policy discourse generally (Rowe & Frewer, 2000), within science and technology policy specifically (Macnaghten & Chilvers, 2014; Stirling, 2008), and even within AI policy (Buhmann & Fieseler, 2022; Stark et al., 2021). Yet, while this focus on public participation is emphasized in current technology and AI policy discourse, these calls also reflect the renewal of arguments in the 21st century for multi-stakeholder governance of emerging

[3]For example, the Office of Management and Budget's (OMB) Memorandum to Executive Agencies on Guidance for Regulation of AI Activities (2020, p. 3) places public participation as its second pillar, arguing that "that public participation...will improve agency accountability and regulatory outcomes, as well as increase public trust and confidence" and that "agencies must provide ample opportunities for the public to provide information and participate in all stages of the rulemaking process."

technologies in contrast to traditional top-down government (Pierre, 2000), and indeed follow a long history of political thought on the value of public participation (Arnstein, 1969; Dahl, 1978). Evidently, advocates for more public involvement do not feel their ambition has been achieved, a concern this study reiterates.

A few further notes are helpful in making sense of this study's use of policy paradigms to understand the AI policy agenda. First, while the competing approaches do seem to set up disparate expectations in terms of underlying goals, policy instruments, and the role of stakeholders, the summaries above are simplifications. Approaches to innovation policy vary along more than the two key dimensions emphasized here: there is variation in terms of breadth and depth of scope, how the roles of government and industry are imagined, how the innovation process is understood, and beyond (Boekholt, 2010; Borrás & Edler, 2020; Diercks et al., 2019). As such, it is best to understand the two paradigms presented here as "ideal types" (Weber, 1949) or heuristics that can provide valuable interpretive and explanatory insight despite simplifying a more complex reality (Aronovitch, 2012). Second, the relationship between these two paradigms is, to some extent, unsettled. A newer paradigm could replace the current one in part, or in whole (Hall, 1993). Competing paradigms could also be layered into one another or find another form of synthesis. This study does not presuppose theoretically how this relationship should unfold, but instead takes this question as a meaningful area to contribute knowledge to.

In summary, an important scholarly and policy question is therefore whether technology governance is indeed characterized by such a shift, and what conditions have enabled or failed to enable this. AI policy—and the U.S. federal AI policy agenda in particular—constitutes a ripe and novel case for exploring this question, given that AI policy discourse reflects an unusual degree of contestation between various economic, social, and ethical goals and because robust public participation has been called for extensively. In short, if such a shift is likely to take place in technology policy, it should be in this domain if nowhere else. Along these lines, Table 2.1 draws on the specified components of

Table 2.1: Competing predictions for AI policy based on elements of the MSF

	Traditional Paradigm	Transformative Paradigm
Key Indicators and Focusing Events	Key indicators and focusing events discussed in the U.S. AI policy agenda surround strategic economic and international competition objectives and concerns. They are less salient, accessible, and related to public concerns, and reflect an expert orientation.	Key indicators and focusing events discussed in the U.S. AI policy agenda surround societal and ethical objectives and concerns. They are relatively more salient, accessible, and related to public concerns.
Policy Problems, Solutions, and Issue Frames	Policy problems, solutions, and frames discussed in the U.S. AI policy agenda emphasize economic objectives, such as supply-side strategies for STEM education and R&D, targeted at economic growth and national competitiveness. A pro-innovation bias is present.	Policy problems, solutions, and frames discussed in the U.S. AI policy agenda emphasize societal and ethical objectives, such as reducing inequality and serving vulnerable groups. Innovation is viewed more skeptically and with more precaution advocated.
Role of the Public and Experts	Experts, acting as policy entrepreneurs or policy community members, play an outsized role in shaping policy problems, solutions, and frames discussed in the U.S. AI policy agenda. The need for technical expertise, and increasing the access, supply, and training of experts are emphasized.	Members of the broader public play an outsized role in shaping policy problems, solutions, and frames discussed in the U.S. AI policy agenda. The need for diverse and inclusive participation, increased access, and public engagement and opinion in policy are emphasized.

the MSF and concept of competing paradigms, and presents competing, provisional predictions (or rough hypotheses). These derived predictions establish preliminary analytical grounds upon which to evaluate the research questions.[4]

[4]The entities, relationships, and attributes selected here served as organizing principles, but subject to evolution as part of the analytical and theory-building process when confronted with the data and case at hand (Bhaskar, 1979).

2.3 Methodology

2.3.1 Overview and Epistemology

The methodology for this chapter, driving both data selection and analysis, is the single, holistic case study approach (Yin, 2018) realized through qualitative and quantitative document and content analysis. The unit of analysis is the agenda-setting process as reflected in 63 key AI policy documents curated by the U.S. federal government and published between January 2016 and December 2020. The rationale for the selection of this case is that AI policy is a new, revelatory, and empirically important emerging technology policy domain. Given clear evidence that the AI agenda is subject to contestation over ethical, social, and economic dimensions, and because this case was previously inaccessible to study, it is arguably a strong critical test of whether technology governance is evolving.

A case study of the U.S. federal AI policy agenda in particular is valuable for several reasons. First, it promotes structured exploration of AI agenda-setting in a scholarly and policy context otherwise marked by diverse and sometimes piecemeal interdisciplinary conversation (Taeihagh, 2021). As such, it offers the possibility of integrating disparate knowledge on the conditions, influences, and actors relevant to AI agenda-setting using a prominent framework and detailed empirical evidence. Further, it offers descriptive, explanatory, and exploratory insights and establishes a helpful foundation for identifying more specific and future research questions with respect to the substantive and theoretical context studied in the paper.

The ontological and epistemological approach employed here is most akin to critical realism. Critical realism acknowledges the existence of objective entities and mechanisms working in the world, but appreciates that these are filtered through human experience and subjectivity (Bhaskar, 1979; Wynn & Williams, 2012). In short, the observable world is understood as theory-laden, but not theory-determined (Fletcher, 2017; Hoddy, 2019). This is appropriate for the context of agenda-setting, given how objective features of the socioeconomic context and of AI as a set of technologies are filtered through policy images, narratives, issue frames, and so on as part of the deliberative process. This broader reality can only be indirectly accessed, in this case by interpreting the discourse present in the U.S. AI policy documents.

2.3.2 Data and Inclusion Criteria

The dataset is composed of 63 "key strategic documents" for AI that apply at the federal or national level and are hosted at AI.gov (National AI Initiative Office, Office of Science and Technology Policy, 2022).[5] These include a diversity of document types, including budgetary documents, requests for information, scientific and technical reports, strategy documents, ethical principles, and international agreements. Importantly, as these documents were explicitly curated by the federal government to reflect AI policy priorities, they serve as a logical starting place to understand the deliberative process influencing the emerging AI policy agenda. These official policy documents can thus be understood as "vehicles of messages, communicating or reflecting official intentions, objectives, commitments, proposals, 'thinking', ideology and responses to external events" (Freeman & Maybin, 2011, p. 157). Moreover, the content therein constitutes "the outcome of a political process" influenced by multiple actors with competing discourses, and thus reflect the developing perspectives and preferences of policy actors over different alternatives with which they have been presented (Diercks et al., 2019, p. 887). As such, these documents are especially helpful for exploring to what extent certain problems, solutions, couplings, indicators of national mood, and so on, are influential in shaping the policy agenda.

Given the intentional choice for the government to curate these specific documents, this study approaches inclusion/exclusion with strong deference to inclusion in the final analytical dataset. However, the final sample excludes three documents from the original 66

[5] For accessibility, documents in the dataset referenced in the paper are identified by authoring organization and publication year and month. Readers can cross-reference Table A.1 and Table A.2 in the Appendix to find the original sources.

over the specified time period of 2016-2020: one document is a duplicate and two contain very little discussion of AI.[6] Figure 2.2 presents basic information about the dataset while Table A.1 in Appendix A.1 includes the complete list of analyzed documents.

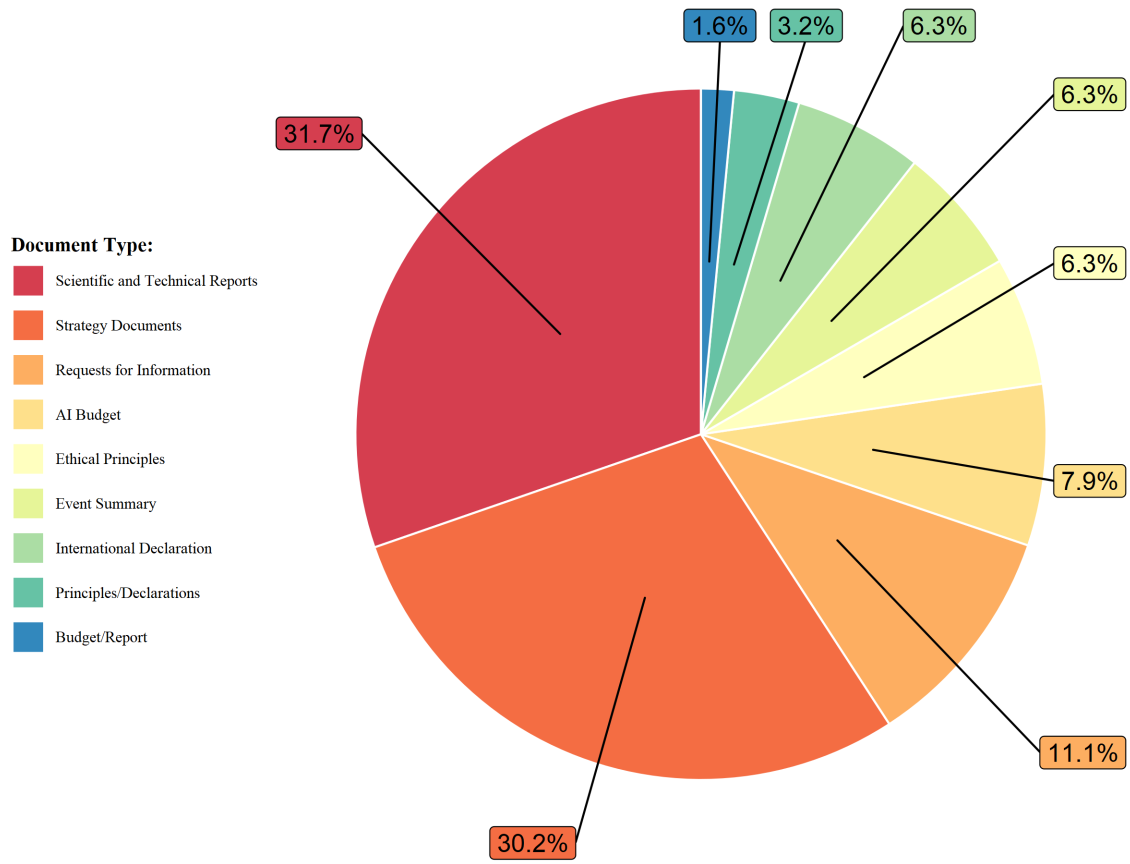

Figure 2.2: Analytical sample: U.S. federal AI policy documents (2016-2020)

Note: Author's calculations from AI.gov, $n = 63$

An important feature of this analysis is that the dataset is centered on executive branch documents, and thus reflects only a subset of the broader agenda-setting process. For example, legislative hearings, private one-on-one meetings, civil society or private sector lobbying efforts, and legislator communications to constituents are not studied apart from mention in the executive agency documents. This scope condition implies limitations with

[6]The excluded documents are the Department of Transportation/Federal Aviation Administration (DOT/FAA) Strategic Plan for FY2019-2022 (2018), the G7 Science and Technology Ministers' Declaration on COVID-19 (2020), and the near duplicate version of the Networking and Information Technology Research and Development Program (NITRD) Artificial Intelligence and Cybersecurity workshop report (2020).

respect to the completeness and generalizability of the findings.

However, there are several key reasons for viewing these documents as a valuable source to understand agenda-setting. First, while document text is often 'strategic' in the sense that underlying attitudes are at least partially hidden, the executive branch of government is thought to be more neutral in its approach than the legislature (H. T. Miller, 2015; Stivers, 2015). Second, the executive branch plays a critical role not only in informing the preliminary stages of agenda-setting by producing analyses often explicitly requested by other policymakers, but also in ultimately shaping and implementing policy. Finally, policy problems and solutions described via lengthy technical reports, for example, are more likely to have the necessary amount of detail to make inferences with confidence as compared to legislative stump speeches. Thus, while it is difficult to understand the complete set of actors and documents involved in AI policy agenda-setting in the context of a single study, these executive agency documents are amongst the best sources to understand the issues studied here.

2.3.3 Analysis Approach

The analysis primarily relies on an a priori (or directed) coding methodology, drawing from the literature on AI policy and the study's conceptual framework to create an initial codebook (Hsieh & Shannon, 2005; Miles et al., 2013). For instance, the codebook includes a set of high-level domains and sub-codes articulating specific AI policy problems, policy solutions, and issue frames, as well as focusing events, indicators of problems, stakeholders, and other areas.[7] Importantly however, the initial codebook was substantially modified, with additions, deletions, along with merging and separation of existing codes, and speci-

[7]Only a subset of the domains coded are discussed in this paper: In particular, the paper omits most results related to the Political Stream (e.g., party ideology) and approaches to Regulation (e.g., precautionary, pro-innovation).

fication of more robust definitions (Hoddy, 2019).[8]

Table 2.2 provides a summary of the key domains analyzed and Table A.2 in the Appendix provides definitions and representative quotes for each code. Table 2.2 organizes codes by domain, and for some domains includes categories for 'traditional,' 'transformative,' and 'hybrid,' a classification approach explained in more detail in the Results section.

The coding process thus sought to answer questions like:

- *Are focusing events and indicators referenced, and if so, which seem to play an important role?*
- *What are the key policy problems, solutions, and issue frames proposed?*
- *What kinds of actors are considered important in shaping the AI agenda (e.g., the public, experts)?*
- *To what extent do socio-ethical considerations translate into concrete policy solutions?*

The application of the codebook involved identifying all relevant codes that applied at the level of individual paragraphs, helpful for capturing surrounding context and analyzing the co-occurrence of codes (e.g., of experts and particular policy solutions). Of course, this approach involves trade-offs, as documents vary widely and structural differences mean that the coding strategy is both under and overinclusive at times.[9] Another limitation is that the codebook development and coding was a single-coder process,[10] which means that, for example, inter-rater reliability measures are not available. However, for this kind of qualitative analysis, other measures of rigor may be more appropriate (Lincoln & Guba, 1986), such as trustworthiness, authenticity, and confirmability, facilitated by transparency in methodology and access to the underlying data sources and codebook. The coding resulted in over 4,100 individual paragraphs coded, with codes applied to around 2,300 pages

[8] After coding an initial sample of documents, I recoded a clean subset of documents and compared results to improve coding reliability, important because this is a single-coder effort. Nevertheless, the choice and parsing of codes is at least partially subjective and other researchers may have other preferences. For further information about the addition, removal, and merging of codes, contact the author.

[9] Of note, the coding process generally excluded appendices, footnotes, generic introductory text (e.g., author or organization biographies), and in cases, sections focused on topics other than AI (e.g., for a document emphasizing AI and quantum computing, it excluded analysis of the latter section).

[10] Feedback from experts in policy process theory and AI policy was helpful in improving the codebook however.

Table 2.2: Final codebook

Coding Domains (# of Codes)	Code
Focusing Events (7)*	**Traditional**: · Expert Concerns · Games · Geopolitics & Military · Industry & Government Adoption · Technical Performance & Advances **Transformative**: · Protests · Scandals & Disasters
Indicators (7)	**Traditional**: · Economic & Workforce · Education · Expert Concerns · Geopolitics & Military · Technical Performance & Advances **Transformative**: · Poverty, Harm, & Fairness · Scandals & Disasters
Issue Frames (3)	**Traditional**: · Geopolitics · Innovation **Transformative**: · Ethics
Problems (16)	**Traditional**: · Data Quality & Access · Misuse & Hostile Actors · Performance & Reliability · Realizing Benefits · Security & Military · Workforce & Education **Hybrid**: · Accountability & Responsibility · Privacy · Safety · Transparency · Trust **Transformative**: · Fairness & Bias · Human & Civil Rights · Inequality & Inclusion · Value Alignment · Vulnerable Populations
Solutions (16)	**Traditional**: · Cooperation & Dialogue · Data Quality & Access · Pilot Projects & Testbeds · R&D & Adoption · Workforce & Education **Hybrid**: · Build Trust · Diverse Participation · Grants & Procurement · Human-AI Teaming · Impact & Risk Assessment · Monitoring & Reporting · Standards/Best Practices · Transparency **Transformative**: · Public Engagement · Social/Ethical Consideration · Technical Fixes for Ethics
Stakeholders (2)	· Experts · Public

* Note that while Birkland (1998) and Birkland and Warnement (2016) define focusing events as sudden, relatively uncommon, typically harmful, and restricted to a particular location, the definition used here deviates, adopting a broader working definition to capture ‘events’ as reflected in AI policy discourse, often less discrete and region-specific.

of text across the 63 documents.

Following the qualitative coding process, the study draws on quantitative content analysis techniques (White & Marsh, 2006). For example, to understand the importance of policy problems, I extract statistics capturing the absolute prevalence of each problem (number of coded paragraphs per document and overall) and binary presence of each problem per document (present or absent). Importantly, I also determine the 'normalized' importance of each problem, by restricting attention to only policy problems (or solutions, or focusing events, etc.), and by calculating percentage representation of each problem per document. This allows me to rank and quantify the salience of given policy problems, solutions, issue frames, and so on, while addressing concerns like the possibility that very large documents with many codes or strong emphases would skew results in a given direction.[11] Finally, I apply an abductive (or retroductive) explanatory approach to synthesize results. This involves employing pattern matching in light of the conceptual framework and provisional predictions (Bouncken et al., 2021; Sinkovics, 2018), and associating what is empirically observed with plausible generative mechanisms (Avenier & Thomas, 2015; Gioia et al., 2013). By redescribing the empirical data in this way, the study thus seeks to confront and elaborate theory to provide a more robust explanatory account of AI policy agenda-setting.

2.4 Results

2.4.1 Focusing Events and Indicators: Dominance of Traditional Concerns

The first results address the relative attention paid to different types of focusing events and indicators, understood as key elements of the problem stream that lead societal conditions to be recognized as pressing policy problems that demand action (DeLeo et al., 2021)). Each focusing event or indicator is classified as 'traditional' or 'transformative' as per Table 2.2.[12] Figure 2.3 shows coverage of focusing events and indicators, arranged by

[11] For example, a document about bias that mentions bias dozens of times will not exert undue weight.

[12] Again, these classifications are only rough heuristics, as individual instances are invariably more complex.

normalized percentage.[13] Overall, the large majority of focusing events and indicators mentioned in U.S. AI policy documents are traditional in nature. For example, while only 10 documents discuss at least one transformative focusing event, 40 mention at least one traditional focusing event. For indicators, similarly, 13 documents mention a transformative indicator compared to 31 which mention a traditional indicator. Nearly 96% of focusing events and 84% of indicators mentioned are traditional in nature.[14]

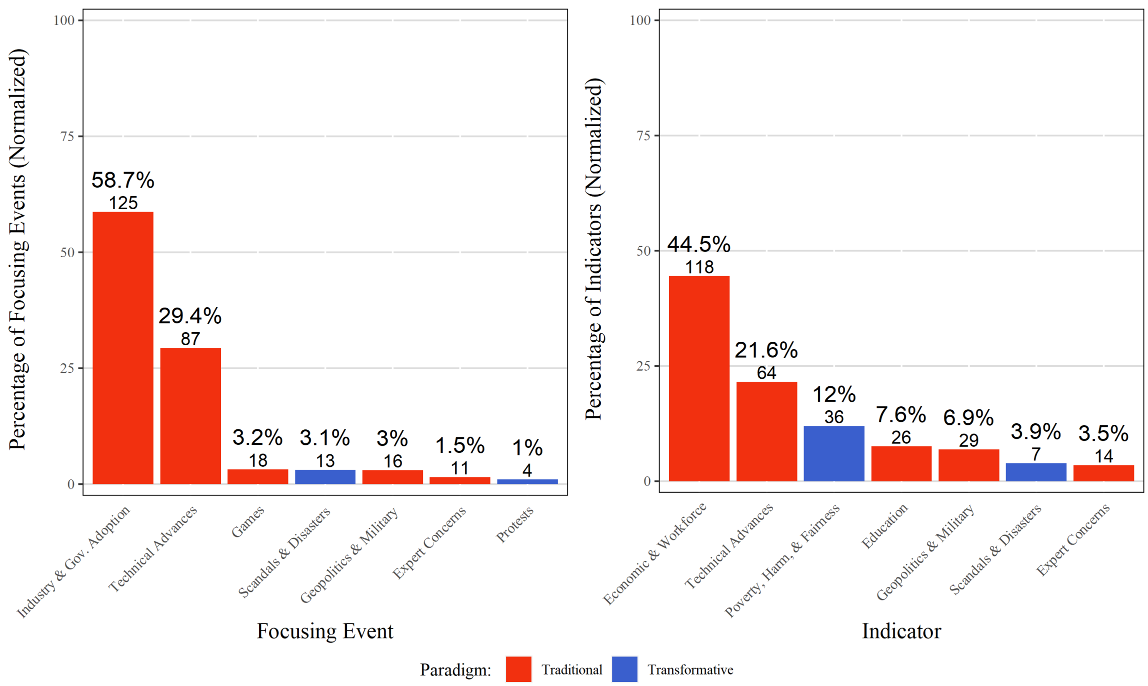

Figure 2.3: Coverage of focusing events and indicators in U.S. AI policy documents

Note: Percentages of all focusing events (total $n = 274$) and indicators (total $n = 294$) in documents. Normalized percentages with absolute totals below.

The slant of these results is to some degree surprising. Despite widespread attention

[13]The normalization process involves reweighing each document to have the same number of elements within a domain (e.g., focusing events or indicators or issue frames). The total number of elements over all documents is then determined along with the percent attention to each respective element. Note that presenting results based an absolute or normalized measures produces largely similar results across all analyses.

[14]For example, a prototypical statement reflective of which indicators and focusing events are commonly emphasized comes from the Office of the Director of National Intelligence (January 2019, p. iv): "Artificial intelligence (AI)...has shown dramatic advances in autonomous systems, computer vision, natural language processing, and game playing. These AI systems can perform tasks significantly beyond what was possible only recently (e.g., autonomous systems) and in some cases even beyond what humans can achieve (e.g., chess and Go)."

to AI ethics topics in media generally (Chuan et al., 2019; Perrault et al., 2019), including ostensibly high-profile scandals surrounding biased facial recognition, wrongful arrests, sexism in hiring algorithms, self-driving car crashes, and more, very few of these incidents appear in U.S. AI policy documents. Further, despite prominent media and AI stakeholder discourse surrounding expressions of expert concern by individuals like Stephen Hawking and Elon Musk (Galanos, 2019; Neri & Cozman, 2020), these concerns are hardly present. Thus while there has been debate about the possibility that some ethics-related focusing events have been important (Ouchchy et al., 2020; Stix & Maas, 2021), this study suggests the impact of these influences is limited. Indeed, even recent efforts to catalogue hundreds incidents of social and ethical harm by the Partnership on AI (McGregor, 2020) in its new AI Incidents Database or to catalogue protests via the Collective Action in Tech database (Tarnoff et al., 2021) are barely reflected in AI policy documents.

Instead, the dominance of traditional problem indicators comports with the focus of many global AI indicator dashboards that assess data infrastructure, research productivity, and human capital, such as the Global Cities AI Readiness Index (2020), Global AI Index (2021), and Global AI Vibrancy Index (2021). While some indicator sets like the AI Social Contract Index (2020), the OECD AI Policy Observatory database (2022), and Government AI Readiness Index (2019; 2020) have more recently incorporated elements surrounding topics like ethics and diversity, the metrics overall are overwhelmingly traditional, an observation supported in a recent extensive analysis of global metrics of AI (Erkkilä, working paper).

A possible explanation for this skew is that the social and ethical indicators and focusing events have influenced public opinion or the attitudes of legislators, but with less visible downstream implications, and that executive agencies are institutionally less prone to discuss these kinds of elements. Yet even in this charitable case, the finding that actors across more than a dozen sectors of government (many of which do articulate social and ethical concerns) overwhelmingly focus on indicators and focusing events surrounding the economy, workforce, and race for AI innovation generally is telling.

2.4.2 Issue Frames: Synthesis with Subsumption of Ethics into Innovation

This paper reviews discussion of three issue frames surrounding AI's innovative, ethical, and geopolitical dimensions, respectively, issues also found to be prominent in AI policy discourse by other researchers (Kim, working paper; Köstler & Ossewaarde, 2022; Ulnicane et al., 2022). Results indicate that issue (or policy) frames surrounding both innovation and ethics feature heavily in U.S. AI policy documents. 59 of 63 documents discuss the innovation frame at some point, while almost as many documents, 57, discuss the ethics frame, with only 37 discussing the geopolitics frame. Moreover, the majority of documents discuss multiple frames. As displayed in Figure 2.4, 19 documents discuss both innovation and ethics, and a full 36 discuss all three issue frames. This provide compelling evidence that there has been some degree of convergence or synthesis in AI policy discourse, and indicates that these policy frames can be mutually reinforcing as opposed to mutually exclusive. In fact, some documents take very explicit efforts to justify the mutually reinforcing nature of these issues.

For example, conveying the overlap between geopolitics and ethics, the Congressional Research Service (CRS) (November 2020, p. 21) states that "it may be important for Congress to understand the state of rival AI development—particularly because U.S. competitors may have fewer moral, legal, or ethical qualms about developing military AI applications." As an example of the merger of innovation and competition frames, a major piece of U.S. legislation is literally entitled the "U.S. Innovation & Competition Act" (2021). As stated by the National Science and Technology Council (NSTC) (September 2019, p. 3), "sustained Federal R&D investment in emerging technologies is critical to promoting and protecting American innovation and international leadership." As an example of merging all three frames, the Organization for Economic Cooperation and Development (OECD) (May 2019, p. 6) argues that "trustworthiness of AI systems is a key factor for the diffusion

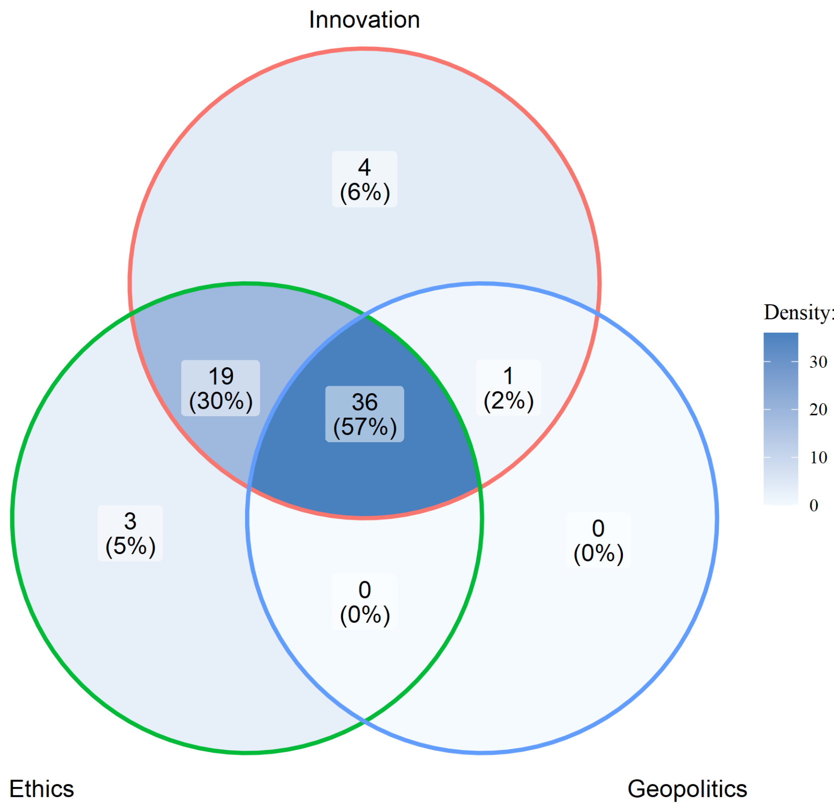

Figure 2.4: Coverage of issue frames in U.S. AI policy documents

Note: Issue frames in documents. Absolute totals with percentages below.

and adoption of AI... essential to fostering adoption of trustworthy AI in society, and to turning AI trustworthiness into a competitive parameter in the global marketplace."

Yet this apparent synthesis risks overstating the relative importance of issue frames other than innovation. When examining the relative amount of attention to each issue frame, allowing for multiple mentions per document, the innovation frame continues to dominate. Treating the dominant frame in a document as the frame mentioned more than all other frames, innovation is dominant in 40 of 63 documents (63.5%), ethics in 12 of 63 documents (19.0%), and geopolitics in only 4 of 63 documents (6.3%).

Importantly, a key related finding is that ethics discourse, while strikingly common, is often subsumed into innovation-oriented goals. This ability to strategically merge and subsume the ethics frame in this way may result from AI's complexity and ambiguity as a general purpose technology with numerous implications, leading to epistemic uncertainty and subsequent interpretive flexibility (Goyal et al., 2021). There are several ways in which this pattern of subsumption is reflected. Ethical issue frames are often featured heavily in introductions, mission statements, and motivations, with less attention in more detailed portions of documents (e.g., policy solutions). Relatedly, ethical issues are often treated as more exploratory, something to be "understood" or concerned with generally. Ethics topics, even when they show up formally in the strategic priority lists of documents, are often further down the list compared to other goals. Finally, even when ethical issues are treated seriously, they are often rendered in the ultimate service of other goals, and thus treated as secondary goals or as means to other ends.

For example, the Executive Order on Maintaining American Leadership in AI (June 2019, p. 20) "emphasizes that maintaining American leadership in AI requires a concerted effort to promote advancements in technology and innovation, *while* protecting civil liberties, privacy, and American values" (emphasis added). The American AI Initiative (September 2019, p. 5) echoes that "in all of these actions, the Initiative emphasizes the importance of advancing AI innovation, while fostering public trust and confidence in AI technologies." Relatedly, ethics is even seen as a barrier to innovation. As the Department of Commerce (DOC) notes (August 2020, p. 1) with respect to explainable AI, "suspicions that the system is biased or unfair…may slow societal acceptance and adoption of the technology" while the G20 notes similarly (June 2019, p. 5) that securing the trust of the uncertain public "is essential for enabling the benefits of the global digital economy." Numerous statements echo these examples.

Yet it is important to note that such a conclusion remain provisional as the agenda continues to develop, and that numerous documents and agencies do put forward strong statements surrounding ethics, sometimes with concrete action. Interestingly, some branches of the military are particularly strong in this respect. The Department of Homeland Security (DHS) and Department of Defense (DOD) (February 2020) and Office of the Director of National Intelligence (ODNI) (February and July 2020) have been unusually active in articulating ethical concerns, and the Air Force (September 2019, p. 2) even conveys a precautionary tone to innovation, noting that "artificial intelligence is not the solution to every problem" and that "its adoption must be thoughtfully considered in accordance with our ethical, moral, and legal obligations to the Nation." Other agencies like the National Highway Traffic Safety Administration (NHTSA) seem to have constructively synthesized goals around ethics (namely, safety, privacy, and access) with goals surrounding innovation. Ultimately however, the evidence supports the concern that prominent discussion of AI ethics remains substantially rhetorical, and that this discourse is being layered into the traditional paradigm of technology governance rather than transforming it.

2.4.3 Policy Problems: Traditional Emphasis Followed by 'Hybrid' Challenges

The next results show the analysis of policy problems in U.S. AI policy documents. The analysis includes only policy problems that received an average of more than one quotation per document.[15] The remaining top 16 problems presented in Figure 2.5 are classified as traditional, transformative or 'hybrid.' This latter category reflects an interesting complexity in AI policy discourse. Namely, while certain topics might be formally associated with AI ethics discourse, such as trust or transparency, the ways in which these topics are discussed often highlight innovation or economic adoption concerns rather than human-centered social and ethical concerns.

For example, 'trust' is often rendered as a means to promote innovation, as exemplified previously. Along similar lines, the DHS (December 2020, p. 14) notes that building trust in the American public can "protect and guard against reputational impacts to the Department" while the DOC (August 2020, p. 4) notes that some forms of transparency are "designed to generate trust and acceptance by society," which according to the DHS (p. 16) should "result in an educated and supportive American public." Other topics with an

[15]For example, problems such as artificial general intelligence, concerns about monopolies, risks surrounding misinformation, and organizational barriers in AI adoption are excluded with less than 63 mentions each.

ostensibly ethical focus, like privacy and safety, are likewise rarely discussed in terms of human impacts, and are often framed in terms of strategic geopolitical or economic goals. Even topics like diverse participation can be rendered as, on the one hand, inclusion of vulnerable subgroups of the public in decision-making, or alternatively, the need for a diverse array of *experts* across different government agencies and industry bodies. Thus in analyzing topics like privacy, transparency, safety, and diverse participation, each topic is treated as on-balance closest to 'hybrid' in its relationship to traditional versus transformative innovation policy.

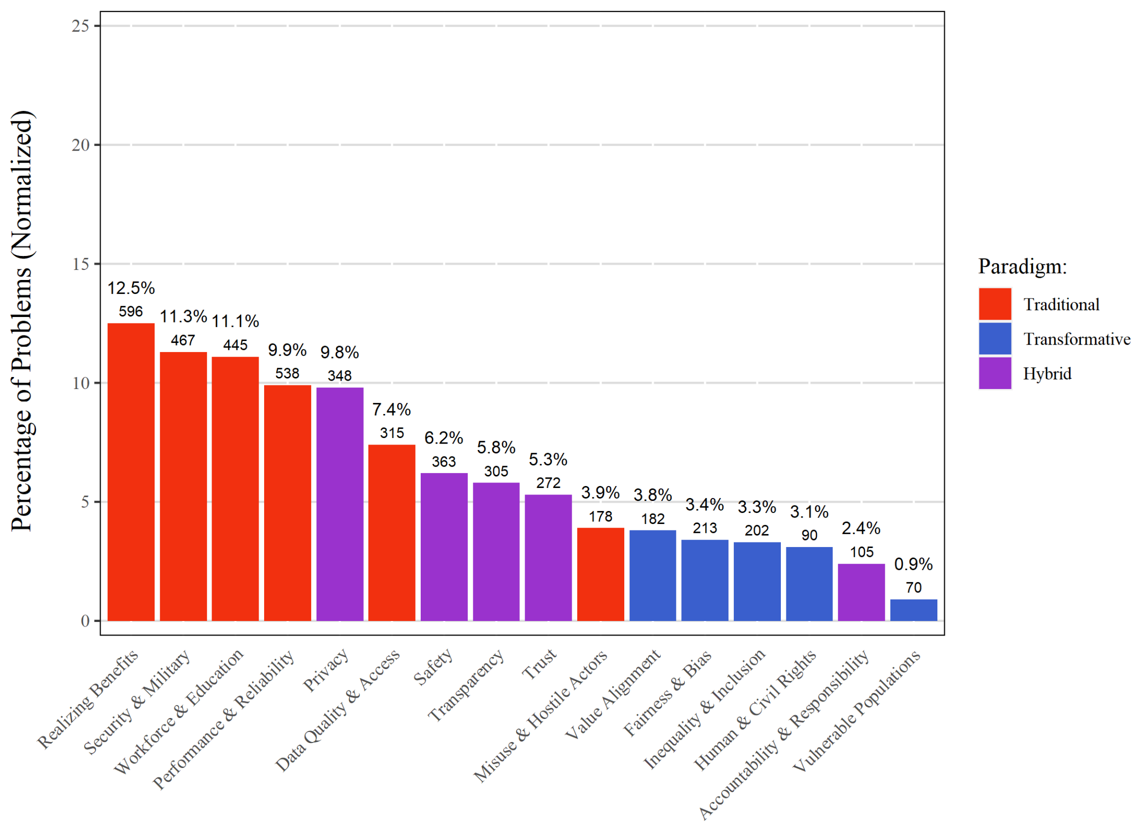

Figure 2.5: Coverage of policy problems in U.S. AI policy documents

Note: Percentages of all policy problems (total $n = 4,689$) in documents. Normalized percentages with absolute totals below.

Based on this classification approach, the top several problems are quite traditional in nature. Interestingly, the most common policy 'problem' overall is the need to realize the benefits of AI, in light of the opportunity cost of not doing so. This interesting sentiment reflects a strong innovation motive and potentially unique hallmark of technology agenda-setting discourse with respect to problem identification. Meanwhile, problems surrounding the expertise or capacity of the workforce and the need to build better AI systems or attend to security concerns are classically traditional. Overall, a full 44% of problems mentioned fall into the top four categories displayed in Figure 2.5. In contrast, the transformative problems receive the least attention, such as those surrounding human rights and vulnerable populations. Jointly, fairness and bias, inequality and inclusion, value alignment, human and civil rights, and vulnerable populations constitute only around 16% of total problems discussed in the U.S. AI policy agenda.

Viewed differently, however, social and ethical problems do show up to a surprising degree. For example, while 49 documents mention workforce and education, and 54 mention security or military, as many as 25 documents mention human or civil rights, 28 mention inequality or inclusion, and 41 mention fairness and bias. Yet the distinction between the mere coverage by a document of an ethical problem and the degree of attention overall is meaningful, a distinction perceivable due to the methodology applied here. One interpretation is that, while ethical problems show up in many documents, they are often treated casually as part of a long list of problems without as detailed attention or consideration. Examining solutions offered in response to these problems is another way of determining the seriousness of policymakers with respect to these problems

2.4.4 Policy Solutions: Traditional Solutions with Hybrid Possibilities

The findings with respect to policy solutions echo the patterns above but suggest new possibilities as well. The top four coded policy solutions, jointly reflecting 53% of all solutions coded, are calls for cooperation (overwhelmingly government and industry), research and development with increased adoption, increasing access to data, and building the STEM and AI workforce. Notably, some of these solutions follow quite linearly from the associated problems: A deficit of high skilled workers implies a need for more training, and the same is true of access to high-quality data.

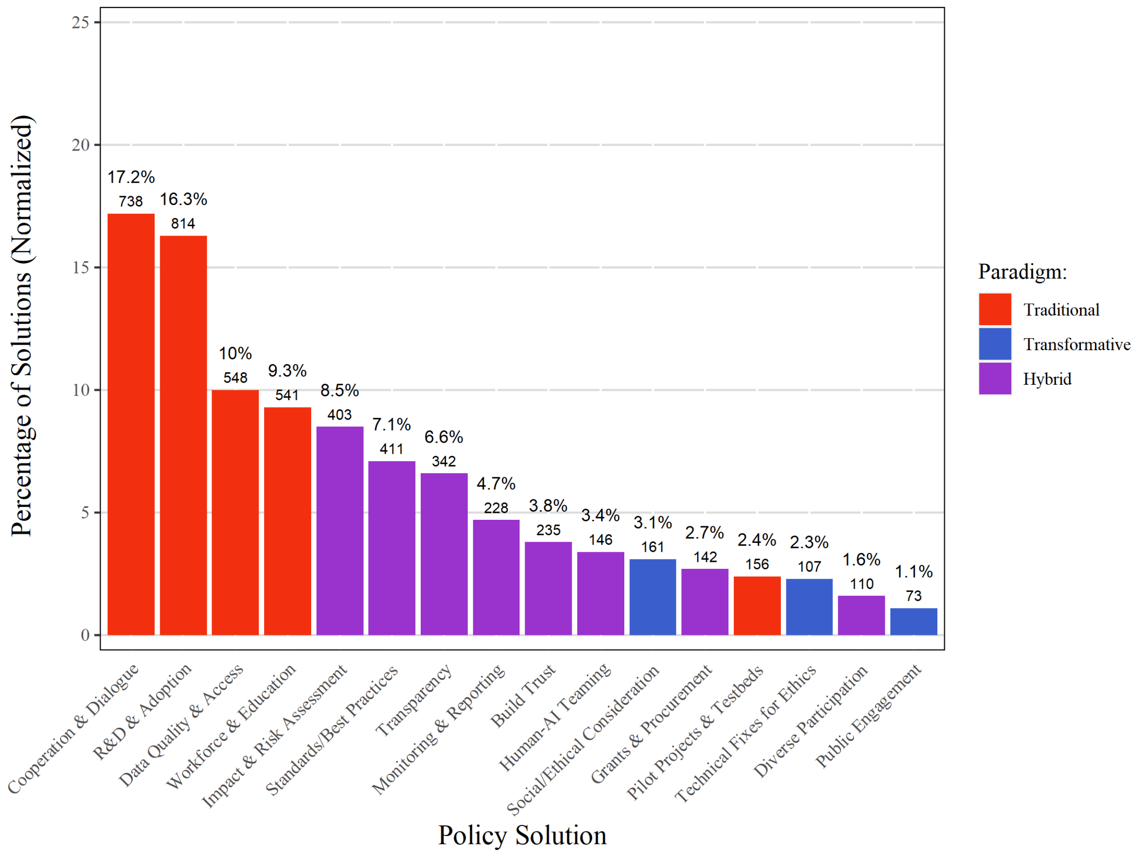

Figure 2.6: Coverage of policy solutions in U.S. AI policy documents

Note: Percentages of all policy solutions (total $n = 5,155$) in documents. Normalized percentages with absolute totals below.

However, these traditional solutions are followed by a set of practices classified as hybrid here. Numerous documents call for efforts to evaluate the impacts and risks of AI systems, to promote standards and best practices, to increase transparency, and so on. Many of these solutions are theoretically and sometimes explicitly responsive to both traditional and transformative type problems. Thus while calls for formal social and ethical consideration in AI development or public engagement as policy solutions are relatively sparse, there are potential inroads to address transformative type policy concerns. To the extent to which impact assessments, standards, reporting practices, and so on, can be designed so as to foster attention to broader kinds of social and ethical goals, the door remains open to serious action on these dimensions.

2.4.5 Roles of the Public and Experts: Strongly Expert-Dominated

Finally, Figure 2.7 reports results related to the role of the public versus that of experts in the U.S. AI policy agenda. The Sankey diagram depicts the overlap between each stakeholder and the 16 policy solutions presented previously. Because the co-occurrence of stakeholders and policy solutions was coded at the level of individual paragraphs, this provides only a rough proxy. Paragraphs often involve long lists of problems and solutions, and solutions and stakeholders may be mentioned in different sections of documents. Nevertheless, Figure 2.7 provides a reasonable approximation, and if anything, overstates the role of the public.

In particular, the public receives relatively less attention and takes on a very modest role overall. The public is mentioned in 46 of the 63 documents, while experts are mentioned in 57 documents. Meanwhile, concrete calls for public engagement as a policy solution appear in only 24 documents, and public opinion is mentioned as a factor of interest in only 13 documents. These findings reflect that, while mentions of public good or public involvement show up quite commonly, these calls for action are rarely realized through concrete policy solutions. Members of the public are most often referenced with respect to general dialogue, and even much of this communication is generally one directional: Policymakers discuss taking action to serve the public, to build the public's trust, to educate the public about AI, to foster public adoption of AI, and so on, but offer few concrete plans for two-sided public engagement.

To a lesser extent, members of the public are thought helpful in providing input, at least in theory. Yet few metrics of public opinion are ever presented, suggesting that document writers have spent little time gathering information from public opinion polls, for example. Further, the idea of the 'public' is often construed more broadly than imagined in the

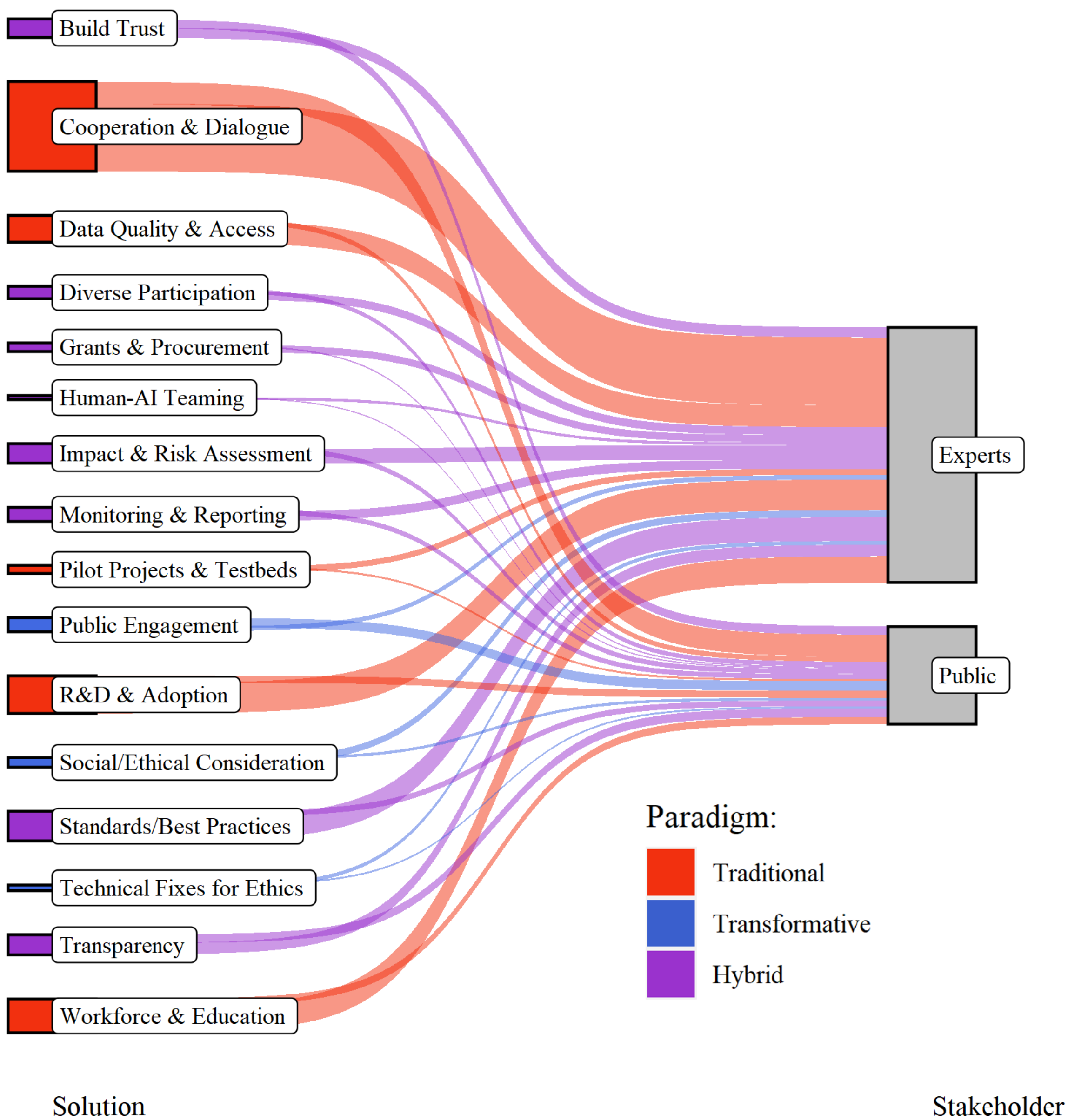

Figure 2.7: Role of the public and experts in U.S. AI policy solutions

transformative paradigm, e.g., with a significant emphasis on members of industry who might comment on regulatory proposals. In contrast, proposals to involve experts from government and industry in AI policy often involve detailed specifications of actors, roles, institutions, objectives, and even timelines. The discrepancy is extremely sharp. This may be explained by beliefs that "industry and academia are the primary sources for emerging AI technologies," (NSTC, October 2016, p. 34), and that "developing technical expertise will provide the basis for" advances in AI (NSTC, October 2019, p. 39). In combination, is quite clear that efforts to foster public participation in AI policy (Buhmann & Fieseler, 2022; Crockett et al., 2021) are not being heeded in practice.

2.5 Discussion

This study sought to provide theoretical and empirical insight into the development of the U.S. AI policy agenda in light of agenda-setting theory and competition between alternative paradigms of technology governance. However, there are key limitations to this study worth considering. The time period under examination is limited to 2016-2020, whereas many important developments are occurring beyond this time window, including new legislation, strategic documents, discourse around geopolitical competition, and possible transformative ethics-related reforms advancing through agencies like the National Institute of Standards and Technology (NIST) and the Federal Trade Commission (FTC) (2021). Moreover, the complexity, nuance, and variation across documents requires some level of interpretation and inference, performed by a single researcher. Other researchers may identify different emphases or ways to aggregate and parse topics of interest. Finally, the analysis is largely limited to executive agency documents at the federal level, with some external evidence invoked in a supporting fashion. Future research should probe sources like hearings, speeches, legislative text, and documents produced by the public, private, and non-governmental sectors to add insight on the agenda-setting process.

Yet reviewing the focusing events, indicators, policy problems and solutions, issue frames, and role of stakeholders in these key strategic AI documents helps to paint an overall picture. Initially, the evidence suggests a striking level of attention to ethics discourse in the U.S. AI policy agenda. Despite the reputation of the United States as comparatively skewed towards innovation compared to, for example, the European Union, numerous government agencies articulate ethical priorities and goals and several even identify their own ethics principles and frameworks. This focus is especially evident in some of the most 'important' AI policy documents (identified in Table A.1)—those with a broad horizontal scope affecting many sectors of government. Developments like the U.S. AI Bill of Rights, the explicit classification of 'ethics documents' by the White House, and participation by the U.S. in ethical statements made by the OECD and G20 are all signs of genuine interest and openness. Yet those invested in the transformative dimensions of AI are likely to be disappointed as these high-level calls for ethics are underserviced when documents move beyond broad mission statements, a concern raised in prior literature (Schiff et al., 2021; Ulnicane et al., 2021).

The MSF provides two compelling theoretical explanations for lack of full translation of ethics into the policy agenda: lack of value acceptability and lack of technical feasibility. Regarding value acceptability, it may be the case that some calls for ethical action go beyond the tolerance of policymakers. For example, more narrowly-tailored ethical solutions surrounding issues like privacy or algorithmic bias may be admissible within the bounds of current U.S. values towards technology. Meanwhile, more sweeping calls (Waelen, 2022) to address structural inequality, reform societal power dynamics, and strictly regulate AI from a precautionary perspective may be too outside the bounds for the current U.S. regulatory mode. For instance, while calls for upskilling the workforce are pervasive across the AI agenda, a vanishingly small number of documents call for reforms to the social safety net or for redistributive tax reforms.[16] The results here thus provide empirical evidence to support concerns by other scholars (Erman & Furendal, 2022; Wong et al., 2022) that only a subset of AI ethics issues are currently tolerated and translated into practice.

The second explanation suggested by the MSF surrounds technical (and financial) feasibility. Simply, some AI ethics issues are thought to be addressable through technical fixes. It is imagined that new technical practices and standards will help developers avoid unfair, opaque, and privacy-violating AI systems. Indeed, many of the frameworks promoted to address AI ethics (Morley et al., 2021) center around these technical solutions, while

[16]These solutions were so sparse that they were eliminated from the analysis of policy solutions.

broader socio-technical solutions and calls for economic and social reform are scoped out of attention. This sentiment with respect to technical feasibility is expressed clearly by the DOC and NIST (August 2019, p. 16): "While stakeholders... expressed broad agreement that societal and ethical considerations must factor into AI standards, it is not clear how that should be done and whether there is yet sufficient scientific and technical basis to develop those standards provisions."

A third and related explanation is the lack of appropriate venues to facilitate more expansive transformative goals in light of institutional norms and constraints (Justo-Hanani, 2022). This is evidenced in how calls for ethical action—especially prominent in government-wide AI policy documents—become narrowed when translated to individual government agencies. For example, addressing vehicle safety and cybersecurity related to autonomous vehicles is comfortable in light of the NHTSA's typical policy rationales and instruments, but addressing widespread societal inequality is not. The same is true of military entities, who are concerned with issues like the trust of autonomous weapons operators and civilian casualties. To the extent to which individual agencies interpret their ethical charges within the bounds of traditional institutional constraints, this seems to delimit the potential for broader transformative change.

Another surprising finding is the prevalence of various 'hybrid' problems and solutions, where topics like privacy and safety become 'rote' engineering and software practices, perhaps even justified with respect to strategic economic and geopolitical goals. Yet that these hybrid proposals often become rhetorically divorced from human-centered impacts may ironically be a sign of progress or synthesis (constructive or otherwise). That is, as abstract ethical goals are operationalized into concrete technical proposals surrounding problems like privacy, this act of sensemaking and reduction from rhetoric to action may reflect a necessary normalization of transformative concerns into concrete ways of working (af Malmborg, working paper). Thus, it is not entirely clear if the transformative agenda is being fully—or destructively— subsumed into the traditional paradigm. The emerging agenda may ultimately reflect some degree of layering and productive synthesis.

These considerations point to areas for future action. Stakeholders concerned with AI's social and ethical implications seem to have had substantial success in promoting an associated issue frame, and even strategically merging it with other issue frames. To move from rhetoric to action, however, AI policy entrepreneurs will likely need to find ways to operationalize and translate their problems into workable solutions, especially within policy sectors, and then socialize these ideas to policymakers in relevant government agencies. This could mean working to ensure that emerging proposals surrounding dialogue, public engagement, impact assessment, and standards incorporate social-ethical concerns and strategies. Indeed, the attention of AI standards like IEEE 7010-2020 (Schiff et al., 2020) and NIST's AI Risk Management Framework (2021) to socio-technical dimensions of AI seem to be indicators of willingness to push beyond traditional technology governance, though these 'public goods' focused standards efforts also face special challenges (von Ingersleben, working paper). Yet for larger scale structural and social reforms where technical fixes seem unlikely, the barriers to transformation may be even greater. Attempting to deeply change the relationship of the United States to innovation may require long-term socialization, replacement of decision-makers over time, or even an unprecedented crisis that resurfaces socio-ethical concerns.

2.6 Conclusion

This study contributes to the literature in several ways. To scholars of the policy process, it constitutes one of very few studies addressing technology policy or AI policy in particular, despite the importance of these domains. Methodologically, through both qualitative and quantitative content analysis, it demonstrates how policy documents can be fruitfully analyzed and interpreted to reveal subtle features of the agenda-setting process. While the large majority of work on the MSF offers qualitative insight based on process tracing and document analysis, this study builds on these methods with extensive mixed methods analysis, showing how the absolute and relative prevalence of elements like focusing events and issue frames can be studied.

For scholars of technology governance and innovation policy, the findings help to answer whether the 21st century is unfolding with respect to a new governance paradigm, as well as how ethics relates to theories of innovation policy. Overall, the evidence for such a grand shift is limited. The rhetorical attention and even willingness to think more broadly about technology's mixed implications suggest a meaningful degree of openness to change, but constraints like technical feasibility and value acceptability in the U.S. context demarcate modern limits. An initial reading of the evidence is that elements of social and ethical concern have been layered into, or are beginning to be synthesized into, the traditional paradigm. Meanwhile, other elements of the transformative agenda, like increased public engagement, have a long way to go.

For scholars and stakeholders of AI policy in particular, the findings also begin to answer key questions like the extent to which ethical concerns in AI are translated into policy, and how the United States may function as an international policy actor. In light of the importance of the United States as a global AI policy leader, the findings offer some notes of promise and some notes of concern. Dismissing the United States as beholden to a neoliberal governance paradigm is too hasty. Indeed, many of the key policy entrepreneurs in the AI ethics community and ongoing work identifying associated policy problems and solutions is occurring in the United States. Moreover, there is (bipartisan) openness to the kinds of socio-ethical perspectives centered in the European Union, reflecting a greater degree of convergence than imagined. How AI governance will unfold in the next decades will thus depend on ongoing efforts to carefully frame problems, develop concrete and feasible solutions, and to identify novel venues and strategies for placing more atypical and transformative strategies on the global policy agenda.

2.7 References for Chapter 2

af Malmborg, F. A. R. (working paper). Between disruption and stability – narrative dynamics of the European Commission's AI policy.

Arnstein, S. R. (1969). A ladder of citizen participation. *Journal of the American Institute of Planners*, *35*(4), 216–224. https://doi.org/10.1080/01944366908977225

Aronovitch, H. (2012). Interpreting Weber's ideal-types. *Philosophy of the Social Sciences*, *42*(3), 356–369. https://doi.org/10.1177/0048393111408779

Atkinson, R. D., & Mayo, M. J. (2010). *Refueling the U.S. innovation economy: Fresh approaches to science, technology, engineering and mathematics (STEM) education* (SSRN Scholarly Paper No. ID 1722822). Social Science Research Network. Rochester, NY. https://papers.ssrn.com/abstract=1722822

Avenier, M.-J., & Thomas, C. (2015). Finding one's way around various methodological guidelines for doing rigorous case studies: A comparison of four epistemological frameworks. *Systemes d'information management*, *20*(1), 61–98. https://doi.org/10.3917/sim.151.0061

Baumgartner, F. R. (2014). Ideas, paradigms and confusions. *Journal of European Public Policy*, *21*(3), 475–480. https://doi.org/10.1080/13501763.2013.876180

Bhaskar, R. (1979). *The possibility of naturalism: A philosophical critique of the contemporary human sciences* (First edition). Harvester Press.

Birkland, T. A. (1998). Focusing events, mobilization, and agenda setting. *Journal of Public Policy*, *18*(1), 53–74. https://www.jstor.org/stable/4007601

Birkland, T. A., & Warnement, M. K. (2016). Refining the idea of focusing events in the Multiple-Streams Framework. In R. Zohlnhöfer & F. W. Rüb (Eds.), *Decision-making under ambiguity and time constraints: Assessing the Multiple-Streams Framework* (pp. 91–108). ECPR Press.

Boekholt, P. (2010). The evolution of innovation paradigms and their influence on research, technological development and innovation policy instruments. In R. E. Smits, S. Kuhlmann, & P. Shapira (Eds.), *The theory and practice of innovation policy*. Edward Elgar Publishing.

Borrás, S., & Edler, J. (2020). The roles of the state in the governance of socio-technical systems' transformation. *Research Policy*, *49*(5), 103971. https://doi.org/10.1016/j.respol.2020.103971

Bouncken, R. B., Qiu, Y., & García, F. J. S. (2021). Flexible pattern matching approach: Suggestions for augmenting theory evolvement. *Technological Forecasting and Social Change*, *167*, 120685. https://doi.org/10.1016/j.techfore.2021.120685

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

Cairney, P., & Jones, M. D. (2016). Kingdon's Multiple Streams Approach: What is the empirical impact of this universal theory? *Policy Studies Journal*, *44*(1), 37–58. https://doi.org/10.1111/psj.12111

Chuan, C.-H., Tsai, W.-H. S., & Cho, S. Y. (2019). Framing artificial intelligence in american newspapers. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 339–344. https://doi.org/10.1145/3306618.3314285

CIFAR. (2020). *Building an AI world: Report on national and regional AI strategies, second edition*. CIFAR. Toronto, Canada. https://cifar.ca/wp-content/uploads/2020/10/building-an-ai-world-second-edition.pdf

Cohen, M. D., March, J. G., & Olsen, J. P. (1972). A garbage can model of organizational choice. *Administrative Science Quarterly*, *17*(1), 1–25. https://doi.org/10.2307/2392088

Crockett, K., Colyer, E., & Latham, A. (2021). The ethical landscape of data and artificial intelligence: Citizen perspectives. *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, 1–9. https://doi.org/10.1109/SSCI50451.2021.9660153

Dahl, R. A. (1978). Pluralism revisited. *Comparative Politics*, *10*(2), 191–203. https://doi.org/10.2307/421645

Daigneault, P.-M. (2014). Reassessing the concept of policy paradigm: Aligning ontology and methodology in policy studies. *Journal of European Public Policy*, *21*(3), 453–469. https://doi.org/10.1080/13501763.2013.834071

DeLeo, R. A., Taylor, K., Crow, D. A., & Birkland, T. A. (2021). During disaster: Refining the concept of focusing events to better explain long-duration crises. *International Review of Public Policy*, *3*(1). https://doi.org/10.4000/irpp.1868

de Saille, S. (2015). Innovating innovation policy: The emergence of 'Responsible Research and Innovation'. *Journal of Responsible Innovation*, *2*(2), 152–168. https://doi.org/10.1080/23299460.2015.1045280

Dierecks, G., Larsen, H., & Steward, F. (2019). Transformative innovation policy: Addressing variety in an emerging policy paradigm. *Research Policy*, *48*(4), 880–894. https://doi.org/10.1016/j.respol.2018.10.028

Edler, J., & Fagerberg, J. (2017). Innovation policy: What, why, and how. *Oxford Review of Economic Policy*, *33*(1), 2–23. https://doi.org/10.1093/oxrep/grx001

Erkkilä, T. (working paper). Global indicators and AI policy: Metrics, policy scripts and narratives.

Erman, E., & Furendal, M. (2022). The global governance of artificial intelligence: Some normative concerns. *Moral Philosophy and Politics*. https://doi.org/10.1515/mopp-2020-0046

Federal Trade Commission. (2021). Aiming for truth, fairness, and equity in your company's use of AI. https://www.ftc.gov/news-events/blogs/business-blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai

Fischer, S.-C., Leung, J., Anderljung, M., O'Keefe, C., Torges, S., Khan, S. M., Garfinkel, B., & Allan Dafoe. (2021). *AI policy levers: A review of the U.S. government's tools to shape AI research, development, and deployment*. Centre for the Governance of AI, Future of Humanity Institute, University of Oxford. Oxford, England. https://www.fhi.ox.ac.uk/wp-content/uploads/2021/03/AI-Policy-Levers-A-Review-of-the-U.S.-Governments-tools-to-shape-AI-research-development-and-deployment-%E2%80%93-Fischer-et-al.pdf

Fletcher, A. J. (2017). Applying critical realism in qualitative research: Methodology meets method. *International Journal of Social Research Methodology*, *20*(2), 181–194. https://doi.org/10.1080/13645579.2016.1144401

Freeman, R., & Maybin, J. (2011). Documents, practices and policy. *Evidence & Policy: A Journal of Research, Debate and Practice*, *7*(2), 155–170. https://doi.org/10.1332/174426411X579207

Galanos, V. (2019). Exploring expanding expertise: Artificial intelligence as an existential threat and the role of prestigious commentators, 2014–2018. *Technology Analysis & Strategic Management*, *31*(4), 421–432. https://doi.org/10.1080/09537325.2018.1518521

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia Methodology. *Organizational Research Methods*, *16*(1), 15–31. https://doi.org/10.1177/1094428112452151

Goyal, N., Howlett, M., & Taeihagh, A. (2021). Why and how does the regulation of emerging technologies occur? Explaining the adoption of the EU general data protection regulation using the multiple streams framework. *Regulation & Governance*, *15*(4), 1020–1034. https://doi.org/10.1111/rego.12387

Greer, S. (2016). John W. Kingdon, agendas, alternatives, and public policies. In M. Lodge, E. C. Page, & S. J. Balla (Eds.), *The Oxford handbook of classics in public policy and administration*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199646135.013.18

Grillitsch, M., Hansen, T., & Madsen, S. (2018). *How novel is transformative innovation policy?* (No. 2020/08). Centre for Innovation, Research and Competence in the Learning Economy (CIRCLE). Lund, Sweden.

Hall, P. A. (1993). Policy paradigms, social learning, and the state: The case of economic policymaking in Britain. *Comparative Politics*, *25*(3), 275–296. https://doi.org/10.2307/422246

Herweg, N. (2016). Clarifying the concept of policy-communities in the Multiple-Streams Framework. In R. Zohlnhöfer & F. W. Rüb (Eds.), *Decision-making under ambiguity and time constraints: Assessing the Multiple-Streams Framework*. ECPR Press.

Herweg, N., Huß, C., & Zohlnhöfer, R. (2015). Straightening the three streams: Theorising extensions of the Multiple Streams Framework. *European Journal of Political Research*, *54*(3), 435–449. https://doi.org/10.1111/1475-6765.12089

Herweg, N., Zahariadis, N., & Zohlnhöfer, R. (2017). The Multiple Streams Framework: Foundations, refinements, and empirical applications. In C. M. Weible & P. A. Sabatier (Eds.), *Theories of the policy process* (Fourth edition). Westview Press.

Hoddy, E. T. (2019). Critical realism in empirical research: Employing techniques from grounded theory methodology. *International Journal of Social Research Methodology*, *22*(1), 111–124. https://doi.org/10.1080/13645579.2018.1503400

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Huber-Stearns, H. R., Schultz, C., & Cheng, A. S. (2019). A multiple streams analysis of institutional innovation in forest watershed governance. *Review of Policy Research*, *36*(6), 781–804. https://doi.org/10.1111/ropr.12359

Imbrie, A., Gelles, R., Dunham, J., & Aiken, C. (2021). *Contending frames: Evaluating rhetorical dynamics in AI*. Center for Security and Emerging Technology. Washington, D.C. https://doi.org/10.51593/20210010

Jones, M. D., Peterson, H. L., Pierce, J. J., Herweg, N., Bernal, A., Raney, H. L., & Zahariadis, N. (2016). A river runs through it: A Multiple Streams meta-review. *Policy Studies Journal*, *44*(1), 13–36. https://doi.org/10.1111/psj.12115

Justo-Hanani, R. (2022). The politics of artificial intelligence regulation and governance reform in the European Union. *Policy Sciences*, *55*(1), 137–159. https://doi.org/10.1007/s11077-022-09452-8

Kim, J. (working paper). Traveling imaginary of artificial optimism and global constitutionalism: A comparison between South Korea and France.

Kingdon, J. W. (1984). *Agendas, alternatives, and public policies* (First edition). Little, Brown.

Knaggård, Å. (2015). The Multiple Streams Framework and the problem broker. *European Journal of Political Research*, *54*(3), 450–465. https://doi.org/10.1111/1475-6765.12097

Köstler, L., & Ossewaarde, R. (2022). The making of AI society: AI futures frames in german political and media discourses. *AI & Society*, *37*(1), 249–263. https://doi.org/10.1007/s00146-021-01161-9

Kuhlmann, S., & Rip, A. (2018). Next-generation innovation policy and grand challenges. *Science and Public Policy*, *45*(4), 448–454. https://doi.org/10.1093/scipol/scy011

Lincoln, Y. S., & Guba, E. G. (1986). But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. *New Directions for Program Evaluation*, *1986*(30), 73–84. https://doi.org/10.1002/ev.1427

Macnaghten, P., & Chilvers, J. (2014). The future of science governance: Publics, policies, practices. *Environment and Planning C: Government and Policy*, *32*(3), 530–548. https://doi.org/10.1068/c1245j

Mazzucato, M. (2018). Mission-oriented innovation policies: Challenges and opportunities. *Industrial and Corporate Change*, *27*(5), 803–815. https://doi.org/10.1093/icc/dty034

McGregor, S. (2020). Preventing repeated real world AI failures by cataloging incidents: The AI Incident Database. *arXiv:2011.08512 [cs]*. http://arxiv.org/abs/2011.08512

Michael Dukakis Institute for Leadership and Innovation. (2020). *The AI Social Contract Index 2020*. Center for AI and Digital Policy, Michael Dukakis Institute for Leadership and Innovation.

Miles, M. B., Huberman, A. M., & Saldana, J. (2013). *Qualitative data analysis*. Sage.

Miller, H., & Stirling, R. (2019). *Government AI Readiness Index 2019*. Oxford Insights. https://www.oxfordinsights.com/ai-readiness2019

Miller, H. T. (2015). Introduction to the symposium, part 2: Interrogating neutral public administration. *Administrative Theory & Praxis*, *37*(4), 223–226. https://doi.org/10.1080/10841806.2015.1083821

Mintrom, M., & Norman, P. (2009). Policy entrepreneurship and policy change. *Policy Studies Journal*, *37*(4), 649–667. https://doi.org/10.1111/j.1541-0072.2009.00329.x

Morlacchi, P., & Martin, B. R. (2009). Emerging challenges for science, technology and innovation policy research: A reflexive overview. *Research Policy*, *38*(4), 571–582. https://doi.org/10.1016/j.respol.2009.01.021

Morley, J., Kinsey, L., Elhalal, A., Garcia, F., Ziosi, M., & Floridi, L. (2021). Operationalising AI ethics: Barriers, enablers and next steps. *AI & Society*. https://doi.org/10.1007/s00146-021-01308-8

National AI Initiative Office, Office of Science and Technology Policy. (2022). National Artificial Intelligence Initiative documents. https://www.ai.gov/documents/

National Institute of Standards and Technology. (2021). *Draft - taxonomy of AI risk*. National Institute of Standards and Technology. Washington, D.C. https://www.nist.gov/system/files/documents/2021/10/15/AI%20RMF_RFI%20Summary%20Report.pdf

Neri, H., & Cozman, F. (2020). The role of experts in the public perception of risk of artificial intelligence. *AI & Society*, *35*(3), 663–673. https://doi.org/10.1007/s00146-019-00924-9

OECD. (2022). OECD AI's live repository of over 260 AI strategies & policies. https://www.oecd.ai/dashboards

OECD. (2021). *State of implementation of the OECD AI principles: Insights from national AI policies* (No. 311). OECD. Paris, France. https://doi.org/10.1787/1cd40c44-en

Office of Management and Budget. (2020). *Memorandum for the heads of executive departments and agencies: Guidance for regulation of artificial intelligence applications* (OMB-2020-0003-0001). Office of Management and Budget. Washington, D.C. https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf

Ouchchy, L., Coin, A., & Dubljević, V. (2020). AI in the headlines: The portrayal of the ethical issues of artificial intelligence in the media. *AI & Society*, *35*(4), 927–936. https://doi.org/10.1007/s00146-020-00965-5

Owen, R., Stilgoe, J., Macnaghten, P., Gorman, M., Fisher, E., & Guston, D. (2013). A framework for responsible innovation. *Responsible Innovation* (pp. 27–50). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118551424.ch2

Perrault, R., Shoham, Y., Brynjolfsson, E., Clark, J., Etchemendy, J., Grosz, B., Lyons, T., Manyika, J., Mishra, S., & Niebles, J. C. (2019). *Artificial Intelligence Index 2019 annual report*. Human-Centered AI Institute, Stanford University. Stanford, CA. https://hai.stanford.edu/sites/g/files/sbiybj10986/f/ai_index_2019_report.pdf

Perry, B., & Uuk, R. (2019). AI governance and the policymaking process: Key considerations for reducing AI risk. *Big Data and Cognitive Computing*, *3*(2), 26. https://doi.org/10.3390/bdcc3020026

Pierre, J. (Ed.). (2000). *Debating governance: Authority, steering, and democracy*. OUP Oxford.

Princen, S., & 't Hart, P. (2014). Putting policy paradigms in their place. *Journal of European Public Policy*, *21*(3), 470–474. https://doi.org/10.1080/13501763.2013.876177

Roberts, N. C., & King, P. J. (1991). Policy entrepreneurs: Their activity structure and function in the policy process. *Journal of Public Administration Research and Theory: J-PART*, *1*(2), 147–175. https://www.jstor.org/stable/1181732

Rowe, G., & Frewer, L. J. (2000). Public participation methods: A framework for evaluation. *Science, Technology, & Human Values*, *25*(1), 3–29. https://doi.org/10.1177/016224390002500101

Schiff, D. (2021). Education for AI, not AI for education: The role of education and ethics in national AI policy strategies. *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-021-00270-2

Schiff, D., Ayesh, A., Musikanski, L., & Havens, J. C. (2020). IEEE 7010: A new standard for assessing the well-being implications of artificial intelligence. *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2746–2753. https://doi.org/10.1109/SMC42975.2020.9283454

Schiff, D., Borenstein, J., Laas, K., & Biddle, J. (2021). AI ethics in the public, private, and NGO sectors: A review of a global document collection. *IEEE Transactions on Technology and Society*, *2*(1), 31–42. https://doi.org/10.1109/TTS.2021.3052127

Schot, J., & Steinmueller, W. E. (2018). Three frames for innovation policy: R&D, systems of innovation and transformative change. *Research Policy*, *47*(9), 1554–1567. https://doi.org/10.1016/j.respol.2018.08.011

Shearer, E., Stirling, R., & Pasquarelli, W. (2020). *Government AI Readiness Index 2020*. Oxford Insights. https://www.oxfordinsights.com/ai-readiness2019

Sinkovics, N. (2018). Pattern matching in qualitative analysis. *The SAGE handbook of qualitative business and management research methods: Methods and challenges* (pp. 468–484). SAGE Publications Ltd. https://doi.org/10.4135/9781526430236.n28

Smits, R., & Kuhlmann, S. (2004). The rise of systemic instruments in innovation policy. *International Journal of Foresight and Innovation Policy*, *1*(1/2), 4. https://doi.org/10.1504/IJFIP.2004.004621

Soete, L. (2007). From industrial to innovation policy. *Journal of Industry, Competition and Trade*, *7*(3-4), 273–284. https://doi.org/10.1007/s10842-007-0019-5

Stanford Institute for Human-Centered Artificial Intelligence. (2021). Global AI Vibrancy Tool. https://aiindex.stanford.edu/vibrancy/

Stark, L., Greene, D., & Hoffmann, A. L. (2021). Critical perspectives on governance mechanisms for AI/ML systems. In J. Roberge & M. Castelle (Eds.), *The cultural life of machine learning: An incursion into critical AI studies* (pp. 257–280). Springer International Publishing. https://doi.org/10.1007/978-3-030-56286-1_9

Stirling, A. (2008). “Opening up” and “closing down”: Power, participation, and pluralism in the social appraisal of technology. *Science, Technology, & Human Values*, *33*(2), 262–294. https://doi.org/10.1177/0162243907311265

Stirling, A. (2016). Addressing scarcities in responsible innovation. *Journal of Responsible Innovation*, *3*(3), 274–281. https://doi.org/10.1080/23299460.2016.1258946

Stivers, C. (2015). Rule by nobody: Bureaucratic neutrality as secular theodicy. *Administrative Theory & Praxis*, *37*(4), 242–251. https://doi.org/10.1080/10841806.2015.1083825

Stix, C. (2021). Actionable principles for artificial intelligence policy: Three pathways. *Science and Engineering Ethics*, *27*(1), 15. https://doi.org/10.1007/s11948-020-00277-3

Stix, C., & Maas, M. (2021). Bridging the gap: The case for an ‘incompletely theorized agreement’ on AI policy. *AI and Ethics*. https://doi.org/10.1007/s43681-020-00037-w

Taeihagh, A. (2021). Governance of artificial intelligence. *Policy and Society*, *40*(2), 137–157. https://doi.org/10.1080/14494035.2021.1928377

Tarnoff, B., Redwine, C., Tan, J., Sheets, K., Nedzhvetskaya, N., & Rajasekaran, S. (2021). Collective action in tech. https://data.collectiveaction.tech/

Tortoise. (2021). The Global AI Index. https://www.tortoisemedia.com/intelligence/global-ai/

Ulnicane, I. (2016). 'Grand challenges' concept: A return of the 'big ideas' in science, technology and innovation policy? *International Journal of Foresight and Innovation Policy*, *11*(1-3), 5–21. https://doi.org/10.1504/IJFIP.2016.078378

Ulnicane, I., Eke, D. O., Knight, W., Ogoh, G., & Stahl, B. C. (2021). Good governance as a response to discontents? Déjà vu, or lessons for AI from other emerging technologies. *Interdisciplinary Science Reviews*, *46*(1-2), 71–93. https://doi.org/10.1080/03080188.2020.1840220

Ulnicane, I., Knight, W., Leach, T., Stahl, B. C., & Wanjiku, W.-G. (2020). Framing governance for a contested emerging technology: Insights from AI policy. *Policy and Society*, *0*(0), 1–20. https://doi.org/10.1080/14494035.2020.1855800

Ulnicane, I., Knight, W., Leach, T., Stahl, B. C., & Wanjiku, W.-G. (2022). Governance of artificial intelligence: Emerging international trends and policy frames. *The global politics of artificial intelligence*. Chapman and Hall/CRC.

United States Innovation and Competition Act of 2021, S. 1260, 117th Cong. (2021). https://www.congress.gov/bill/117th-congress/senate-bill/1260

Vesnic-Alujevic, L., Nascimento, S., & Pólvora, A. (2020). Societal and ethical impacts of artificial intelligence: Critical notes on European policy frameworks. *Telecommunications Policy*, *44*(6), 101961. https://doi.org/10.1016/j.telpol.2020.101961

von Ingersleben, N. (working paper). Competition and cooperation in artificial intelligence standard-setting: Prospects for the emergence of global standards.

Vu, K., Hanafizadeh, P., & Bohlin, E. (2020). ICT as a driver of economic growth: A survey of the literature and directions for future research. *Telecommunications Policy*, *44*(2), 101922. https://doi.org/10.1016/j.telpol.2020.101922

Waelen, R. (2022). Why AI ethics is a critical theory. *Philosophy & Technology*, *35*(1), 9. https://doi.org/10.1007/s13347-022-00507-5

Warnke, P., Koschatzky, K., Dönitz, E., Zenker, A., Stahlecker, T., Som, O., Cuhls, K., & Güth, S. (2016). *Opening up the innovation system framework towards new actors* *and institutions* (Working Paper No. 49). Fraunhofer ISI Discussion Papers - Innovation Systems and Policy Analysis. https://www.econstor.eu/handle/10419/129191

Weber, M. (1949). *The methodology of the social sciences* (First edition). Free Press.

White, M. D., & Marsh, E. E. (2006). Content analysis: A flexible methodology. *Library Trends*, *55*(1), 22–45. https://doi.org/10.1353/lib.2006.0053

Wong, R. Y., Madaio, M. A., & Merrill, N. (2022). Seeing like a toolkit: How toolkits envision the work of AI ethics. *arXiv:2202.08792 [cs]*. http://arxiv.org/abs/2202.08792

Wyman, O. (2020). Global Cities AI Readiness Index. https://www.oliverwymanforum.com/city-readiness/global-cities-ai-readiness-index-2019.html

Wynn, D., & Williams, C. K. (2012). Principles for conducting critical realist case study research in information systems. *MIS Quarterly*, *36*(3), 787–810. https://doi.org/10.2307/41703481

Yin, R. K. (2018). *Case study research and applications: Design and methods* (Sixth edition). SAGE.

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

CHAPTER 3
IS TECHNOLOGY GOVERNANCE CHANGING? FRAMING CONTESTATION AND PUBLIC PARTICIPATION IN U.S. AI POLICY

3.1 Introduction

Technology governance is at a crossroads. Given the actual or perceived role of advanced information technology in economic transformation and vitality, emerging technologies have increasingly become pillars of national innovation strategy (Branscomb, 1992; Soete, 2007; Vu et al., 2020). Yet simultaneously, recognition of the power accruing to industrial and national technology leaders has inspired arguably unprecedented calls for attention to technology's social and ethical risks. This is nowhere more true than with respect to artificial intelligence (AI), a strategically important general purpose technology often construed as the pillar of 21st century innovation (Schwab, 2016), but marked by substantial alarm about a wide array of potential social and ethical harms (D. Schiff et al., 2021), and situated in the middle of a global technology race (Ulnicane, 2022).

In the context of these tensions, a multisectoral consensus has emphasize the need for the public to exert agency in shaping the emerging AI policy agenda (Buhmann & Fieseler, 2022). This call for diverse participation in what might otherwise be an expert-dominated policy domain also aligns with newfound attention to notions of shared governance (Minkkinen et al., 2022; Pierre, 2000). Indeed, while the bulk of AI research and development occurs in the private sector, policy actors across the public, private, and non-governmental sectors agree that members of the public should play a critical role in weighing the benefits and risks of AI and informing ultimate decision-making (Crockett et al., 2021). In particular, the public is argued to have a special stake in informing government and industry actors about their concerns and tolerance with respect to social and ethical risks of AI. How AI governance unfolds, then, may also depend on how AI is understood—or framed—as a policy issue (Imbrie et al., 2021).

Yet, notwithstanding this apparent normative consensus, what is less clear is whether the public actually has meaningful opportunities to shape the AI policy agenda. In turn, this chapter seeks to answer the following in the context of the United States (U.S.), a global leader in AI research and development and key policy actor:

- *Does the public play a substantive role in influencing policymakers' attention to AI, despite the fact that technology policy is traditionally expert-dominated and not highly salient to the public?*
- *Given the public's special stake in social and ethical issues, is the public especially influential in shaping policymaker attention when AI is framed in terms of its social and ethical implications?*

To assess these questions, I create datasets aimed at capturing attention to several issue frames related to AI policy as invoked by the public, federal policymakers, and news media in the United States. The primary data are social media messages by members of the public and 115th and 116th Congresses, as well as articles by the New York Times, all covering the period of 2017 through 2019. Using quantitative text analysis techniques, I extract actor-specific time series datasets representing attention to three issue frames: These frames address potentially competing concerns about AI's prospects for innovation, social and ethical dimensions, and implications for geopolitical competition. Next, to analyze whether the public shapes policymaker attention to AI in general or only with respect to certain issues frames, I apply time series methods including autoregressive integrated moving average (ARIMA) and vector autoregression (VAR) modeling.

Results, consistent across a variety of specifications, indicate that the public has a heightened but ultimately limited role in shaping policymaker attention with respect to the emerging AI policy agenda. Notably, public attention to AI *does* predict policymaker attention to AI, and no other relationship between members of the public, media, and policymakers exhibits this degree of influence. Yet while the public does lead policymaker attention to AI, this ostensible influence only occurs when the public discusses AI generally or with respect to its role in economic innovation. That is, *the public plays no special role in driving attention to AI's social and ethical implications*, potentially challenging assumptions about the public's priorities or the nature of policymakers' responsiveness to them. Relatedly, the innovation frame is the most dominant frame through which AI is understood, though there is a striking increase in attention to AI's ethical implications amongst policymakers, and relatively little emphasis on AI's geopolitical dimensions over the study's time period.

Overall, the study contributes by advancing the use of text analysis and time series methods to assess questions in policy process theory and innovation policy, particularly related to the emergence of of the U.S. AI policy agenda. Through empirical study of *sub-issue* attention to issue frames and the role framing contestation plays in agenda-setting, this research advances the understanding of agenda-setting dynamics for strategic and emerging technology policy domains, and provides key insight into whether the public plays a meaningful role in technology governance. The results reveal that the public does matter, but perhaps not in the way expected or imagined.

3.2 Theoretical Approach

3.2.1 New Directions in Technology Governance

The governance of science and technology may be changing in the 21st century. Limitations of traditional science and technology policy, such as its failure to solve major societal challenges like inequality and climate change (Ulnicane, 2016), have led scholars to propose new approaches to innovation policy altogether. For example, movements like transformative innovation policy, mission-oriented innovation policy, and Responsible Research and Innovation (de Saille, 2015; Diercks et al., 2019; Grillitsch et al., 2018; Mazzucato, 2018; von Schomberg, 2013) deviate from a more economic, expert-driven, and firm-centered logic for innovation. While these approaches are evolving, remain contested, and skirt the line between normative and descriptive (Schot & Steinmueller, 2018), they share certain key features and can be jointly understood to reflect a new aspirational paradigm (Hall, 1993; Hogan & Howlett, 2015) in technology policy.[1]

In particular, such a paradigm emphasizes attention to social rather than purely economic goals for innovation (Kuhlmann & Rip, 2018). Along these lines, it acknowledges the possibility that innovation can lead to ethical and societal harms, such that a ubiquitous pro-innovation bias may be problematic. Further, it implies a more inclusive and participatory societal policy agenda, for which it is important to incorporate a broader array of actors beyond firms, academia, and government, such as civil society and especially the public (Warnke et al., 2016). Along these lines, an important question for the future of technology governance is therefore whether critical emerging policy domains like AI are in fact reflective of elements of this new paradigm, namely an increased reliance on public participation and a focus on social and ethical dimensions of technology.

3.2.2 Broadening Participation in AI Policy

The central question for this study is whether—and under what conditions—the public shapes the technology policy agenda, specifically in terms of public influence on federal policymaker discourse in U.S. AI policy. Normatively speaking, there is a substantial multi-sector and international consensus on the value of public participation in AI governance (Stix, 2021; Ulnicane et al., 2021; Vesnic-Alujevic et al., 2020). Members of the public are argued to be deeply affected by the social transformations resulting from AI, implying that they have a special stake and that policymakers need to be responsive to public concerns (Stahl, 2021). Further, members of the public may function as "citizen experts" (F. Fischer, 2000) given first-hand experience of AI-based harms, such as faulty automated decision systems used by government, labor displacement resulting from automation, and algorithmic surveillance (Eubanks, 2017; D. S. Schiff et al., 2021; Stark et al., 2021).

[1]This article draws on the idea of policy paradigms as a simple heuristic for understanding competing goals and values in technology governance. For a broader discussion of policy paradigms, including the conditions and theories surrounding paradigm change, see Carstensen (2015) and Daigneault (2014).

Of course, reflection on the importance of public participation in policy is not new and extends from a long history of political thought (Arnstein, 1969; Dahl, 1978) which is reflected in broader 21st century approaches to governance rather than top-down government alone (Pierre, 2000; Rowe & Frewer, 2000). Yet, in the context of science and technology policy, the perceived heightened risks of technological impacts and the gap between subject expert and public understanding (Dryzek & Pickering, 2017) have led to urgent calls for public participation in technology governance generally (Macnaghten & Chilvers, 2014; Stirling, 2008), and in AI policy specifically (Cihon et al., 2021; König & Wenzelburger, 2021; Stark et al., 2021; Stray, 2020). In particular, actors in the public, private, and non-governmental sectors have called for increased public scrutiny, public discourse, and participation of specialized mini-publics (Setälä & Smith, 2018) to inform public and private sector AI governance (Buhmann & Fieseler, 2022; D. Schiff et al., 2021; Tsamados et al., 2022).[2]

Yet, evidence on the willingness, ability, and capacity of policymakers to engage the public and the ultimate influence of public participation on policy agendas is decidedly mixed (Abelson & Gauvin, 2006; Lodge & Wegrich, 2015; Yackee & Yackee, 2006). For example, the Multiple Streams Framework (MSF) considers public opinion (or national mood) to be important in shaping the policy agenda (Kingdon, 1984). Yet, Barberá et al. (2019) find in a recent study of 100 policy topics that the general public does not lead issue priorities for Congress. In the AI space in particular, some have argued that calls for participation may constitute mere "participation-washing," (Sloane et al., 2020) and it remains unclear whether policymakers know how to effectively solicit public opinion even when the intention to do so is genuine. This challenge is exacerbated by value trade-offs, lack of concrete evidence of AI's implications (Whittlestone et al., 2019), and the technical

[2]For example, Ouchchy et al. (2020) find that "encouraging public involvement" is the most common recommendation made in media coverage of AI, while the Office of Management and Budget's (2020) key guidance to federal agencies on AI places public trust and public participation is its first and second pillars. Meanwhile, the IEEE's Global Initiative on the Ethics of AI (2019) makes clear throughout its magnum opus document, Ethically Aligned Design, that active participation of "a diverse set of stakeholders" is critical.

complexities of AI (O'Shaughnessy et al., 2022), which may lead policymakers to doubt the ability of the public to provide useful insight.

This study contributes to scholarly efforts in policy process theory and innovation policy to understand the role of the public in agenda-setting by considering whether public attention to AI (i.e., within the public or systemic agenda) is echoed by policymakers (i.e., in the institutional or Congressional agenda) for emerging technology policy domains.[3] If the calls for public participation in AI are bearing out in practice, we should expect to see increased attention by policymakers to the concerns of the general public, diverting somewhat from recent findings (Barberá et al., 2019). Thus, a first hypothesis is that:

Public Agenda-Setting Hypothesis: *Issue attention by members of the public to AI policy generally will predict issue attention by policymakers.*

However, it may be the case that public attention only predicts policymaker attention when AI is conceived of in certain ways, i.e., with respect to specific *issue frames*. For example, policymakers may follow public attention with respect to AI's ethical and social issues, but rely more on the expertise of military or economic specialists in closed-door settings when international competition and technological leadership is at stake. Further, given the potentially increased salience of ethical and social issues for the public and how benefits and costs of AI policy decisions are concentrated on members of the public with respect to these issues, policymakers may feel greater pressure to engage with and be responsive to public pressure here. To explore these dynamics, I introduce issue framing below as a crucial facet of the agenda-setting process.

[3] A note of causal caution is prudent. The methods in this study may only be able to nod towards causation in a limited fashion; in that sense, public-policy issue attention correlation can be understood in terms of "issue congruence" (Jones & Baumgartner, 2004) rather than strict causation.

3.2.3 Contested Issue Frames in AI Agenda-Setting Discourse

At their core, issue frames attempt to capture an ongoing discourse and interpret events and indicators in a way that packages multifaceted issues in terms of a more simplified essence (Chong & Druckman, 2007; Gamson & Modigliani, 1989).[4] In this way, frames structure the underlying categories associated with a topic, strategically emphasizing or excluding certain elements, in order to constrain or expand reasoning in preferred directions (Baumgartner & Jones, 1993; Sharp, 1994). Powerful and effective frames consequently may direct both public and elite attention, legitimizing certain policy narratives and bringing particular policy solutions into the mainstream (Goddard & Krebs, 2015). At one level then, the impact of frames on agenda formation is straightforward: When specific issue frames become dominant, there is a greater chance that policymakers will respond to these signals by putting specific issues and solutions to them on the institutional agenda (Peters & Hogwood, 1985). This channel of influence, recognized in accounts of democratic legitimacy theory as well as in pluralist theory (Dahl, 1978; Peter, 2008), is based on policymaker responsiveness to public pressure and public willingness to hold politicians accountable.[5]

In the context of AI policy, this study examines three issue frames prominent in AI discourse. The first frame is the "innovation frame." In line with traditional technology and innovation policy, this frame emphasizes AI's potential for economic transformation in terms of fostering industrial productivity, economic efficiency gains, high-tech growth acceleration, entrepreneurial activity, and national GDP expansion (Edler & Fagerberg, 2017; Soete, 2007). Advocates of this issue frame for AI typically emphasize the need for supply-side reforms, such as education and training efforts to increase the size of the AI workforce, increased research and development funding, and regional and national

[4]Issue framing can be understood as closely connected to, but broader than, problem definition in that it involves defining both problems and solutions and potentially coupling them.

[5]Note that 'public' may be construed broadly, as public opinion may be channeled through media, interest groups, civil society, and even private sector actors.

industry-government-university coordination (Atkinson & Mayo, 2010; S.-C. Fischer et al., 2021; Smits & Kuhlmann, 2004). Given the perception that AI and automation may be the driving technologies in the Fourth Industrial Revolution, national and international AI policy strategy documents indeed manifest a substantial emphasis on AI's innovative potential (D. Schiff, 2021; Ulnicane et al., 2020). For example, the United States Innovation and Competition Act (2021), American AI Initiative (2019), and National AI Research and Development Strategic Plan (2016) deeply embody this thinking. Unsurprisingly then, this frame is an explicit object of scholarly attention, termed by Imbrie et al. (2020, p. 10) as the "economic gold rush" frame.

A second "often invoked" frame emphasizes social and ethical implications of AI—an "ethics frame" (Ulnicane et al., 2020, p. 2). In particular, a large body of work by civil society, epistemic communities of academic experts and practitioners, and the private sector has reviewed numerous ethical problems associated with AI (Fjeld et al., 2020; Morley et al., 2019; Shneiderman, 2022) surrounding AI's implications for inequality, racial and gender bias, transparency and accountability to the public, human rights, safety, and more (D. Schiff et al., 2021). In turn, proposed solutions include technical "fixes" to algorithms and organizational processes to address issues like privacy and bias (Mökander et al., 2021; Ryan & Stahl, 2020), as well as policy-level solutions such as banning harmful applications like facial recognition (Access Now, 2021) or requiring impact assessments to assess risks and harms (Salgado-Criado & Fernández-Aller, 2021; D. Schiff et al., 2020). References to social and ethical implications are now common in policy discourse (D. Schiff et al., 2021). Proponents of this frame are likely more prone to advocate for stricter regulation and precaution in policymaking. Yet, there is uncertainty and debate about whether this frame is likely to influence policy in a meaningful way (Hickok, 2021; Leung, 2020; Morley et al., 2021; Taeihagh, 2021) given numerous barriers to operationalizing AI ethics and concerns about private and public sector "ethics-washing" and "ethics shirking" (Bietti, 2020; de Laat, 2021; Rességuier & Rodrigues, 2020).

The third frame is the "competition frame."[6] This frame, from the perspective of the United States, largely highlights the "significant economic and national security threat to the United States" of losing the AI contest to China (Future of Defense Task Force, 2020, p. 5). There is a remarkable bipartisan consensus on this issue, with entities like the Future of Defense Task Force, the Office of Science and Technology Policy, the Congressional AI Caucus, and the National Security Commission on AI in its key report (2021) expressing support for the competition frame in their headline statements. Indeed, Stix and Maas (2021, p. 3) note that "public and global framings of AI in recent years have seemed to drift towards narratives of competition and 'arms races'" and Imbrie et al. (2020, p. 8) argue that "the competition frame has diffused widely in public discourse and become shorthand for understanding the larger geopolitical context of investments in AI." This frame for AI is now manifested clearly in the U.S. policy agenda, for example, via the U.S. Congress's Innovation and Competition Act (2021) which mentions China hundreds of times. However, there is also substantial pushback to this approach to framing in AI policy in acknowledgment of the harmful effects of great powers competition discourse and arms races in the past (Ulnicane, 2022; Zwetsloot et al., 2018). As Franke (2021, p. 13) notes, the arms race narrative has "become so commonplace that a whole academic subfield has emerged to fight this framing." Figure 3.1 provides examples of the three AI issue frames, as invoked by U.S. federal legislators.

Regarding the relative importance of these three frames, there are some reasons to believe that the ethics frame will be less influential than arguably more traditional frames

[6]The choice of frames is, to some extent, subjective and constrained by which audiences and time period are considered (Fast & Horvitz, 2017). For example, the Center for Security and Emerging Technology considers four frames including a "World without Work" frame and a "Killer Robots" frame (Imbrie et al., 2021) and the Department of Homeland Security (2017) considers seven benefit- and threat-focused frames including "Threat to Humanity" and "Fueling the Surveillance Machine" frames. The choice of frames for this analysis is based on review of scholarly and policy literature along with expert opinion and a desire for parsimony. Approaches like discourse analysis and topic modeling can be fruitfully applied to study coverage of AI in media and other sources by those who wish to identify, combine, and parse frames in different ways (Cave et al., 2018). Yet, Neuman et al. (1992) identify human impact, economics, and conflict as three common issue frames used in the news media; arguably, then, the three frames pursued in this study have universal relevance.

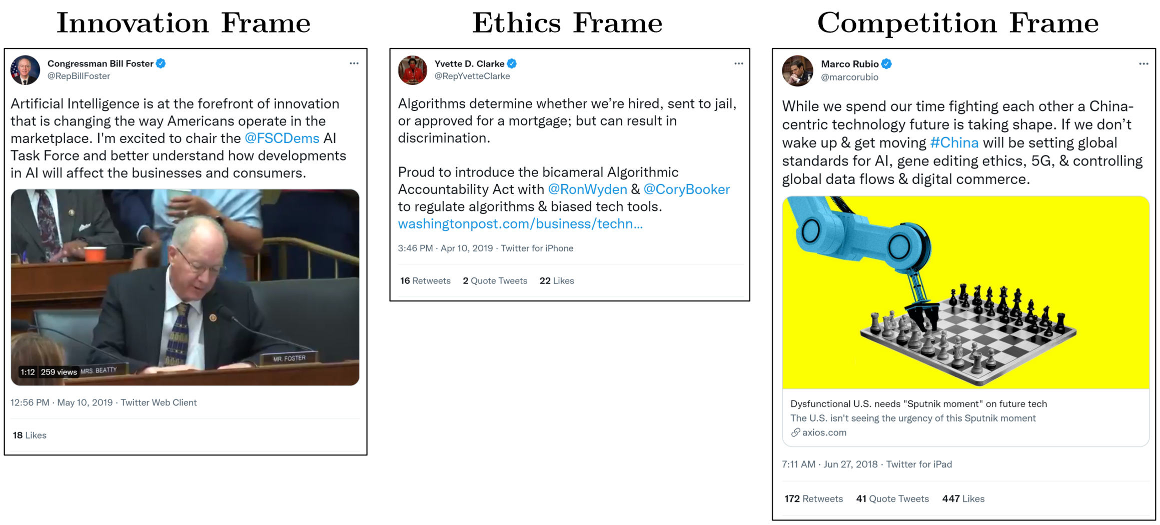

Figure 3.1: Examples of AI issue frames in U.S. federal policy discourse

invoking economic and security considerations. For instance, Gilardi, Shipan, et al. (2021) find that normative frames are less likely to predict policy adoption, while more 'concrete' frames take their place increasingly over time. Moreover, economic and national security considerations typically rise to the top of public issue priorities (Jones et al., 2009). This in part results from—and reflects—the high levels of status and government access that actors focused on innovation and national security enjoy, compared to, for example, civil society groups.

Yet, the prevalence of the ethics frame in AI may be unprecedented in technology policy: Advances such as aerospace technology and cryptography received minimal ethical and societal scrutiny in the past (Leung, 2020), whereas the urgent attention to AI ethics across public, private, and NGO sectors may surpass even ethical scrutiny applied to more publicly prominent areas like biotechnology and nuclear policy. Indeed, according to the European Commission (2021, p. 8), ethics in AI is a "widespread and common approach, as evidenced by a plethora of ethical codes and principles developed by many private and public organisations...that AI development and use should be guided by certain essential value-oriented principles." In sum, these considerations offer competing expectations with respect to the dominance of competing AI issue frames and their dynamics over time:

Ethical Framing Hypothesis: *Compared to the innovation and competition frames, the ethics frame will find more purchase with the public, media, and policymakers, and is likely to become more dominant over time.*

Traditional Framing Hypothesis: *Compared to the ethics frame, the innovation and competition frames will find more purchase with the public, media, and policymakers, and are likely to become more dominant over time.*

3.2.4 Mutual Interaction of Public Participation and Framing Dynamics

The ascendance of issue frames that capture social and ethical considerations would provide some evidence for transformation in technology governance, as would increased attention to public concerns. *Yet it is the combination of the two that is arguably most indicative of a potential policy paradigm shift*. In particular, it is not through sheer volume alone that policymakers are persuaded to change their policy preferences. Instead, frames may alter which actors have a stake in, expertise relevant to, or even formal decision-making authority with respect to certain issues, as recognized in the literature on policy typologies (Wilson, 1973) and the adage that "policies determine politics" (Schattschneider, 1935).

Particularly complex issues with low public salience admit to "board room politics" (Gormley, 1986), where citizens are disconnected from policy discussions while bureaucratic elites and regulatory capture characterize decision-making. In light of this, Lowi (1964, p. 707) recognized that "one of the most important strategies in any controversial issue is to attempt to define it in redistributive terms in order to broaden the base of opposition or support." Thus, issue framing can be used strategically to contain or expand issues and alter the actors that have influence, potentially facilitating or diminishing the influence of the public (Daviter, 2011; Steinberger, 1980).

In sum, if technology governance is shifting, we might expect that the increased risk aversion and precaution members of the public attach to emerging technologies (Dryzek et al., 2009; Sturgis & Allum, 2004) is perceived as increasingly legitimate and worthy of serious consideration (de Saille, 2015). Given wide calls for public participation in AI policy and evidence that the public is more skeptical of the use of AI than experts (O'Shaughnessy et al., 2022), a key question is thus whether policymakers are indeed especially likely to listen to the public's social and ethical concerns regarding AI—issues that might otherwise be neglected and for which the public has arguably a greater stake and perhaps even more knowledge. If this is the case, we should expect:

Special Role of the Public Hypothesis: *Issue attention by members of the public to the ethics frame as compared to the innovation or competition frames will more strongly predict issue attention by policymakers.*

3.3 Methodology

3.3.1 Data Sources and Rationale

There are three primary sources of data used in the study. To measure public attitudes towards AI, I collect a dataset of approximately 4.9 million Twitter messages sent between January 2017 and December 2019 that reference "#AI," the standard hashtag for general AI interest. To measure policymaker discourse about AI, I collect all Twitter messages sent by members of the 115th and 116th U.S. Congresses[7] during the same time period

[7] I am able to identify Twitter accounts for 639 of 645 members of the Senate and House of Representatives—99% of those in office between 2017 and 2019—and collect tweets only during the specific terms when members were in office. As members of Congress often have multiple Twitter accounts (e.g., campaign accounts, personal accounts, official office accounts) (Siddique, 2019), I select for each person the account with the most followers, as this is the account with arguably the most importance for their messaging efforts. While 612 members of Congress tweeted approximately 1.1 million times over the period, only about half of them (291) tweeted about AI.

before extracting messages specifically about AI.[8] Finally, while the relationship between the public and policymakers is of primary concern here, I also consider the media as a meaningful third actor, treated in this paper primarily as a proxy for public opinion or national mood. To measure media attention to AI, I collect all New York Times (NYT) articles that mention the phrase "artificial intelligence"[9] over the time period using the Nexis Uni database.[10]

A few questions are worth considering in light of the data collection strategy here. First, to what extent does studying the sources above (namely social media) and actors allow for valid evaluations of issue (frame) priorities and agenda-setting influence? According to a growing body of work, these research questions are increasingly tractable. Twitter is now almost universally used by members of Congress, for purposes such as providing information, credit-claiming, and indicating responsiveness to the public, along with other goals (Golbeck et al., 2010; Shapiro et al., 2018). Research indicates that *citizens* also act strategically through these platforms to influence policy agendas (not limited to "shouting" at politicians) (Hemphill & Roback, 2014), and that policymakers in turn use social media to expand issue attention beyond elites as well (Fazekas et al., 2021). Perhaps surprisingly,

[8]To identify AI-relevant messages, I investigate a set of 100 AI keywords (e.g., neural network, machine translation) used by Imbrie et al. (2020) for a similar task, along with additional keywords identified when investigating Congress's messages (e.g., autonomous vehicles, deepfakes). For each keyword, I investigate its frequency of occurrence, and retain keywords that appear at least 10 times in the corpus and for which at least 75% of the associated messages reflect a reference to AI, to my understanding. The full dictionary used to extract AI relevant messages for the 115th and 116th Congresses is available in Appendix B.1.

[9]The choice of the different primary keywords that dictate the inclusion criteria for data collection for the three actors is based on evaluation of the coverage of those terms, in light of differences in the length, context, and style of content.

[10]The NYT is often used in media studies as a proxy for media coverage broadly as it is considered the "paper of record" in the United States: It covers a wide range of topics and is widely read by the public and political elites (Ringel, 2021; Weaver & Bimber, 2008). In the agenda-setting literature, the media plays a complex and important role (McCombs & Shaw, 1972). It may serve to synthesize societal debate and reflect overall public opinion (Hopkins et al., 2017; Ripberger, 2011), or may indeed *construct* public opinion (McCombs, 2004; Zaller, 1992). Further, media outlets are not only responsive to the needs of their customers, but are also responsive to political elites, both by reporting on their activity and providing crucial information to them (Schnell, 2001; Shoemaker & Reese, 1991; Van Aelst & Walgrave, 2016). Given the importance of public opinion and national mood in the agenda-setting literature, I consider the media here primarily as a proxy or alternative source that policymakers look to understand public opinion. Indeed, other research on AI policy has observed the prominent role of the media in presenting issue frames (Imbrie et al., 2021) and has considered media coverage in outlets like the NYT as a proxy for public opinion (Fast & Horvitz, 2017; Ouchchy et al., 2020).

nearly half of U.S. politicians engage in conversation on Twitter; about one-quarter of this engagement is directly with private citizens (Tromble, 2018).

A related question is whether issue attention on social media is reflective of issue attention more broadly. Regarding issue congruence across sources, evidence indicates that issue attention by policymakers on Twitter is indeed substantially similar to their issue priorities in other formats such as on Facebook, via press releases, and in parliament (Casas & Morar, 2019; Peeters et al., 2021). These priorities are also congruent with issue coverage in the NYT specifically, suggesting that the NYT is a useful source for studying elite-media issue attention relationships (Shapiro & Hemphill, 2017). Thus, there are good reasons to think that issue attention, issue congruence, and responsiveness between the public, policymakers, and media can be plausibly evaluated using the identified datasets, even though the platforms, contexts, and rationales for each actor's messaging efforts are somewhat distinct. In sum, there is growing evidence of the importance of social media in agenda-setting dynamics.[11] Social media may serve as a direct channel in influence efforts, and—given issue congruence across platforms—it may also serve as a meaningful *proxy* of public and policymaker opinion. This paper considers both channels to be viable. However, much remains unknown about the role of social media in agenda-setting, especially given the novelty of the platforms and the still-evolving political communication dynamics surrounding them. This paper helps to advance this literature.

A distinct question important in the context of the study is what is meant by "public." Public may refer to any actors outside of government (e.g., industry bodies and academic experts who reply to requests for public comment); it may refer to society or citizens generally; or it may refer to special subsets of the public (Hallahan, 2000). For example, various taxonomies distinguish between "strong" and "weak" publics, between "inattentive" and "attentive" or "engaged" publics (Fraser, 2021; Stoker, 2014), and scholars dispute whether there is such a thing as a general public or rather multiple mini-publics. Notably,

[11] See Barberá et al. (2019) and Gilardi, Gessler, et al. (2021) for further discussion.

many definitions focus on the shared stake in and potentially action surrounding particular social or policy issues (Barnes et al., 2003), meaning that public is often defined as explicitly policy-engaged. Along these lines, this paper considers the relevant public to be individuals on social media who have some heightened personal stake or interest in AI and are more likely to engage politically on those issues.[12]

3.3.2 Identification of Issue Frames: Text Analysis

Within each of these three datasets, I use text analysis (or quantitative content analysis) methods, primarily a dictionary approach, to identify the prevalence of issue frames (Boräng et al., 2014).[13] To identify the ethics frame, I draw on keywords from three studies used to comprehensively identify AI ethics concepts in AI conferences and journals (Prates et al., 2018), in public, private, and NGO sector AI policy and ethics documents (D. Schiff et al., 2021), and in media sources (Fjeld et al., 2020; Zhang et al., 2021). For the competition and innovation frames, I rely on keywords used by Imbrie et al. (2020) and Imbrie et al. (2021) in their extensive study of AI frames in media coverage. For all three frames, I engage in snowball searches of the tweets/articles through an iterative process to identify and evaluate other potentially applicable terms. Finally, I employ word embeddings via the word2vec family of algorithms (Mikolov et al., 2013) to identify other candidate terms.

In constructing the three dictionaries, for each possible term, I examine possible variations of the associated stem of that term to determine relevance. For example, the terms "moral" and "develop" capture ethics and innovation concepts effectively, while the terms "morale" and "developers" do not. I deem a keyword worthy of inclusion if at least 75% of the usages of it in a random sample of messages accurately reflect the relevant con-

[12] Notably, this Twitter audience is likely to skew male, be younger, more educated, live in urban areas, and is potentially more politically liberal than the general population (Barberá & Rivero, 2015; Mellon & Prosser, 2017; Mislove et al., 2011). Yet, such a public is not dissimilar from individuals more likely to contact political representatives, vote, attend hearings, etc. (Vaccari et al., 2015). That is, a public so defined is a highly relevant audience for considering political accountability, responsiveness, and agenda-setting influence.

[13] I begin with standard cleaning and pre-processing steps for text analysis, including removing duplicate tweets/articles, text segmentation, tokenization and lemmatization, removal of punctuation and numbers, case conversion, and collapsing of some n-word terms to individual tokens (Jacobi et al., 2016).

cept in AI, based on my substantive understanding of the field.[14] I perform this analysis for both the Twitter and NYT datasets, as the nature of term usage depends on the length and context of the medium (Grimmer & Stewart, 2013), and find the three dictionaries perform adequately overall compared to similar classification efforts (Ghosh & Loustaunau, 2021).[15] A message (tweet or article) that contains one or more of the frame-specific keywords is then indicated, non-exclusively, as portraying that respective frame. The full dictionaries are available in Appendix B.1 and the sample sizes for each corpus and issue frame are as displayed in Table 3.1.

Table 3.1: Counts of issue frames used by public, policymakers, and media

Dataset (2017-2019)	All AI Messages	Ethics Frame	Innovation Frame	Competition Frame
Public Tweets	4,893,581	344,802	1,257,420	203,710
Congress Tweets	1,041	310	356	127
NYT Articles	2,895	2,388	2,762	2,175

3.3.3 Measuring Agenda-Setting Influence: Time Series Analysis

After constructing the datasets, the main analysis strategy used to study between-actor influence is time series analysis, particularly autoregressive integrated moving average (ARIMA) and vector autoregression (VAR) modeling approaches. These methods offer advantages such as modeling temporal order to support stronger causal claims, and are designed to deal with associated challenges such as autoregression in time series data and residuals as well as non-stationarity and conditional heteroscedasticity (Vliegenthart &

[14]The relevant sampling thresholds vary depending on the corpus size and prevalence of the specific term.

[15]Note that this initial classification strategy only identifies term-specific true positives and false positives, but does not provide a holistic sense of false negatives or other classification criteria. As such, I hand code a subset of documents and compare accuracy against the dictionary-based classifications. I report results in Appendix B.2. Note that techniques like supervised classification could provide an alternative (Wilkerson & Casas, 2017). However, a study using a wide variety of state-of-the-art ML methods, including various doc2vec and self-attention based classifiers to specifically study AI frames in media documents, led to a maximum F1 classification score for the positive class of 0.81 for the best classifier, with precision and recall of 0.76 and 0.87 for the positive class (Ghosh & Loustaunau, 2021). For this reason and because these techniques have not been applied to study shorter Twitter content for AI issue framing, I stick with the simpler dictionary approach, which Gentzkow et al. (2019, p. 24) expect to "remain the optimal choice in many settings."

Walgrave, 2008). ARIMA methods have been recommended for use in media and communication studies (Hester & Gibson, 2003; Hollanders & Vliegenthart, 2008), and more recently in policy and political science scholarship (Boef & Keele, 2008), including in studies of agenda-setting using Twitter data specifically (Barberá et al., 2019), and in correlating media mentions with policy issue attention (Howlett, 1998).

Preparation for the ARIMA analysis includes examining the time series data and associated autocorrelation and partial autocorrelation functions. I examine the data for trends and seasonality, finding evidence of the former but not the later, and after evaluating different aggregation strategies, opt to aggregate the data weekly. Of note, there are a total of 12 time series in the analysis (three datasets x mentions for three issue frames plus total number of AI mentions). Next, I iteratively search over possible stationary, non-seasonal ARIMA(p, d, q) specifications using maximum likelihood estimation and a small sample consistent AIC to balance predictive fit against model parsimony. Results suggest that models such as ARIMA$(0, 1, 1)$ (simple exponential smoothing) and ARIMA$(0, 1, 2)$ (damped trend exponential smoothing) may be appropriate (McKenzie & Gardner, 2010).[16]

However, to extend this analysis to bivariate regressions, it is necessary to model multiple time series simultaneously. The approach used here is ordinary least squares (OLS) regression using ARIMA to model errors, and including the predictor time series as an exogenous regressor. This strategy is permissible when the two series in a given regression are co-integrated, which is confirmed in all cases by the Phillips-Ouliaris test. Thus the basic specification is OLS regression with two actors' AI issue attention represented as either the independent or dependent variable, modeling a simple contemporaneous bivariate causal relationship within the course of a week. A benefit of this approach is that it allows for simple OLS-style interpretation. If the Public Agenda-Setting Hypothesis is correct, we would thus expect an increase in public messaging about AI to predict an increase in

[16]While Barberá et al. (2019) and Gilardi, Gessler, et al. (2021) opt for a more complicated 7-day lag structure, I prefer a more parsimonious model choice.

policymaker messaging.[17] Appendix B.4 provides further details about the preparation and justification for modeling choices.

While ARIMA is particularly suitable for bidirectional relationships where the order of causation is well known, VAR analysis allows for multidirectional influence through simultaneous estimation of multiple equations. While typically applied in macroeconomic analysis and forecasting, VAR has also been recommended and applied to analyze media and agenda-setting effects where causal relationships are less obvious and endogeneity is a concern (Gilardi, Gessler, et al., 2021; Liu et al., 2011): This is particularly important in policy scholarship given that policy theory often highlights multi-actor influence and complex feedback effects (Edwards & Wood, 1999; Granato & Krause, 2000; Wolfe et al., 2013). However, while VAR analysis enables modeling of all three actors simultaneously, it emphasizes lagged rather than contemporaneous relationships, and interpretation can be more complex. I therefore follow the advice of Vliegenthart (2014) by considering the two methods as complementary approaches that may reveal different dynamics associated with issue framing and agenda-setting influence.

Importantly, coefficients cannot generally be interpreted straightforwardly for VAR. A few alternative methods are normally used. First, forecast error variance decomposition estimates indicate, for each time series, the amount of variation over time that can be attributed to its own lagged values as well as to the lagged values of the *other* endogenous variables (in this case, the other actors). Second, impulse response functions provide a graphical depiction of the movement of a response variable to a unit shock in the impulse variable. Appendix B.5 provides further details on the modeling approach and decisions for the VAR analysis.

[17] Further, we would expect the converse (policymakers influencing the public) to not be true, as this would raise concerns about endogeneity.

3.4 Results

3.4.1 Issue Frame Prominence and Trends

Figure 3.2 displays the proportion of all AI tweets (or articles) per actor that address each issue frame over the 2017 to 2019 time period. As such, it depicts the *relative attention* to each issue frame over time out of all AI messages (which need not sum to 100 percent). A few patterns are clear. For the public, policymakers, and NYT, innovation appears to be the most prevalent frame, with a consistently high and stable level of attention from the beginning of the time period. In any given week, around 20-30% of public tweets, 20-50% of Congress tweets, and more than 90% of NYT articles which address AI discuss some aspects of innovation. The consistent attention across all actors to AI's implications for innovation is highly supportive of the Traditional Framing Hypothesis, though with respect to economic rather than geopolitical dynamics.

Indeed, and in contrast, the competition frame receives the least attention overall. Yet it appears that competition-related discourse is growing amongst members of Congress, increasing steadily from around 0% to around 25% of AI mentions between 2017 and 2019. A ramp-up of geopolitical discourse about AI—including outside of the time window examined here—seems quite plausible given the centrality of US-China competition in post-2019 legislation such as the US Innovation and Competition Act (2021). That the increase in attention to geopolitical dimensions seems particularly marked amongst members of Congress may also serve as a preliminary indication that the emerging policy agenda is forming somewhat independently from public influence (potentially contradicting the Public Agenda-Setting Hypothesis). The next subsections test this more formally.

The most striking dynamics however, surround the ethics frame. Indeed, attention to the ethics frame appears to be increasing substantially over time, especially for the media and policymakers. Relative attention to the ethics frame steadily grows from around 70% of media discourse to more than 90% over the three year time period. In Congress, it increases

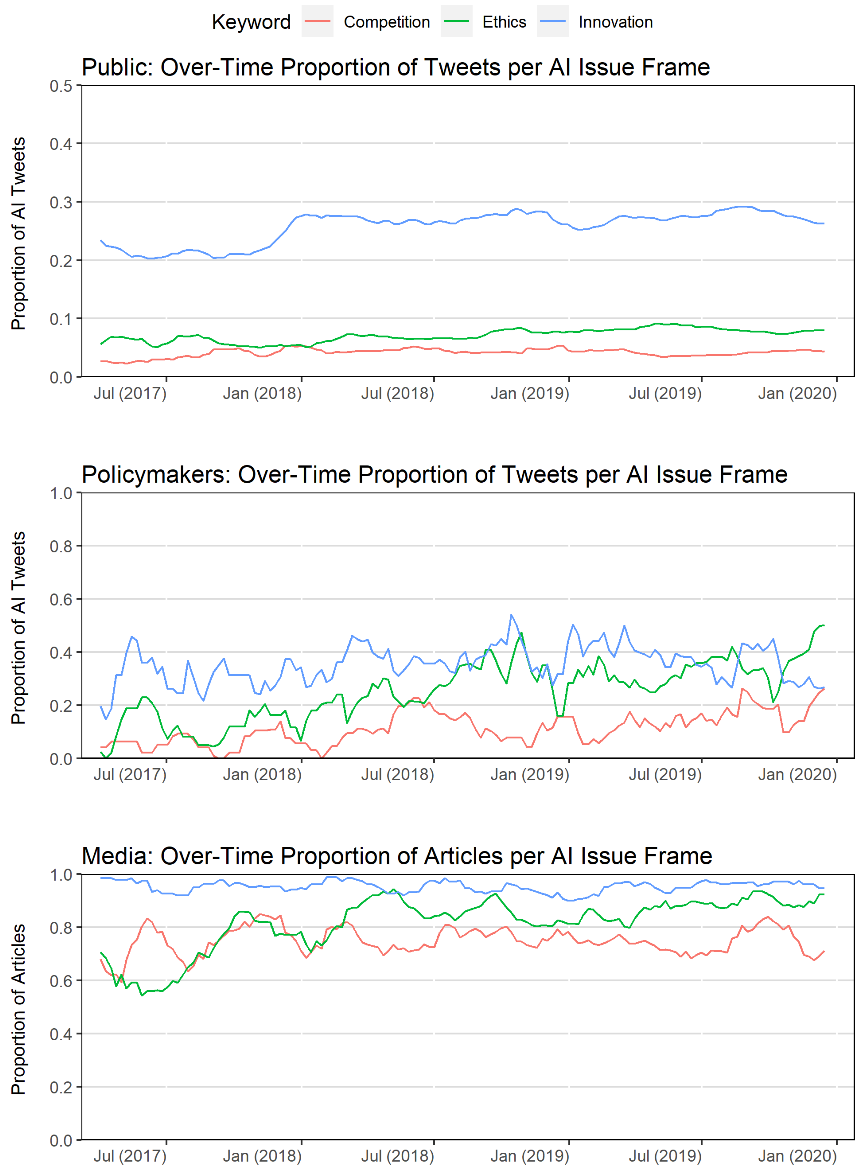

Figure 3.2: Issue frame prevalence by actor over time

Note: The data cover January 2017 through December 2019. However, I use 2-month rolling averages to stabilize trends, which means the first data points depicted are from February 2017. Sources are 1) tweets from the public using #AI, 2) tweets from the 115th and 116th Congresses on AI, and 3) NYT article coverage of 'artificial intelligence.'

from essentially 0% of AI discourse to approximately 50%, even surpassing attention to innovation frame. This constitutes meaningful evidence that, notwithstanding the absence of attention to AI ethics during the early years of AI policy agenda-setting, ethics appears to have taken on a major role. Research extending into future time periods would be necessary to confirm these trends. Yet attention to AI ethics does appear to be manifesting through legislation and other governance efforts like the Algorithmic Accountability Act (2022), the AI Bill of Rights (Rodrigo, 2021), and standards-related AI ethics efforts in the National Institute of Standards and Technology (Schwartz et al., 2022).

However, growth in Congressional attention to AI ethics does not seem to be tracking public attention, though there is modest growth in public attention to the ethics frame, from around 5-6% to 8%. The public's greater interest in AI innovation likely results from the particular composition of social media users who discuss AI broadly, who likely differ from the specialized mini-public (e.g., academia and civil society) focused on AI ethics. These results could signal an elite-public (and media-public) divide, implying that advocates of a new paradigm of technology governance may overstate or at least not closely reflect the current priorities of the public, particularly regarding social and ethical implications of AI. Alternatively these trends could equally serve as a call for action to better inform and engage the public.

Overall then, the findings offer strong support for the Traditional Framing Hypothesis, particularly with respect to innovation rather than competition during this time period. Interestingly, the results also provide some support for the Ethical Framing Hypothesis given the striking growth in the ethics frame over time, which could have meaningful impacts on the AI policy agenda. However, there are early indications that the Special Role of the Public Hypothesis may not bear out.

3.4.2 ARIMA Analysis

The ARIMA analysis involves modeling relationships between each pair of two actors separately for the three issue frames plus overall AI attention. I search through different ARIMA(p, d, q) specifications using a small sample consistent AIC to identify the best-performing model, subsetting to non-seasonal models only. Next, I perform a set of statistical and visual tests of goodness-of-fit for plausible models. This includes examining autocorrelation and partial autocorrelations of residuals along with applying the Augmented Dickey Fuller test, Kwiatkowski–Phillips–Schmidt–Shin test, and Ljung-Box test, in order to confirm stationarity of the time series and independence of residuals. On balance given these diagnostics, I identify a preferred model. I opt for simpler models over complex ones if they perform similarly, as it is preferable for no more than one of the p or q terms to be larger than 1 to avoid overfitting (Brandt & Williams, 2006).[18]

Table 3.2 displays the main ARIMA results from bivariate regressions of the three actors' issue and issue framing attention datasets. The first column indicates the actor (Y) potentially influenced by another leading actor (X). The second column indicates the selected ARIMA parameters based on information criteria and modeling diagnostics. The third column indicates the effect in terms of the *number of predicted additional tweets or articles per week* from Y that result from a one standard deviation increase in messages from X. The fourth column and fifth column assist with interpretation by presenting this change in terms of a percentage change compared to the baseline average weekly value. For example, the results indicate that a one standard deviation increase in public tweeting behavior in a given week corresponds with a 22.4% increase over the baseline average of policymaker messaging that week. The final column indicates significance, noting that some statistically significant effects are too modest to be substantively important.

To assist with interpretation, Figure 3.3 present the main patterns of influence across the three actors and four issue frames. Larger magnitude effects are represented with bolder lines, and statistically significant effects at the 0.05 level are represented with solid rather than dashed lines. *Overall, the results suggest that the public does influence policymakers, both overall and with respect to the innovation frame, leading to an approximate 20-22%*

[18]Note that this strategy means I identify several different preferred p, d, q specifications for various bivariate relationships, when it could be argued that the same underlying processes should manifest across issue frames if agenda-setting dynamics are indeed part of a stable underlying system, with the same actors and communication channels in play. As such, Appendix B.6 presents key results using alternative ARIMA and fixed effects specifications as robustness checks. The fixed effects specifications differ in considering the impact of public tweets on *individual* Congressmember behavior at the weekly level, but nevertheless lead to highly consistent results. This provides some reassurance in the robustness of the main models presented.

Table 3.2: ARIMA results: Mutual influence of public, policymakers, and media

X influence on Y	ARIMA Model	Effect Size	Baseline Value	% Change	p-value
Public on Congress					
All	(0,1,1)	1.49	6.66	22.4%	0.003
Ethics	(0,1,1)	0.26	1.98	13.1%	0.149
Innovation	(0,1,1)	0.46	2.28	20.2%	0.015
Competition	(0,1,1)	0.03	0.81	3.7%	0.733
Congress on Public					
All	(0,1,1)	753	31,322	2.4%	0.002
Ethics	(1,1,1)	60	2,205	2.8%	0.161
Innovation	(1,1,1)	136	8,046	1.7%	0.042
Competition	(1,1,1)	17	1,303	1.4%	0.694
NYT on Congress					
All	(0,1,1)	0.87	6.66	13.1%	0.024
Ethics	(1,1,3)	0.27	1.98	13.6%	0.079
Innovation	(0,1,1)	0.41	2.28	18.0%	0.012
Competition	(0,1,1)	0.03	0.81	3.7%	0.736
Congress on NYT					
All	(1,0,1)	1.57	18.56	8.5%	0.01
Ethics	(0,1,1)	0.94	15.3	6.1%	0.119
Innovation	(2,0,3)	1.59	17.71	9.0%	0.003
Competition	(0,1,1)	0.17	13.95	1.2%	0.74
NYT on Public					
All	(0,1,2)	-30	31,322	-0.1%	0.896
Ethics	(1,1,1)	42	2,205	1.9%	0.269
Innovation	(1,1,1)	24	8,046	0.3%	0.709
Competition	(1,0,2)	-12	1,303	-0.9%	0.774
Public on NYT					
All	(0,1,1)	0.17	18.56	0.9%	0.828
Ethics	(0,1,1)	0.50	15.3	3.3%	0.384
Innovation	(3,0,0)	1.18	17.71	6.7%	0.091
Competition	(0,1,1)	-0.28	13.95	-2.0%	0.594

Note: ARIMA Model refers to the preferred (p, d, q) specification. Effect Size refers to the number of additional AI messages by the influenced actor (Y) in a given week resulting from a one standard deviation increase of messaging from the influencer (X) during that week. Baseline Value refers to the average number of AI messages from the influenced actor (Y) per week. % Change refers to the increase in AI messaging over the baseline resulting from a one standard deviation increase of messaging from the influencer during that week. Statistical significant results at the 5% level are highlighted. Note that some digits are trimmed for legibility.

increase in Congressional messaging over normal baseline attention. In contrast, even when the results are technically statistically significant, Congress exerts very minor influence on the public generally, on the order of 1-3%.[19] Notably, attention by media to AI is also associated with increased attention by Congress on the order of 13-18%, indicating that this channel of influence is active, while attention by Congress leads to additional media attention (8-9%) as well. This confirms prior work recognizing the role of media in shaping and being shaped by policymaker attention (Wolfe et al., 2013; Yanovitzky, 2002) To the extent that the media is treated as a reflection of public opinion, this provides further evidence of the Public Agenda-Setting Hypothesis.

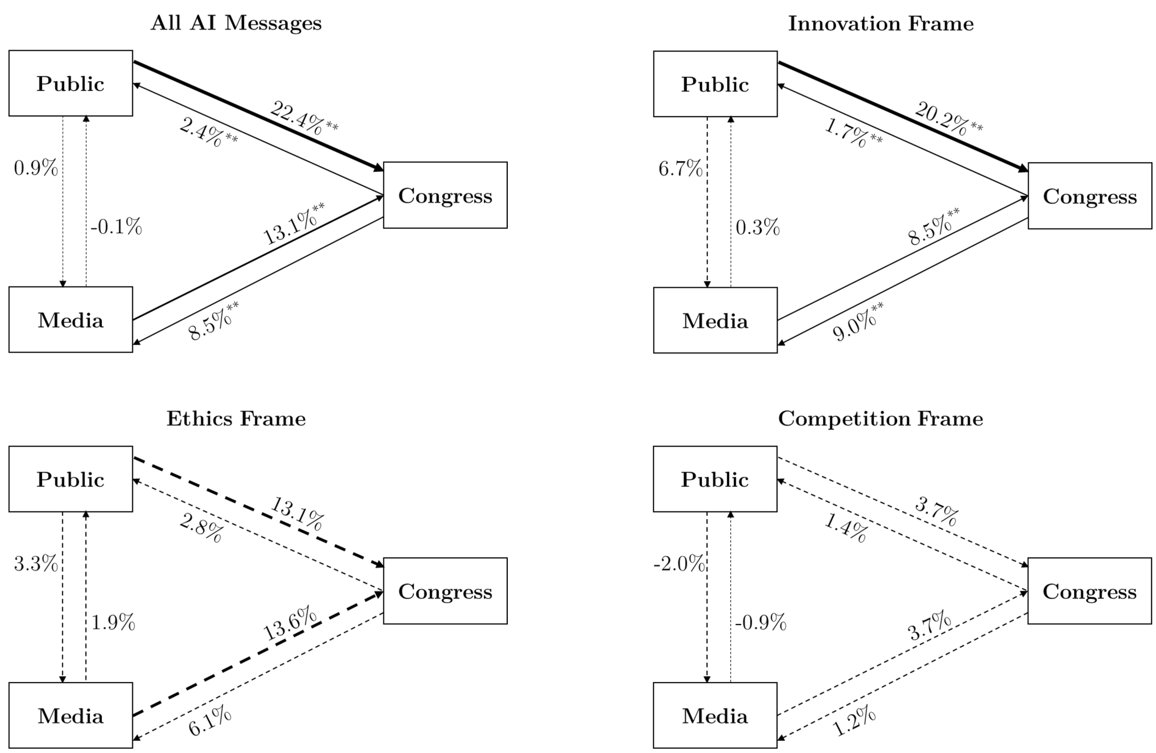

Figure 3.3: Results from ARIMA analysis: Suggested bivariate relationships

Note: Bivariate relationships per issue frame between public, policymakers, and media. Effect sizes listed correspond to percentage increase over weekly baseline resulting from one standard deviation increase in messaging from the influencer. Solid lines indicate significant results at the 5% level and line width corresponds to magnitude.

[19] Indeed, note that for sets of bivariate relationships with less stable ARIMA parameters, generally insignificant results, and smaller effect sizes (e.g., Congress effects on public), this suggests that it is less like that there is a stable and meaningful predictive relationship that can be captured.

Critically, both channels of public influence are present for all AI messages and for the innovation frame alone. In contrast—and across all actors—there are no significant patterns of influence for the ethics frame or for the competition frame at the 5% level. It is true, however, that the magnitude of effects for the ethics frame is substantially larger and borderline significant at the 10% level in some cases, suggesting that the public could have some influence now or in the future. Meanwhile the completely insignificant and marginal effects for the competition frame suggests the public has little role in shaping these dimensions of AI policy. It may be the case that policymakers are strictly looking to experts in geopolitical and military dimensions of AI on these issues. Overall then, while there is support for the Public Agenda-Setting Hypothesis with respect to the public's role generally, *the Special Role of the Public Hypothesis with respect to the ethics frame is not supported*. Instead, it appears policymakers may listen to the public (and media) only when the economic and innovation-oriented dimensions of AI are emphasized.

While these results align to some degree with the descriptive trends, they should nevertheless be interpreted with caution. First, as they are unidirectional relationships that omit the influence of of third party actors or other influences (such as focusing events), the results cannot safely be interpreted causally. Evidence of mutual influence in some cases could indeed indicate that both time series in a given relationship (e.g., public and Congress) are reflecting omitted external factors. Yet, it is plausible that there are not many additional external influences that would bring AI to the attention of policymakers or the public beyond what they learn from the media and one another. Further, because some of the relationships (again, for example, public and Congress) seem to strongly convey one-sided influence dynamics, it is less likely that confounders play a major role for these relationships. Another potential limitation is that the results depict contemporaneous or instantaneous influence within the course of a week, while other temporal modeling structures are possible.[20] To address some of these concerns, especially those related to endogeneity, I next present results from the three-actor VAR analysis.

3.4.3 VAR Analysis

Importantly, modeling with VAR allows for multidirectional influence, and for more than two actors at a time. However, unlike with ARIMA, VAR measures lagged influence rather than instantaneous influence. Thus it is important to identify an appropriate number of weekly lags: A comparison of information criteria and prediction errors suggests two weekly lags is most appropriate (Pfaff, 2008). The VAR models therefore include all three actors (public, policymakers, and media) simultaneously, and I again perform analyses for the three issue frames plus AI attention overall.[21]

To display associated results, I focus on impulse response functions depicting the impact of the *influenced* actor's attention to AI resulting from a one standard deviation shock in the *leading* actor's behavior. The results in Figure 3.4 provide evidence that *public attention does influence policymakers, while Congress does not appear to influence the public*. The NYT also appears to have borderline influence on Congress, with a lesser magnitude compared to the public's influence, though bootstrap confidence intervals cross zero. Notably, results from the forecast error variance decomposition (available in Appendix B.5) also confirm these patterns, with the public explaining about 9-10% of the variance in Congressional AI attention overall, while Congress exerts no such influence on the public. Overall, there is further evidence of the Public Agenda-Setting Hypothesis, reiterating the

[20] For example, a strong bidirectional relationship between two actors could indicate that the agenda-setting process is even shorter term and better modeled at the level of days, for example. I have not modeled the time series at the daily level due to sparsity and to avoid having to specify an overly complex lag structure. Appendix B.4 discusses temporal modeling choices in more detail.

[21] In order for VAR to be applied fruitfully, the underlying time series data should satisfy some conditions. After binding together the two main time series—public AI tweets and Congress AI tweets—and detrending them, I perform associated tests. In particular, the Phillips-Perron test ($p < 0.01$) confirm the main time series are independently stationary and, as before, the Phillips-Ouliaris test ($p < 0.01$) confirms they can be treated as co-integrated. Further, the small-sample adjusted Edgerton-Shukur test ($p = 0.55$) indicates that there are not serially correlated errors (Edgerton & Shukur, 1999), the ARCH Lagrange Multiplier test ($p = 0.25$) confirms a lack of heteroskedasticity (Sjölander, 2011), and a stability test confirms a lack of visible structural breaks in the data (Zeileis et al., 2002).

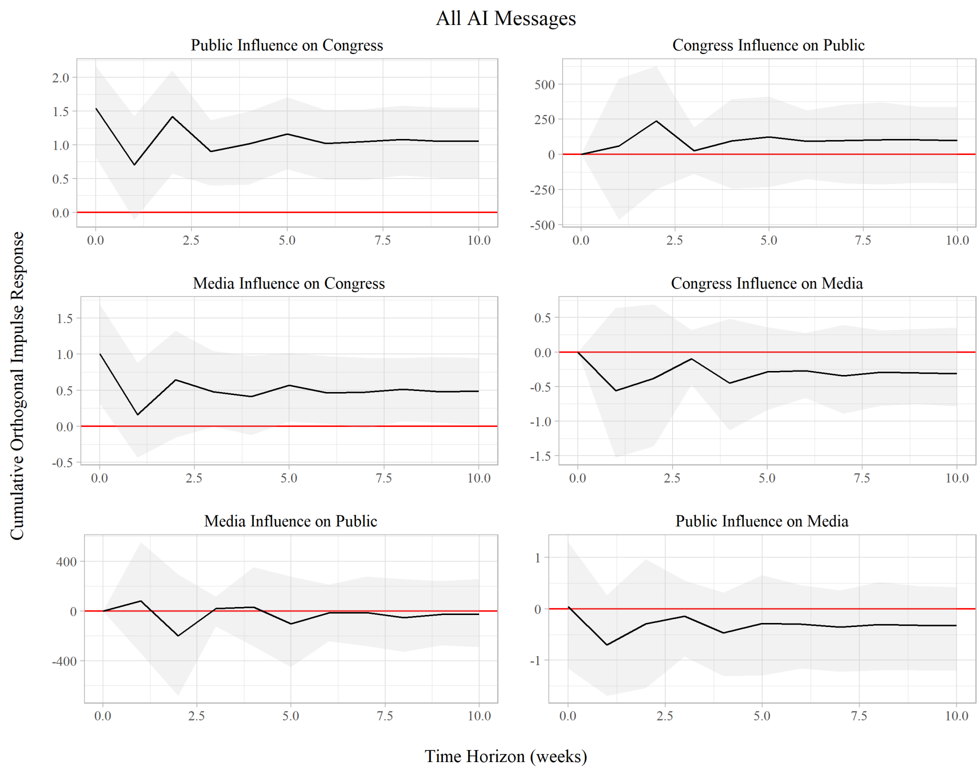

Figure 3.4: Mutual influence of public, policymakers, and media: All AI messages

Note: Cumulative orthogonal impulse response functions indicating impact of a one standard deviation shock in the leading actor's attention to AI with 95% confidence intervals. Impacts measured in number of additional tweets or articles per week are along the Y axis; time in weeks is along the X axis.

prior ARIMA results.

However, the next key question is whether the public influences Congress with respect to particular issue frames. To address this question, I repeat the impulse response analysis for each issue frame and AI messages overall. The patterns in Figure 3.5 roughly mirror the ARIMA results: *The public appears to have influence on Congress with respect to AI messages overall and for the innovation frame, but not for the ethics and competition frames.*[22] Complete results for each set of impulse response functions per issue frame across the three actors are presented in Appendix B.5. In sum, there is little evidence for the Special Role of the Public Hypothesis.

[22]Note that while the magnitudes are small, e.g., 0.25 to 1.5 additional policymaker tweets about AI per week, this corresponds to around 10-20% increases over baseline Congressional messaging, as described in Table 3.2.

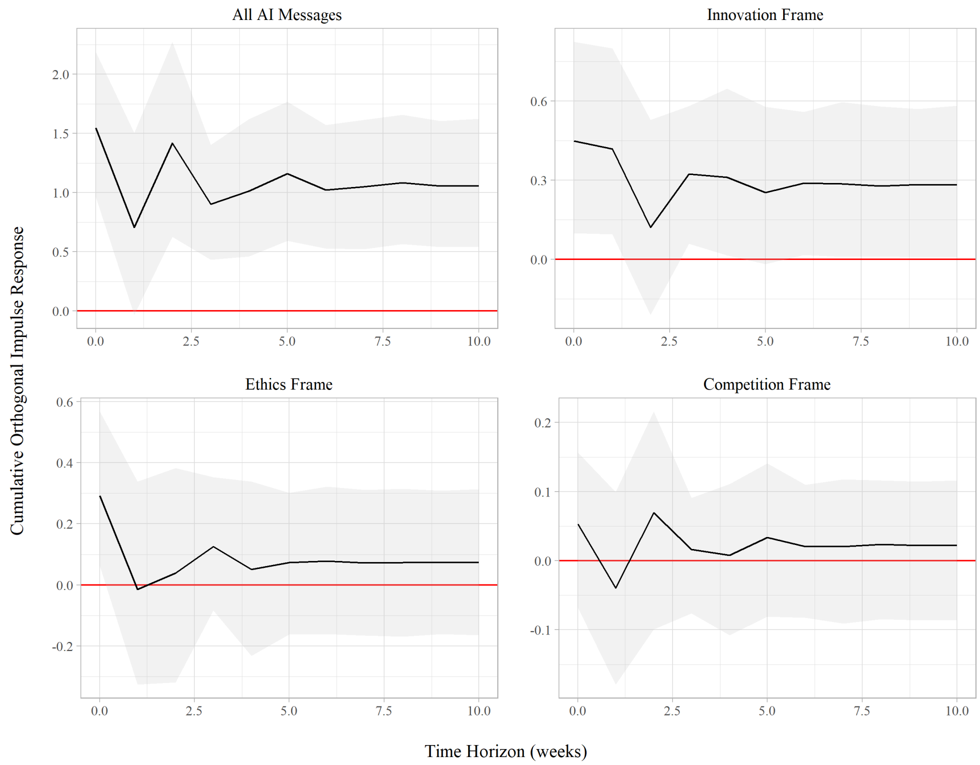

Figure 3.5: Public influence on policymakers per issue frame

Note: Cumulative orthogonal impulse response functions indicating impact of a one standard deviation shock in public attention to AI with on policymaker attention with 95% confidence intervals. Impacts measured in number of additional tweets per week are along the Y axis; time in weeks is along the X axis.

3.5 Implications and Conclusion

As policymakers increasingly wrestle with how to govern complex, high impact, strategic technologies, it is critical to understand both what kinds of issues will shape the policy agenda and which actors have influence in this agenda-setting process. In the context of AI policy, this paper considered whether the emerging AI agenda is developing along a traditional approach—emphasizing expert participation and strategic economic and geopolitical goals— or in line with a new 'paradigm' surfacing social and ethical dimensions of technology and calling for increased public participation.

Findings here reveal a story that, while consistent, is also dynamic and mixed with respect to the role of issue framing contestation and public participation in AI policy. The descriptive results suggest that policymaker attention to the ethics frame for AI is indeed growing, even surpassing attention to geopolitical and economic dimensions of AI. These patterns may indeed be evidence that concerted efforts to govern technology with more proactive consideration of social and ethical dimensions are finding success. Further, across a variety of modeling approaches, results indicate that the public does lead policymaker attention to AI. This both contradicts and and elaborates on prior work (Barberá et al., 2019) that did not fund such a role for the general public across policy domains, while it did find a modest role for more attentive publics including aligned partisan publics, a finding this study cannot evaluate. Increased public attention to AI is consistently associated with increased policymaker attention, with substantive effect sizes, while the converse is not true. Though less novel, the paper also finds that increased media attention, somewhat less consistency and to a lesser degree, is also associated with increased policymaker attention. The fact that public issue attention, directly, or through media distillation, might influence the policy agenda even in the case of a highly complex and plausibly expert-dominated technology policy domain is striking.

Yet the prospects for meaningful public participation in AI governance also appear to be significantly circumscribed. Results demonstrate that the public influences policymaker attention to AI only when AI is discussed in terms of its implications for economic growth and innovation. Meanwhile, and despite the activism of numerous actors in civil society and academia, the public does not appear to be an influence channel specifically with respect to social and ethical implications of AI. This could indicate that policymakers are engaged in a kind of confirmation bias, only listening to the messages they deem fit (Butler & Dynes, 2016). Alternatively, it could indicate that the public is relatively more invested in AI innovation than worried about AI's ethical risks, a finding reiterated in some public opinion research that shows stronger public support for adoption of AI than for its regulation (O'Shaughnessy et al., 2022). In either case, increased public participation and increased ethical consideration may not go hand-in-hand in a straightforward fashion, and imagining the public to serve as the 'voice' of social and ethical consideration may be a simplification of a more complex and evolving role for the public. Additional strategies, like increased public education or facilitation of public involvement in policy via specialized fora may be necessary to better understand and advance the public's role on these issues (Buhmann & Fieseler, 2022).

Importantly, there are several key limitations with respect to the study. First, the selection of issue frames and strategy for classification involve various subjective determinations: While studies aimed at identifying AI frames are broadly consistent with respect to general topics of importance (Imbrie et al., 2021; Ouchchy et al., 2020) other approaches to measuring issue framing could lead to different results. Second, though the three year time period under study is a critical one in light of the emergence of AI policy discourse at this time, it cannot reveal longer-term dynamics. For example, it seems quite possible that competition discourse has increased in the time period exceeding the scope of the study, and it could be the case, for example, that the public is gradually taking on a greater role as society becomes more familiarized with AI. Research in future years and decades will be better positioned to determine if the dynamics identified here are stable or fleeting.

A third limitation surrounds the prospects for generalizing the findings here to other policy domains or settings. The results could indeed be signs of a paradigmatic shift in technology governance or agenda-setting generally, but this paper provides little direct evidence to that effect. Indeed, it is known that agenda-setting dynamics differ across policy domains (Kingdon, 1984; Liu et al., 2011), and AI could be a relatively unique issue given its technical complexity, widespread impact as a general purpose technology, and mixed benefits and risks. The United States is also uniquely situated with respect to AI policy, as well as distinctive in terms of its agenda-setting institutions, policy goals, concerns, and so forth, meaning the exact findings may not translate to different regions and governance scales. Fourth, the time series models applied here do not explicitly take into account other factors like electoral context (Vliegenthart & Walgrave, 2008), or severity of problem indicators and focusing events (Liu et al., 2011). Given the short time period, subjectivity involved in time series modeling, and relatively sparse data in some cases (e.g., a small number of policymaker messages per week), future research will be needed to confirm or challenge the results here.

Overall, this study can benefit research in policy process theory, political communication, and AI policy with implications for the scholarly understanding of the roles of the public, media, and issue framing in the agenda-setting process. Through application of text-as-data methods and the creation of AI issue frame dictionaries, it also helps to concretize more preliminary and conceptual work concerned with AI issue frames and narratives, providing grounds for further study. Further, the application of time series methods to study not merely issue attention, but sub-issue issue frame attention by different policy actors represents a new approach to understanding the evolution of contested policy agendas. Finally, the paper helps to answer some concrete and pressing questions in technology and AI governance surrounding how AI is framed and whether the public has influence in the emerging agenda. While much remains to be known about the direction that technology governance will take, indications in the early years of AI policy suggest that a full transformation has not yet occurred.

3.6 References for Chapter 3

Abelson, J., & Gauvin, F.-P. (2006). *Assessing the impacts of public participation: Concepts, evidence and policy implications*. Centre for Health Economics and Policy Analysis (CHEPA). Hamilton, Canada,

Access Now. (2021). *Open letter calling for a global ban on biometric recognition technologies that enable mass and discriminatory surveillance*. Access Now, Amnesty International, European Digital Rights (EDRi), Human Rights Watch, Internet Freedom Foundation (IFF), Instituto Brasileiro de Defesa do Consumidor (IDEC). Brooklyn, New York. https://www.accessnow.org/ban-biometric-surveillance/

Algorithmic Accountability Act of 2022, H.R. 6580, 117th Cong. (2022). https://www.congress.gov/bill/117th-congress/house-bill/6580/text

Arnstein, S. R. (1969). A ladder of citizen participation. *Journal of the American Institute of Planners*, *35*(4), 216–224. https://doi.org/10.1080/01944366908977225

Atkinson, R. D., & Mayo, M. J. (2010). *Refueling the U.S. innovation economy: Fresh approaches to science, technology, engineering and mathematics (STEM) education* (SSRN Scholarly Paper No. ID 1722822). Social Science Research Network. Rochester, NY. https://papers.ssrn.com/abstract=1722822

Barberá, P., Casas, A., Nagler, J., Egan, P. J., Bonneau, R., Jost, J. T., & Tucker, J. A. (2019). Who leads? Who follows? Measuring issue attention and agenda setting by legislators and the mass public using social media data. *American Political Science Review*, *113*(4), 883–901. https://doi.org/10.1017/S0003055419000352

Barberá, P., & Rivero, G. (2015). Understanding the political representativeness of twitter users. *Social Science Computer Review*, *33*(6), 712–729. https://doi.org/10.1177/0894439314558836

Barnes, M., Newman, J., Knops, A., & Sullivan, H. (2003). Constituting ‘the public’ in public participation. *Public Administration*, *81*(2), 379–399. https://doi.org/10.1111/1467-9299.00352

Baumgartner, F. R., & Jones, B. D. (1993). *Agendas and instability in American politics* (First edition). University of Chicago Press.

Bietti, E. (2020). From ethics washing to ethics bashing: A view on tech ethics from within moral philosophy. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 210–219. https://doi.org/10.1145/3351095.3372860

Boef, S. D., & Keele, L. (2008). Taking time seriously. *American Journal of Political Science*, *52*(1), 184–200. https://doi.org/10.1111/j.1540-5907.2007.00307.x

Boräng, F., Eising, R., Klüver, H., Mahoney, C., Naurin, D., Rasch, D., & Rozbicka, P. (2014). Identifying frames: A comparison of research methods. *Interest Groups & Advocacy*, *3*(2), 188–201. https://doi.org/10.1057/iga.2014.12

Brandt, P. T., & Williams, J. T. (2006). *Multiple time series models*. SAGE Publications.

Branscomb, L. M. (1992). America's emerging technology policy. *Minerva*, *30*(3), 317–336. https://www.jstor.org/stable/41820881

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

Butler, D. M., & Dynes, A. M. (2016). How politicians discount the opinions of constituents with whom they disagree. *American Journal of Political Science*, *60*(4), 975–989. https://doi.org/10.1111/ajps.12206

Carstensen, M. B. (2015). Bringing ideational power into the paradigm approach: Critical perspectives on policy paradigms in theory and practice. In J. Hogan & M. Howlett (Eds.), *Policy paradigms in theory and practice: Discourses, ideas and anomalies in public policy dynamics* (pp. 295–318). Palgrave Macmillan UK. https://doi.org/10.1057/9781137434043_14

Casas, A., & Morar, D. (2019). *Different channel, same strategy? Filling empirical gaps in Congress literature* (SSRN Scholarly Paper No. ID 3395307). Social Science Research Network. Rochester, NY. https://doi.org/10.2139/ssrn.3395307

Cave, S., Craig, C., Dihal, K., Dillon, S., Montgomery, J., Singler, B., & Taylor, L. (2018). *Portrayals and perceptions of AI and why they matter*. The Royal Society. https://doi.org/10.17863/CAM.34502

Chong, D., & Druckman, J. N. (2007). Framing theory. *Annual Review of Political Science*, *10*(1), 103–126. https://doi.org/10.1146/annurev.polisci.10.072805.103054

Cihon, P., Schuett, J., & Baum, S. D. (2021). Corporate governance of artificial intelligence in the public interest. *Information*, *12*(7), 275. https://doi.org/10.3390/info12070275

Crockett, K., Colyer, E., & Latham, A. (2021). The ethical landscape of data and artificial intelligence: Citizen perspectives. *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, 1–9. https://doi.org/10.1109/SSCI50451.2021.9660153

Dahl, R. A. (1978). Pluralism revisited. *Comparative Politics*, *10*(2), 191–203. https://doi.org/10.2307/421645

Daigneault, P.-M. (2014). Reassessing the concept of policy paradigm: Aligning ontology and methodology in policy studies. *Journal of European Public Policy*, *21*(3), 453–469. https://doi.org/10.1080/13501763.2013.834071

Daviter, F. (2011). Policy framing analysis. In F. Daviter (Ed.), *Policy framing in the European Union* (pp. 26–53). Palgrave Macmillan UK. https://doi.org/10.1057/9780230343528_2

de Laat, P. B. (2021). Companies committed to responsible AI: From principles towards implementation and regulation? *Philosophy & Technology*. https://doi.org/10.1007/s13347-021-00474-3

Department of Homeland Security, Office of Cyber and Infrastructure Analysis, United States. (2017). *Narrative analysis research paper: Artificial intelligence risk to critical infrastructure*. Department of Homeland Security, Office of Cyber and Infrastructure Analysis, United States. Washington, D.C. https://info.publicintelligence.net/OCIA-ArtificialIntelligence.pdf

de Saille, S. (2015). Innovating innovation policy: The emergence of 'Responsible Research and Innovation'. *Journal of Responsible Innovation*, *2*(2), 152–168. https://doi.org/10.1080/23299460.2015.1045280

Diercks, G., Larsen, H., & Steward, F. (2019). Transformative innovation policy: Addressing variety in an emerging policy paradigm. *Research Policy*, *48*(4), 880–894. https://doi.org/10.1016/j.respol.2018.10.028

Dryzek, J. S., Goodin, R. E., Tucker, A., & Reber, B. (2009). Promethean elites encounter precautionary publics: The case of GM foods. *Science, Technology, & Human Values*, *34*(3), 263–288. https://doi.org/10.1177/0162243907310297

Dryzek, J. S., & Pickering, J. (2017). Deliberation as a catalyst for reflexive environmental governance. *Ecological Economics*, *131*, 353–360. https://doi.org/10.1016/j.ecolecon.2016.09.011

Edgerton, D., & Shukur, G. (1999). Testing autocorrelation in a system perspective testing autocorrelation. *Econometric Reviews*, *18*(4), 343–386. https://doi.org/10.1080/07474939908800351

Edler, J., & Fagerberg, J. (2017). Innovation policy: What, why, and how. *Oxford Review of Economic Policy*, *33*(1), 2–23. https://doi.org/10.1093/oxrep/grx001

Edwards, G. C., & Wood, B. D. (1999). Who influences whom? The president, congress, and the media. *American Political Science Review*, *93*(2), 327–344. https://doi.org/10.2307/2585399

Eubanks, V. (2017). *Automating inequality: How high-tech tools profile, police, and punish the poor* (First edition). St. Martin's Press.

European Commission. (2021). *Proposal for a regulation on a European approach for artificial intelligence* (COM(2021) 206 final). European Commission. Brussels, Belgium. https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence

Fast, E., & Horvitz, E. (2017). Long-term trends in the public perception of artificial intelligence. *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, 963–969.

Fazekas, Z., Popa, S. A., Schmitt, H., Barberá, P., & Theocharis, Y. (2021). Elite-public interaction on twitter: EU issue expansion in the campaign. *European Journal of Political Research*, *60*(2), 376–396. https://doi.org/10.1111/1475-6765.12402

Fischer, F. (2000). *Citizens, experts, and the environment: The politics of local knowledge*. Duke University Press. http://site.ebrary.com/id/10236528

Fischer, S.-C., Leung, J., Anderljung, M., O'Keefe, C., Torges, S., Khan, S. M., Garfinkel, B., & Allan Dafoe. (2021). *AI policy levers: A review of the U.S. government's tools to shape AI research, development, and deployment*. Centre for the Governance of AI, Future of Humanity Institute, University of Oxford. Oxford, England. https://www.fhi.ox.ac.uk/wp-content/uploads/2021/03/AI-Policy-Levers-A-Review-of-the-U.S.-Governments-tools-to-shape-AI-research-development-and-deployment-%E2%80%93-Fischer-et-al.pdf

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI* (SSRN Scholarly Paper No. ID 3518482). Berkman Klein Center for Internet & Society. Rochester, NY. https://papers.ssrn.com/abstract=3518482

Franke, U. (2021). *Artificial intelligence diplomacy: Artificial intelligence governance as a new external policy tool*. European Parliament, Special Committee on Artificial Intelligence in a Digital Age. Luxembourg.

Fraser, N. (2021). Rethinking the public sphere: A contribution to the critique of actually existing democracy. *Public space reader* (pp. 34–41). Routledge.

Future of Defense Task Force. (2020). *Future of Defense Task Force report 2020*. House Armed Services Committee. Washington, D.C. https://armedservices.house.gov/2020/9/future-of-defense-task-force-releases-final-report

Gamson, W. A., & Modigliani, A. (1989). Media discourse and public opinion on nuclear power: A constructionist approach. *American Journal of Sociology*, *95*(1), 1–37. https://www.jstor.org/stable/2780405

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, *57*(3), 535–574. https://doi.org/10.1257/jel.20181020

Ghosh, S., & Loustaunau, P. (2021). RheFrameDetect: A text classification system for automatic detection of rhetorical frames in AI from open sources. *arXiv:2112.14933 [cs]*. http://arxiv.org/abs/2112.14933

Gilardi, F., Gessler, T., Kubli, M., & Müller, S. (2021). Social media and political agenda setting. *Political Communication*, 1–22. https://doi.org/10.1080/10584609.2021.1910390

Gilardi, F., Shipan, C. R., & Wüest, B. (2021). Policy diffusion: The issue-definition stage. *American Journal of Political Science*, *65*(1), 21–35. https://doi.org/10.1111/ajps.12521

Goddard, S. E., & Krebs, R. R. (2015). Rhetoric, legitimation, and grand strategy. *Security Studies*, *24*(1), 5–36. https://doi.org/10.1080/09636412.2014.1001198

Golbeck, J., Grimes, J. M., & Rogers, A. (2010). Twitter use by the U.S. Congress. *Journal of the American Society for Information Science and Technology*, *61*(8), 1612–1621. https://doi.org/10.1002/asi.21344

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

Granato, J., & Krause, G. A. (2000). Information diffusion within the electorate:: The asymmetric transmission of political–economic information. *Electoral Studies*, *19*(4), 519–537. https://doi.org/10.1016/S0261-3794(99)00027-X

Grillitsch, M., Hansen, T., & Madsen, S. (2018). *How novel is transformative innovation policy?* (No. 2020/08). Centre for Innovation, Research and Competence in the Learning Economy (CIRCLE). Lund, Sweden.

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, *21*(3), 267–297. https://www.jstor.org/stable/24572662

Hall, P. A. (1993). Policy paradigms, social learning, and the state: The case of economic policymaking in Britain. *Comparative Politics*, *25*(3), 275–296. https://doi.org/10.2307/422246

Hallahan, K. (2000). Inactive publics: The forgotten publics in public relations$_*$. *Public Relations Review*, *26*(4), 499–515. https://doi.org/10.1016/S0363-8111(00)00061-8

Hemphill, L., & Roback, A. J. (2014). Tweet acts: How constituents lobby Congress via Twitter. *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, 1200–1210. https://doi.org/10.1145/2531602.2531735

Hester, J. B., & Gibson, R. (2003). The economy and second-level agenda setting: A time-series analysis of economic news and public opinion about the economy. *Journalism & Mass Communication Quarterly*, *80*(1), 73–90. https://doi.org/10.1177/107769900308000106

Hickok, M. (2021). Lessons learned from AI ethics principles for future actions. *AI and Ethics*, *1*(1), 41–47. https://doi.org/10.1007/s43681-020-00008-1

Hogan, J., & Howlett, M. (2015). Reflections on our understanding of policy paradigms and policy change. In J. Hogan & M. Howlett (Eds.), *Policy paradigms in theory and practice: Discourses, ideas and anomalies in public policy dynamics* (pp. 3–18). Palgrave Macmillan UK. https://doi.org/10.1057/9781137434043_1

Hollanders, D., & Vliegenthart, R. (2008). Telling what yesterday's news might be tomorrow: Modeling media dynamics. *33*(1), 47–68. https://doi.org/10.1515/COMMUN.2008.003

Hopkins, D. J., Kim, E., & Kim, S. (2017). Does newspaper coverage influence or reflect public perceptions of the economy? *Research & Politics*, *4*(4), 2053168017737900. https://doi.org/10.1177/2053168017737900

Howlett, M. (1998). Predictable and unpredictable policy windows: Institutional and exogenous correlates of Canadian federal agenda-setting. *Canadian Journal of Political Science / Revue canadienne de science politique*, *31*(3), 495–524. https://www.jstor.org/stable/3233042

Imbrie, A., Dunham, J., Gelles, R., & Aiken, C. (2020). *Mainframes: A provisional analysis of rhetorical frames in AI*. Center for Security and Emerging Technology. Washington, D.C. https://doi.org/10.51593/20190046

Imbrie, A., Gelles, R., Dunham, J., & Aiken, C. (2021). *Contending frames: Evaluating rhetorical dynamics in AI*. Center for Security and Emerging Technology. Washington, D.C. https://doi.org/10.51593/20210010

Jacobi, C., van Atteveldt, W., & Welbers, K. (2016). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism*, *4*(1), 89–106. https://doi.org/10.1080/21670811.2015.1093271

Jones, B. D., & Baumgartner, F. R. (2004). Representation and agenda setting. *Policy Studies Journal*, *32*(1), 1–24. https://doi.org/10.1111/j.0190-292X.2004.00050.x

Jones, B. D., Larsen-Price, H., & Wilkerson, J. (2009). Representation and American governing institutions. *The Journal of Politics*, *71*(1), 277–290. https://doi.org/10.1017/S002238160809018X

Kingdon, J. W. (1984). *Agendas, alternatives, and public policies* (First edition). Little, Brown.

König, P. D., & Wenzelburger, G. (2021). The legitimacy gap of algorithmic decision-making in the public sector: Why it arises and how to address it. *Technology in Society*, *67*, 101688. https://doi.org/10.1016/j.techsoc.2021.101688

Kuhlmann, S., & Rip, A. (2018). Next-generation innovation policy and grand challenges. *Science and Public Policy*, *45*(4), 448–454. https://doi.org/10.1093/scipol/scy011

Leung, J. (2020). *Who will govern artificial intelligence?* (Doctoral dissertation). Oxford University. Oxford, England. https://ora.ox.ac.uk/objects/uuid:ea3c7cb8-2464-45f1-a47c-c7b568f27665

Liu, X., Lindquist, E., & Vedlitz, A. (2011). Explaining media and Congressional attention to global climate change, 1969-2005: An empirical test of agenda-setting theory. *Political Research Quarterly*, *64*(2), 405–419. https://doi.org/10.1177/1065912909346744

Lodge, M., & Wegrich, K. (2015). Crowdsourcing and regulatory reviews: A new way of challenging red tape in british government? *Regulation & Governance*, *9*(1), 30–46. https://doi.org/10.1111/rego.12048

Lowi, T. J. (1964). American business, public policy, case-studies, and political theory. *World Politics*, *16*(4), 677–715. https://doi.org/10.2307/2009452

Macnaghten, P., & Chilvers, J. (2014). The future of science governance: Publics, policies, practices. *Environment and Planning C: Government and Policy*, *32*(3), 530–548. https://doi.org/10.1068/c1245j

Mazzucato, M. (2018). Mission-oriented innovation policies: Challenges and opportunities. *Industrial and Corporate Change*, *27*(5), 803–815. https://doi.org/10.1093/icc/dty034

McCombs, M. E. (2004). *Setting the agenda: The mass media and public opinion*. Polity; Blackwell Pub.

McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly*, *36*(2), 176–187. https://doi.org/10.1086/267990

McKenzie, E., & Gardner, E. S. (2010). Damped trend exponential smoothing: A modelling viewpoint. *International Journal of Forecasting*, *26*(4), 661–665. https://doi.org/10.1016/j.ijforecast.2009.07.001

Mellon, J., & Prosser, C. (2017). Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Research & Politics*, *4*(3), 2053168017720008. https://doi.org/10.1177/2053168017720008

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, *26*. https://proceedings.neurips.cc/paper/2013/hash/9aa42b31882ec039965f3c4923ce901b-Abstract.html

Minkkinen, M., Zimmer, M. P., & Mäntymäki, M. (2022). Co-shaping an ecosystem for responsible AI: Five types of expectation work in response to a technological frame. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-022-10269-2

Mislove, A., Lehmann, S., Ahn, Y.-Y., Onnela, J.-P., & Rosenquist, J. (2011). Understanding the demographics of Twitter users. *Proceedings of the International AAAI Conference on Web and Social Media*, *5*(1), 554–557. https://ojs.aaai.org/index.php/ICWSM/article/view/14168

Mökander, J., Morley, J., Taddeo, M., & Floridi, L. (2021). Ethics-based auditing of automated decision-making systems: Nature, scope, and limitations. *Science and Engineering Ethics*, *27*(4), 44. https://doi.org/10.1007/s11948-021-00319-4

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2019). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics*, *26*(4), 2141–2168. https://doi.org/10.1007/s11948-019-00165-5

Morley, J., Kinsey, L., Elhalal, A., Garcia, F., Ziosi, M., & Floridi, L. (2021). Operationalising AI ethics: Barriers, enablers and next steps. *AI & Society*. https://doi.org/10.1007/s00146-021-01308-8

National Science and Technology Council. (2016). *The national artificial intelligence research and development strategic plan*. National Science and Technology Council (NSTC), Networking and Information Technology Research and Development Subcommittee, United States. Washington, D.C. https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/national_ai_rd_strategic_plan.pdf

National Security Commission on Artificial Intelligence. (2021). *Final report*. National Security Commission on Artificial intelligence. Washington, D.C. https://www.nscai.gov/wp-content/uploads/2021/01/NSCAI-Draft-Final-Report-1.19.21.pdf

Neuman, W. R., Neuman, R. W., Just, M. R., & Crigler, A. N. (1992). *Common knowledge: News and the construction of political meaning*. University of Chicago Press.

Office of Management and Budget. (2020). *Memorandum for the heads of executive departments and agencies: Guidance for regulation of artificial intelligence applications*. Office of Management and Budget. Washington, D.C. https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf

Office of Science and Technology Policy. (2019). *The American AI Initiative*. Office of Science and Technology Policy. Washington, D.C. https://trumpwhitehouse.archives.gov/articles/accelerating-americas-leadership-in-artificial-intelligence/

O'Shaughnessy, M., Schiff, D. S., Varshney, L. R., Rozell, C. J., & Davenport, M. A. (2022). What governs attitudes toward artificial intelligence adoption and governance? *Working paper*. https://doi.org/10.31219/osf.io/pkeb8

Ouchchy, L., Coin, A., & Dubljević, V. (2020). AI in the headlines: The portrayal of the ethical issues of artificial intelligence in the media. *AI & Society*, *35*(4), 927–936. https://doi.org/10.1007/s00146-020-00965-5

Peeters, J., Van Aelst, P., & Praet, S. (2021). Party ownership or individual specialization? A comparison of politicians' individual issue attention across three different agendas. *Party Politics*, *27*(4), 692–703. https://doi.org/10.1177/1354068819881639

Peter, F. (2008). *Democratic legitimacy*. Routledge. https://doi.org/10.4324/9780203401392

Peters, B. G., & Hogwood, B. W. (1985). In search of the issue-attention cycle. *The Journal of Politics*, *47*(1), 238–253. https://doi.org/10.2307/2131074

Pfaff, B. (2008). VAR, SVAR and SVEC models: Implementation within R package **vars**. *Journal of Statistical Software*, *27*(4). https://doi.org/10.18637/jss.v027.i04

Pierre, J. (Ed.). (2000). *Debating governance: Authority, steering, and democracy*. OUP Oxford.

Prates, M., Avelar, P., & Lamb, L. C. (2018). On quantifying and understanding the role of ethics in AI research: A historical account of flagship conferences and journals. *arXiv:1809.08328 [cs]*, 188–173. https://doi.org/10.29007/74gj

Rességuier, A., & Rodrigues, R. (2020). AI ethics should not remain toothless! A call to bring back the teeth of ethics. *Big Data & Society*, *7*(2), 2053951720942541. https://doi.org/10.1177/2053951720942541

Ringel, S. (2021). Digitizing the paper of record: Archiving digital newspapers at the new york times. *Journalism*, 14648849211023849. https://doi.org/10.1177/14648849211023849

Ripberger, J. T. (2011). Capturing curiosity: Using internet search trends to measure public attentiveness. *Policy Studies Journal*, *39*(2), 239–259. https://doi.org/10.1111/j.1541-0072.2011.00406.x

Rodrigo, C. M. (2021). White House science advisers call for AI 'bill of rights'. *TheHill*. https://thehill.com/policy/technology/576011-white-house-science-advisers-call-for-ai-bill-of-rights

Rowe, G., & Frewer, L. J. (2000). Public participation methods: A framework for evaluation. *Science, Technology, & Human Values*, *25*(1), 3–29. https://doi.org/10.1177/016224390002500101

Ryan, M., & Stahl, B. C. (2020). Artificial intelligence ethics guidelines for developers and users: Clarifying their content and normative implications. *Journal of Information, Communication and Ethics in Society*, *19*(1), 61–86. https://doi.org/10.1108/JICES-12-2019-0138

Salgado-Criado, J., & Fernández-Aller, C. (2021). A wide human-rights approach to artificial intelligence regulation in Europe. *IEEE Technology and Society Magazine*, *40*(2), 55–65. https://doi.org/10.1109/MTS.2021.3056284

Schattschneider, E. E. (1935). *Politics, pressures and the tariff;: A study of free private enterprise in pressure politics, as shown in the 1929-1930 revision of the tariff,* Prentice-Hall.

Schiff, D. (2021). Education for AI, not AI for education: The role of education and ethics in national AI policy strategies. *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-021-00270-2

Schiff, D., Ayesh, A., Musikanski, L., & Havens, J. C. (2020). IEEE 7010: A new standard for assessing the well-being implications of artificial intelligence. *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2746–2753. https://doi.org/10.1109/SMC42975.2020.9283454

Schiff, D., Borenstein, J., Laas, K., & Biddle, J. (2021). AI ethics in the public, private, and NGO sectors: A review of a global document collection. *IEEE Transactions on Technology and Society*, *2*(1), 31–42. https://doi.org/10.1109/TTS.2021.3052127

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

Schnell, K. C., Frauke. (2001). Assessing the democratic debate: How the news media frame elite policy discourse. *Political Communication*, *18*(2), 183–213. https://doi.org/10.1080/105846001750322970

Schot, J., & Steinmueller, W. E. (2018). Three frames for innovation policy: R&D, systems of innovation and transformative change. *Research Policy*, *47*(9), 1554–1567. https://doi.org/10.1016/j.respol.2018.08.011

Schwab, K. (2016). *The fourth industrial revolution* (First U.S. edition). Crown Business.

Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., & Hall, P. (2022). *Towards a standard for identifying and managing bias in artificial intelligence*. National Institute of Standards and Technology. Washington, D.C. https://doi.org/10.6028/NIST.SP.1270

Setälä, M., & Smith, G. (2018). Mini-publics and deliberative democracy. In A. Bächtiger, J. S. Dryzek, J. Mansbridge, & M. E. Warren (Eds.), *Oxford handbook of deliberative democracy*. Oxford University Press.

Shapiro, M. A., & Hemphill, L. (2017). Politicians and the policy agenda: Does use of Twitter by the U.S. Congress direct New York Times content? *Policy & Internet*, *9*(1), 109–132. https://doi.org/10.1002/poi3.120

Shapiro, M. A., Hemphill, L., & Otterbacher, J. (2018). Updates to congressional speech acts on Twitter. *American Political Science Association Annual Confrence*, 23.

Sharp, E. B. (1994). The dynamics of issue expansion: Cases from disability rights and fetal research controversy. *The Journal of Politics*, *56*(4), 919–939. https://doi.org/10.2307/2132067

Shneiderman, B. (2022). *Human-centered AI*. Oxford University Press.

Shoemaker, P. J., & Reese, S. D. (1991). *Mediating the message: Theories of influences on mass media content*. Longman.

Siddique, N. (2019). Creating a data set for 116th Congress Twitter handles. https://ws-dl.blogspot.com/2019/04/2019-04-01-creating-data-set-for-116th.html

Sjölander, P. (2011). A stationary unbiased finite sample ARCH-LM test procedure. *Applied Economics*, *43*(8), 1019–1033. https://doi.org/10.1080/00036840802600046

Sloane, M., Moss, E., Awomolo, O., & Forlano, L. (2020). Participation is not a design fix for machine learning. *arXiv:2007.02423 [cs]*. http://arxiv.org/abs/2007.02423

Smits, R., & Kuhlmann, S. (2004). The rise of systemic instruments in innovation policy. *International Journal of Foresight and Innovation Policy*, *1*(1/2), 4. https://doi.org/10.1504/IJFIP.2004.004621

Soete, L. (2007). From industrial to innovation policy. *Journal of Industry, Competition and Trade*, *7*(3-4), 273–284. https://doi.org/10.1007/s10842-007-0019-5

Stahl, B. C. (2021). Addressing ethical issues in AI. In B. C. Stahl (Ed.), *Artificial intelligence for a better future: An ecosystem perspective on the ethics of AI and emerging digital technologies* (pp. 55–79). Springer International Publishing. https://doi.org/10.1007/978-3-030-69978-9_5

Stark, L., Greene, D., & Hoffmann, A. L. (2021). Critical perspectives on governance mechanisms for AI/ML systems. In J. Roberge & M. Castelle (Eds.), *The cultural life of machine learning: An incursion into critical AI studies* (pp. 257–280). Springer International Publishing. https://doi.org/10.1007/978-3-030-56286-1_9

Steinberger, P. J. (1980). Typologies of public policy: Meaning construction and the policy process. *Social Science Quarterly*, *61*(2), 185–197. https://www.jstor.org/stable/42860711

Stirling, A. (2008). "Opening up" and "closing down": Power, participation, and pluralism in the social appraisal of technology. *Science, Technology, & Human Values*, *33*(2), 262–294. https://doi.org/10.1177/0162243907311265

Stix, C. (2021). Actionable principles for artificial intelligence policy: Three pathways. *Science and Engineering Ethics*, *27*(1), 15. https://doi.org/10.1007/s11948-020-00277-3

Stix, C., & Maas, M. (2021). Bridging the gap: The case for an 'incompletely theorized agreement' on AI policy. *AI and Ethics*. https://doi.org/10.1007/s43681-020-00037-w

Stoker, K. (2014). Defining public in public relations: How the 1920s debate over public opinion influenced early philosophies of public relations. *Pathways to public relations*. Routledge.

Stray, J. (2020). Aligning AI optimization to community well-being. *International Journal of Community Well-Being*, *3*(4), 443–463. https://doi.org/10.1007/s42413-020-00086-3

Sturgis, P., & Allum, N. (2004). Science in society: Re-evaluating the deficit model of public attitudes. *Public Understanding of Science*, *13*(1), 55–74. https://doi.org/10.1177/0963662504042690

Taeihagh, A. (2021). Governance of artificial intelligence. *Policy and Society*, *40*(2), 137–157. https://doi.org/10.1080/14494035.2021.1928377

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2019). *Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems, first edition*. IEEE. Piscataway, New Jersey. https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/%20autonomous-systems.html

Tromble, R. (2018). The great leveler? Comparing citizen–politician Twitter engagement across three western democracies. *European Political Science*, *17*(2), 223–239. https://doi.org/10.1057/s41304-016-0022-6

Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., & Floridi, L. (2022). The ethics of algorithms: Key problems and solutions. *AI & Society*, *37*(1), 215–230. https://doi.org/10.1007/s00146-021-01154-8

Ulnicane, I. (2016). 'Grand challenges' concept: A return of the 'big ideas' in science, technology and innovation policy? *International Journal of Foresight and Innovation Policy*, *11*(1-3), 5–21. https://doi.org/10.1504/IJFIP.2016.078378

Ulnicane, I. (2022). Against the new space race: Global AI competition and cooperation for people. *AI & Society*. https://doi.org/10.1007/s00146-022-01423-0

Ulnicane, I., Eke, D. O., Knight, W., Ogoh, G., & Stahl, B. C. (2021). Good governance as a response to discontents? Déjà vu, or lessons for AI from other emerging technologies. *Interdisciplinary Science Reviews*, *46*(1-2), 71–93. https://doi.org/10.1080/03080188.2020.1840220

Ulnicane, I., Knight, W., Leach, T., Stahl, B. C., & Wanjiku, W.-G. (2020). Framing governance for a contested emerging technology: Insights from AI policy. *Policy and Society*, *0*(0), 1–20. https://doi.org/10.1080/14494035.2020.1855800

United States Innovation and Competition Act of 2021, S. 1260, 117th Cong. (2021). https://www.congress.gov/bill/117th-congress/senate-bill/1260

Vaccari, C., Valeriani, A., Barberá, P., Bonneau, R., Jost, J. T., Nagler, J., & Tucker, J. A. (2015). Political expression and action on social media: Exploring the relationship between lower- and higher-threshold political activities among Twitter users in italy. *Journal of Computer-Mediated Communication*, *20*(2), 221–239. https://doi.org/10.1111/jcc4.12108

Van Aelst, P., & Walgrave, S. (2016). Information and arena: The dual function of the news media for political elites. *Journal of Communication*, *66*(3), 496–518. https://doi.org/10.1111/jcom.12229

Vesnic-Alujevic, L., Nascimento, S., & Pólvora, A. (2020). Societal and ethical impacts of artificial intelligence: Critical notes on European policy frameworks. *Telecommunications Policy*, *44*(6), 101961. https://doi.org/10.1016/j.telpol.2020.101961

Vliegenthart, R. (2014). Moving up. Applying aggregate level time series analysis in the study of media coverage. *Quality & Quantity*, *48*(5), 2427–2445. https://doi.org/10.1007/s11135-013-9899-0

Vliegenthart, R., & Walgrave, S. (2008). The contingency of intermedia agenda setting: A longitudinal study in Belgium. *Journalism & Mass Communication Quarterly*, *85*(4), 860–877. https://doi.org/10.1177/107769900808500409

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

Vu, K., Hanafizadeh, P., & Bohlin, E. (2020). ICT as a driver of economic growth: A survey of the literature and directions for future research. *Telecommunications Policy*, *44*(2), 101922. https://doi.org/10.1016/j.telpol.2020.101922

Warnke, P., Koschatzky, K., Dönitz, E., Zenker, A., Stahlecker, T., Som, O., Cuhls, K., & Güth, S. (2016). *Opening up the innovation system framework towards new actors and institutions* (Working Paper No. 49). Fraunhofer ISI Discussion Papers - Innovation Systems and Policy Analysis. https://www.econstor.eu/handle/10419/129191

Weaver, D. A., & Bimber, B. (2008). Finding news stories: A comparison of searches using LexisNexis and Google News. *Journalism & Mass Communication Quarterly*, *85*(3), 515–530. https://doi.org/10.1177/107769900808500303

Whittlestone, J., Nyrup, R., Alexandrova, A., Dihal, K., & Cave, S. (2019). *Ethical and societal implications of algorithms, data, and artificial intelligence: A roadmap for* *research*. Nuffield Foundation. London. http://lcfi.ac.uk/media/uploads/files/Ethical-and-Societal-Implications-of-Data-and-AI-report-Nuffield-Foundat_tG6yy08.pdf

Wilkerson, J., & Casas, A. (2017). Large-scale computerized text analysis in political science: Opportunities and challenges. *Annual Review of Political Science*, *20*(1), 529–544. https://doi.org/10.1146/annurev-polisci-052615-025542

Wilson, J. Q. (1973). *Political organizations*. Basic Books.

Wolfe, M., Jones, B. D., & Baumgartner, F. R. (2013). A failure to communicate: Agenda setting in media and policy studies. *Political Communication*, *30*(2), 175–192. https://doi.org/10.1080/10584609.2012.737419

Yackee, J. W., & Yackee, S. W. (2006). A bias towards business? Assessing interest group influence on the U.S. bureaucracy. *The Journal of Politics*, *68*(1), 128–139. https://doi.org/10.1111/j.1468-2508.2006.00375.x

Yanovitzky, I. (2002). Effects of news coverage on policy attention and actions: A closer look into the media-policy connection. *Communication Research*, *29*(4), 422–451. https://doi.org/10.1177/0093650202029004003

Zaller, J. (1992). *The nature and origins of mass opinion*. Cambridge University Press.

Zeileis, A., Leisch, F., Hornik, K., & Kleiber, C. (2002). Strucchange: An R package for testing for structural change in linear regression models. *Journal of Statistical Software*, *7*, 1–38. https://doi.org/10.18637/jss.v007.i02

Zhang, D., Saurabh, M., Brynjolfsson, E., Etchemendy, J., Ganguli, D., Grosz, B., Lyons, T., Manyika, J., Niebles, J. C., Sellitto, M., Shoham, Y., Clark, J., & Perrault, R. (2021). *Artificial Intelligence Index report 2021*. AI Index Steering Committee, Human-Centered AI Institute, Stanford University. Stanford, CA. https://aiindex.stanford.edu/wp-content/uploads/2021/03/2021-AI-Index-Report_Master.pdf

Zwetsloot, R., Toner, H., DingNovember 16, J., & 2018. (2018). Beyond the AI arms race: America, China, and the dangers of zero-sum thinking. https://www.foreignaffairs.com/reviews/review-essay/2018-11-16/beyond-ai-arms-race

CHAPTER 4

REASON AND PASSION IN AGENDA-SETTING: EXPERIMENTAL EVIDENCE ON STATE LEGISLATOR ENGAGEMENT WITH AI POLICY

4.1 Introduction

The politics of influence play a central role in theories of agenda-setting. According to the most prominent framework of agenda-setting (M. D. Jones et al., 2016), the Multiple Streams Framework (MSF), policy problems and solutions to those problems are brought together by skilled policy entrepreneurs through careful strategic problem definition, networking, and provision of information to policymakers (Kingdon, 1984). Yet, recent stock-taking of the agenda-setting literature and its impact over the past decades has encouraged heightened attention to hypothesis formulation and empirical testing beyond qualitative case study in order to support long-term theory building (Cairney & Jones, 2016). Further, recent calls emphasize the need to unpack the inner workings or 'black box' of policy processes and identify clear findings that "inform the work of policymakers and issue advocates" (Anderson et al., 2020, p. 590). For example, in the case of policy entrepreneurs, Petridou and Mintrom (2021) call for research that more carefully measures the impact of policy entrepreneurs including with respect to specific strategies they employ.

In response, scholars have widened their attention to other theories in the policy process literature in order to better conceptualize, refine, and test theories. One such area of fruitful intersection is between the agenda-setting literature and the relatively recent Narrative Policy Framework (NPF) (M. D. Jones et al., 2016; McBeth et al., 2007). The NPF embraces the use of stories in policy persuasion, involving characters, context, plot, and moral, and offers hypotheses along micro, meso, and macro levels of analysis (McBeth et al., 2014). It has been increasingly embraced as a promising explanatory framework bringing post-positivist elements into policy change theory (McBeth et al., 2014; Weible & Schlager, 2016). For example, Birkland and Warnement (2016) suggest the utility of the NPF in explaining focusing events, McBeth and Lybecker (2018) argue that narratives can better explain the role of policy entrepreneurs in problem definition and coupling problems and solutions, and Petridou and Mintrom (2021) argue that policy entrepreneurs can serve as "policy marketers" who promote narratives to streamline complex policy issues.

This chapter embraces this direction and responds to these calls through incorporation of elements of narrative strategy into a study of agenda-setting and policy entrepreneur influence. In particular, I consider whether policy entrepreneurs can effectively use narratives to influence policymakers even in highly technical, complex policy domains where the provision of technical information is traditionally considered essential. The context for this work is one such emerging and understudied technical policy domain, that of artificial intelligence (AI). AI policy is a valuable testbed for policy process and agenda-setting research given its sweeping social, ethical, and economic implications across policy domains, and because a wide diversity of policy entrepreneurs are now drawing on a variety of strategies including framing to set the terms of debate (Cave et al., 2018; Minkkinen et al., 2022). In short, study of AI policy in the near future provides the opportunity to closely examine agenda-setting taking place for a novel, complex, and important policy domain for which the agenda has not yet been established.

This study examines agenda-setting influence in AI policy by observing the behavior of United States (U.S.) state legislators in response to differing influence strategies of policy entrepreneurs. In partnership with a leading AI policy think tank, I perform a information-provision field experiment— specifically an audit or correspondence experiment— and randomly assign more than 7,300 state legislators to receive different email communications about AI policy. In line with the study's pre-registered design, legislators receive messages providing either expert technical information, a persuasive narrative story, or more generic policy entrepreneur outreach (a control message). Separately, as part of the experiment's factorial design, the emails emphasize a social and ethical frame surrounding AI policy, or an economic and competition-based frame. To measure policymaker engagement with these messages as an indication of influence, I determine whether policymakers click on links to fact sheets or stories embedded in the emails, reply to the email, and click to visit a registration page for or attend a webinar on AI policy developed as part of the study.

I find that both technical information and narratives are statistically more engaging than more generic policy entrepreneur outreach messages: Legislators are around 30 percent more likely to engage in activities like clicking on a fact sheet or story about AI policy. Strikingly, despite AI's noted technical complexity, narratives are just as effective as expert information in engaging policymakers. This result holds whether policymakers receive a frame emphasizing social and ethical issues or economic and geopolitical dimensions of AI. Moreover, despite significant attention to AI's role for innovation and competition, I find that policymakers are at least as drawn to an ethics-based frame. Finally, I find that legislators in states with little prior experience in AI policymaking are especially interested in engaging not only with expert information, but also with narratives. Overall, the results reiterate the value of understanding narratives in the policy process, and indicate that 'passion' can be just as important as 'reason' in policy influence efforts, even for highly technical domains.

4.2 Theory

4.2.1 The Role of Narratives in Agenda-Setting

Following the post-positivist turn in policy process theory (Fischer, 1998), scholars have increasingly recognized subjectivity, value-laden reasoning, and politics as important components in policymaking and agenda-setting more specifically, while also emphasizing the need for empirical testing of these interpretive elements. This movement in policy scholarship permeates prominent policy process theories through concepts such as focusing events (Birkland & Warnement, 2016), policy images (Baumgartner & Jones, 1991), beliefs (Sabatier, 1988), social constructions (Schneider & Ingram, 1993), and recently, narratives (McBeth et al., 2014). Narratives arguably play a role in shaping policy images and social constructions as well as beliefs, by helping to establish what the relevant policy 'story' and its context are, who the story's heroes and villains are, and what moral lessons can be learned. Successful narratives can influence agenda-setting by shaping perceptions about the construction of costs and benefits, expanding or containing issues, and reducing uncertainty and perceived risk (McBeth et al., 2007).

While narratives are often discussed in the context of shaping public opinion, the meso-level of analysis in the NPF recognizes "how groups use policy narratives to try and influence public policy" and policymakers more specifically (McBeth & Lybecker, 2018, p. 170). Narratives may be influential for policymakers for a variety of reasons: Not only can they provide a valuable tool for a policymaker's own messaging efforts, they may also directly persuade a policymaker of the legitimacy and feasibility of a proposed policy (Anderson et al., 2020). With respect to the former, especially at the outset of a new policy question or domain, policymakers may be in need of messages, images, and narratives used to craft their own political identity, persuasive messaging, and public and constituent communications. In terms of their direct persuasive impacts, a body of research (Brysk, 1994; Small & Loewenstein, 2003) shows that 'personal' narratives, particularly those discussing the plight of individuals, may be "very effective at eliciting an emotional reaction, personalizing the issue, making it more salient, and making people feel a greater need to act" (Mcentire et al., 2015, p. 14).

In the context of agenda-setting, it is policy entrepreneurs—in part functioning as policy marketers—who play the key role as the crafters and promoters of narratives through the act of problem definition, or elevating public conditions to the status of policy problems through careful and strategic shaping of indicators and focusing events (Mcbeth & Shanahan, 2004; Mintrom & Norman, 2009). Given the central role that problem (and indeed solution) definition plays in the activities of policy entrepreneurs and in frameworks like the MSF, narratives should arguably be incorporated as prominent or even dominant considerations in making sense of influence dynamics in agenda-setting. As McBeth and Lybecker (2018) argue, public policy may in fact be increasingly driven by narratives.

4.2.2 Policy Entrepreneurship and Expertise in Technical Policy Domains

This theoretical contribution of the NPF can thus be appropriately encapsulated by attending to the activities of policy entrepreneurs in agenda-setting processes. Policy entrepreneurs are known to play a key role in problem definition, as well as in activities like coalition-building (Mintrom & Norman, 2009), and in the provision of expertise, amongst other influence strategies (Capano & Galanti, 2018; Frisch Aviram et al., 2020). Indeed, in recent and closely-related research, Anderson et al. (2020) demonstrate empirically that state legislators are especially reliant on the provision of information and evidence, exceeding even the importance of traditional policy entrepreneurship activities like coalition-building. Policy entrepreneurs may provide critical information about policy problems and solutions that are poorly understood (serving as 'information entrepreneurs' or 'expert entrepreneurs' (Crow, 2010) or 'knowledge brokers' (Knaggård, 2014)) by empirically demonstrating the severity of problems or the feasibility of solutions (Knaggård, 2015). This expertise provision may reduce the perceived riskiness of policymaking decisions in light of uncertainty (Dewulf & Biesbroek, 2018; Knaggård, 2014), and can even be used as a justificatory strategy for policymakers who desire scientific or technical credibility.

In particular, the provision of expertise should be especially important in policy areas that are technically complex or 'hard' issues (Carmines & Stimson, 1980; Gormley, 1986), such as environmental policy (Knaggård, 2014) and AI. Given the difficulty of understanding associated policy issues and the resulting uncertainty (Zito, 2001), policymakers are especially in need of 'hard' evidence to inform decisionmaking. Further, in the case of these highly complex or technical issues, policymaking is typically dominated by powerful elites, expert bureaucrats and professionals, and associated business interests. This is especially the case for issues of low salience—jointly leading to 'board room' politics—where the public has little access (Gormley, 1986). In these cases, public attention and electoral accountability is also less critical for politicians (Eshbaugh-Soha, 2006). As such, policymakers will arguably be less in need of narratives for the purpose of stump speeches and credit-claiming.

Emerging technologies are especially good candidates of highly technical and complex domains, and AI is a paradigmatic example of this. Notably, a major concern nationally is building up competence in the government to understand and address AI issues and policies, demonstrated through the high prioritization and urgency evidenced across government AI policy strategies. For example (such examples abound), U.S. Executive Order 13859, Maintaining American Leadership in Artificial Intelligence (2019) includes the launch of a government-wide AI community of practice while the National Security Commission on AI (NSCAI), a prominent and early-moving actor in shaping U.S. AI policy, features more than 50 pages on educating, recruiting, and training "AI Talent" in its most recent report (National Security Commission on Artificial Intelligence, 2021). These concerns are persistent across state, national, and international AI policy discourse (Schiff, 2021), demonstrating the clear perceived importance of enhancing government expertise with respect to this emerging policy domain.

4.2.3 Competing Influence Dynamics in Agenda-Setting

These theoretical developments in policy entrepreneurship, the NPF, and agenda-setting offer potentially competing answers to the question of what modes of policy entrepreneurship are most influential to policymakers in the case of complex, emerging technologies. On the one hand, initial scholarship bridging the NPF and agenda-setting literatures would suggest that narratives should be effective as policy entrepreneurs work to define problems and couple them with solutions to persuade policymakers. McBeth and Lybecker (2018, p. 871) have argued that they "do not expect significant use of evidence in efforts of policy entrepreneurs to define the focusing event" and suggest that even when evidence is used, it "will often be embedded within a narrative element." This is consistent with the argument from Zahariadis (2014), who argues that while the provision of information alone may reduce uncertainty, it will not dissolve the ambiguities that must be resolved for agenda-setting to take place.

On this view, carefully-crafted narratives, especially of a personal nature, could be just as effective as expert information (Mcentire et al., 2015). This should hold especially if AI is salient to the public or influential public messengers such as the media and civil society groups. It is also more likely if policy entrepreneurs have successfully framed AI in terms of implications that might capture the public's attention, such as impacts on ethics, racial injustice, inequality, or job displacement. These kinds of framings (Iyengar, 1994), made possible by AI's sweeping impacts across social and economic life, would constitute successful issue expansion aimed at expanding one's coalition and public attention in turn. Notably, these kinds of efforts are under study by scholars in AI as well, exemplified by work at CSET (Imbrie et al., 2021) and the Global AI Narratives Project out of the Leverhulme Centre for the Future of Intelligence and the Royal Society (Cave et al., 2018).

Yet, while there is a plethora of evidence that various policy entrepreneurs have indeed tried to define AI in terms of public issue considerations, science and technology policy typically rests at the bottom of public issue priorities (B. D. Jones et al., 2009). In the current environment, despite its acknowledged widespread impact, AI is likely overshadowed by crises in public health, environment, racial and political cohesion, and the economy. For example, a single presidential candidate in the 2020 election (Andrew Yang) placed automation among their top issue priorities, but public attention was not long-lasting. Further, given the established importance of expert information for domains like the environment (Knaggård, 2014; Michaels, 2009), there are reasons to expect that a potentially even more complex domain like AI will be just as reliant on expert information in policy entrepreneurship if not more so. Along these lines, the competing prediction suggests that it is reason, and not passion, that should be most influential in agenda-setting. That is, for a highly complex technological domain like AI, characterized by great uncertainty and little public (or policymaker) understanding, we might expect the provision of expertise to be far more valuable and influential to policymakers.

4.2.4 Issue Frames

Scholars of political communication and media have also examined how policy issues may be framed in different ways, to emphasize alternative sub-issues or dimensions involved (Chong & Druckman, 2007; Iyengar, 1990). For example, Neuman et al. (1992) identify human impact, economics, and conflict as three common issue frames used in the news media. These contrasting issue frames not only have the ability to influence public opinion: They also may influence policy entrepreneurs and policymakers (McBeth et al., 2014). Policy entrepreneurs can thus strategically take advantage of different available issue (or policy) frames (Mcbeth & Shanahan, 2004; Mintrom & Luetjens, 2017) to expand or contain issues into favorable policy venues and to help construct a preferred policy image (Baumgartner & Jones, 1991).[1]

Indeed, there is ample evidence that key stakeholders in AI are already conceiving of challenges and solutions of AI policy in terms of these kinds of issue frames. In terms of frames emphasizing human impact, a large number of organizations have produced ethical codes, frameworks, and principles (Schiff et al., 2021) that have filtered into national and international policy strategies and agreements, such as the OECD's (2019) Principles on Artificial Intelligence. Alternatively, policymakers have also paid substantial attention to the economic and technological competitiveness dimensions of AI, often contrasting the success of U.S. AI policy and development against that of China strategically, economically, and even militarily (Castro & McLaughlin, 2021; Ulnicane, 2022). In this case,

[1] A possible source of confusion is the distinction between a narrative and a frame. Simply, narratives can contain or promote one of many possible issue frames, while issue frames need not be accompanied by a narrative, as the experimental design in this study makes clear.

the economic and conflict frames often appear to be merged in practice, as the economic standing of the United States in terms of technology and innovation is considered a pillar of its national security, and vice versa. For example, the NSCAI (2021) devotes significant attention to how U.S. talent development is important for securing its competitive position against that of China and key AI policy is being advanced through legislation literally entitled the U.S. Innovation and Competition Act (2021).

I thus employ issue frames surrounding ethical implications of AI or, alternatively, the economic and technological competitiveness implications of AI, as a separate dimension of the experimental design for three reasons. One, as there is arguably no way for policy entrepreneurs to present a 'neutral' frame of AI policy (Elder & Cobb, 1984), it is important to identify prominent and realistic frames to build knowledge about issue framing as well as support external validity in the AI policy domain. Relatedly, the use of multiple frames helps ensure that findings about policy entrepreneur influence strategies are less likely to merely reflect esoteric aspects of AI policy. Finally, the use of contrasting frames also allows us to examine the relationship between framing and policy entrepreneur influence efforts (Petridou & Mintrom, 2021), for example whether certain influence strategies are more or less effective when used in tandem with particular issue frames.

4.2.5 Hypotheses

In the following experimental tests of these theoretical predictions, I evaluate three main hypotheses and two exploratory hypotheses. These hypotheses are evaluated in terms of the effectiveness of different email messages in engaging U.S. state legislators, as a proxy of policy entrepreneur agenda-setting influence. I provide details on how engagement is measured in the next section.

Main Hypotheses:

Policy Entrepreneur Effectiveness Hypothesis: *When policy entrepreneurs provide narratives or expertise as part of their influence efforts, it will increase policymaker attention to and engagement with the policy issue at hand.*

Dominance of Narratives Hypothesis: *The provision of narratives will induce greater policymaker engagement than the provision of expertise.*

Dominance of Expertise Hypothesis: *The provision of expertise will induce greater policymaker engagement than the provision of narratives.* This is simply the converse.

Issue Framing Hypothesis: *Policymakers will exhibit greater average levels of engagement in response to issue frames emphasizing the economic and technological competitiveness dimensions of AI, as compared to issue frames emphasizing the ethical and social dimensions of AI.* This likely reflects the under-valuing of ethics in policy and the influence of greater constituent attention to economic and security issues, especially in the context of technology policy (B. D. Jones et al., 2009; Schiff et al., 2021).

Exploratory Hypotheses:

Strategies by Issue Framing Hypothesis: *Policymakers will respond with greater engagement to narratives when they are provided issue frames emphasizing the ethical and social dimensions of AI as compared to issue frames emphasizing the economic and technological competitiveness dimensions of AI.* I expect that the provision of narratives and expertise might influence legislators differently in the context of the issue frames of ethics and economic competition. While legislators may engage with ethics at a lower rate overall, I expect that they might find narratives particularly appealing or useful when AI is framed in terms of its ethical dimensions.

Prior Experience Hypothesis: *Compared to legislators with greater prior experience in AI policymaking, legislators with less experience with AI will respond with greater engagement to the expertise treatment.* I expect that there might be meaningful differences in how legislators respond to narratives and expertise based on their prior experience with AI as measured by the degree of prior state-level policymaking activity. This results from AI's status as a complex, technical policy area that has raised significant concern about the lack of government and policymaking expertise (White House, 2019). As such, I expect legislators without prior knowledge of AI will seek out expertise at a higher rate.

4.3 Experimental Design

4.3.1 Benefits of a Field Experiment

The research design used to answer the hypotheses is a field experiment, specifically an informational audit or correspondence experiment that takes place over email. The experimental design has several advantages over other methods used for studying policy entrepreneur-policymaker relationships. First, policy entrepreneur-policymaker engagement patterns, including the kinds of influence strategies and issue frames deployed and their subsequent effectiveness, may be endogenously determined. That is, policymakers may actively solicit information from, and be targeted by, certain policy entrepreneurs because of a policymaker's prior expertise and engagement with a policy issue.[2] As such, policymaker responsiveness to policy entrepreneur influence efforts may reflect self-selection (as a function of prior experiences, ideology, and institutional roles) rather than solely the effectiveness of the influence strategy itself (Bennett & Iyengar, 2010), making causal identification difficult. An experimental design helps to avoid this problem (Druckman et al., 2012) by randomly assigning policymakers to different influence strategies, enabling valid comparisons of the effects of different influence strategies across groups of policymakers

[2] For example, while it may be possible to examine observationally and descriptively which policy entrepreneurs submit comments in response to government requests, testify at hearings, or meet with policymakers, it is difficult to disentangle why these influence pathways were opened and whether they are effective.

who are, on average, similar across both observable and unobservable characteristics (and thus equally likely to self-select into different messages).

Second, policy entrepreneurs may employ multiple influence methods concurrently (e.g., in-person meetings versus public persuasion efforts), making it difficult to distinguish which if any are effective in a broader influence campaign. Identifying a particular type of policy entrepreneur contact (i.e., email communication) allows us to isolate effects for this channel with confidence. Third, and importantly, while previous studies have used online surveys of state legislators to study agenda-setting influence, a field experiment also offers the benefit of heightened external validity by ensuring that the treatments, context, and outcomes accurately reflect real-world dynamics of interest (Gerber & Green, 2012). Finally, the use of narratives juxtaposed with facts in the experimental design may itself constitute an innovation, and one necessary for empirically measuring and comparing the effectiveness of key policy entrepreneur influence strategies as called for by Petridou and Mintrom (2021). As Haaland et al. (2021, p. 17) argue in a recent review, "experiments systematically studying the role of stories, anecdotal evidence and narratives are still very scarce, and...a fruitful area for future research."

The study is done in partnership with a leading AI think tank (McGann, 2020), The Future Society, which is especially beneficial as I seek to measure authentic engagement with policy entrepreneur strategies in the context of policymakers going about their everyday job activities. The Future Society is a nonprofit think-and-do tank focused on AI policy that consults with and provides resources for policymakers, amongst other activities (The Future Society, 2022). Coordination with this organization helps to ensure not only the authenticity of the treatments, but also enhances the utility of the study for policymakers and their staffs by providing access to a leading organization with which they can connect further in the future. Specifically, the email correspondences come from one of the organization's email accounts and use the organization's branding. Additionally, as part of the study, the organization hosted resources on their website and co-developed and hosted a webinar on AI policy targeted at informing state policymakers.

4.3.2 Study Sample

The study sample includes 7,356 U.S. state legislators, or approximately all legislators[3] with email addresses that were available through official state legislative websites as of May, 2021.[4] There are a few reasons for studying state legislators. In contrast to federal legislators, state legislators may have a greater need for information about AI, as state-level AI policymaking is (generally) at a more early stage of development, and state legislature capacity and professionalism is generally lower (Squire, 1992). Further, policy process scholars have argued for the importance of expanded attention to local policymaking in order to validate and build theory (Ridde, 2009). Finally, this body of policymakers constitutes a relatively robust sample size, providing more power to evaluate the hypotheses. Table C.1 in Appendix C.1 presents key descriptive statistics about the sample.

4.3.3 Randomization and Treatment Assignment

I randomly assign state legislators within blocks to email treatments, following the procedure of Butler and Broockman (2011).[5] Specifically, I block randomize by state, legislative chamber, and political party (198 blocks total, as Nebraska has a unicameral legislature).[6]

[3]Note that by sending email communications to state legislators' email addresses, I am effectively treating the legislators' *offices*, as it is possible that staff members rather than the legislators themselves would receive the treatment. I consider this to be a *feature* of the normal environment in which policy entrepreneurs attempt to influence legislators (rather than a bug), and consider engagement by members of a legislative office to be indicative of the policymaker's activities and priorities.

[4]Of the total of 7,383 state legislative seats, according to the National Conference on State Legislators, I am able to identify 7,358 individuals with associated email addresses (99.7%). For the 25 remaining legislators, approximately 15 of the seats were vacant at the time of data collection, while around 10 did not have publicly-available email addresses. In addition, I exclude two Oklahoma legislators who were mistakenly associated with the same email address. Note that regular resignations, retirements, and special elections mean the sample was likely not fully up-to-date at the time of administration.

[5]Block randomization has several benefits here. It ensures balance across key covariates, avoiding the possibility of aberrant randomizations and allowing for simple estimation of subgroup effects. More importantly, to the extent that the blocking variables are correlated with outcomes, blocking decreases variance, effectively increasing power.

[6]Given the small number of independents (67 of the total 7,356), I randomly assign these to either a Democratic or Republican block.

Policymakers within blocks are then randomly assigned to one of six treatment or control groups. They receive an email communication that draws on either the 1) expertise strategy or 2) narrative strategy or 3) a more generic control message employing neither strategy. The messages also incorporate one of two issue frames, emphasizing either the a) social and ethical or b) economic and technological leadership dimensions of AI. This produces an overall 3x2 factorial design with six total treatment/control groups.

The wording of the email treatments is designed to emulate the issue frames and language used in real news media stories and policy entrepreneur communications with policymakers. In particular, the emails discuss either ethical and social harms associated with facial recognition,[7] or economic and technological leadership implications of U.S. competition with China, prominent topics discussed for the respective issue frames.[8] Moreover, when constructing the narrative strategy treatments, I adopt a personal narrative approach (Mcentire et al., 2015) and include the core elements of narrative structure according to the NPF: a setting, characters (a victim and a villain), a plot, and a policy moral (McBeth et al., 2014). The emails are also designed to emulate newsletters sent out by the partner organization, and were vetted by several members of the staff. Figure 4.1 presents a draft example with additional formatting elements, and the full text and template for the emails is available in Appendix C.5.

The email messages contain links to extended versions of information presented, in the

[7]For the expertise strategy employing the ethics frame, I adopt language from policy entrepreneur comments about facial recognition (Saleh, 2020) in response to a request for comments from the Office of Management and Budget (2020) regarding key guidance for U.S. agencies about AI regulation, as well as language from advocacy statements in the U.S. and internationally put out by Access Now (2021) and European Digital Rights (EDRi) (2021), collectively signed by a coalition of over 175 civil society groups, also regarding facial recognition. For the narrative strategy employing the ethics frame, I rely on language in news articles (Hill, 2020a, 2020b) and an advocacy statement by the ACLU (2021) and numerous other signatories about the wrongful arrest of individuals because of facial recognition systems.

[8]For the expertise strategy employing the competition frame, I adopt language from a report by the Center for Data Innovation about international competition over AI (Castro & McLaughlin, 2021). For the narrative strategy employing the competition frame, I rely on language in news articles (Vincent, 2017) about how AlphaGo's defeat of Chinese Go champion Ke Jie launched a "Sputnik-like" moment for China. Critically, the issue frames and arguments made in these documents are consistent with language used by a wide range of policy entrepreneurs and other actors in different contexts.

Artificial Intelligence - What to Know as a Legislator in Alabama

TS The Future Society
Tue 10/26/2021 1:36 PM
To: Schiff, Daniel S

Dear Representative Schiff,

We at The Future Society are reaching out to you to share a compelling story about the **important social and ethical implications of artificial intelligence (AI)** and to **invite you to a webinar** on what you need to know about AI as a state legislator. We believe that state legislators such as yourself have an important role to play in shaping Alabama's response to these critical social and ethical issues.

When Robert Julian-Borchak Williams went to work in his office at an automotive supply company in Detroit, he had no idea he would be handcuffed and arrested later that day in front of his wife and two young daughters. That day, Robert became one of the first Americans wrongfully arrested because of a false match of a facial recognition algorithm, an example of how faulty or misused AI algorithms can go awry. To hear more about how AI went wrong in this case and what policymakers can do about the social and ethical implications of AI, please **read more about Robert's story**.

How can The Future Society support you?

- **RSVP for our webinar** on Monday, December 13 at 2PM ET / 11AM PT on **what you need to know about AI as a state legislator**, with speakers from The Future Society, Georgia Tech, and the National Conference of State Legislatures.
- **Share your thoughts**: What is the top concern or hope you have about AI policy? Let us know in a reply to this email and we'll try to address your comments in our webinar!

Thank you for your time and consideration of these important issues,

The Future Society

THE
FUTURE
SOCIETY

Figure 4.1: Sample legislator email: Narrative + ethics treatment condition

form of fact sheets (for the expertise conditions), stories (for the narrative conditions),[9] or a link to the organization website (for the control conditions). The fact sheets and stories are crafted based on the materials mentioned and additional research and reporting on these topics, and were also vetted by several members of the partner organization. The email messages additionally contain an encouragement to reply to the email and a link to RSVP for a webinar on AI for state legislators, planned and conducted in conjunction with the partner organization in December 2021 to benefit study participants.

[9]The full fact sheets and stories are hosted on the organization website and available by request. See Appendix C.5.

4.3.4 Outcome Measures

The outcomes of interest are measures of policymaker engagement with the email communications. Because policymaker time is limited, policy entrepreneurs must act strategically and compete to gain their attention in order to influence the policy agenda. Therefore, the willingness of policymakers to devote attention to the information or arguments provided by policy entrepreneurs is an important indication of their interest, engagement, and ultimately the effectiveness of policy entrepreneurs influence strategies.

I evaluate policymaker engagement with the emails by measuring willingness to partake in a series of actions that demand increasing levels of effort, time, and attention: clicking on links to additional information discussed in the email (a fact sheet, story, or the organization's website), clicking on a link to sign up for a webinar, replying to the email, and attending the webinar,[10] all behaviors that are reflective of real-world policymaker-policy entrepreneur influence dynamics. As the primary outcome measure, I construct a binary variable which indicates whether legislators participated in at least one of these activities.[11] Table 4.1 presents the average rate of engagement for each outcome of interest in the control group, as a baseline for understanding legislator behavior and the effectiveness of the various treatments.

Table 4.1: Average legislator engagement in control group

Outcome	**Rate**
Resource Click Rate	10.0%
Webinar Click Rate	9.7%
Email Reply Rate	0.1%
Webinar Attendance Rate	0.2%
Combined Engagement (Binary)	10.9%

Importantly, the study involves non-compliance as not all legislators opened and read the contents of the assigned email given limited time and resources. Further, for a portion

[10]Measuring both intent to participate and participation is a strategy also employed by McClendon (2014).

[11]This measurement approach deviates slightly from the pre-registration, which proposed using a count measure. I chose to make this alteration because the binary outcome is easier to interpret and very highly correlated with the count measure. I present additional results for different outcomes separately in Appendix C.3.

of the sample, emails bounced or were otherwise undeliverable, as is typical for email campaigns. Legislators who *do* open the email are identified as compliers with treatment, which is important for calculating the complier average causal effect (CACE), the primary treatment effect of interest for this study.

4.3.5 Administration

Emails were sent through the partner organization's email service (Google Suite) using a third party platform, Saleshandy, to track email open rates and link clicks. Given daily email rate limits and to successfully deliver more emails to primary inboxes, I sent out emails over several weeks. After block randomizing state legislators into treatment groups, I further randomized the state legislators within blocks to receive their treatment emails on specific days and at specific times. I sent out two follow-up emails (one 10 days after the initial email and another 5 days after the first follow-up email) to individuals that had not opened the email, in order to increase response rates. After all reminders were sent, 48.4% of the 7,356 legislators emailed opened the email (increasing to 50.0% amongst legislators for whom the email did not bounce), exceeding typical open rates of around 29% for email campaigns in the government sector.[12]

4.3.6 Ethical Considerations

Given the ethical implications of the research design, I took care to follow best practices from other correspondence studies in political science, policy, and public administration. In particular, the study aimed to satisfy three main ethical considerations: reducing deception, minimizing harm, and minimizing burden (Butler & Broockman, 2011). I reduced deception involved in the study by partnering with a real organization that has expertise and a vested interest in this policy domain and the study's research questions. Policymakers received accurate information about the issue frames and topics, representative of

[12]Source: 2019 Mailchimp email benchmarks available at https://mailchimp.com/resources/email-marketing-benchmarks.

typical AI policy discourse, through the email messages, links, and during the webinar. Next, because the information is carefully-researched and designed to be an accurate depiction of key problems and solutions in AI policy, I expect that the study benefited many policymakers by providing useful information, potential contacts, and follow-up resources curated explicitly for legislators.[13] Finally, the time burden on legislators and their staffs resulting from opening or reading the initial email was minimal and a standard part of work activities. Beyond this, participants voluntarily opted into additional activities, such as the webinar, likely due to genuine interest and a desire to learn more about this increasingly important policy domain.

4.4 Analysis Strategy

4.4.1 Estimands and Regression Models

Due to randomization, treatment and control groups are identical in expectation before treatment. Thus, causal effects can be identified straightforwardly by assessing outcome differences between treatment groups. However, as the study seeks to evaluate the effectiveness of influence strategies that *actually reach* policymakers, I prefer CACEs as the estimands of interest, referring to average treatment effects identified for those who actually receive treatment (i.e., those who open the emails).[14] To recover the CACEs, the study employs an instrumental variable (IV) approach with two-stage least squares (2SLS) regression, where random assignment to treatment is an instrument for opening and reading the treatment email (Angrist et al., 1996).

Below, I describe the analysis approach used to evaluate each hypothesis. For all associ-

[13]One additional ethical question surrounds the study's use of a frame invoking technological competition between the United States and China. While this frame no doubt accurately reflects prominent policymaker AI discourse, I worked with the partner organization to minimize possible harms related to the promotion of an 'AI arms race' (Ulnicane, 2022; Zwetsloot et al., 2018), including by modifying language and presenting balanced content in the additional resources and webinar.

[14]In Appendix C.3, I also report key intent-to-treat effects (ITTs) as a measure of whether policy entrepreneur messages are influential in a broader sense, given that non-compliance is common in real-world settings. CACEs are defined as the ITT effects of treatment assignment on the outcomes of interest divided by the proportion of compliers.

ated regression specifications, I use two-tailed *t*-tests and Huber-White heteroskedasticity-robust standard errors, with statistical significance assessed primarily at the 5% level.[15] Additional information about covariates and power can be found in Appendix C.2 and Appendix C.4.

Policy Entrepreneur Effectiveness Hypothesis

For the Policy Entrepreneur Effectiveness Hypothesis, I compare legislator engagement for legislators in the narrative and expertise treatment conditions (both pooled and individually) to those in the control groups. The main estimands of interest are:

$$E[Engagement(D = narrative) - Engagement(D = control)|complier]$$

$$E[Engagement(D = expertise) - Engagement(D = control)|complier]$$

These CACE estimands represent treatment effects among the compliers, where the compliers are those who took up treatment (opened and read the treatment emails) when assigned to treatment. As the control group, by design, does not and cannot receive the treatment emails, this context involves the potential for only one-sided non-compliance. The associated regression specifications are thus of the following form:

$$\text{CACE 1st stage: } \textit{Read narrative} = \beta_0 + \beta_1 narrative + \gamma X + \epsilon$$

$$\text{CACE 2nd stage: } Engagement = \beta_0 + \beta_1 \widehat{\textit{read narrative}} + \gamma X + \epsilon$$

$$\text{CACE 1st stage: } \textit{Read expertise} = \beta_0 + \beta_1 expertise + \gamma X + \epsilon$$

$$\text{CACE 2nd stage: } Engagement = \beta_0 + \beta_1 \widehat{\textit{read expertise}} + \gamma X + \epsilon$$

where $Engagement$ refers to the outcome measure of interest, **X** refers to a vector of the covariates, and ϵ refers to the error.[16] The covariates included are legislator party,

[15] Another minor deviation from the pre-registration is that I registered some one-tailed and some two-tailed *t*-tests. I opt for all two-tailed *t*-tests here to both be more conservative and preserve easy comparability.

[16] Note that much of this analysis involves pooling across the issue frames.

chamber, gender, and tenure, as well as state-level prior experience with AI policy and legislature professionalism. For the 2SLS IV models, *read narrative* and *read expertise* represent whether legislators opened and read the respective treatment emails (were treated)[17], and the predicted values of those variables from the first stage are then used to estimate policymaker engagement in the second stage, where the β_1 coefficients are the CACEs of interest. I expect the β_1 estimates from the second stage 2SLS regressions to be positive, indicating that the treatments increased policymaker engagement.

To assess the relative effectiveness of strategies invoking narratives or expertise, I use a z-test to evaluate whether the CACE estimates for the narrative and expertise treatments are equivalent. A significant p-value implies that we can reject the null hypothesis that the two treatments produced equivalent effects. In this case, the relatively more influential strategy is the one with the larger treatment effect magnitude.

Issue Framing Hypothesis

For the Issue Framing Hypothesis, I compare legislator engagement for legislators in the ethics treatment group to those in the competition treatment group. I expect that the competition issue frame will produce stronger effects. The associated regression specifications are similar to the 2SLS models presented above, but using issue frames instead of influence strategies.

Strategies by Issue Framing Hypothesis

For the exploratory Strategies by Issue Framing Hypothesis, I compare legislator engagement for each of the six distinct treatment groups to assess whether the ethics narrative treatment effect is larger than the competition narrative treatment effect. To do this, I use interactive model specifications to estimate ITTs for each unique treatment as it is more complicated to define compliance for both issue frames and strategies simultaneously when

[17] Similar to McClendon (2014), I consider opening the email to be a proxy for reading the email.

using a 2SLS approach (note that ITT estimates will generally have the same significance as CACE estimates). I then use z-tests to assess treatment effect differences.

Prior Experience Hypothesis

For the exploratory Prior Experience Hypothesis, I interact the expertise treatment with legislature prior experience with AI policy, and expect to find that policymakers in states with less prior work on AI policy have greater need for expertise. To investigate this, I use interactive 2SLS model specifications, again similar to inform to those shown above, to estimate CACEs for legislators in states with high and low prior AI experience.

4.5 Results

4.5.1 Does the Provision of Expertise or Narrative Influence Legislators?

I first consider whether the Policy Entrepreneur Effectiveness Hypothesis holds. In short, does the use of narrative or expert information by policy entrepreneurs successfully engage policymakers as compared to more generic outreach? Table 4.2 presents the results for the main legislator sample based on use of either strategy (pooled) and for both strategies separately.[18] The outcome of interest is the CACE based on a binary indicator of whether legislators engaged in at least one of the possible activities: opening a fact sheet provided in the email, clicking to register for the webinar, etc.[19] A corresponding table of ITT results is presented in Appendix C.3. Note that for the primary sample, I exclude legislators from Indiana, as the unusual click behavior (100% click rate on resources, and 0% click rate on webinar registration page) that I observed during administration suggested that an automated email system (perhaps for security reasons) was in use.

[18]For all three model specifications, p-values for the weak instrument and Wu-Hausman tests are all less than 0.001, confirming strong instruments and the suitability of an IV approach.

[19]In practice, almost all legislators who engaged did so through one of these two lower-effort activities, such that the binary outcome essentially reflects engagement with the fact sheet (or story) or interest in the webinar. In line with expectations, relatively few legislators replied to the email or attended the webinar. Inclusion or exclusion of those more costly activities thus has essentially no effect on study results, as shown in Table C.4 in Appendix C.3.

Table 4.2: Impact of policy entrepreneur strategies on legislator engagement

	Legislator Engagement		
	(1)	(2)	(3)
Either Strategy	0.096^{***}		
	(0.016)		
Expertise		0.092^{***}	
		(0.019)	
Narrative			0.100^{***}
			(0.019)
N	7,206	7,206	7,206
Robust SEs	Yes	Yes	Yes
Covariates	Yes	Yes	Yes

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

I find that the use of both strategies by policy entrepreneurs is indeed effective, confirming the Policy Entrepreneur Effectiveness Hypothesis. When compared to a control message of a similar style and length which provides more general information about the organization—including the same invitation to join a webinar and reply to the email—both expert information and narratives significantly increase policymaker interest. The corresponding effect size of the CACE is 0.096 ($p < 0.01$) for those who received either strategy compared to control, with effects slightly larger in the narrative group (0.100) versus in the expertise group (0.092). Thus, the probability that a legislator in one of the treatment groups who opened and read the email took at least one action is 9-10 percentage points than for legislators in the control group.

Figure 4.2 presents these results using a coefficient plot with 95% confidence intervals, and includes alternative model specifications as robustness checks. Namely, in addition to the 1) primary specification, I present results 2) without covariate adjustment, 3) when including Indiana, and 4) when excluding states for which I observed identical click rates for the resource and webinar.[20] Across specifications, results are nearly identical.[21]

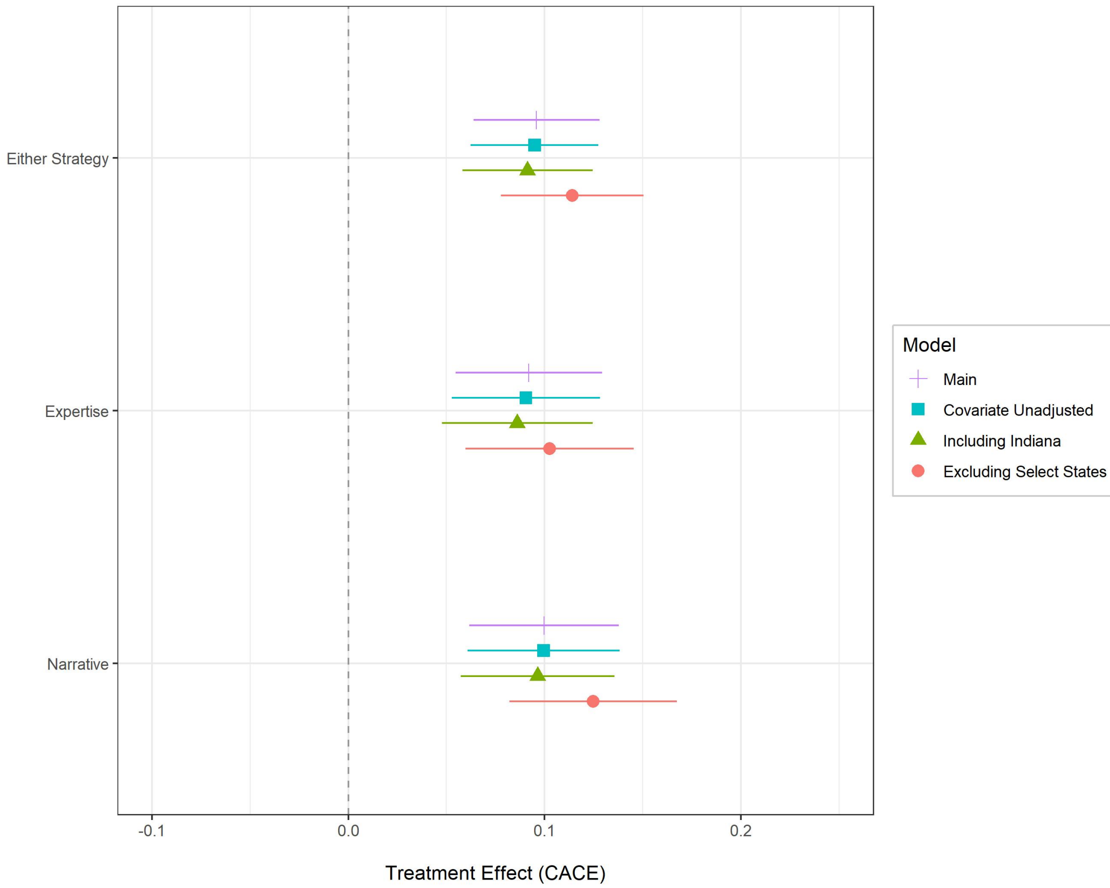

Figure 4.2: Impact of policy entrepreneur strategies on legislator engagement

More concretely, subsetting to individuals who opened their assigned emails, approximately 24.3% of legislators in the control groups took at least one action upon receiving the emails compared to 31.2% of legislators in the expertise treatment group and 32.6% of legislators in the narrative treatment group.[22] This corresponds to a 28.3 percent increase in engagement for those who received expert information, and a 34.2 percent increase in

[20]Similar to the logic for excluding Indiana, it is possible that these states employ automated email systems to screen links for security reasons. This select sample retains about two-thirds of the full sample, with $n = 5,268$ total legislators.

[21]I also find the same results when using the alternative count outcome measure as when using the binary outcome measure. See Table C.4 in Appendix C.3.

[22]Note these metrics for the control group differ from those reported in Table 4.1, which presents engagement rates for individuals regardless of whether they actually opened their assigned emails (akin to ITTs). The metrics here are amongst those individuals who opened either their assigned control or treatment emails (akin to CACEs), and are thus arguably more meaningful representations of differential behavior.

engagement for those who received narratives.[23]

While both influence strategies employed by policy entrepreneurs are statistically effective, they are not statistically distinct. Though the narrative treatment leads to a larger magnitude increase in legislator engagement as compared to the expertise treatment, a z-test indicates these two strategies are not statistically distinguishable ($p = 0.77$). Thus, while there are theoretical reasons to expect either narratives or expertise to be more influential strategies (as per the Dominance of Expertise and Narrative Hypotheses), I instead find that these strategies are similarly, and meaningfully, impactful. Nevertheless, it is most striking overall that for a highly technical policy domain that has been inundated with calls for expertise-building, such as massive efforts to promote training of STEM workers and PhD researchers, narratives are at least as critical in engaging policymakers vis-à-vis the emerging AI policy agenda.

4.5.2 How do Issue Frames Impact Legislator Engagement?

Next, I turn to the question of whether use of distinct issue frames by policy entrepreneurs affects engagement by policymakers. I anticipated as per the Issue Framing Hypothesis that an issue frame (or policy frame) emphasizing economic and technological leadership dimensions would be more influential for policymakers, given substantial focus on AI's innovative potential in policy discourse, and because policymakers may already be inclined to emphasize economic-type issues over ethical ones.

To my surprise, results point in the opposite direction, if anything. Specifically, according to the covariate-adjusted IV regression results, legislators receiving the ethics frame via email were about 2 percentage points more likely to take at least one action then legislators receiving the economic frame, though this difference is not statistically significant ($p = 0.24$). To put this into context, the covariate-unadjusted group differences indicate

[23] When including the entire sample, not just those who opened their emails, 10.9% of individuals in the control group engaged in at least one action, compared to 15.4% in the expertise treatment group and 15.7% in the narrative group, translating to engagement increases of 40.7% and 43.7%, respectively.

that around 29.0% of legislators who opened the email in the competition group took at least one action as compared to around 30.0% of legislators in the ethics group. Thus, despite ample and arguably growing pressure to emphasize AI's economic, competitive, and geopolitical dimensions in policy discourse, legislators were at least as likely, if not more likely, to express interest in AI's ethical implications.

I also considered whether specific issue frames interacted differently with various policy entrepreneur strategies. For example, I hypothesized as per the Strategies by Issue Framing Hypothesis that the use of a narrative strategy would be more naturally coherent and impactful when joined with an issue frame emphasizing social and ethical issues, as compared to joining a narrative strategy with an economic frame. Table 4.3 displays the engagement rates for combinations of strategies and frames, along with corresponding p-values of differences across these combinations. While I hypothesized that narratives would be differentially engaging based on the issue frames employed, I find that narratives are equally engaging across the ethics and competition issue frames ($p = 0.91$). Moreover, I find no clear evidence of differences between issue frames for the control ($p = 0.56$) and expertise ($p = 0.16$) policy entrepreneur strategies as well. Instead, the greatest visible differences are between each influence strategy and the control group (generic outreach message that does not employ either strategy).

Table 4.3: Engagement rates for distinct strategy and issue frame combinations

	Ethics	**Competition**	p-value of diff.
Control	0.11	0.11	0.56
Narrative	0.16	0.16	0.91
Expertise	0.16	0.14	0.16
p-value (Narr. vs. Control)	0.00	0.00	
p-value (Exp. vs. Control)	0.00	0.01	
p-value (Narr. vs. Exp.)	0.60	0.32	

To evaluate the stability of this pattern, I draw on an additional source of data. Through the use of tracking links tied to each individual strategy x issue frame combination, I was able to track through the registration platform for the webinar, Eventbrite, how many individuals in each treatment group sought out information about the webinar.[24] The data in Table 4.4 reveal a very similar pattern. There were minimal differences in page views *across* issue frames, but distinct differences for both narratives and expertise relative to control *within* each issue frame. For example, for the control group, a very similar number of individuals visited from the ethics frame group (376) as from the competition frame group (366). In the expertise group, a larger number of individuals visited overall, with a highly similar number of visitors from the ethics group (421) and competition group (433). The narrative group acted similarly, with the highest number of visitors overall and a nearly identical number from the ethics group (465) and competition group (466).

Table 4.4: Webinar registration views by treatment group

	Ethics	**Competition**
Control	376	366
Narrative	465	466
Expertise	421	433

That these patterns are highly stable regardless of the use of quite distinct types of issue frames suggests that *it is the strategies themselves that are most salient in policy entrepreneur influence efforts*. Moreover, that the number of visitors from the narrative group was substantially larger than the control group (90-100 more views) as well as larger than the number of visitors from the expertise group (30-40 more views), provides further evidence about the relative effectiveness of the influence strategies compared to control. It even provides suggestive evidence about the potentially heightened effectiveness of narratives versus expertise. However, given more limited data about the source of these numbers (e.g., uncertainty around how many of these views are repeat visitors), and because the pre-registration did not emphasize this data source, we cannot safely conclude that narratives are more effective.

[24]The data captured by EventBrite on the number of individuals visiting the registration page differ from those captured through SalesHandy based on direct link clicks from the emails. Eventbrite is able to track any use of these links to visit the site, accounting for the larger overall number of visits to the registration page. I rely on the SalesHandy data for the primary results, given the pre-registered plan, but present the Eventbrite data as a helpful addition.

4.5.3 Engagement by Legislative Prior Experience and Capacity

Finally, I consider whether certain legislative characteristics moderate the effectiveness of policy entrepreneur influence strategies. I hypothesized as per the Prior Experience Hypothesis that legislators with less prior experience in AI policy would be especially in need of expert information to advance their ability to work effectively in this policy domain. As a proxy for legislator experience with AI policy, I use data from the National Conference of State Legislatures (NCSL) indicating the number of proposed and/or passed pieces of legislation in each state between 2019 and 2021.[25] I binarize these count data such that the top 50% of states are considered to have high prior experience with AI policy, with the rest having low experience. This essentially divides states into those who have considered or passed at least one piece of AI legislation, and those who have not. Note that this means the experience variable is measured at the state level, and is thus only a rough proxy for individual legislator experience.

Table 4.5: Impact of policymaker AI experience on engagement with expertise

	Legislator Engagement
Expertise	0.040
	(0.029)
Low AI Experience	-0.052^{***}
	(0.014)
Expertise x Low AI Experience	0.103^{***}
	(0.038)

Note $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$
With robust SEs, including covariates

Table 4.5 reports the results of interacting the expertise influence strategy with legislator prior AI policymaking experience. The coefficient on the interaction term is positive (coef $= 0.103$) and highly significant ($p < 0.01$), indicating that policymakers in states with low AI experience are indeed more likely to pursue expert information than policymakers on

[25] This time period is appropriate given the relative recency of AI policy efforts at the state level.

states with greater AI experience.

While these results comported with expectations, further exploration revealed an additional surprising finding. Low experience legislators were not only more likely to seek out expertise; they were also more likely to seek out narratives. Figure 4.3 displays the heterogeneous effects of both policy entrepreneur strategies for low versus high experience legislators. Compared to high experience legislators, low experience legislators were 10.3 percentage points more likely to take one activity in response to an email providing expert information, but also 12.1 percentage points more likely to act in response to receipt of a narrative about AI policy.

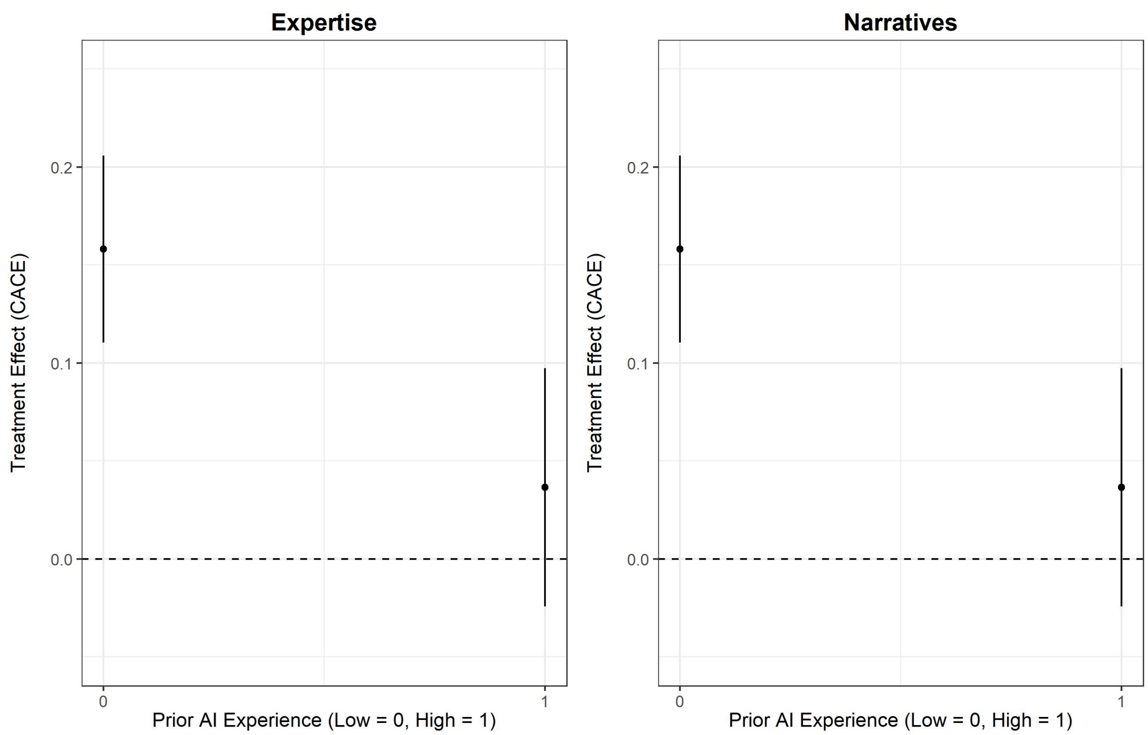

Figure 4.3: Prior legislature experience impacts on seeking expertise and narratives

I also explored a related but distinct issue, which is how legislative professionalism (or capacity) moderates the effectiveness of policy entrepreneur strategies. I use the adjusted Squire Index as the measure of state legislative capacity, a continuous measure ranging from 0 to 1. Figure 4.4 displays heterogeneous effects for a model interacting influence strategy with legislature capacity, in this case using the ITT effects rather than CACEs. Unsurprisingly, the main effect in the model for the Squire Index indicates that legislative capacity drives significant increases in AI policy engagement, with around a 50 percentage point increase in taking at least one action in response to the emails.

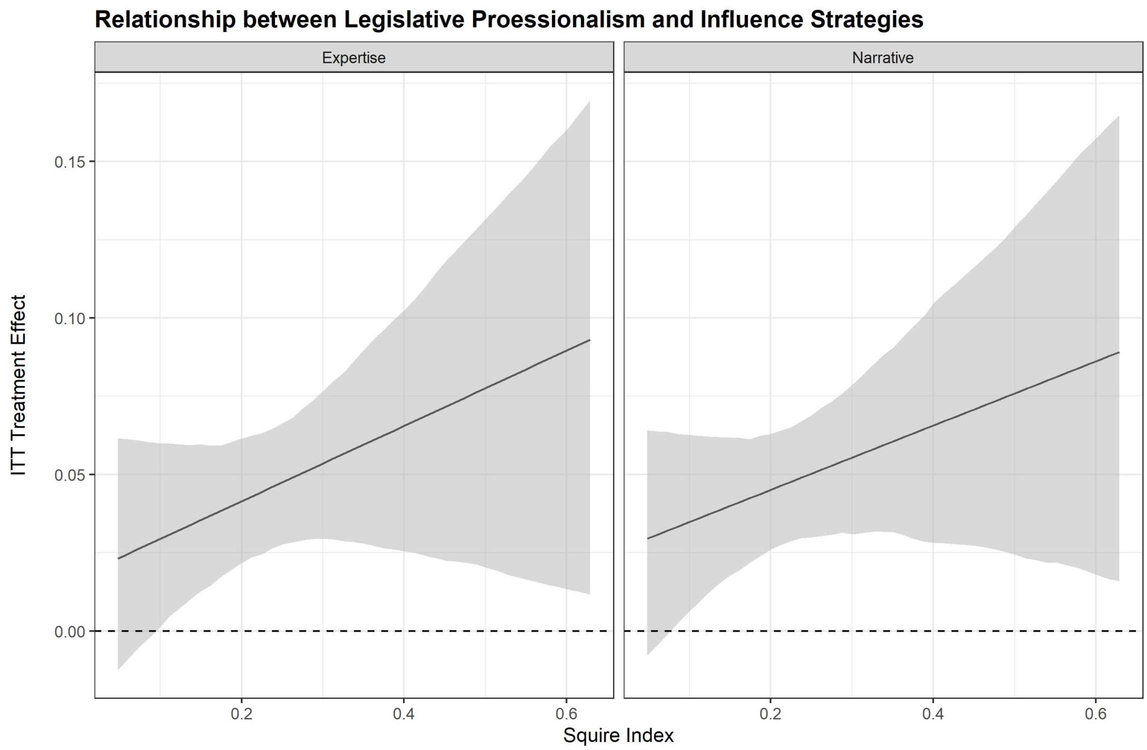

Figure 4.4: Relationship between legislative capacity and influence strategy effectiveness

The figure and coefficients on the interaction terms suggest that legislators working in professionalized legislatures are also more likely to be engaged by both the expertise and narrative strategies. Yet, while professionalism seems to drive increased engagement with narratives (a 10.5 percentage point increase in taking action) and expertise (a 12.1 percentage point increase in taking action), the variance is too high to support a clean conclusion (p-value for expertise $= 0.17$; p-value for narratives $= 0.24$).[26] Nevertheless, Figure 4.4 shows that the overall patterns for legislators who received the expertise treatment are very similar to the patterns for legislators who received narratives, another indication of the

[26] The power analyses in Appendix C.4 suggested some of the more highly-specified estimands might be underpowered, though the study is limited to the sample of all U.S. legislators ultimately.

importance of narratives in the agenda-setting process.

The additional exploratory and associational findings in this subsection provide helpful context for evaluating potential policy entrepreneur influence efforts. Higher-capacity legislatures are unsurprisingly better positioned to engage with policy entrepreneurs and to be responsive to information about critical policy issues (though we cannot conclude statistically that the expertise and narrative strategies are differentially effective for these policymakers compared to more generic outreach). More strikingly, legislators in states with little prior AI policy work are especially likely to engage with policy entrepreneurs, and they are receptive to both expert information and narrative strategies. Considering the study results in their totality, we see one especially surprising and consistent pattern: Narratives appear to be at least as influential as expert information, even for a technology like AI known for its complexity and technical opacity.

4.6 Discussion

Highly technical, complex policy domains have typically been expert-dominated spaces. This is true nowhere if not with respect to AI policy. Actors in the public, private, and non-governmental sectors — in the United States and globally — have urgently called for expert information, training of more STEM researchers and PhDs, and capacity-building in government. Yet despite the seeming ubiquity of the expert orientation in AI policy, I find that persuasive narratives are at least as effective in engaging state legislators as they work to formulate the early AI agenda.

The findings demonstrate that, while provision of expert information by policy entrepreneurs remains influential, the provision of narratives was at least as likely to gain policymaker attention. Legislators were around 30 percent more likely to take costly actions such as clicking on an additional fact sheet or story when they received either expert information or a persuasive narrative. Moreover, this pattern of influence does not appear to be isolated to policy discourse circumscribed around narrow aspects of AI, such as only applying to issue frames that emphasize social and ethical harms. Indeed, whether policy entrepreneurs emphasized social and ethical dimensions of AI, or implications related to economic and technological competitiveness, narratives remained equally influential.

Further and surprisingly, despite the traditional association of science and technology policy with high-level concerns surrounding economic growth and innovation, an issue frame promoting ethical consideration was at least as effective as an economic-style frame. This finding speaks to the literature on issue framing and the relative effectiveness of disparate issue frames, particularly for highly technical policy domains. A topic for further study is how these results might extend to other policy domains, technical or otherwise: The findings here could in part reflect esoteric aspects of AI as a policy object, or they could reflect a broader paradigm shift in how technology is governed.

Relatedly, a possible limitation of the study and common challenge with domain-specific research is the extent which it can be generalized to other policy domains. A further threat to inference for this study is the possibility of spillover effects due to legislators talking to each other about the emails. Yet, most outcomes are actions likely performed immediately after reading the email, such as clicking on links, replying to the email, and registering for the webinar. Finally, an open question is the extent to which initial policymaker engagement translates to meaningful agenda-setting influence and policy change. That approximately a thousand legislators took the costly action of clicking on links to additional resources suggests the potential for influence is more than minimal.[27] Overall, the use of a field experiment approach, conducted in partnership with the leading AI think tank, buttresses the study's internal and external validity.

One ambition of policy scholarship is to provide actionable insights to guide policy actors (Anderson et al., 2020). I find here that policy entrepreneurs, in this case a nonpartisan civil society organization, can make effective use of both expert information and narratives in their influence and advocacy efforts. Further, the results suggest that legisla-

[27] An additional if anecdotal indication is that the study resulted in the author being contacted by state legislators to advise in adoption of AI policy by government and to help advise on writing AI legislation.

tors in states without much prior policy experience are especially inclined to seek out expert information and narratives. Future work is needed to better understand why policymakers seek out these narratives, and which features of the narratives are appealing. It may be that policymakers are drawn to narratives as a way of developing their own messaging for political deliberation and constituent communications. It may also be that policymakers are directly persuaded by narratives much in the way that members of the public are. The results do caution that legislators in less professionalized legislatures may struggle with limited capacity to engage with policy entrepreneurs, meaning that expanded efforts are needed to reach, and understand the needs of, these legislative bodies, an ongoing problem in state-level policymaking (Fortunato & Parinandi, 2022).

More broadly, this study helps to realize ambitions in policy scholarship to unpack the 'black box' common in complex policy processes, to integrate NPF theory with agenda-setting dynamics, and to rigorously measure and compare the effectiveness of different policy entrepreneur influence strategies (McBeth & Lybecker, 2018; Petridou & Mintrom, 2021). In particular, this study demonstrates that narratives can be successfully applied to better understand policy entrepreneurship in the context of agenda-setting, and play a meaningful role even for the highly technical and complex domain of AI policy. Indeed, the complexity, uncertainty, and ambiguity associated with AI policy may help to explain why policymakers appear susceptible to diverse influence strategies and framing efforts despite an ostensible need for 'hard' evidence. Finally, while much of the literature on narratives focuses on their use in media or to influence the public, this study demonstrates that narratives can also be used by policy entrepreneurs to influence policymakers. Just as narratives can support the translation of policy conditions into problems in the eye of the public, so too can they shape the perspectives of policymakers who are wrestling with contested visions around policy agendas.

4.7 Conclusion

As policy agendas begin to form to address the prospects and risks associated with AI, it is critical to understand how influence strategies employed by interested policy entrepreneurs may shape policymaking going forward. This study employs an experimental approach to assess whether different influence strategies may be more or less successful, and in doing so contributes to the literature on agenda-setting especially in the context of highly technical domains and for general purpose technologies. These findings address topics central to policy process theory including policy entrepreneurship, issue framing, and influence dynamics in policy entrepreneur-policymaker relationships.

Based on a field experiment of over 7,300 U.S. state legislators, I find that policy entrepreneurs who employ narratives or expert information can successfully engage state legislators in taking costly actions like clicking on a fact sheet or resource. I find narratives are just as effective whether AI policy is framed as an economic or ethical issue. The results further indicate that legislators with less AI policy experience, and in highly professionalized legislatures, are most likely to engage with policy entrepreneurs. In the context of the complex, critical, and emerging domain of AI policy, the results reveal that passion may be just as likely as reason to shape the minds of policymakers as they establish the ultimate policy agenda.

4.8 References for Chapter 4

Access Now. (2021). *Open letter calling for a global ban on biometric recognition technologies that enable mass and discriminatory surveillance*. Access Now, Amnesty International, European Digital Rights (EDRi), Human Rights Watch, Internet Freedom Foundation (IFF), Instituto Brasileiro de Defesa do Consumidor (IDEC). Brooklyn, New York. https://www.accessnow.org/ban-biometric-surveillance/

ACLU Coalition Signatories. (2021). Coalition letter to President Biden on use of facial recognition technology. https://www.aclu.org/letter/coalition-letter-president-biden-use-facial-recognition-technology

Anderson, S. E., DeLeo, R. A., & Taylor, K. (2020). Policy entrepreneurs, legislators, and agenda setting: Information and influence. *Policy Studies Journal*, *48*(3), 587–611. https://doi.org/10.1111/psj.12331

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*(434), 444–455. https://www.tandfonline.com/doi/abs/10.1080/01621459.1996.10476902

Baumgartner, F. R., & Jones, B. D. (1991). Agenda dynamics and policy subsystems. *The Journal of Politics*, *53*(4), 1044–1074. https://doi.org/10.2307/2131866

Bennett, W. L., & Iyengar, S. (2010). The shifting foundations of political communication: Responding to a defense of the media effects paradigm. *Journal of Communication*, *60*(1), 35–39. https://doi.org/10.1111/j.1460-2466.2009.01471.x

Birkland, T. A., & Warnement, M. K. (2016). Refining the idea of focusing events in the Multiple-Streams Framework. In R. Zohlnhöfer & F. W. Rüb (Eds.), *Decision-making under ambiguity and time constraints: Assessing the Multiple-Streams Framework* (pp. 91–108). ECPR Press.

Brysk, A. (1994). *The politics of human rights in Argentina: Protest, change, and democratization*. Stanford University Press.

Butler, D. M., & Broockman, D. E. (2011). Do politicians racially discriminate against constituents? A field experiment on state legislators. *American Journal of Political Science*, *55*(3), 463–477. https://doi.org/10.1111/j.1540-5907.2011.00515.x

Cairney, P., & Jones, M. D. (2016). Kingdon's Multiple Streams Approach: What is the empirical impact of this universal theory? *Policy Studies Journal*, *44*(1), 37–58. https://doi.org/10.1111/psj.12111

Capano, G., & Galanti, M. T. (2018). Policy dynamics and types of agency: From individual to collective patterns of action. *European Policy Analysis*, *4*(1), 23–47. https://doi.org/10.1002/epa2.1031

Carmines, E. G., & Stimson, J. A. (1980). The two faces of issue voting. *The American Political Science Review*, *74*(1), 78–91. https://doi.org/10.2307/1955648

Castro, D., & McLaughlin, M. (2021). *Who is winning the AI race: China, the EU, or the United States? — 2021 update*. Center for Data Innovation. Washington, D.C. https://www2.datainnovation.org/2021-china-eu-us-ai.pdf

Cave, S., Craig, C., Dihal, K., Dillon, S., Montgomery, J., Singler, B., & Taylor, L. (2018). *Portrayals and perceptions of AI and why they matter*. The Royal Society. https://doi.org/10.17863/CAM.34502

Chong, D., & Druckman, J. N. (2007). Framing theory. *Annual Review of Political Science*, *10*(1), 103–126. https://doi.org/10.1146/annurev.polisci.10.072805.103054

Crow, D. A. (2010). Policy entrepreneurs, issue experts, and water rights policy change in Colorado. *Review of Policy Research*, *27*(3), 299–315. https://doi.org/10.1111/j.1541-1338.2010.00443.x

Dewulf, A., & Biesbroek, R. (2018). Nine lives of uncertainty in decision-making: Strategies for dealing with uncertainty in environmental governance. *Policy and Society*, *37*(4), 441–458. https://doi.org/10.1080/14494035.2018.1504484

Druckman, J. N., Fein, J., & Leeper, T. J. (2012). A source of bias in public opinion stability. *The American Political Science Review*, *106*(2), 430–454. https://www.jstor.org/stable/41495086

Elder, C. D., & Cobb, R. W. (1984). Agenda-building and the politics of aging. *Policy Studies Journal*, *13*(1), 115–127. https://doi.org/10.1111/j.1541-0072.1984.tb01704.x

Eshbaugh-Soha, M. (2006). The conditioning effects of policy salience and complexity on American political institutions. *Policy Studies Journal*, *34*(2), 223–243. https://doi.org/10.1111/j.1541-0072.2006.00167.x

European Digital Rights (EDRi). (2021). Open letter: Civil society call for the introduction of red lines in the upcoming European Commission proposal on artificial intelligence. https://edri.org/our-work/civil-society-call-for-ai-red-lines-in-the-european-unions-artificial-intelligence-proposal/

Fischer, F. (1998). Beyond empiricism: Policy inquiry in post positivist perspective. *Policy Studies Journal*, *26*(1), 129–146. https://doi.org/10.1111/j.1541-0072.1998.tb01929.x

Fortunato, D., & Parinandi, S. C. (2022). Social welfare returns to legislative capacity: Evidence from the opioid epidemic, 19.

Frisch Aviram, N., Cohen, N., & Beeri, I. (2020). Wind(ow) of change: A systematic review of policy entrepreneurship characteristics and strategies. *Policy Studies Journal*, *48*(3), 612–644. https://doi.org/10.1111/psj.12339

Gerber, A. S., & Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation* (First edition). W. W. Norton.

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

Haaland, I., Roth, C., & Wohlfart, J. (2021). Designing information provision experiments. *Journal of Economic Literature*. https://doi.org/10.1257/jel.20211658

Hill, K. (2020a). Another arrest, and jail time, due to a bad facial recognition match. *The New York Times*. https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html

Hill, K. (2020b). Wrongfully accused by an algorithm. *The New York Times*. https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html

Imbrie, A., Gelles, R., Dunham, J., & Aiken, C. (2021). *Contending frames: Evaluating rhetorical dynamics in AI*. Center for Security and Emerging Technology. Washington, D.C. https://doi.org/10.51593/20210010

Iyengar, S. (1990). Framing responsibility for political issues: The case of poverty. *Political Behavior*, *12*(1), 19–40. https://www.jstor.org/stable/586283

Iyengar, S. (1994). *Is anyone responsible?: How television frames political issues*. University of Chicago Press.

Jones, B. D., Larsen-Price, H., & Wilkerson, J. (2009). Representation and American governing institutions. *The Journal of Politics*, *71*(1), 277–290. https://doi.org/10.1017/S002238160809018X

Jones, M. D., Peterson, H. L., Pierce, J. J., Herweg, N., Bernal, A., Raney, H. L., & Zahariadis, N. (2016). A river runs through it: A Multiple Streams meta-review. *Policy Studies Journal*, *44*(1), 13–36. https://doi.org/10.1111/psj.12115

Kingdon, J. W. (1984). *Agendas, alternatives, and public policies* (First edition). Little, Brown.

Knaggård, Å. (2015). The Multiple Streams Framework and the problem broker. *European Journal of Political Research*, *54*(3), 450–465. https://doi.org/10.1111/1475-6765.12097

Knaggård, Å. (2014). What do policy-makers do with scientific uncertainty? The incremental character of Swedish climate change policy-making. *Policy Studies*, *35*(1), 22–39. https://doi.org/10.1080/01442872.2013.804175

McBeth, M. K., Jones, M. D., & Shanahan, E. A. (2014). The narrative policy framework. In P. A. Sabatier & C. M. Weible (Eds.), *Theories of the policy process* (Third, pp. 225–266). Westview Press. https://doi.org/10.4324/9780429494284-14

McBeth, M. K., & Lybecker, D. L. (2018). The narrative policy framework, agendas, and sanctuary cities: The construction of a public problem. *Policy Studies Journal*, *46*(4), 868–893. https://doi.org/10.1111/psj.12274

Mcbeth, M. K., & Shanahan, E. A. (2004). Public opinion for sale: The role of policy marketers in greater yellowstone policy conflict. *Policy Sciences*, *37*(3), 319–338. https://doi.org/10.1007/s11077-005-8876-4

McBeth, M. K., Shanahan, E. A., Arnell, R. J., & Hathaway, P. L. (2007). The intersection of narrative policy analysis and policy change theory. *Policy Studies Journal*, *35*(1), 87–108. https://doi.org/10.1111/j.1541-0072.2007.00208.x

McClendon, G. H. (2014). Social esteem and participation in contentious politics: A field experiment at an LGBT pride rally. *American Journal of Political Science*, *58*(2), 279–290. https://doi.org/10.1111/ajps.12076

Mcentire, K. J., Leiby, M., & Krain, M. (2015). Human rights organizations as agents of change: An experimental examination of framing and micromobilization. *American Political Science Review*, *109*(3), 407–426. https://doi.org/10.1017/S0003055415000295

McGann, J. (2020). *2019 Global Go To Think Tank Index report*. Think Tanks and Civil Societies Program. Philadelphia, PA. https://repository.upenn.edu/think_tanks/17

Michaels, S. (2009). Matching knowledge brokering strategies to environmental policy problems and settings. *Environmental Science & Policy*, *12*(7), 994–1011. https://doi.org/10.1016/j.envsci.2009.05.002

Minkkinen, M., Zimmer, M. P., & Mäntymäki, M. (2022). Co-shaping an ecosystem for responsible AI: Five types of expectation work in response to a technological frame. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-022-10269-2

Mintrom, M., & Luetjens, J. (2017). Policy entrepreneurs and problem framing: The case of climate change. *Environment and Planning C: Politics and Space*, *35*(8), 1362–1377. https://doi.org/10.1177/2399654417708440

Mintrom, M., & Norman, P. (2009). Policy entrepreneurship and policy change. *Policy Studies Journal*, *37*(4), 649–667. https://doi.org/10.1111/j.1541-0072.2009.00329.x

National Security Commission on Artificial Intelligence. (2021). *Final report*. National Security Commission on Artificial intelligence. Washington, D.C. https://www.nscai.gov/wp-content/uploads/2021/01/NSCAI-Draft-Final-Report-1.19.21.pdf

Neuman, W. R., Neuman, R. W., Just, M. R., & Crigler, A. N. (1992). *Common knowledge: News and the construction of political meaning*. University of Chicago Press.

OECD. (2019). *OECD principles on artificial intelligence*. OECD. Paris, France. https://www.oecd.org/going-digital/ai/principles/

Office of Management and Budget. (2020). *Memorandum for the heads of executive departments and agencies: Guidance for regulation of artificial intelligence applications*. Office of Management and Budget. Washington, D.C. https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf

Petridou, E., & Mintrom, M. (2021). A research agenda for the study of policy entrepreneurs. *Policy Studies Journal*, *49*(4), 943–967. https://doi.org/10.1111/psj.12405

Ridde, V. (2009). Policy implementation in an African state: An extension of Kingdon's Multiple-Streams Approach. *Public Administration*, *87*(4), 938–954. https://doi.org/10.1111/j.1467-9299.2009.01792.x

Sabatier, P. A. (1988). An advocacy coalition framework of policy change and the role of policy-oriented learning therein. *Policy Sciences*, *21*(2), 129–168.

Saleh, K. (2020). *Comments to the OMB: Draft memorandum to the heads of executive departments and agencies, guidance for regulation of artificial intelligence applications* (Docket No. OMB_FRDOC_0001-0261). Zest.AI. Burbank, CA.

Schiff, D. (2021). Education for AI, not AI for education: The role of education and ethics in national AI policy strategies. *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-021-00270-2

Schiff, D., Borenstein, J., Laas, K., & Biddle, J. (2021). AI ethics in the public, private, and NGO sectors: A review of a global document collection. *IEEE Transactions on Technology and Society*, *2*(1), 31–42. https://doi.org/10.1109/TTS.2021.3052127

Schneider, A., & Ingram, H. (1993). Social construction of target populations: Implications for politics and policy. *American Political Science Review*, *87*(2), 334–347. https://doi.org/10.2307/2939044

Small, D. A., & Loewenstein, G. (2003). Helping a victim or helping the victim: Altruism and identifiability. *Journal of Risk & Uncertainty*, *26*(1), 5–16. https://doi.org/10.1023/A:1022299422219

Squire, P. (1992). Legislative professionalization and membership diversity in state legislatures. *Legislative Studies Quarterly*, *17*(1), 69–79. https://doi.org/10.2307/440081

The Future Society. (2022). The Future Society. https://thefuturesociety.org

Ulnicane, I. (2022). Against the new space race: Global AI competition and cooperation for people. *AI & Society*. https://doi.org/10.1007/s00146-022-01423-0

United States Innovation and Competition Act of 2021, S. 1260, 117th Cong. (2021). https://www.congress.gov/bill/117th-congress/senate-bill/1260

Vincent, J. (2017). China and the US are battling to become the world's first AI superpower. *The Verge*. https://www.theverge.com/2017/8/3/16007736/china-us-ai-artificial-intelligence

Weible, C. M., & Schlager, E. (2016). The Multiple Streams Approach at the theoretical and empirical crossroads: An introduction to a special issue. *Policy Studies Journal*, *44*(1), 5–12. https://doi.org/10.1111/psj.12143

White House. (2019). Executive Order on Maintaining American Leadership in Artificial Intelligence. https://www.federalregister.gov/documents/2019/02/14/2019-02544/maintaining-american-leadership-in-artificial-intelligence

Zahariadis, N. (2014). Ambiguity and multiple streams. In P. A. Sabatier & C. M. Weible (Eds.), *Theories of the policy process* (Third edition, pp. 26–58). Westview Press.

Zito, A. R. (2001). Epistemic communities, collective entrepreneurship and european integration. *Journal of European Public Policy*, *8*(4), 585–603. https://doi.org/10.1080/13501760110064401

Zwetsloot, R., Toner, H., DingNovember 16, J., & 2018. (2018). Beyond the AI arms race: America, China, and the dangers of zero-sum thinking. https://www.foreignaffairs.com/reviews/review-essay/2018-11-16/beyond-ai-arms-race

CHAPTER 5
CONCLUSION

5.1 Review of Key Findings

This dissertation represents an effort to better understand the substance, development, and implications of the emerging U.S. AI policy agenda as of the early 2020s. Namely, as rapid advancements in AI research in the 2010s were followed by increased industrial and public sector adoption, questions surrounding AI governance began to loom large in the later half of the 2010s and will continue to be major topics of concern in the 2020s and beyond. Yet the identification and construction of a policy agenda is no little thing, particularly in light of the expansive, significant, and mixed impacts associated with AI.

The decision-makers who formulate the agenda must implicitly or explicitly decide on numerous questions: What is AI and what is it not? What are the most pressing policy problems and opportunities associated with AI? Which solutions are feasible and aligned with societal values and governmental priorities, now or in the future? Which actors in the public or private sector, in civil society and academia, and in the broader public, are best suited to advise on and govern AI? Should regulation be centralized or decentralized, assigned to a few entities or many? What is the appropriate relationship between formal and informal regulation, between hard and soft law, between self-regulation and external enforcement? How should the U.S. act in light of its desire for international cooperation with allies while also desiring competitive advantage with respect to both allies and adversaries? What balance should be struck between innovation and risk, between too much precaution and too little? How can governance be designed to be nimble and flexible, and how quickly should regulation be put forth? For the matter, who has the authority, perspective, or expertise to speak to these questions?

These questions—amongst many others—face the decision-makers of today and tomorrow as they work out AI policy in general and with respect to specific sectors and contexts. Moreover, many such decisions must be made in the context of great uncertainty, while the capabilities of AI are evolving, the impacts unclear, and the priorities contested. Policymakers must make sense of the AI policy agenda while simultaneously attending to public opinion and broader policy priorities, attitudes towards the private sector and big technology companies, appetites for regulation and innovation, pressures from advocacy and lobbying, social and political conditions, and concerns surrounding the economy, racial and social justice, and environmental sustainability.

Importantly, the act of agenda-setting itself is also marked by both rational and interpretive elements, an interplay increasingly recognized as essential to understanding policymaking and embraced in this dissertation. 'Hard' technical and economic possibilities and limits necessarily interact with ideas, beliefs, narratives, focusing events, and sociotechnical imaginaries. The agenda must speak to present facts and future predictions, to fears, aspirations, hopes, and mere guesses. In its own right then, the AI policy agenda as examined in this work is an entity with its own character, discourse, and evolution.

In Chapter 1, I began by defining AI, presenting its historical and modern context, and reviewing key 21st century questions surrounding timing, centralization, self-regulation, coordination, and precaution in AI policy. I next sought to describe AI as an object of policy study, an important step in opening AI policy up to empirical examination, ensuring that AI is properly situated in its present and historical scholarly and policy context. I argued that while AI can be viewed as partially unique in light of its expansive impacts, complexity, and capacity for autonomy, learning, and prediction, much about AI policy should also not be viewed as a wholesale departure from the past. Instead, it is prudent to approach AI policy with awareness of the characteristics of other similar policy domains and how they have shaped agenda-setting processes in the past. As such, scholarship on policy process theory, political communication, technology governance, innovation policy, framing, narratives, policy entrepreneurship, and more can shed light on what we might expect to see in AI policy, and whether we are likely to see continuities or divergences.

To begin such an effort to characterize AI policy, I drew on the scholarly literature surrounding general purpose technologies, emerging technologies, and strategic technologies. Deriving from this understanding, I emphasized two central aspects of AI as an object of policy study, that it is marked by AI's prominent and pervasive multi-sector impact, and that it is subject to significant uncertainty and ambiguity. Jointly, these conditions lead to a state of what I called *high-stakes policy ambiguity*. Such a conceptualization implies that many actors will have a stake in AI policy, that its impacts will be uncertain (yet still subject to human agency rather than mere technological determinism), and that strategic competition over the agenda is likely to play a prominent role in the agenda-setting process. Because of this, the dissertation examined a few core topics: which fundamental ideas are being centered in the emerging AI policy agenda, what conditions appear to have led to this result, whether actors like members of the public are having a say in the policy process as compared to experts, and how policy entrepreneurs are working to influence the agenda.

Chapter 2 continued this examination through performance of an extensive case study of the federal U.S. AI policy agenda. I relied on 63 strategic AI policy documents published between 2016 and 2020, and identified and curated by the federal government, as a key starting point for understanding the emerging agenda. In this study, I drew on the MSF, a leading framework for understanding agenda-setting processes, which directed my attention towards components like policy problems, solutions, focusing events, and indicators. Further, I considered these components in light of a ostensible contest between competing visions, ideas, or approaches towards AI and technology governance. In particular, I asked whether the state of the U.S. AI policy agenda better reflected a traditional 'paradigm' reflective of expert leadership and strategic economic and geopolitical goals, or a 'transformative' one reflecting increased attention to social and ethical issues with a greater role for the public. I applied qualitative coding and quantitative content analysis techniques to study the policy problems, solutions, focusing events, indicators, issue frames, and roles of the public and experts as described in these documents. Through reference to the underlying conceptual frameworks, this chapter was able to provide not only extensive empirical evidence, but also scholarly insights about the development of the U.S. AI policy agenda.

The results indicated that the emerging AI agenda has striking and unusual characteristics. Arguably unlike any other complex technological policy domain in the past, reference to ethical concepts and an ethical frame is pervasive throughout policy documents, including the elaboration of explicit ethical principles and frameworks for AI. Yet this prominent rhetoric does not appear to be mirrored in terms of concrete action and is surpassed by references to the traditional innovation frame for technology policy. For example, despite numerous ethical scandals in AI and evidence of harms, the overwhelming majority of indicators and focusing events discussed referenced AI's technical performance surpassing that of humans, of economic possibilities, or of geopolitical competition. Further, while reference to ethical problems was quite common, particularly surrounding issues like fairness, trust, and transparency, these problems were often not matched by concrete solutions, while problems surrounding economic and technical needs were. The failure to translate ethical rhetoric into action may be significantly explained by the lack of concrete ethical solutions or evidence of their technical feasibility, by the incompatibility of sweeping calls for social change against American neoliberal values for innovation policy, and by institutional inertia and scope limitations leading individual governmental agencies little understanding of how to translate ethics into practice.

In light of the high-stakes strategic ambiguity of AI as a policy object, a particularly interesting finding was how ostensibly ethics-focused issues like transparency and trust took on mixed strategic and ethical characteristics. Trust, for example—often perceived as a linchpin in linking AI's ethical dimensions with its appropriate usage—took on a mode that I described as 'hybrid' in nature. No longer simply about avoiding ethical harms, issues like transparency or fairness were repeatedly rendered as barriers to consumer adoption, investment, and innovation. These framing efforts indicate that policymakers either assumed or perceived that response to social and ethical concerns needed to be presented in light of traditional economic or geopolitical goals. This suggests that normative goals in AI policy ultimately play a secondary role not only in terms of the development of associated policy solutions, but even in the policy rationales and justifications thought important. In a sense, there is less evidence of a genuine transformation of the U.S. approach to technology governance but rather a synthesis and partial subsumption of an ethics lens into traditional innovation policy.

A particularly unambiguous finding was that, despite nearly ubiquitous calls for public and diverse participation, the public is hardly referenced with respect to concrete policy proposals. While experts are referenced with respect to specific committees, timelines, deliverables, and responsibilities, the public tends to be mentioned only generically at the mission statement level. Relatedly, there is very little mention of specific public concerns or public opinion about AI, neglecting research that has been done on these issues. Despite the existence of many frameworks that have been developed to support participatory policymaking, there appears to be little notion in these documents of how to actually engage the broader public. The U.S. AI policy documents instead reflect a body of policymakers more comfortable working in traditional ways, highlighting traditional strategic economic and geopolitical concerns and goals, and engaging experts in working out policy solutions.

Nevertheless, the rhetoric suggests a genuine interest in engaging with the social and ethical implications of technology. Further, there is evidence that some ethical issues like privacy have been concretized to a sufficient extent such that policy solutions are laid out clearly, and to the point that these topics 'lose' their more abstract framing as human-centered ethical concerns. Policy stakeholders invested in safeguarding against AI social and ethical risks may still have an open doorway if they can demonstrate the technical feasibility of their proposed solutions. Indeed, the common deference of U.S. AI policy documents which urge a need for solutions to *other* actors responsible for developing these future standards, best practices, impact assessments, and risk management approaches indicates that more work can be done to formally embed social and ethical consideration in technology governance. Yet some proposals, like regulating AI to reduce structural inequities in the U.S. or globally, may simply be functionally outside of the bounds of U.S. policy values for the time being.

While Chapter 2 sought to descriptively review U.S. AI policy documents directly to expose explanatory insights on the nature of the emerging agenda and agenda-setting processes, Chapter 3 turned to U.S. federal policymakers themselves. In particular, it examined the AI policy discourse of members of the 115 and 116 Congresses, the attention of these policymakers to issue frames surrounding AI's economic, ethical, or geopolitical dimensions, and whether public attention to AI meaningfully influences policymaker attention. In this way, Chapter 3 continued the focus on whether AI policy is marked by traditional or transformative notions of technology governance, including an emphasis on ethical versus strategic dimensions, and on expert versus public participation. To study AI issue attention and influence, I drew on traditional and social media messages from these federal policymakers, the public, and the media, over a three year period, starting from the effective beginning of U.S. federal AI policy discourse in 2017 through the end of 2019. I employed text-as-data techniques to construct time series data capturing attention to AI issue frames over time, and used time series analysis methods to evaluate whether members of the public seem to influence U.S. policymakers both in general and with respect to specific issue frames.

Across a range of models and specifications, Chapter 3 provided consistent evidence that the public does seem to lead policymaker attention to AI. Heightened public attention to AI within a given week is associated with fairly sizable increases in policymaker attention, while policymaker attention conversely does not seem to substantively alter public attention. I also considered news media, measured through issue attention in New York Times articles, as an alternative source for conveying the national mood to policymakers.

The evidence indicates that media does lead policymaker attention to AI as well, suggesting that the media is another important channel in agenda-setting influence for AI policy in particular. Relatedly, the chapter also considered which issue frames for AI are dominant in the early agenda-setting stage, and whether certain frames are becoming more prominent over time. I find that the innovation frame is consistently the most prominent across all three categories of actors, though there is evidence of increased attention to the ethics frame particularly amongst policymakers.

Yet this shift in focus does not appear to be driven by public concerns. Indeed, relative to other issue frames, attention to the ethics frame in AI is not especially increasing amongst the public. Moreover, while the public does lead policymaker attention to AI generally, it does not with respect to the ethics frame, suggesting that policymakers concerned with these aspects of AI policy are drawing their encouragement from different sources. Instead, the public and media only appear to influence policymakers when they discuss the innovative aspects of AI. These findings may demonstrate that a certain portrayal of the role of the public in technology policy is too simplistic. While some policy stakeholders concerned with AI ethics and public participation imagine the public to have a higher stake and more expertise with respect to social and ethical dimensions of AI, is not clear that the public's priorities actually reflect a heightened focus on these concerns.

One possibility is that the public is not yet substantively informed about AI, much less its complex social and ethical dimensions. More work may be needed to genuinely engage and inform the public, if these are truly considered requisites in shaping the agenda as expressed by many actors. Indeed, the public opinion scholarship on AI indicates that the public still has limited understanding of AI in general. Moreover, members of the public, especially the subset of the public engaged on AI issues on social media, may simply be more concerned with economic implications. For example, despite the strong advocacy of civil society organizations against the use of predictive AI technologies employing biometrics, such as in surveillance and policing, it is unclear that the public shares these sentiments. Members of the public appear to be much more in favor of AI applications such as predictive policing than expert groups such as AI researchers and non-governmental organizations. To the extent that these patterns hold over time, it may not be a case that policymakers are ignoring public issue priorities, but rather that AI ethics advocates have an idealized conception or even a misconception about the public's priorities.

In line with the second chapter then, Chapter 3 reveals a mixed picture when confronting theoretical and normative expectations with data. U.S. policymakers are paying increased attention to the ethical dimensions of AI, with these topics even rising to the most prominent issue frame for AI by the end of 2019. Yet, given that the general public does not appear to be the source of this change, it may be the case that specialized elite groups such as civil society organizations and academics are more directly responsible for this shift in policy priorities. If correct, these findings might indicate that AI policy is truly a battleground amongst elites, with policymakers and other elite stakeholders calling for public participation, but only listening to public opinion when it confirms their prior conceptions. With respect to policymakers, this could lead them to be especially responsive only when the public attends to AI's implications for innovation, a possibility buttressed by the findings in Chapter 2 that reveal few genuine policy channels for incorporating public opinion—much less participation. With respect to civil society and academia, this may reflect something of an implicit technocratic disposition of elites who claim to speak for the public but do not truly reflect current public priorities.

This dissertation, of course, does not settle what the 'correct' issue frames or priorities are for AI policy. It does indicate, however, that issue frame attention and public participation are not necessarily deeply integrated. Public influence on the policy agenda can manifest without increasing attention to social and ethical dimensions of technology, while these socio-ethical dimensions of technology may also become growing policy priorities independent of the public's influence. Overall then, this chapter confirms the prior chapter's findings that some dimensions of the AI policy agenda are quite novel, such as the surprising degree of focus on AI ethics, but it does not suggest a systemic transformation in the role of the public in terms of shared governance or even a real diminution of traditional strategic technology policy goals surrounding geopolitics and economic innovation.

In light of these mixed dynamics as the emerging AI policy agenda develops, Chapter 4 looks much more deliberately at what strategies influence policymakers, a question only loosely addressed in prior chapters. To provide insight here, Chapter 4 turns to the role of policy entrepreneurs, key actors in the agenda-setting process, and looks in particular at how members of civil society act to shape state-level AI policy. I perform a field experiment—specifically an information provision or correspondence experiment—in partnership with a leading AI policy think tank to see what kinds of policy entrepreneur influence strategies most effectively gain the attention of state policymakers. In particular, policymakers were exposed to email messages providing either technical expertise about AI, traditionally thought most important for technical policy domains, or persuasive narratives about AI, reflective of growing attention in policy scholarship to the role of ideas and stories in policymaking. Chapter 4 follows the prior chapters in incorporating issue framing as well, as policymakers are provided emails that emphasize either AI's social and ethical dimensions or AI's implications for economic and technological leadership globally. By measuring to what extent policymakers took actions like clicking on links provided in the policy entrepreneur emails and viewing information about a webinar on AI policy for state legislators, this chapter provided clear empirical evidence to isolate the effectiveness of different influence strategies.

I find, quite surprisingly, that policymakers were just as likely to engage in policy entrepreneur influence efforts centered on the provision of narratives as they were with technical and expert information about AI, and that policymakers were much more likely to engage with both strategies than when presented with an alternative control message that did not emphasize either such strategy. Thus, notwithstanding the traditional association of hard policy domains with technical complexity and uncertainty, narratives about AI's impact on individuals were equally likely if not more likely to lead a policymaker to click on a fact sheet or story, or to consider attending a webinar about AI policy. Relatedly, a surprising finding was that state policymakers were just as engaged with correspondence emphasizing the social and ethical dimensions of AI as they were with messages surrounding technological and economic leadership considerations. In the context of the other chapters, this finding suggests that the social and ethical dimensions of AI make for, minimally, conversation starters. Policymakers appear to be interested in and concerned enough about these issues that it leads them to take costly action in terms of their limited time and attention. However, as illuminated in the prior chapters, while interest in these issues may be necessary to place them on the emerging policy agenda, it may not be sufficient towards realization of concrete and effective policy solutions.

The chapter provides a few additional insights on policy entrepreneur influence efforts and what kinds of policymakers are receptive to them. First, is clear that state policy actors in high-capacity legislators have much more ability to engage in AI policy, a finding with implications for which policymakers and U.S. states are likely to be leaders in AI policy and which may need additional assistance and targeted effort to reach. Second, while I anticipated that actors with less experience in AI policy would be especially prone to engaging with expert information, the findings revealed that these policymakers and legislatures inexperienced in AI were equally likely to seek out narratives. Such a finding encourages continued scholarly attention to the role of narratives in policy and agenda-setting specifically. If even a notoriously complex and opaque domain like AI policy can be shaped just as easily by persuasive narratives as by expert information, then more research should be undertaken to see what aspects of narratives are influential in technology policy and how actors have invoked them in practice to shape the agenda.

5.2 Limitations

No set of three papers or single dissertation can provide definitive evidence about a topic as multifaceted as agenda-setting in the case of AI policy. To help contextualize the contributions of these studies and future research priorities, therefore, it is important to make sense of this dissertation's limitations.

A first limitation surrounds the dissertation's temporal scope. Arguably, the U.S. AI policy agenda began developing no sooner than 2016 or 2017, a starting point that is encapsulated well by the dissertation. However, the ending point for the AI policy agenda, if there is indeed such a clear demarcation, is not yet in sight. The agenda might solidify to a significant degree over the next few years or decade, or perhaps continue to evolve for several decades as new issues, AI capabilities, and contextual factors play out, and as prior policy decisions feed forward into future ones. This means that the prevalence of issue frames, the nature of influence efforts, the feasibility of policy solutions, and the ultimate synthesis reflected by the policy agenda may very well evolve in directions not captured by the studies performed here. For example, the three issue frames studied most prominently here are already merely a subset of broader scholarly, public, and policy discourse over the 2010s-2020s and even prior decades; the public still knows relatively little about AI; scholars have yet to make sense of many of AI's impacts; and policymakers are only now developing the in-house expertise to weigh in on these issues.

Many more AI policy changes will come into play in the near future and subsequent years. Policy-related activities in the early 2020s such as the forthcoming U.S. National AI Policy Strategy, NIST's AI Risk Management Framework and other standards efforts, individual agency responses to the OMB Guidance for Regulation of AI Applications, the developing AI Bill of Rights, efforts by the National AI Advisory Committee, National AI Initiative and Office, increased regulatory scrutiny by the FTC, and proposed legislation like the Algorithmic Accountability Act are just a few of the dozens of developments that will characterize—and shape—the U.S. AI policy agenda. Nevertheless and notwithstanding this temporal limitation, this dissertation does benefit from being situated in this fluid stage. Some insights that seem obvious in the future may only be so in retrospect, while the instability and ambiguity of AI policy in the early stages becomes lost. Through immersion in various AI policy discourses, communities, and data sources, this dissertation helps to create a historical record beneficial for future scholarship.

A second set of limitations surrounds the data sources and associated methodologies used to make sense of the U.S. AI policy agenda. The case study relies on federal AI policy documents in the executive branch of the U.S. government, while the subsequent chapter looks at social media messages from the public and Congress, and the field experiment looks at state legislator responsiveness to a fairly discrete set of influence strategies. These sources of evidence are incomplete. Agenda-setting takes place through a broad array of communities and channels beyond those studied directly here, including through governmental hearings, academic research, constituent communications, closed-door industry and civil society lobbying efforts, and interconnected national and international negotiations. While some of these agenda-setting activities are visible, others are not, and many require interpretation given their strategic construction. Further, attention in frameworks like the MSF to numerous aspects of the policy process such as indicators, focusing events, policy feedback, balance of political interests, venue shopping, and more, imply that agenda-setting is a deeply complex and fluid process. Each chapter individually, and the dissertation jointly, therefore, omit important data sources, actors, and processes that constitute the totality of agenda-setting.

However, this should not be taken to mean that no empirical or theoretical insights about AI policy and agenda-setting can be secured with some degree of confidence. Indeed, one of the main goals of this dissertation is to help concretize otherwise expansive and amorphous policy process scholarship invoking many actors, elements, and processes. To do so, I drew on a range of empirical contexts and methodological approaches, ranging from a thorough but descriptive and primarily qualitative analysis in Chapter 2, to a causally precise but scope-limited experimental study in Chapter 4. I aimed to provide some insight on the AI policy agenda with respect to numerous actors: federal and state policymakers, members of the federal bureaucracy, the public, news media, and policy entrepreneurs such as members of civil society. Moreover, the analysis attempted to invoke not only hard facts and quantitative indicators, but also the role of ideas, narratives, persuasion, and the interplay between 'reason' and 'passion' in agenda-setting. The use of qualitative and quantitative methodologies and the convergence of several key findings across contexts and studies show that it is possible to examine concrete research questions and test hypotheses with implications for policy process research and other domains of scholarship.

A third limitation is that the dissertation is strongly attentive to U.S. AI policymaking and its surrounding institutions, processes, values, and so on, while the AI policy agenda is ultimately shaped by local, national, and international factors. Indeed, despite early U.S. leadership in AI policy, the U.S. was quickly surpassed by more rapid policy activity in the late 2010s and early 2020s, especially by the EU, but also arguably in countries like Brazil and China. Numerous international bodies now play a role in shaping the AI policy agenda, ranging from long-standing institutions like the EU, UN, and OECD, to new institutions like the Global Partnership on AI and the EU-U.S. Trade and Technology Council. The fact that the U.S. has signed on to several documents related to AI policy such as those emerging from the OECD and G20, and that the EU has taken leadership in creating its Proposal for AI Regulation, imply that many important agenda-setting conversations are happening abroad and at the international scale. For reasons of scope feasibility, this dissertation pays little direct attention to these developments beyond how they indirectly manifest in U.S. legislator discourse or federal AI policy documents, with some extra attention devoted to geopolitical dynamics between the U.S. and China. Ongoing scholarship focused on international dimensions of the AI policy agenda and comparative work contrasting the U.S. with other regions will be needed to make sense of the broader picture.

It is also certainly the case that the U.S. constitutes a unique context for studying technology governance, given its current political composition, traditional association with pro-innovation policy, cultural and public values, and specific array of AI-engaged actors in industry, academia, and civil society. As such, some of the findings here, for example about the balance between ethics and innovation, the role of the public and experts, or how policy actors are negotiating various tensions in the emerging agenda, may not translate straightforwardly to other national and international contexts. Yet, targeted attention to U.S. AI policy is still valuable in its own right. The U.S. is a global leader—if not the global leader—in AI research and development, and its policy activities have a direct bearing on many leading private sector organizations and on many international partners and even competitors. Understanding the U.S. AI policy agenda is thus important for understanding AI policy broadly beyond U.S. borders. Further, the findings here can be theoretically generalized, contextualized, contrasted, and adapted to make sense of AI policy and agenda-setting processes in other contexts, locales, and time periods.

5.3 Contributions to Theory, Methodology, and Practice

This dissertation offers various contributions to scholars, as well as novel evidence and associated implications for a broader set of AI policy practitioners and stakeholders. While the research here is situated most directly with respect to policy process theory and agenda-setting theory, it cuts across various fields and subfields including innovation policy, technology governance, political communication, media studies, AI ethics, and more.

For scholars of policy process theory and especially those interested in agenda-setting, this dissertation draws on a range of methods and data sources to reveal new findings. I provide clear empirical evidence that policy entrepreneurs can be effective in engaging state-level policymakers, that narratives may play an increasingly powerful role in policy entrepreneur influence efforts, and that federal policymakers are at least partially responsive to the public during the agenda-setting process. While some of these findings have been supported in prior research, others are marked by mixed findings or isolated anecdotes and case studies, and are thus buttressed by the work done here. Further, this dissertation establishes these findings clearly in the context of the significantly understudied domain of AI policy, representing arguably a pinnacle of highly complex technical policy domains. As such, it represents one of few efforts to concretely apply agenda-setting theory to understand technology policy and AI policy in particular, with special implications for how technical and policy complexity shape agenda-setting behaviors and processes—for example, with these factors leading to heightened ambiguity, framing contestation, and fluid participation. Overall, through the conceptual efforts to define AI as an object of policy study, and the subsequent findings, this research can thus serve as a baseline for future research efforts in AI policy and policy process theory research broadly.

Beyond this core audience, this research also has implications for an array of other fields of scholarship and practice. For scholars and other stakeholders of innovation policy and technology governance, the findings caution that movements like transformative innovation policy and Responsible Innovation may thus far have limited purchase in the U.S. context, despite calls for increased socio-ethical consideration and public participation including in technology policy specifically. The results reveal a complex interplay between competing issue frames and visions of technology policy, such that scholars should be attentive to implicit deprioritization of certain policy problems and solutions, the contrast between rhetoric and action, and the strategic subsumption or synthesis of new paradigms into old ones. Further, while ethics is not typically incorporated into formulations of innovation policy, the conceptualization and findings here reveal how ethics does and does not play a role, a facet future innovation scholars may wish to consider extending. For practitioners and scholars interested in shifting the deepest rationales of innovation policy, much more work will be needed to create effective venues for meaningful two-directional public engagement, and to understand the limits and opportunities for incorporating meaningful ethical consideration into technology policy.

Along related lines, the dissertation has significant implications for scholarship and practice on AI ethics. Numerous commentators have noted the juxtaposition between the stated importance of AI ethics and the likelihood of it being taken seriously, especially in the private sector. The findings here provide a partial source of comfort as AI ethics discourse does manifest to a surprising degree across federal U.S. AI policy documents, is just as likely to engage state-level policymakers as a more traditional strategic issue frame, and even surpasses economic and geopolitical discourse for federal policymakers as of the end of 2019. Yet, ethics appears to be a conversation starter, able to engage policymakers but not quite able to change the underlying orientation of technology policy. Thus while many have worried about the translation of AI ethics into practice in the private sector, this dissertation also provides some notes of caution about the translation of ethics into *policy*.

There are several pathways forward. Realistically speaking, some ethical concerns may be simply excised from the scope of acceptable policy solutions barring more dramatic political change. Other ethical concerns may be reinterpreted as means to secure AI adoption, innovation, and geopolitical advantage. Still others may find genuine purchase in terms of policy solutions if they are able to be formulated in terms of actionable strategies. AI ethics stakeholders in the U.S. will have to decide which of these strategic options to endorse, such as whether it is viable to try to shift the deepest rationales invoked in technology policy towards ethics, or whether it is simpler to identify a concrete set of best practices for a limited number of ethical issues and formulate them in terms of the traditional language of innovation policy. As a first pass, AI ethics stakeholders should pay serious attention to demonstrating the technical feasibility of preferred AI ethics solutions, and to identifying the appropriate venues, such as standards organizations and impact assessments, in order to successfully socialize these proposals.

For scholars in political communication, media studies, and related fields, this research also provides insight about how members of the public, policy entrepreneurs, and policymakers engage in agenda-setting discourse and influence efforts. While much attention has focused on how the media frames issues to influence the public, this dissertation demonstrates quite distinctly how members of the public, media, and civil society can frame issues to influence policymakers. Further, the results provide further if inconclusive evidence that issue attention can be shown to correspond across traditional and social media sources and across different actors, and that social media may constitute a meaningful source for understanding influence in agenda-setting processes. As scholars work to make sense of still novel data sources like social media and developing techniques like text analysis, the research here contributes through building on a small number of related studies showing that study of phenomena like issue attention, framing, and agenda-setting influence using these sources and methods is increasingly viable.

Meanwhile, scholars of political behavior and other stakeholders interested in shared governance and democratic responsiveness in public policy may benefit from the dissertation's focus on public participation. The findings of some public influence, but circumscribed to economic dimensions of AI only, and with genuine public engagement largely ignored in terms of concrete policy solutions, show there is much still to be understood (and normatively, desired). Calls for public participation in AI policy appear to be significantly unrealized, despite decades of attention to the importance of the public in policy and an array of frameworks designed to facilitate such engagement. This may be in part because the public has limited knowledge on AI policy as a complex and non-salient issue, because elite actors impute their own assumptions or preferences on the public, or because key decision-makers lack the know-how on approaches to effectively engage the public. If public participation is truly to become a meaningful feature in technology governance, much more work needs to be done to test and socialize best practices, and to understand barriers and how to overcome them. In the meantime, there is a stark divide between how the public's role is articulated or imagined and how it is realized.

This dissertation also offers contributions which are methodological in nature. While approaches to studying the policy process like the MSF have been marked by their disproportionate use in qualitative case studies and absence of hypothesis testing—in cases leading to a lack of theoretical development—this dissertation demonstrates that the MSF can be fruitfully and rigorously applied to understand agenda-setting and to evaluate competing expectations using both qualitative and quantitative methods. In particular, Chapter 2 exemplifies how MSF components like indicators, focusing events, and policy problems and solutions can be measured, contextualized, and interpreted to provide scholarly insights. The quantitative content analysis methods employed, such as comparing normalized relative attention to policy problems and issue frames, depart from more impressionistic process tracing studies which informally review a few key documents or policy developments. This mixed methods approach to studying and formalizing the MSF thus represents an additional set of tools in the policy scholar's toolkit, important as policy scholars attempt to refine their frameworks and methods and build knowledge.

Chapters 3 also offers new methodological contributions, building on a still nascent body of text analysis techniques, towards more carefully characterizing issue attention and sub-issue attention and allowing for rigorous study of policy influence. Foremost, the dictionaries constructed and the processes used to construct them can serve as helpful benchmarks for AI policy scholars and other scholars concerned with, for example, issue frames in the policy process. While much remains to be learned about robust use of qualitative content analysis to analyze large amounts of data, this chapter represents a sustained effort to construct, validate, apply, and interpret policy texts using text analysis, with successes and limitations that can inform future work. A related theoretical and methodological contribution is the focus on within-issue issue frames, and the use of time series techniques to evaluate agenda-setting influence. While scholars have studied competing issue frame dynamics, and competition in issue attention generally across policy areas, the joint use of text and time series analysis techniques here represents a novel attempt to make tractable the study of framing contestation and between-actor influence in agenda-setting. Advancing this research is critical to making sense of the complex dynamics of agenda-setting.

Chapter 4 similarly offers several methodological contributions to policy scholarship. While the use of experiments is gradually growing in policy research, the use of field experiments is still quite limited. Chapter 4 in particular demonstrates how experimental study of policy processes, such as policy entrepreneur efforts in agenda-setting, can be done in an authentic fashion to provide compelling causal evidence in an otherwise complex space. While not all of the outcomes captured in the study played a key role in the ultimate findings, the quantification of policymaker behavior in terms of clicks, email replies, and webinar registration and attendance also suggests a novel approach to measurement in field experiment and elite research. The experiment is also relatively unique in applying narratives to an experimental context, as well as contrasting them against expert information. In combination, performing a field experiment in partnership with a leading AI policy organization provides authenticity, legitimacy, and minimizes ethical burdens often associated with audit studies. Expanded use of experiments like the one performed here in policy process research can help address concerns about endogeneity as well as the conceptual overspecification of entities and relationships that often render policy research intractable.

Finally, one of the goals of the dissertation was to demonstrate the utility and necessity of mixed methods research to make sense of complex topics like agenda-setting and the role of ethics in policy. Both within individual chapters and collectively across chapters, the dissertation drew on qualitative, quantitative, and conceptual dimensions to guide its empirical exploration. Part of this effort, and again arguably necessary to understand AI policy, is that elements like normative claims, narratives, and frames play a key role in the policy process. Yet these elements must be made amenable to empirical study, rigorous characterization, and hypothesis testing, goals which this dissertation attempted to realize.

Through the study of policy issues, actors, and influence efforts in the emerging U.S. AI policy agenda, this dissertation has sought to bring clarity to and spark interest in an important issue. My hope is that future scholars and stakeholders will feel empowered to better understand and—in turn—shape AI policy for the benefit of all.

Appendices

APPENDIX A

SUPPORTING INFORMATION FOR CHAPTER 2

A.1 AI Policy Document Details

Table A.1: Key U.S. strategic AI policy documents, 2016-2020

Document Author and Title	Document Type	Date	Pages
(NSTC) Preparing for the Future of Artificial Intelligence	Scientific and Technical Reports*	10/1/2016	58
(NSTC) The National Artificial Intelligence Research and Development Strategic Plan	Strategy Documents*	10/2/2016	48
(EOP) Artificial Intelligence, Automation, and the Economy	Scientific and Technical Reports	12/1/2016	55
(GAO) Face Recognition Technology: DOJ and FBI Need to Take Additional Actions to Ensure Privacy and Accuracy	Scientific and Technical Reports	3/22/2017	24
(DOT) Automated Driving Systems 2.0: A Vision for Safety	Strategy Documents	8/12/2017	36
(DOT/FAA) Strategic Plan, FY2019-2022	Strategy Documents	2/1/2018	42
(GAO) Technology Assessment: Artificial Intelligence – Emerging Opportunities, Challenges, and Implications	Scientific and Technical Reports*	3/1/2018	100
(OSTP) Summary of the 2018 White House Summit on Artificial Intelligence for American Industry	Event Summaries	5/10/2018	15
(Treasury) A Financial System That Creates Economic Opportunities – Nonbank Financials, Fintech, and Innovation	Scientific and Technical Reports	7/1/2018	222
(HHS/NIH) National Institutes of Health Workshop: Harnessing Artificial Intelligence and Machine Learning to Advance Biomedical Research	Event Summaries	7/23/2018	19
(CRS) Artificial Intelligence (AI) and Education	Scientific and Technical Reports	8/1/2018	3
(USAID) Reflecting the Past, Shaping the Future: Making AI Work for International Development	Scientific and Technical Reports	9/5/2018	98
(NSF) Request for Information on Update to the 2016 National Artificial Intelligence Research and Development Strategic Plan	Requests for Information	9/26/2018	1
(DOT) Preparing for the Future of Transportation: Automated Vehicles 3.0	Strategy Documents	10/1/2018	80
(CRS) U.S. Ground Forces Robotics and Autonomous Systems (RAS) and Artificial Intelligence (AI): Considerations for Congress	Scientific and Technical Reports	11/20/2018	47
(NSTC) Charting a Course for Success: America's Strategy for STEM Education	Strategy Documents	12/1/2018	48
(ODNI) The AIM Initiative: A Strategy for Augmenting Intelligence Using Machines	Strategy Documents	1/1/2019	25
(DoD) Department of Defense Artificial Intelligence Strategy	Strategy Documents	2/1/2019	17
(DOE) Workshop Report on Basic Research Needs for Scientific Machine Learning: Core Technologies for Artificial Intelligence	Scientific and Technical Reports	2/10/2019	109
(OSTP) Artificial Intelligence and Quantum Information Science Research and Development Summary: Fiscal Years 2020-2021	AI Budget	2/11/2019	6
(DOJ) Data Strategy for the U.S. Department of Justice	Strategy Documents	2/26/2019	13
(USAID) AI in Global Health: Defining a Collective Path Forward	Scientific and Technical Reports	4/1/2019	42
(DOC/NIST) Artificial Intelligence Standards	Requests for Information	5/1/2019	3
(OECD) Recommendation on AI	Ethical Principles; International Declarations*	5/1/2019	11
(G20) AI Principles	International Declarations*	6/1/2019	14
(NSTC) National AI R&D Strategic Plan: 2019 Update	Strategy Documents*	6/21/2019	50
(DOE) AI for Science	Scientific and Technical Reports	7/1/2019	224
(OMB) Identifying Priority Access or Quality Improvements for Federal Data and Models for Artificial Intelligence Research and Development and Testing	Requests for Information	7/10/2019	2
(G7) Biarritz Strategy for an Open, Free and Secure Digital Transformation	International Declarations	8/1/2019	2

(DOC/NIST) U.S. Leadership in AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools	Strategy Documents	8/10/2019	52
(DOC/USPTO) Request for Comments on Patenting Artificial Intelligence Inventions	Requests for Information	8/27/2019	1
(DoD/Air Force) Air Force Artificial Intelligence Annex to DoD AI Strategy	Strategy Documents	9/1/2019	6
(OSTP) Summary of the 2019 White House Summit on Artificial Intelligence in Government	Event Summaries	9/9/2019	7
(NSTC) NITRD Supplement to the President's FY 2020 Budget	AI Budget	9/10/2019	52
(DOC/NIST) A Taxonomy and Terminology of Adversarial Machine Learning	Scientific and Technical Reports	10/1/2019	36
(DOC/USPTO) Request for Comments on Intellectual Property Protection for Artificial Intelligence Innovation	Requests for Information	10/30/2019	2
(DOC/NOAA) Public Comment for the Four Draft NOAA Science and Technology Strategies: NOAA Unmanned Systems, Artificial Intelligence, 'Omics, and Cloud Strategies	Requests for Information	11/14/2019	2
(NSTC) 2016–2019 Progress Report: Advancing Artificial Intelligence R&D	AI Budget	11/20/2019	48
(DOT) Raising Awareness of Artificial Intelligence for Transportation Systems Management and Operations	Scientific and Technical Reports	12/1/2019	76
(DOT) Ensuring American Leadership in Automated Vehicle Technologies 4.0	Strategy Documents	1/1/2020	56
(OMB) Request for Comments on a Draft Memorandum to the Heads of Executive Departments and Agencies, "Guidance for Regulation of Artificial Intelligence Applications"	Requests for Information*	1/13/2020	1
(DOC/NOAA) Artificial Intelligence Strategy	Strategy Documents	2/1/2020	8
(OSTP) American Artificial Intelligence Initiative: Year One Annual Report	Strategy Documents*	2/1/2020	36
(DoD) Ethical Principles for Artificial Intelligence	Ethical Principles	2/24/2020	1
(ODNI) Principles of Artificial Intelligence Ethics for the Intelligence Community	Ethical Principles	2/24/2020	1
(NSTC) Artificial Intelligence and Cybersecurity: Opportunities and Challenges: Technical Workshop Summary Report	Event Summaries	3/2/2020	15
(OMB) Federal Data Strategy: 2020 Action Plan	Strategy Documents	5/14/2020	52
(G7) Science and Technology Ministers' Declaration on COVID-19	International Declarations	5/28/2020	6
(NITRD) Artificial Intelligence and Cybersecurity: A Detailed Technical Workshop Report	Scientific and Technical Reports	6/2/2020	23
(DOC/NOAA) Data Strategy: Maximizing the Value of NOAA Data	Strategy Documents	7/1/2020	12
(ODNI) Artificial Intelligence Ethics Framework for the Intelligence Community	Ethical Principles	7/1/2020	6
(GAO) Facial Recognition Technology: Privacy and Accuracy Issues Related to Commercial Uses	Scientific and Technical Reports	7/13/2020	66
(DOC/NIST) Four Principles of Explainable Artificial Intelligence	Scientific and Technical Reports	8/1/2020	30
(NITRD) AI R&D Dashboard, Networking and Information Technology R&D Program	AI Budget	8/14/2020	5
(NSTC) NITRD Supplement to the President's FY2021 Budget	AI Budget	8/14/2020*	40
(DoD) DoD Data Strategy	Strategy Documents	9/1/2020	16
(GAO) Facial Recognition: CBP and TSA are Taking Steps to Implement Programs, but CBP Should Address Privacy and System Performance Issues	Scientific and Technical Reports	9/2/2020	111
(DOE/ASCAC) Final Report of the Subcommittee on AI/ML, Data-intensive Science and High-Performance Computing	Scientific and Technical Reports	9/23/2020	62
(State) Declaration of U.S. and UK Cooperation in AI R&D	International Declarations	9/25/2020	4
(DOC/USPTO) Inventing AI: Tracing the diffusion of artificial intelligence with U.S. patents	Scientific and Technical Reports	10/1/2020	19
(FOC) FOC Joint Statement on Artificial Intelligence and Human Rights	Ethical Principles; International Declarations	11/5/2020	6
(CRS) Artificial Intelligence and National Security	Scientific and Technical Reports	11/10/2020	43
(NSTC) Recommendations for Leveraging Cloud Computing Resources for Federally Funded Artificial Intelligence Research and Development	AI Budget; Scientific and Technical Reports	11/17/2020	14
(DHS) U.S. Department of Homeland Security Artificial Intelligence Strategy	Strategy Documents	12/3/2020	18
(White House) Executive Order Promoting the Use of Trustworthy AI in the Federal Government	Ethical Principles*	12/8/2020	5
(DOJ) Artificial Intelligence Strategy for the U.S. Department of Justice	Strategy Documents	12/15/2020	13

* indicates documents with special importance and application across sectors of government

A.2 Codebook Details

Table A.2: Final codebook with definitions, examples, and sources

Domain and Code	Brief Definition	Example Quote	Source
Focusing Events			
Focusing Events: Expert Concerns	Predictions and concerns raised by prominent experts or organizations that gain enough attention to constitute focusing events, including statements conveyed through workshops, research publications, and news media.	"In the past three years, newly established as well as longstanding conferences, workshops, and task forces have prioritized human-AI collaboration broadly. For example, the Conference on Human Computation and Crowdsourcing has grown from a workshop to a major international conference that fosters research in the intersection of AI and human-computer interaction (HCI)."	06.21.19 (NSTC)
Focusing Events: Games	The success of AI systems in demonstrating performance through achievement in games or similar activities, e.g., Deep Blue, AlphaGo, Watson in the context of playing chess, Go, or Jeopardy.	"Artificial intelligence (AI), especially its sub-discipline machine learning (ML), has shown dramatic advances in autonomous systems, computer vision, natural language processing, and game playing. These AI systems can perform tasks significantly beyond what was possible only recently (e.g., autonomous systems) and in some cases even beyond what humans can achieve (e.g., chess and Go)."	01.01.19 (ODNI)
Focusing Events: Geopolitics & Military	Concerns raised with respect to foreign states taking action on AI research, implementation and policy.	"Recognizing the importance of these technologies, in 2017 the Chinese government reportedly stated its goal of being the world's premier artificial intelligence innovation center by 2030, with Russian President Vladimir Putin stating, "Whoever becomes the leader in this sphere [AI] will become ruler of the world."	11.20.18 (CRS)
Focusing Events: Industry & Government Adoption	References to increased adoption of AI systems, typically as a signal to urgency and AI's potential.	"Significant advances in robotic technologies over the last decade are leading to potential impacts in a multiplicity of applications, including manufacturing, logistics, medicine, healthcare, defense and national security, agriculture, and consumer products."	06.21.19 (NSTC)
Focusing Events: Protests	Protests by employees or members of the public in response to policies or activities where AI is used.	"Moreover, some companies are refusing to work with DOD due to ethical concerns over the government's use of AI in surveillance or weapon systems. Notably, Google canceled existing government contracts for two robotics companies it acquired—Boston Dynamics and Schaft— and prohibited future government work for DeepMind, a Google-acquired AI software startup. In May 2018, Google employees successfully lobbied the company to withdraw from Project Maven and refrain from further collaboration with DOD."	11.10.20 (CRS)
Focusing Events: Technical Performance & Advances	Discussion of recent technical achievements and benchmarks demonstrating AI's potential, e.g., ImageNet.	"NHTSA recognizes that the accelerating pace of technological change, especially in the development of software used in ADS-equipped vehicles, requires a new approach to the formulation of the Federal Motor Vehicle Safety Standards (FMVSS)."	10.01.18 (DOT)
Focusing Events: Scandals & Disasters	Examples of AI incidents causing harm or raising alarm, e.g., autonomous vehicle crashes.	"For example, a facial recognition start-up company is currently facing a number of lawsuits alleging it used web scraping to amass a data set of 3 billion facial images from millions of websites without obtaining the consent of individuals in the images or the companies whose websites were scraped."	07.13.20 (GAO)
Indicators			
Indicators: Economic & Workforce	References to indicators related to economic growth, challenges, labor disruption, etc.	"PricewaterhouseCoopers estimated that by 2030, AI technologies could increase North American gross domestic product (GDP) by $3.7 trillion and global GDP in $15.7 trillion."	07.01.18 (Treasury)
Indicators: Education	References to indicators of educational achievement and gaps.	"In addition, China faces challenges in recruiting and retaining AI engineers and researchers. Over half of the data scientists in the United States have been working in the field for over 10 years, while roughly the same proportion of data scientists in China have less than 5 years of experience. Furthermore, fewer than 30 Chinese universities produce AI-focused experts and research products."	11.10.20 (CRS)
Indicators: Expert Concerns	References to indicators surfaced via the opinion or prediction of experts.	"several use-cases appear poised for more wide-spread adoption. Within the banking industry, for example, large percentages of U.S. banks report either current or planned AI deployment within the next 18 months across the following use-cases: more than 60% in biometrics, about 60% in fraud & security detection, about 55% in chatbots or robo-advisers; and about 35% in voice assistants."	07.01.18 (Treasury)
Indicators: Geopolitics & Military	References to indicators of adoption of AI in the military context or in the context of concern about foreign state actors.	"According to one estimate, China is on track to possess 30% of the world's share of data by 2030."	11.10.20 (CRS)
Indicators: Poverty, Harm, & Fairness	References to indicators illuminating social harms or disparities.	"As of the end of February 2020, FTC had received approximately 155 complaints related to facial recognition technology.57 complaints included privacy concerns related to social media companies, technology not working, and fraudulent misuse of the technology."	07.13.20 (GAO)
Indicators: Scandals & Disasters	References to indicators in the context of evidence of harmful AI incidents.	"Driverless vehicles are still far from perfect, as evidenced by crashes with fire trucks… and the recent fatalities of drivers and pedestrians."	12.01.19 (DOT)
Indicators: Technical Performance & Advances	References to indicators of increased performance and adoption of AI systems.	"Since 2017, there has been a profound increase in DL studies, of which 9 percent are in radiology. DL and ML have improved prediction of disease."	07.23.18 (HHS)
Issue Frames			
Issue Frames: Geopolitics	Discussion highlighting competition with foreign actors and the need to maintain technological, economic, or military leadership.	"The opportunity is great; the threat is real; the approach must be bold: Recognizing that strategic advantage is fleeting and fragile, the IC must be willing to rethink or abandon processes and mechanisms designed for an earlier era, establish disciplined engineering and operations practices, and maintain an absolute focus on assuring advantage in an intensely competitive global adversarial environment."	01.01.19 (ODNI)
Issue Frames: Innovation	Discussion emphasizing AI's prospects for economic growth and transformation, efficiency gains, and other innovation-oriented goals.	"These actions reflect the Administration's recognition of the importance of STEM education and training as a driver of American job creation and economic prosperity."	12.01.18 (NSTC)

Issue Frames: Ethics	Discussion featuring AI's ethical and social implications, such as impacts on the public good and ethical risks associated with AI.	"The responsible development and use of AI can be a driving force to help advance the SDGs and to realize a sustainable and inclusive society, mitigating risks to wider societal values. The benefits brought by the responsible use of AI can improve the work environment and quality of life, and create potential for realizing a human-centered future society with opportunities for everyone, including women and girls as well as vulnerable groups."	06.01.19 (G20)
Problems			
Problems: Accountability & Responsibility	Concerns about the need to hold actors responsible and to create accountability mechanisms when AI is used.	"When we look at the challenge of bringing fully safe, largely automated vehicles to the road, we shift risk and return where the automotive manufacturer . . . now bears all the risk for the vehicle's behavior—all of the risk and none of the benefit. And so we have this public-health situation where industry has contributed to significantly improving public health by virtue of improving the fleet, and taking full responsibility for anything that goes wrong."	03.01.18 (GAO)
Problems: Data Quality & Access	Concerns about a lack of access to high-quality data, typically needed to enable more widespread development of AI systems, or of more robust or responsible systems.	"Much of the discourse around ML assumes organizations are awash in data. In international development we often face the opposite problem — we are most concerned about the areas of the world where data availability is lowest. When data are scarce and expensive, then their collection and analysis can be in tension with the primary goals of a development project."	09.05.18 (USAID)
Problems: Fairness & Bias	Concerns about the potential of AI systems to reflect social or technical biases, manifesting in unfair predictions or system outcomes.	"If we aim to change an unjust status quo, predictions based on what happened in the past might be unhelpful, even if they are highly accurate. For example, if women have traditionally faced discrimination in hiring, then an algorithm that scores resumes based on past hiring records will discriminate against women."	09.05.18 (USAID)
Problems: Human & Civil Rights	Concerns that adoption of AI systems may negatively impact human rights or be used to suppress civil rights.	"Their investments threaten to erode U.S. military advantage, destabilize the free and open international order, and challenge our values and traditions with respect to human rights and individual liberties."	02.01.19 (DOD)
Problems: Inequality & Inclusion	Concerns that disparate access to AI systems or downstream consequences of AI adoption may exacerbate inequality or poverty.	"When these tools fail unevenly for different groups of people, the people affected may be unable to use a payment system, singled out for enhanced screening at border crossings, or wrongly called into a police station for questioning. The cumulative burden of this "selective" failure can be substantial, effectively compounding existing marginalization or inequity."	09.05.18 (USAID)
Problems: Misuse & Hostile Actors	Concerns about actors intentionally or deliberately misusing AI systems.	"Adversarial AI techniques represent opportunities and risks. We have highly sophisticated adversaries with access to the same tools, their own data, and experts trained in the same universities as our own people. AI is merely one of the new battlegrounds for a technology-based arms race."	01.01.19 (ODNI)
Problems: Performance & Reliability	Concerns about the accuracy, performance, reliability, and robustness of AI systems.	"The credibility of research based on SciML requires that outcomes come from a process that is not sensitive to perturbations in training data, modeling choice, and/or computational errors. Progress in this PRD will require research for showing that SciML methods and implementations are well-posed, stable, and robust."	02.10.19 (DOE)
Problems: Privacy	Concerns about privacy risks and protections, as well as associated issues such as consent.	"Data privacy issues are particularly important for digital health and AI solutions since health data is generally government owned, raising concerns about private companies gaining access to the data and potentially profiting by leveraging it for their own uses."	04.01.19 (USAID)
Problems: Realizing Benefits	Concerns about the need to develop and adopt AI systems to take advantage of AI's benefits and the associated opportunity costs of not doing so.	"The costs of not implementing this strategy are clear. Failure to adopt AI will result in legacy systems irrelevant to the defense of our people, eroding cohesion among allies and partners, reduced access to markets that will contribute to a decline in our prosperity and standard of living, and growing challenges to societies that have been built upon individual freedoms."	02.01.19 (DOD)
Problems: Safety	Concerns about safety risks and incidents associated with AI, typically involving physical safety.	"NHTSA has established a Vehicle Cybersecurity Response Process for Incidents Involving Safety-Critical Systems. During a significant incident, coordination will be handled through DHS's National Cybersecurity & Communications Integration Center (NCCIC), with NHTSA having an information/advisory role and performing its statutory responsibility under the Safety Act."	01.01.20 (DOT)
Problems: Security & Military	Concerns about security risks and threats, including those implicating military and cybersecurity domains.	"Many experts worry that current law, passed largely before AI became a major policy consideration, is insufficient to address today's cybersecurity threats."	08.01.18 (CRS)
Problems: Transparency	Concerns about the opacity of AI systems and/or of the broader processes involved in creating and implementing them.	"One popular algorithm is so complex and opaque that a corrections official described it as a "giant correctional pinball machine." For an individual case, it can be nearly impossible to point to the precise factors that led to a low or high score."	09.05.18 (USAID)
Problems: Trust	Concerns that users or potential of AI systems, including both professional practitioners and members of the public, may lack sufficient trust in AI systems or that the systems are themselves not trustworthy.	"Although many factors contribute to the production of harms, the disproportionate trust that is often placed in ML-based tools is particularly worrisome. When these tools are not only fallible, but used at scale, they can be sources of significant harm. Excessive trust can be dangerous when it leads to unquestioning acceptance of model results, which can result in misinformed choices when models do get it wrong."	09.05.18 (USAID)
Problems: Value Alignment	Concerns that AI systems may not be compatible with societal values and ethical principles.	"Likewise, the challenge of understanding and designing human-AI ethics and value alignment into systems remains an open research area."	06.21.19 (NSTC)
Problems: Vulnerable Populations	Concerns about the impact of AI systems on vulnerable groups such as children, the elderly, and individuals with disabilities.	"It is vital to be inclusive in teaching CT and CS skills even to individuals who may struggle to master these approaches, including students with disabilities, so that they, too, can fully participate in modern society. "	12.01.18 (NSTC)
Problems: Workforce & Education	Concerns about disruptions to the workforce resulting from AI and the need to reform education and training to keep pace with these disruptions.	"What is already clear and certain is that new technical developments will have a fundamental impact on the global labor market within the next few years, not just on industrial jobs but on the core of human tasks in the service sector that are considered 'untouchable.' Economic structures, working relationships, job profiles and well-established working time and remuneration models will undergo major changes."	11.20.18 (CRS)

Solutions			
Solutions: Build Trust	Calls to promote and facilitate greater trust, comfort, and acceptance of AI systems.	"RECOGNISING that trust is a key enabler of digital transformation; that, although the nature of future AI applications and their implications may be hard to foresee, the trustworthiness of AI systems is a key factor for the diffusion and adoption of AI."	05.01.19 (OECD)
Solutions: Cooperation & Dialogue	Calls to engage in collaboration across entities via dialogue, information sharing, and coordination.	"Include discussion of NOAA AI activity in NOAA executive-level engagement and communications with key stakeholders, particularly focusing on OMB, Congressional members and staff, and counterparts from other federal agencies."	02.01.20 (DOC_NOAA)
Solutions: Data Quality & Access	Calls to make more high-quality datasets available to facilitate research, development, and adoption of AI systems.	"Data is at the heart of AI and ML, so it is important to expand it as a resource to continue innovation"	07.23.18 (HHS)
Solutions: Diverse Participation	Calls for more diverse interdisciplinary, cross sector, or multi-demographic participation of multiple actors or organizations.	"A wide body of research has established that organizations that are diverse in terms of gender, race, socioeconomic status, ethnicity, ability, geography, religion, etc., and provide an inclusive environment that values diversity better retain talent, are more engaged and productive, are more innovative, and generally are higher-performing organizations."	12.01.18 (NSTC)
Solutions: Grants & Procurement	Calls to reform procurement and grantmaking practices to incentivize AI research, development, and adoption.	"Updating acquisition processes across agencies to include specific requirements for AI standards in requests for proposals will encourage the community to further engage in standards development and adoption."	10.10.16 (NSTC)
Solutions: Human-AI Teaming	Calls to facilitate effective sharing of responsibilities between humans and AI systems, or to preserve some degree of human control.	"The walls between humans and AI systems are slowly beginning to erode, with AI systems augmenting and enhancing human capabilities. Fundamental research is needed to develop effective methods for human-AI interaction and collaboration, as outlined in Strategy 2."	06.21.19 (NSTC)
Solutions: Impact Assessment & Testing	Calls to test or evaluate AI systems through conformity assessment, model evaluation, verification, validation, sensitivity analysis, assurance models, performance testing, stress testing, interoperability testing, etc.	"For reliable and credible use of SciML, we need the ability to rigorously quantify ML performance in these outcomes. Performance measurement implies an assessment of quality, as well as a cost measure of computations and/or data preparation and management."	02.10.19 (DOE)
Solutions: Monitoring & Reporting	Calls for developers, users, and especially regulators of AI systems to monitor and report on progress and risks associated with AI.	"For example, the failure of Google Flu Trends was noticed because the Centers for Disease Control maintain and publish accurate records of flu cases. If something like Google Flu Trends were used as a substitute for CDC reporting, its model drift may never have been noticed."	09.05.18 (USAID)
Solutions: Pilot Projects & Testbeds	Calls to initiate testbeds, grand challenges, sandboxes, prototyping, and demonstration projects.	"Provide funding to early-stage AI concepts to pilot or otherwise show proof of concept, for example for funding field tests of AI-enabled ultrasound device in LMICs, and encourage future procurement if validated."	04.01.19 (USAID)
Solutions: Public Engagement	Calls to engage the public in AI policy via sharing or receiving information or through two-way dialogue.	"Public trust will strengthen the Department's use of AI by ensuring the support of the citizens it intends to protect and guard against reputational impacts to the Department. DHS will facilitate this trust by engaging strategically with the American public regarding AI."	12.03.20 (DHS)
Solutions: R&D & Adoption	Calls for more research, development, investment, and adoption of AI systems.	"The research priorities outlined in this AI R&D Strategic Plan focus on areas that industry is unlikely to address on their own, and thus, areas that are most likely to benefit from Federal investment. These priorities cut across all of AI to include needs common to the AI sub-fields of perception, automated reasoning/planning, cognitive systems, machine learning, natural language processing, robotics, and related fields."	06.21.19 (NSTC)
Solutions: Social/Ethical Consideration	Calls for developers of AI systems to adopt human-centered approaches, ethical frameworks, and align with human ethics and values.	"AI actors should respect the rule of law, human rights and democratic values, throughout the AI system lifecycle. These include freedom, dignity and autonomy, privacy and data protection, non-discrimination and equality, diversity, fairness, social justice, and internationally recognized labor rights."	06.01.19 (OECD)
Solutions: Standards & Best Practices	Calls to formulate and adhere to formal standards or informal best practices surrounding AI development and implementation, including the use of benchmarks and standardized metrics.	"Likewise, the use of common metrics and data sources can enable benchmarking of Federal STEM education programs along with those in the private sector. Such benchmarking enables stakeholders in all arenas to assess their performance within a broader context. Ultimately, the adoption of common metrics will be critical to developing the evidence base to permit efficient, rigorous, and integrated assessments of the range of Federal initiatives designed to achieve the goals of this strategic plan."	12.01.18 (NSTC)
Solutions: Technical Fixes for Ethics	Calls to specifically perform research to address ethical concerns associated with AI, such as research surrounding fairness, transparency, and privacy.	"Investing in research and development for resilient, robust, reliable, and secure AI. In order to ensure DoD AI systems are safe, secure, and robust, we will fund research into AI systems that have a lower risk of accidents; are more resilient, including to hacking and adversarial spoofing; demonstrate less unexpected behavior; and minimize bias."	02.01.19 (DOD)
Solutions: Transparency	Calls to enhance transparency in the development and implementation of AI systems, either through promoting technical or process transparency to provide more information to stakeholders.	"AI Actors should commit to transparency and responsible disclosure regarding AI systems. To this end, they should provide meaningful information, appropriate to the context, and consistent with the state of art."	06.01.19 (OECD)
Stakeholders			
Stakeholders: Experts	Individuals in government, industry, and academia with a professional stake in or expertise related to AI.	"The summit brought together over 100 senior government officials, technical experts from top academic institutions, heads of industrial research labs, and American business leaders."	05.10.18 (OSTP)
Stakeholders: Public	Members of the general public, such as customers and citizens.	"At each step of model building, it is important to bring in perspectives of those who are "closer" to the realities in the field. They can help validate model results and lend credibility to the insights gained from the work."	09.05.18 (USAID)

APPENDIX B

SUPPORTING INFORMATION FOR CHAPTER 3

B.1 AI and Issue Frame Dictionaries

This section presents the dictionaries used for extracting AI-related messages and the three issue frames studied in this paper. While there were some minor differences in the initial keywords and validation efforts with respect to the different data sources examined in this study, the resulting dictionaries were substantially similar, so only one dictionary is presented for AI overall and for the three issue frames. Of note, the dictionaries were adjusted throughout the research process as classification improved; however, the main results were not sensitive to these minor changes. Nevertheless, differences in data sources, initial search criteria, length of texts, time period, etc., imply that is important for researchers to verify and potentially adjust these AI dictionaries when applying them to another context or data source (Grimmer & Stewart, 2013). In addition to the dictionaries themselves, the high-level categories listed here can also be used or repurposed to extract content related to specific sub-dimensions of AI discourse that may be of interest to researchers, such as 'discrimination' or 'military.'

Dictionary for identifying AI discourse:
- [AI]: ai*, a.i.*, artificial* intellig*, artificialintelligen*, artificial-intell*
- [automation]: automate, automated, automation, autonomous syst*, autonomous tech*
- [ML]: machine learn*, deep learn*, reinforce* learn*, supervised learn*, unsupervised learn*
- [algorithm]: algorithm*
- [other_techniques]: recommend* system*, speech recognition, computer vision, pattern recognition, natural language, machine translation, recommendation system*, image process*, information retrieval, speech synthesis, handwriting recognition, object recognition
- [AVs]: autonomous veh*, self-driv*, self driv*, autonomous car*, AVs, autonomous navigation, AV START, AV test*, AV tech*
- [facial_recognition]: facial recog*
- [deepfakes]: deepfake*, deep fake*
- [weapons]: autonomous weapon*

Dictionary for identifying AI's economic and innovative dimensions—"innovation frame":
- [innovation]: innovat*, grow, growth, patent*, invent*, invest*, billion*
- [economy]: economy, economies, economic
- [transformation]: transform*, chang*, revolution*, disrupt*, reinvent*, disrupt*, reinvent*, shake
- [acceleration]: accelerat*, drive, drives, driven, enhanc*, boost*, fueled, fueling, enable, enabler*, enables, enabling, optimize, optimizing, optimizes, augment*
- [business]: business, businesses, start-up*, startup*, enterpris*
- [opportunity]: embrac*, opportunity, opportunities, ambitions, potential
- [productivity]: productiv*, efficien*, edge
- [pace]: pace, faster, momentum, rise

Dictionary for identifying AI's social and ethical dimensions—"ethics frame":
- [ethics]: ethic*, moral*, morality, evil
- [rights]: rights, freedom*
- [values]: human values, societal values, democratic values
- [responsiblity]: responsibil*, responsibleAI
- [control]: human control
- [fairness]: fairness*, fairer*, unfair*
- [discrimination]: discrim*, non-discrim*, anti-discrim*, antidiscrim*
- [transparency]: transparenc*, explainab*, interpretab*
- [safety]: safe*
- [accountability]: accountab*
- [privacy]: privac*
- [bias]: bias*, unbias, racis*, sexis*, homophob*, transphob*
- [equality]: equality*, inequal*, unequal*
- [equity]: equita*
- [humanity]: humanity, humane, human-cent*
- [trust]: trust, trustworth*, trusted, untrust*, fear*, afraid*, threat*
- [misuse]: misus*, abus*
- [manipulation]: manip*
- [harm]: harm, harmed, harmful, harming, harms
- [consent]: consent*
- [deception]: deceiv*, decept*, fooled
- [misinformation]: misinform*, propagand*, democracy
- [displacement]: displac*, unemploy*
- [diversity]: diverse, diversity
- [society]: societ*
- [sustainability]: sustainable, sustainability, environmental*

Dictionary for identifying AI's strategic, geopolitical dimensions—"competition frame":
- [competition]: competit*, compete, rivals, rivalry
- [position]: ahead*, left behind*, fall* behind, advantag*, rank, ranks
- [supremacy]: dominance, dominate, dominating, domination, supremacy
- [china_russia]: china*, chinese*, russia*
- [arms_race]: arms race, artificial intelligence race, AI race, global race, #AI race, sputnik
- [war]: war, warfare, battlefield*, weapon*, lethal*
- [military]: national security, militar*, army*, navy*, naval, air force*, marines, armed forces

B.2 Classification Results

To evaluate the quality of the AI issue frame dictionaries, I perform a manual classification on 1,000 randomly-selected messages from the public and separately on 1,000 randomly-selected messages from Congress. The results in Table B.1 report the precision (or positive predictive value) and recall (or sensitivity) for each dataset and issue frame. I emphasize these metrics and the overall F1 score given this study's focus on predicting the positive class, i.e., identifying the issue frames themselves.

Table B.1: Classification results for issue frames: Public and Congress datasets

Issue Frame	**Precision**	**Recall**	**F1 Score**
Public AI Messages			
Ethics	0.96	0.70	0.81
Innovation	0.98	0.95	0.96
Competition	0.86	0.91	0.89
Policymaker AI Messages			
Ethics	0.98	0.87	0.92
Innovation	0.93	0.93	0.93
Competition	0.98	0.79	0.87

Overall, the classification leads to adequate to strong performance for both social media datasets with F1 scores ranging from 0.81 to 0.96.[1] The innovation frame is captured most accurately across datasets, with slightly lower performance for the ethics frame in the public dataset and for the competition frame in the policymaker dataset. Of note, the dictionaries appear to perform slightly better on the policymaker messages, perhaps because these messages are, on average, more carefully-crafted as well as focused on a smaller subset of AI issues.

[1] Note that I do not perform classification on the NYT articles given that they are of secondary importance in this study and because of the more pressing need to evaluate the robustness of the dictionaries in the context of short Twitter messages. For readers interested in classifying news media messages with respect to AI issue frames, Ghosh and Loustaunau (2021) provide some promising approaches.

B.3 Issue Frame Timelines and Descriptive Statistics

Figure B.1 depicts the raw correlations between the three actors' time series data across the three issue frames plus AI messages overall. Across all three actors, there is modest correlation with respect to 'all AI messages' and for the ethics and innovation frames, but little correlation with respect to the competition frame. This provides preliminary evidence that AI issue attention is generally correlated across actors, though these raw correlations do not account for spurious trends and autocorrelation of each time series with itself.

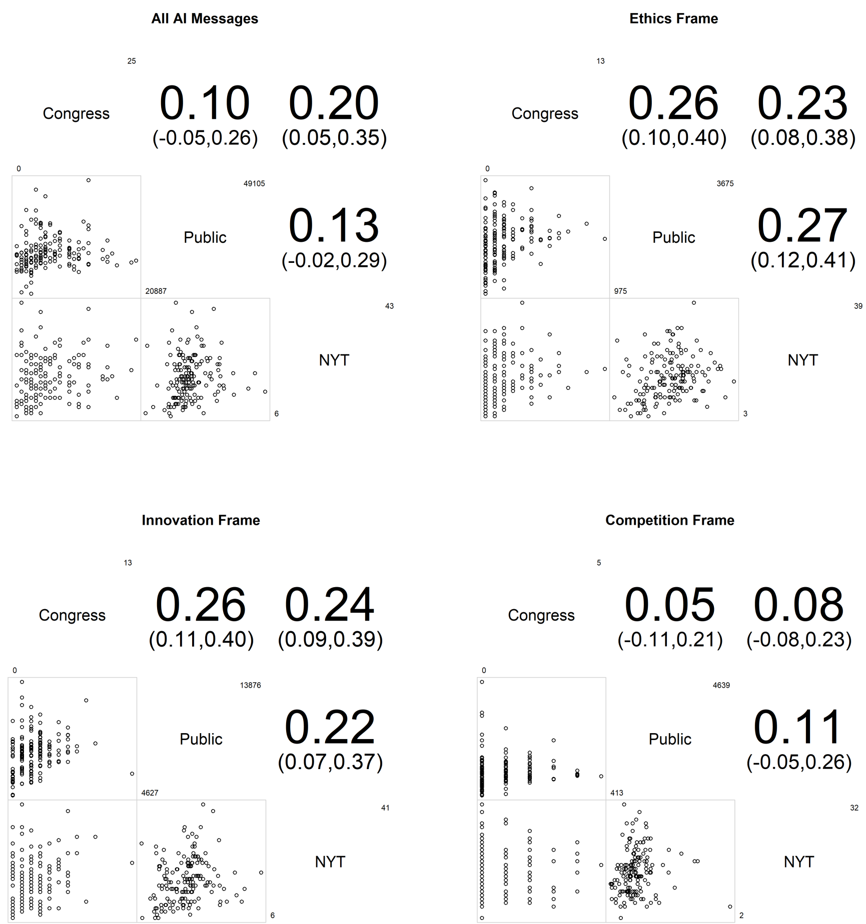

Figure B.1: AI issue frame correlations across public, policymakers, and media

Note: Depicts raw correlations between the 12 time series datasets: three issue frames plus overall AI messages across the three actors during 2017 to 2019. There are small correlations for AI messages overall, and for the innovation and ethics frames.

Beyond the question of whether single frames 'prevail' during a specific time period, frames may also merge over time or become subsumed into one another. Policymakers might attempt to advance both ethics and competition simultaneously, such as by arguing that ethical AI is a strategy to promote competitive advantage (Minkkinen et al., 2022).[2] Additionally, policymakers may employ the ethics and innovation frames simultaneously, such as by arguing that ethical AI is needed to foster trust and support consumer adoption.[3] Gilardi et al. (2021) find evidence that frames tend to become more complex over time—effectively by merging together. These changes may reflect the development of a stable consensus and dominant "policy image" (Baumgartner & Jones, 1991), at least in the near-term.[4]

Along these lines, Figure B.3 depicts the prevalence of issue frames over time for each actor while additionally incorporating hybrid or joint frames, e.g., *messages that discuss both ethics and innovation*. Overall, the vast majority of messages contain just a single issue frame, and there is little evidence of growth for any hybrid frame, particularly amongst the public and Congress. Because NYT articles are longer than tweets, however, they are more likely to reflect both individual and hybrid frames given the dictionary classification approach utilized in the study. However even here, the hybrid frames seem to merely track the growth of individual frames rather than represent a new framing consensus.

[2]For example, the European Commission's Ethical Guidelines for Trustworthy AI (2019, p. 5) state that ethical AI confers "competitive advantage" for individual firms and EU countries: "A trustworthy approach is key to enabling 'responsible competitiveness', by providing the foundation upon which all those affected by AI systems can trust that their design, development and use are lawful, ethical and robust."

[3]For example, according to U.S. Executive Order on Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government (2020, p. 1): "The ongoing adoption and acceptance of AI will depend significantly on public trust. Agencies must therefore design, develop, acquire, and use AI in a manner that fosters public trust and confidence while protecting privacy, civil rights, civil liberties, and American values."

[4]Is important to note that evidence of compounding may not decisively settle whether different frames are treated as similarly important, or whether some are subsumed into others in more superficial or strategic ways. For example, advocates of the ethics frame might adopt innovation language while proposing policies that innovation advocates reject. Alternatively, advocates of the competition frame might put forward ethical notions in a limited or abstract sense, such as by arguing that soldiers must "trust" autonomous weapons technologies. The analysis in this paper cannot fully settle to what extent frame compounding represents genuine synthesis, compromise, or strategic subsumption across frames. For scholars and stakeholders of AI policy, it will thus be important to distinguish "rhetorical frames" from "action frames," (Rein & Schön, 1996; Ulnicane et al., 2020) a determination that will only be possible over time as policies are adopted and implemented.

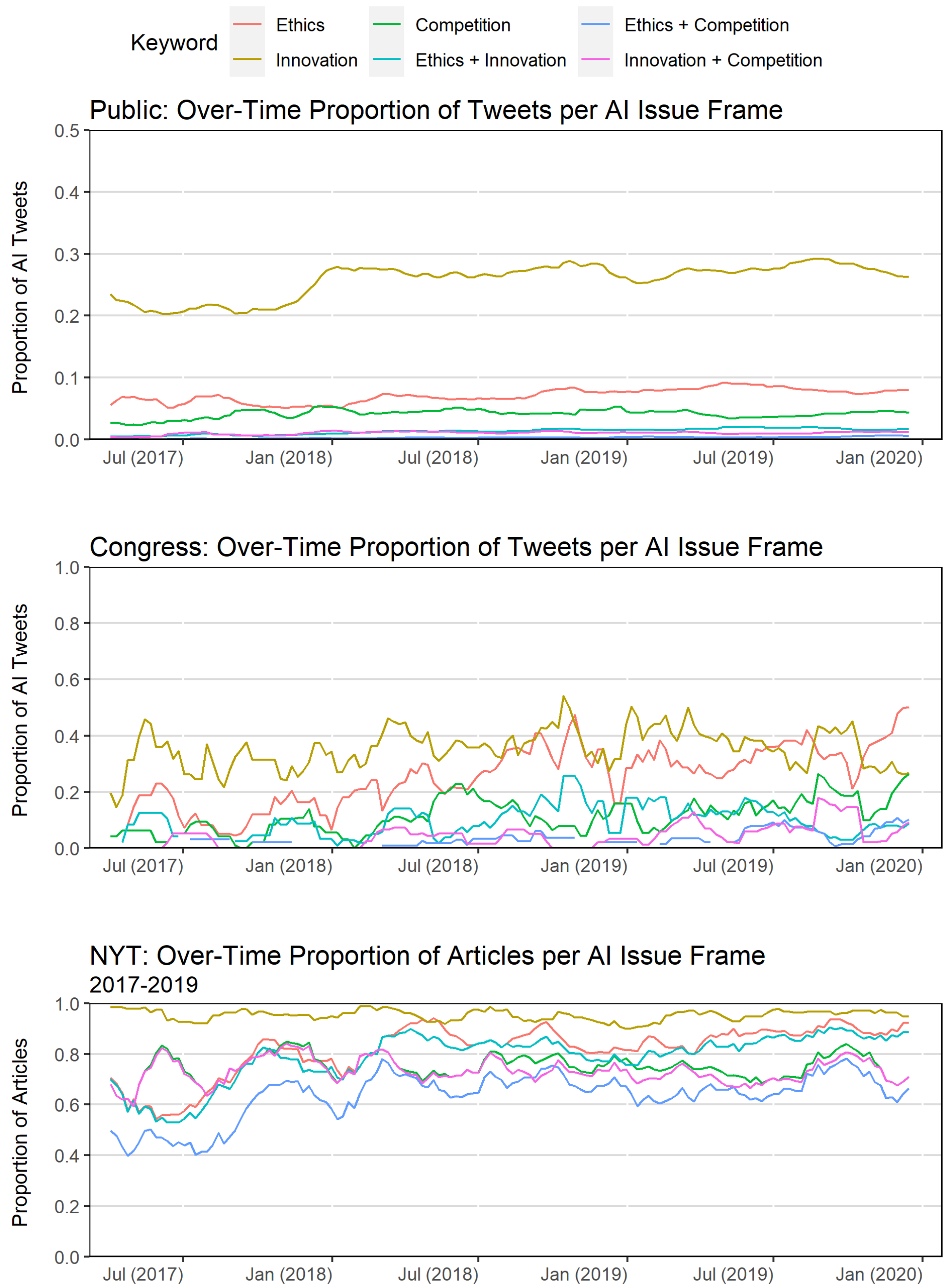

Figure B.3: Issue frame prevalence by actor with joint frames

Note: Single and joint issue frames from January 2017 through December 2019. I use 2-month rolling averages to stabilize trends, which means the first data points depicted are from February 2017. Sources are 1) tweets from the public using #AI, 2) tweets from the 115th and 116th Congresses on AI, and 3) NYT article coverage of 'artificial intelligence.'

B.4 Notes on ARIMA Analysis Approach

The ARIMA(p, d, q) models considered in this analysis take the general form:

$$y_t' = a_0 + \sum_{i=1}^{p} a_i * y_{t-i} + \sum_{i=1}^{q} b_i * \epsilon_{t-i} + \epsilon_t \tag{B.1}$$

where y_t' is a variable that has been differenced from itself one or more times to remove trends or other statistically undesirable properties that might lead to spurious correlations, (e.g., $d = 1$ corresponds to $y_t' = y_t - y_{t-1}$); p refers to the number of autoregressive terms, or lagged values of y; q refers to the number of moving average terms, or lagged prediction errors estimated from previous time periods; and ϵ refers to the unexplained random residual. Selecting the appropriate ARIMA(p, d, q) specification typically requires understanding the autocorrelation and partial autocorrelation functions for a given time series, and examining information criteria such as the Akaike Information Criterion (AIC) and Schwarz Bayesian Criterion (BIC) to identify which model produces stronger goodness of fit, balanced against ensuring parsimony in the number of terms to avoid overfitting. I prefer the use of the AIC or AICc (an AIC with a small sample correction) (Bozdogan, 1987) as underfitting is a more significant problem for small sample sizes (Dziak et al., 2012; Vaida & Blanchard, 2005).[5] A resulting model should have residuals that are uncorrelated and stationary, with constant error variance over time.

There are a total of 12 time series involved in this analysis (three datasets x mentions for each of three issue frames plus total number of mentions). Not all time series may have the same properties, which complicates model identification and choice of analytical method. However, analysis of seasonality using trend and seasonal decompositions suggests that that there are not seasonal patterns or cycles, so I use nonseasonal analyses throughout. Visual evidence and tests (the Augmented Dickey-Fuller Test or ADF test and

[5] Nevertheless, using alternative information criteria like the BIC does not lead to substantively different results across the main analyses.

Kwiatkowski–Phillips–Schmidt–Shin or KPSS test) also suggest at least some time series (e.g., the ethics frame) are characterized by over-time trends, which implies that one level of differencing may be appropriate to remove those trends and render the time series stationary, a prerequisite for time series modeling.

I aggregate the time series data to the level of week for three reasons. First, some times series data in my dataset, particularly policymaker messaging, is relatively sparse at the daily level. Second, messaging behavior indicates visible within-week relationships, such as a tendency to tweet the same day each week or more regularly on Fridays, and aggregation to the week level should smooth out these cycles. Third, theory and prior literature suggests that responsiveness to messages may occur over days-to-weeks rather than more quickly (hours) or slowly (months) (Gonzales, 2018). For example, news media and policymakers may only start to message about new legislation, hearings, or focusing events over several days rather than immediately.

Another question is whether the models should consider the leading actors to influence other actors simultaneously (i.e., within the same week) or with a lag (e.g., only a week or two later). Models performed based on one, two, or three week lags exclusive of contemporaneous influence exhibit decreasing and insignificant effects. Additionally, multi-model selection identifies contemporaneous influence models as the most plausible, and does not recommend including prior week lags at all. Tests of instantaneous and Granger causality using multi-actor VAR modeling also suggest that contemporaneous rather than past influence is important, especially for AI messages overall and for the innovation frame (with no evidence of instantaneous influence for the ethics and competition frames). And as stated, I expect attention to AI issue frames to manifest over fairly short time frames. Thus I adopt a contemporaneous influence modeling approach using ARIMA for the primary specifications, while I consider lagged influence more directly in the VAR analysis.

Using a small-sample adjusted AICc as the primary selection criteria, I search iteratively over possible stationary, non-seasonal models using maximum likelihood estimation

(MLE) to maximize fit. Results suggest that models such as ARIMA$(0, 1, 1)$ (simple exponential smoothing) and ARIMA$(0, 1, 2)$ (damped trend exponential smoothing) may be appropriate (McKenzie & Gardner, 2010) though the preferred model varies by time series. This strategy generally results in stationary models and residuals, i.e., independent errors without unit roots in the moving average or autoregressive terms. This provides some insight about the nature of the time series under study.

However, to extend this analysis to bivariate regressions, it is necessary to model multiple time series simultaneously. I therefore vary the standard ARIMA approach applied to a single time series and depicted in the equation above. The approach used here is ordinary least squares (OLS) regression using ARIMA to model errors, and including the predictor time series as an exogenous regressor. Thus the basic specification is a simple bivariate OLS with two of the three actors represented as independent or dependent variables:

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t \tag{B.2}$$

but with ϵ_t modelled in the style of Equation B.2, and with the x_t and y_t terms potentially subject to differencing to remove trends.

A benefit of this approach is that it allows for simple OLS-style interpretation, while avoiding problems associated with autoregressive errors. The AICc is used to identify the final model for each relevant bivariate regression of two actors, and differencing is applied to both time series variables in almost all of these models. Importantly, it is allowable to apply this approach even to nonstationary time series if some linear combination of those series is stationary, i.e., if the series are 'cointegrated.' Thus I do not engage in extra precleaning steps beyond differencing. The Phillips-Ouliaris (PO) test confirms that the series in the study are cointegrated in all cases. Also note that, while there is information about the past incorporated in each model, the modeling strategy implies that the effects of x on y are instantaneous or contemporaneous.

B.5 Notes on VAR Analysis Approach

In contrast to ARIMA, a VAR model allows for lagged values of the key dependent variable as well as lagged values of separate exogenous variables. The standard or reduced form of a VAR(1) representation (with a single lag) is:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{222} \end{bmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \tag{B.3}$$

such that $y_{1t} = a_{11}y_{1t-1} + a_{12}y_{2t-1} + \epsilon_{1t}$ and $y_{2t} = a_{21}y_{1t-1} + a_{22}y_{2t-1} + \epsilon_{2t}$. Thus, all key time series variables (e.g., public, policymaker, and media messages) are included in both the left- and right-hand sides of the respective equations, in a symmetric fashion.

Diagnostic tests (ADF, KPSS, and PO) provide mixed results regarding whether the relevant time series are stationary. For this reason, differencing (of all series) may be appropriate, though some research cautions that differencing will remove long-term dynamics (Brandt & Williams, 2006) and suggest using Vector Error Correction Models (VECM) instead. As long-term dynamics are not the central focus of this analysis, I stick with VAR analysis of differenced series. Next, I use the AIC to determine the appropriate lag order, p for a VAR(p) process, as AIC is again a preferable information criteria for smaller samples. A total of two to three (week) lags is deemed appropriate depending on whether differencing is applied, and seems sufficiently parsimonious. The results from VAR models are consistent with no autocorrelation or heteroskedasticity in the residuals, as desired.

Figure B.4 presents the forecast error variance decomposition based on all AI messages, indicating what proportion of the variance of a given actor's behavior over time is explained by lagged values of its own time series versus that of other actors. The results indicate that AI attention in the media and public is substantially autonomous: For both the public and NYT, only 1-2% of AI messaging behavior is explained by other actors. This is consistent

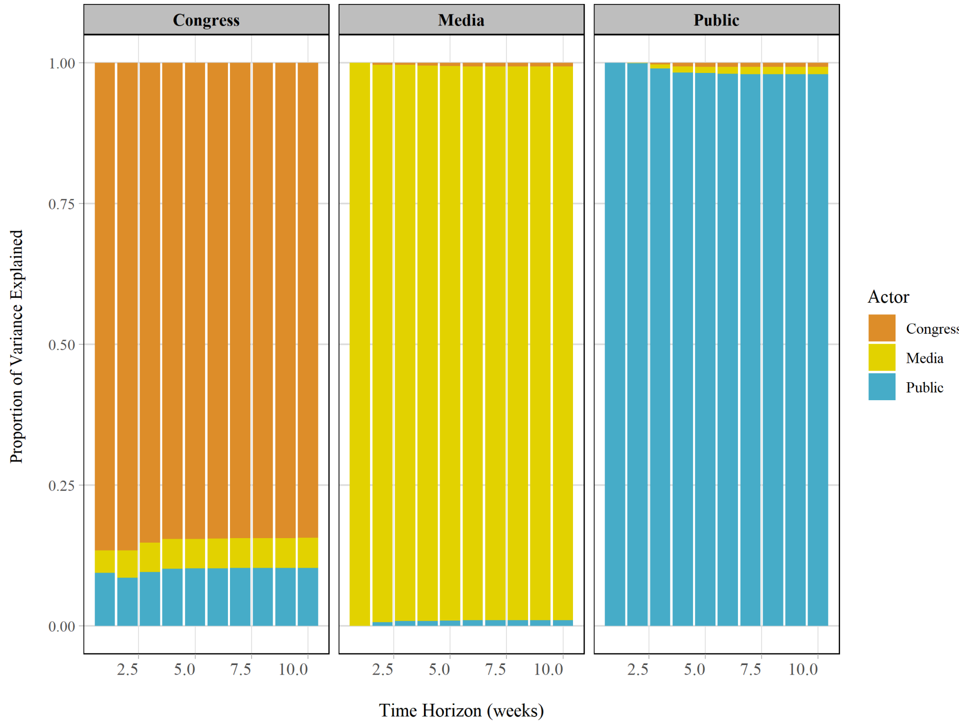

Figure B.4: Forecast error variance decomposition: Mutual influence of actors

with prior scholarship demonstrating a high degree of attentional inertia (Liu et al., 2011; Wood & Peake, 1998). In contrast, approximately 9-10% of policymaker AI messaging is explained by prior public messaging, with some influence (4-5%) from the media as well. Thus, when accounting for multidirectional influence and all three actors, the findings align with main ARIMA results and the Public Agenda-Setting Hypothesis.

B.6 Alternative Specifications: Fixed Effects and ARIMA Models

As an additional set of robustness checks, I alternatively run fixed effects models to assess the impact of public AI issue attention on Congressmember AI issue attention overall and per issue frame. This approach is quite distinct from the ARIMA and VAR models: Rather than aggregating all Congressmember messaging behavior into a single weekly count (representing policymaker attention as a whole), I consider impacts on individual *Congress members per week*. Specifically, I include Congressmember fixed effects in my model specification, and cluster standard errors by week as the "treatment" of public AI issue attention is delivered at the level of weeks. Finally, I include a simple linear time trend to account for the order of weeks.

Table B.2: Public impact on policymaker attention: Fixed effects specifications

	Congress AI Issue Attention			
	All AI Messages	Ethics	Innovation	Competition
	(1)	(2)	(3)	(4)
Public: AI Messages	0.002***			
	(0.001)			
Public: Ethics Frame		0.0002		
		(0.0002)		
Public: Innovation Frame			0.001***	
			(0.0003)	
Public: Competition Frame				0.00003
				(0.0001)
Week Trend	0.0001***	0.00005***	0.00002***	0.00002***
	(0.00001)	(0.00001)	(0.00001)	(0.00000)
Congressmember Fixed Effects	Yes	Yes	Yes	Yes
Observations	100,323	100,323	100,323	100,323
R^2	0.053	0.030	0.034	0.024
Note:				$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

The results in Table B.2 indicate that a one standard deviation increase in public attention to AI is associated with an statistically significant average increase of 0.0017 tweets per Congressmember per week. Multipled by the members of Congress in the dataset, this corresponds to 1.10 additional tweets per week across all of Congress. Similarly, increased public messaging about the innovation frame is associated with about 0.45 additional weekly tweets by Congress related to AI innovation, quite similar to the effect sizes found with the ARIMA models. Yet, also as with the ARIMA results, the public is only influential with respect to all AI messages and the innovation frame. In contrast, public messaging about ethics during a given week is associated with a very modest 0.099 increased tweets by Congress, while messaging about competition leads to a miniscule 0.021 increased tweets by Congress, and both of these latter associations are highly insignificant ($p = 0.47$ and $p = 0.75$, respectively).

The fact that this quite distinct modeling structure leads to similar findings with respect to public influence provides further confidence about the robustness of the ARIMA results. However, I consider two additional robustness checks using unit fixed effects based on Congressmember-week AI attention. First, I subset the sample to members of Congress who are particularly engaged in social media. In particular, I restrict the sample to members of Congress who are above the median in terms of total social media messaging over the study's time period. If public-policymaker engagement over social media indeed constitutes a valid channel for agenda-setting influence, we should expect these social media-engaged legislators should be particularly influenced by increased public attention.

The results in Table B.3 appear to confirm this expectation. Multiplying by the number of engaged legislators, increased public messaging is associated with 6.23 additional weekly tweets about AI across all social media-engaged members of Congress ($p < 0.01$), and 4.18 additional tweets about AI innovation ($p < 0.01$), but has minuscule and insignificant effects for the ethics frame (0.63 additional tweets, $p = 0.72$) and competition frame (0.22 additional tweets, $p = 0.87$).

Finally, I restrict the sample to only AI-engaged legislators, identified as legislators who tweeted about AI 10 or more times during the three year period. These legislators are amongst the most active on AI policy, including members of the House and Senate Congressional AI Caucuses who are often co-sponsors of AI-related legislation and members of committees that discuss AI in Congressional hearings and reports. Again, to the extent that the public does indeed influence policymaker issue attention to AI, we should expect that these members of Congress would be especially attentive and responsive.

The results in Table B.4 provide further confirmation that AI-engaged policymakers are particularly responsive to the public. Across all of Congress, increased public messag-

Table B.3: Public impact on policymaker attention: Social media-engaged policymakers

	Congress AI Issue Attention: Active Tweeters			
	All AI Messages	Ethics	Innovation	Competition
	(1)	(2)	(3)	(4)
Public: AI Messages	0.020***			
	(0.006)			
Public: Ethics Frame		0.002		
		(0.006)		
Public: Innovation Frame			0.014**	
			(0.006)	
Public: Competition Frame				0.001
				(0.005)
Week Trend	0.001***	0.001***	0.001***	0.001***
	(0.0002)	(0.0002)	(0.0001)	(0.0001)
Congressmember Fixed Effects	Yes	Yes	Yes	Yes
Observations	48,042	48,042	48,042	48,042
R^2	0.053	0.031	0.034	0.025
Note:				$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Note: Subset to $n = 306$ legislators who are actively engaged on Twitter, or legislators who tweeted above the median amount of times amongst legislators over the three year time period.

ing is associated with 5.45 additional weekly tweets about AI for all AI-engaged members of Congress ($p < 0.01$), and 6.05 additional tweets about AI innovation ($p < 0.01$), but still has relatively far smaller and insignificant effects for the ethics frame (0.73 additional tweets, $p = 0.64$) and competition frame (0.62 additional tweets, $p = 0.63$). Overall, these additional specifications not only help to confirm the key findings emerging from the time series models; they also provide some additional evidence that public issue attention is especially influential for policymakers who are engaged in AI policy, and suggest that public influence on the policy agenda may indeed be operating directly through social media as a channel.

One concern with respect to the use of time series models to assess multiple bivariate

Table B.4: Public impact on policymaker attention: AI-engaged policymakers

	Congress AI Issue Attention: AI-Engaged Legislators			
	All AI Messages	Ethics	Innovation	Competition
	(1)	(2)	(3)	(4)
Public: AI Messages	0.130***			
	(0.046)			
Public: Ethics Frame		0.017		
		(0.038)		
Public: Innovation Frame			0.144***	
			(0.052)	
Public: Competition Frame				0.015
				(0.031)
Week Trend	0.006***	0.008***	0.002**	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
Congressmember Fixed Effects	Yes	Yes	Yes	Yes
Observations	6,594	6,594	6,594	6,594
R^2	0.037	0.029	0.029	0.023
Note:				$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Note: Subset to $n = 42$ 'AI-engaged' legislators, or legislators who tweeted about AI at least 10 times over the three year time period.

relationships is that the identified "preferred" (p, d, q) model differs in some cases. Yet, to the extent that agenda-setting influence behavior is part of a stable underlying data generating process or processes, we might expect that the same kinds of temporal dynamics should be in play across actors and issue frames. Therefore, Table B.5 reproduces the main results in Table 3.2, but using a consistent time series specification for all actors and issue frames. In particular, I use an ARIMA$(0, 1, 2)$ or damped trend exponential smoothing approach to model the errors.

The results are highly stable here and with respect to other minor adjustments in the ARIMA modeling approach not displayed here. There are no major changes to the significance of the particular between-actor influence relationships, and effect sizes are largely

Table B.5: ARIMA$(0, 1, 2)$ results: Mutual influence of public, policymakers, and media

X influence on Y	ARIMA model	Effect Size	Baseline Value	% Change	p-value
Public on Congress					
All	(0,1,2)	1.48	6.66	22.2%	0.004
Ethics	(0,1,2)	0.26	1.98	13.1%	0.152
Innovation	(0,1,2)	0.47	2.28	20.6%	0.015
Competition	(0,1,2)	0.03	0.81	3.7%	0.728
Congress on Public					
All	(0,1,2)	753	31,322	2.4%	0.002
Ethics	(0,1,2)	60	2,205	2.7%	0.163
Innovation	(0,1,2)	147	8,046	1.8%	0.036
Competition	(0,1,2)	19	1,303	1.5%	0.679
NYT on Congress					
All	(0,1,2)	0.9	6.66	13.5%	0.019
Ethics	(0,1,2)	0.26	1.98	13.1%	0.119
Innovation	(0,1,2)	0.43	2.28	18.9%	0.009
Competition	(0,1,2)	0.03	0.81	3.7%	0.733
Congress on NYT					
All	(0,1,2)	1.41	18.56	7.6%	0.024
Ethics	(0,1,2)	0.94	15.3	6.1%	0.116
Innovation	(0,1,2)	1.4	17.71	7.9%	0.012
Competition	(0,1,2)	0.18	13.95	1.3%	0.73
NYT on Public					
All	(0,1,2)	-30.01	31,322	-0.1%	0.896
Ethics	(0,1,2)	40	2,205	1.8%	0.291
Innovation	(0,1,2)	15	8,046	0.2%	0.813
Competition	(0,1,2)	-16.23	1,303	-1.2%	0.695
Public on NYT					
All	(0,1,2)	0.3	18.56	1.6%	0.702
Ethics	(0,1,2)	0.45	15.3	2.9%	0.441
Innovation	(0,1,2)	0.52	17.71	2.9%	0.563
Competition	(0,1,2)	-0.17	13.95	-1.0%	0.759

Note: ARIMA Model refers to the preferred (p, d, q) specification. Effect Size refers to the number of additional AI messages by the influenced actor (Y) in a given week resulting from a one standard deviation increase of messaging from the influencer (X) during that week. Baseline Value refers to the average number of AI messages from the influenced actor (Y) per week. % Change refers to the increase in AI messaging over the baseline resulting from a one standard deviation increase of messaging from the influencer during that week. Statistical significant results at the 5% level are highlighted. Note that some digits are trimmed for legibility.

similar as well, providing some confidence in the robustness of the main paper results.

B.7 Additional VAR Results

The additional figures presented here depict orthogonal cumulative impulse response functions across the three actors and three main issue frames. Results complement main Figure 3.4 and confirm that there are few relationships between two actors reflecting issue attention influence. That is, in addition to the public's influence with respect to AI issue attention *overall* on Congress, only Figure B.6 depicting the innovation frame shows clear evidence of meaningful influence. This reiterates the study's general findings related to the Public Agenda-Setting Hypothesis and lack of evidence for the Special Role of the Public Hypothesis.

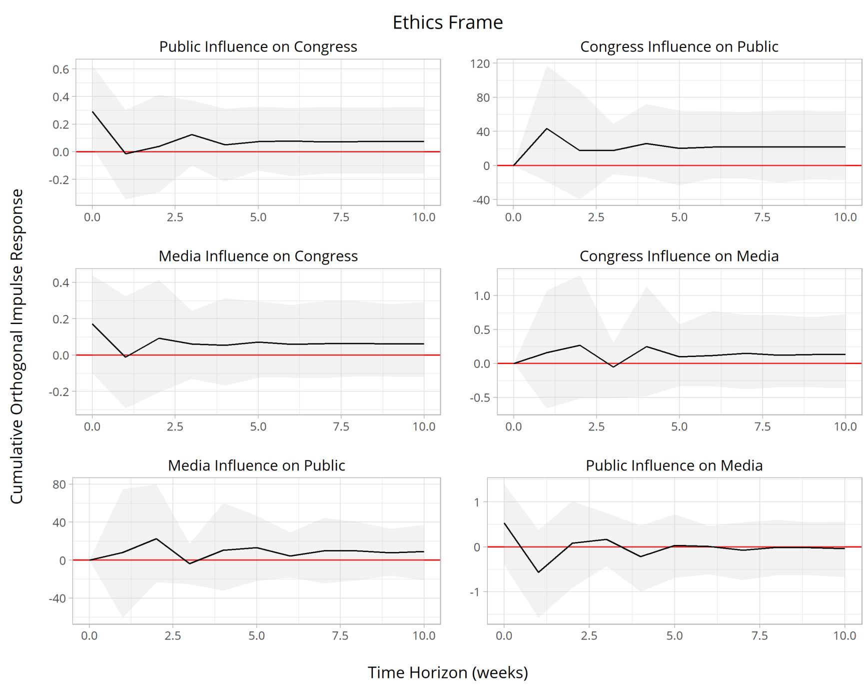

Figure B.5: Cumulative impulse response functions: Ethics frame

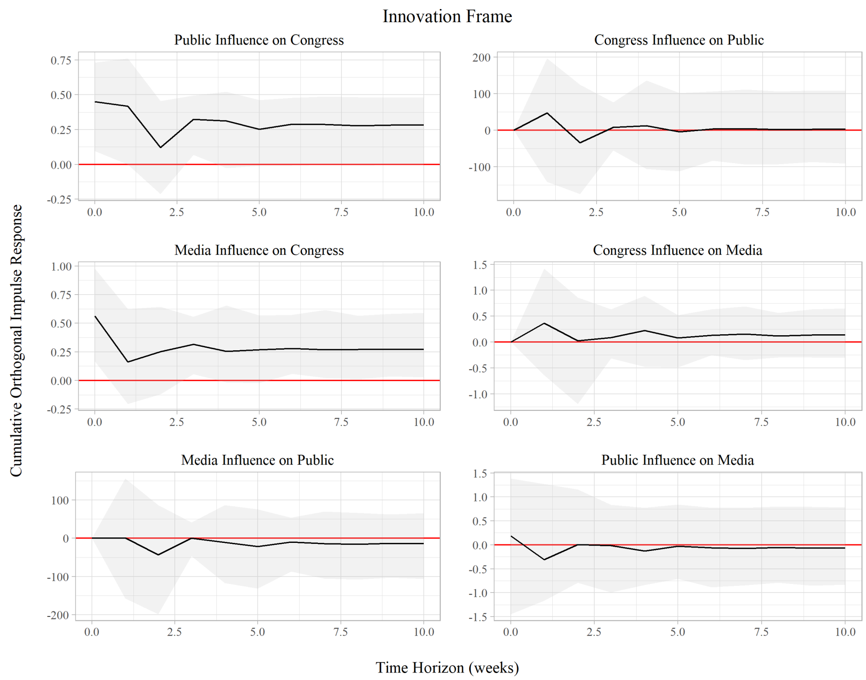

Figure B.6: Cumulative impulse response functions: Innovation frame

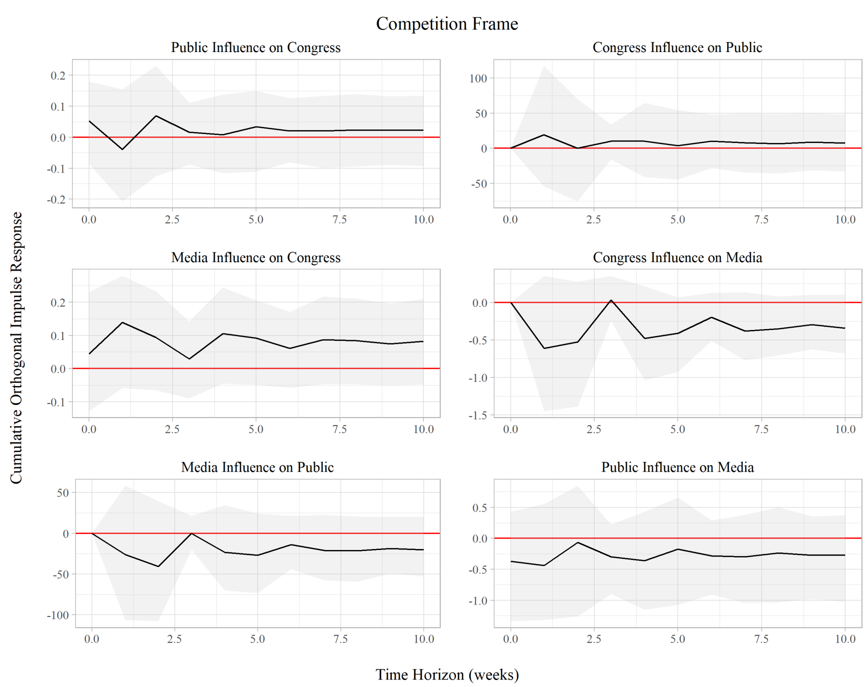

Figure B.7: Cumulative impulse response functions: Competition frame

B.8 Additional References for Appendix B

Baumgartner, F. R., & Jones, B. D. (1991). Agenda dynamics and policy subsystems. *The Journal of Politics*, *53*(4), 1044–1074. https://doi.org/10.2307/2131866

Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, *52*(3), 345–370. https://doi.org/10.1007/BF02294361

Brandt, P. T., & Williams, J. T. (2006). *Multiple time series models*. SAGE Publications.

Dziak, J. J., Coffman, D. L., Lanza, S. T., & Li, R. (2012). *Sensitivity and specificity of information criteria* (No. 12-119). The Methodology Center, The Pennsylvania State University. State College, PA.

European Commission. (2019). *Ethical guidelines for trustworthy AI*. European Commission, High-Level Expert Group on Artificial Intelligence (AI HLEG). Brussels, Belgium. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419

Ghosh, S., & Loustaunau, P. (2021). RheFrameDetect: A text classification system for automatic detection of rhetorical frames in AI from open sources. *arXiv:2112.14933 [cs]*. http://arxiv.org/abs/2112.14933

Gilardi, F., Shipan, C. R., & Wüest, B. (2021). Policy diffusion: The issue-definition stage. *American Journal of Political Science*, *65*(1), 21–35. https://doi.org/10.1111/ajps.12521

Gonzales, A. (2018). Do more with Twitter data. https://twitterdev.github.io/do_more_with_twitter_data/timeseries.html

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, *21*(3), 267–297. https://www.jstor.org/stable/24572662

Liu, X., Lindquist, E., & Vedlitz, A. (2011). Explaining media and Congressional attention to global climate change, 1969-2005: An empirical test of agenda-setting theory. *Political Research Quarterly*, *64*(2), 405–419. https://doi.org/10.1177/1065912909346744

McKenzie, E., & Gardner, E. S. (2010). Damped trend exponential smoothing: A modelling viewpoint. *International Journal of Forecasting*, *26*(4), 661–665. https://doi.org/10.1016/j.ijforecast.2009.07.001

Minkkinen, M., Zimmer, M. P., & Mäntymäki, M. (2022). Co-shaping an ecosystem for responsible AI: Five types of expectation work in response to a technological frame. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-022-10269-2

Rein, M., & Schön, D. (1996). Frame-critical policy analysis and frame-reflective policy practice. *Knowledge and Policy*, *9*(1), 85–104. https://doi.org/10.1007/BF02832235

Ulnicane, I., Knight, W., Leach, T., Stahl, B. C., & Wanjiku, W.-G. (2020). Framing governance for a contested emerging technology: Insights from AI policy. *Policy and Society*, *0*(0), 1–20. https://doi.org/10.1080/14494035.2020.1855800

Vaida, F., & Blanchard, S. (2005). Conditional Akaike Information for mixed-effects models. *Biometrika*, *92*(2), 351–370. https://www.jstor.org/stable/20441193

White House. (2020). Executive Order on Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government. https://www.govinfo.gov/content/pkg/DCPD-202000870/pdf/DCPD-202000870.pdf

Wood, B. D., & Peake, J. S. (1998). The dynamics of foreign policy agenda setting. *American Political Science Review*, *92*(1), 173–184. https://doi.org/10.2307/2585936

APPENDIX C

SUPPORTING INFORMATION FOR CHAPTER 4

C.1 Additional Information about Legislator Sample

Table C.1 presents key descriptive statistics about the full sample, including all legislators contacted through the experiment.

Table C.1: Descriptive statistics for state legislator sample

Gender	
Male	0.69 (0.46)
Female	0.31 (0.46)
Party	
Democrat	0.45 (0.50)
Independent	0.01 (0.09)
Republican	0.54 (0.50)
Tenure	
0-1 year	0.21 (0.41)
2-5 years	0.39 (0.49)
6-10 years	0.25 (0.43)
11-20 years	0.10 (0.30)
20+ years	0.05 (0.23)
Chamber	
Lower Chamber	0.73 (0.44)
Upper Chamber	0.27 (0.44)

Note: $n = 7,356$. Proportions with standard deviations in parentheses.

C.2 Covariates and Balance

I make use of the following demographic covariates as part of the covariate-adjusted regression strategy: legislator gender, party, tenure, and chamber, as well as each state's prior experience with AI policy and legislature professionalism.[1] The covariates are coded as

[1] In a deviation from the pre-analysis plan, I do not include state as a covariate, as it is highly colinear with other state-level variables and also reduces the degrees of freedom considerably.

follows:

- Legislator party is coded as a factor variable with three levels: Democrat, Independent, and Republican.
- Legislator chamber is coded as a binary variable with two levels: upper and lower (e.g., Senate and House).
- Legislator gender is coded as a binary variable with male and female as the two levels.
- Legislator tenure is coded as a count variable based on the number of years served consecutively in the current role.
- Prior experience of legislators with AI policy is coded as a state-level count variable. This variable is constructed by counting the number of proposed and/or passed pieces of legislation in the state since 2019 that address AI according to the NCSL.[2]
- State legislator professionalism is coded as a numeric variable based on the adjusted Squire Index (Squire, 1992).

To assess covariate balance, I perform F-tests of global significance by regressing binary treatment group indicator variables for the components of the factorial design on the specified covariates and extract the F-statistic and associated p-value from each regression. With p-values of 0.18 (control vs. not), 0.91 (narrative vs. not), 0.88 (expertise vs. not), and 0.97 (ethics vs. competition), I find that none of the covariates predict treatment. Results in Table C.2 indicate that there is very strong covariate balance across all six treatment groups.

Table C.2: Covariate balance table

Variable	**Control Comp.**	**Control Ethics**	**Exp. Comp.**	**Exp. Ethics**	**Narr. Comp.**	**Narr. Ethics**
Democrat	0.46	0.45	0.45	0.46	0.45	0.45
Independent	0.01	0.01	0.01	0.01	0.01	0.01
Republican	0.53	0.54	0.54	0.53	0.54	0.54
Lower	0.73	0.74	0.74	0.74	0.74	0.73
Upper	0.27	0.26	0.26	0.26	0.26	0.27
Female	0.33	0.33	0.32	0.28	0.29	0.30
Male	0.67	0.67	0.68	0.72	0.71	0.70
Tenure	6.26	6.05	6.06	5.94	6.06	6.05
Prior Legislation	2.36	2.35	2.35	2.36	2.30	2.31
Squire	0.23	0.23	0.23	0.23	0.22	0.23

[2]For the analysis related to prior AI policy experience, I binarize this variable so that the top 50% of states by this count are identified as having high prior experience and the rest as low experience. This roughly divides states into those who have considered at least one piece of AI legislation and those that have not.

C.3 ITT Results for Policy Entrepreneur Effectiveness Hypothesis

Table C.3: ITT results: Impact of policy entrepreneur strategies on legislator engagement

	Legislator Engagement (Binary Measure)	
	(1)	(2)
Either Strategy	0.047***	
	(0.008)	
Expertise		0.045***
		(0.010)
Narrative		0.048***
		(0.010)
N	7,206	7,206
R^2	0.036	0.036
F Statistic	33.573*** (df = 8; 7,197)	29.849*** (df = 9; 7,196)

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. With robust SEs, including covariates.

Table C.3 reproduces main Table 4.2 but using ITTs instead of CACEs. Table C.4 presents ITT results for additional outcomes evaluated in the study. The outcomes include: clicked on resource, clicked on webinar link, replied to email, attended webinar, binary engagement measure, and count engagement measure.

Table C.4: ITT results: Alternative outcome measures

	Resource	Webinar	Replied	Attended	Binary	Count
	(1)	(2)	(3)	(4)	(5)	(6)
Expertise	0.046***	0.007	0.001	0.001	0.045***	0.055***
	(0.009)	(0.009)	(0.001)	(0.001)	(0.010)	(0.017)
Narrative	0.049***	0.005	0.00004	0.001	0.048***	0.056***
	(0.009)	(0.009)	(0.001)	(0.001)	(0.010)	(0.017)
N	7,206	7,206	7,206	7,206	7,206	7,206
R^2	0.037	0.019	0.001	0.002	0.036	0.029
F Statistic (df = 9; 7,196)	30.907***	15.277***	0.757	1.806*	29.849***	24.003***

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. With robust SEs, including covariates.

C.4 Power Analysis

To assess whether the study is sufficiently powered to detect significant effects for the Policy Entrepreneur Effectiveness Hypothesis, for which there are three treatment/control groups, I performed simulations of the 2SLS IV analysis. To do this, I used hypothesized treatment effects along a range of possible compliance rates, given a fixed original sample size of 7,356. The outcome of interest was any of the binary measures, such as clicking on a link. Significance and power were determined at the standard 5% and 80% levels, using two-sided p-values to be conservative. Compliance rates ranging from 5% to 75% were assessed. I assumed that the proportion of individuals who clicked on a link (or registered for the webinar and so on) would be 10% in the control group, and performed 250 simulations for four scenarios each, assuming that the expertise or narrative treatments would lead to increased click rates of 2.5, 5.0, 7.5, and 10 percentage points, respectively. (Note that these correspond to standardized effect sizes of 0.08, 0.17, 0.25, and 0.33 standard deviations given the standard deviation of the control group.)

According to industry-level metrics provided by Mailchimp, click-to-open rates (i.e., clicking on a link conditional on opening an email) range from between 8.5% for retail to 22.4% for government and politics, while email open rates range from around 12.6% for retail to 26.7% for government and politics messages. Further, given that the population is state legislators rather than the general public, compliance rates from prior studies are also instructive. Fisher and Herrick (2013) found that 11.5% of state legislators responded to a survey, Butler et al. (2012) found that 19% of state legislators responded to a letter asking about policy positions, and Anderson et al. (2020) found that 6% of state legislators started and 4% completed a survey. Given that the tasks involved in this study include both low-effort tasks (clicking on a link) and high-effort tasks (attending a webinar), click rates in the range of 5% to 20% were deemed reasonable. Thus, baseline click rates of around 10-20%, and increases resulting from treatments on the order of a few percentage points to

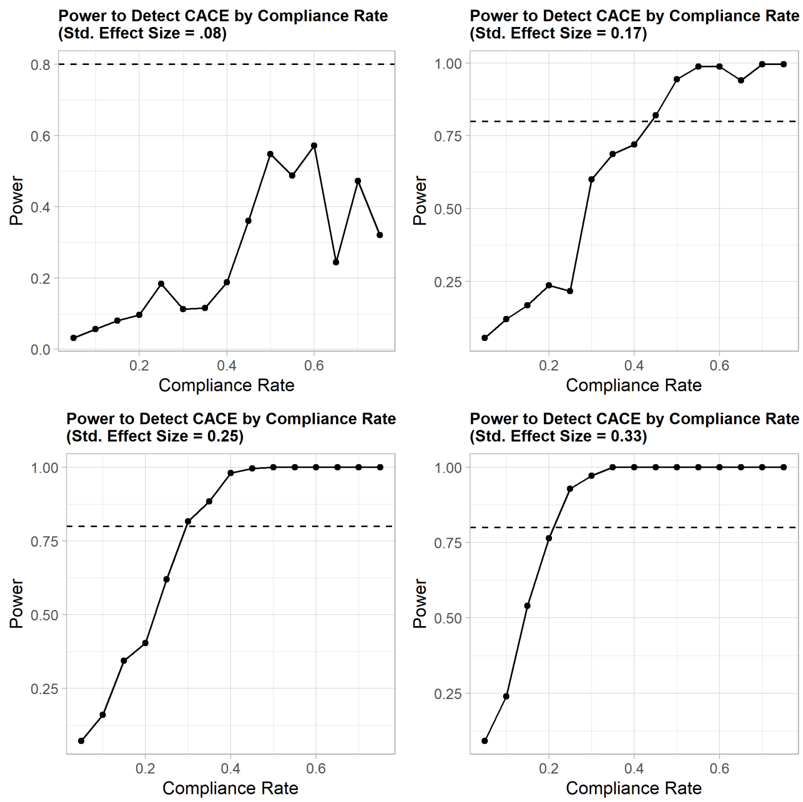

Figure C.1: Power to detect treatment effects for main hypothesis by compliance rate

perhaps 10 percentage points, were deemed reasonable estimates.

Figure C.1 indicates that the study is sufficiently powered if either the compliance rate ($\sim$ 35% email open rate or more) or effect size ($\sim$ 5 percentage point increase or more) is relatively high, but not if both are low. Further, the research design has more power for estimands that involve only two treatment groups, such as the Issue Framing Hypothesis, but likely lacks sufficient power for exploratory estimands that require additional subsetting, such as the Strategies by Issue Framing Hypothesis.

To improve power through the research design, I designed the email messages to strengthen treatment effects, e.g., using professional aesthetics and the partner organization logo to increase credibility. I also strategically prepared and scheduled emails through the email management platform to minimize the number of messages that go to spam folders (for example, by "warming up" the account, avoiding spam buzzwords, and associating the organization domain name with the emails).

C.5 Treatment Wordings and Email Template

Email Sent to State Legislators:

From: daniel.schiff@thefuturesociety.org

To: [Legislator's E-mail Address]

Subject: Artificial Intelligence - What to Know as a Legislator in [State]

Dear [Representative/Senator] [Last Name],

We at The Future Society are reaching out to you today [about / to share a compelling story about / to provide a useful fact sheet about] the **important [social and ethical / economic and technological leadership] implications of artificial intelligence (AI)** and to **invite you to a webinar** on what you need to know about AI as a state legislator. We believe that state legislators such as yourself have an important role to play in shaping [state's] response to these critical [social and ethical / economic and technological leadership] issues.

[Control or Treatment paragraphs: See Table C.5]

How can The Future Society support you?

- **RSVP for our webinar** on Monday, December 13 at 2PM ET / 11AM PT on what you need to know about AI as a state legislator, with speakers from The Future Society, Georgia Tech, and the National Conference of State Legislatures.
- **Share your thoughts**: What is the top concern or hope you have about AI policy? Let us know in a reply to this email and we'll try to address your comments in our webinar!

Thank you for your time and consideration of these important issues,

The Future Society

[Logo]

Table C.5: Treatment wordings in second paragraph of email experiment

	Ethics	Competition
Control	The Future Society is an independent 501(c)(3) nonprofit think-and-do tank. We've engaged over 6,000 senior decision-makers in over 100 countries. We build understanding of AI and its impact, we build bridges between relevant constituents, and we build innovative solutions to help communities and people all over the world enjoy the benefits of AI and avoid its risks. Incorporated since 2016 and funded through a diverse community of donors, we provide an independent and nuanced perspective on the governance of AI. For more about our work, please visit our website.	The Future Society is an independent 501(c)(3) nonprofit think-and-do tank. We've engaged over 6,000 senior decision-makers in over 100 countries. We build understanding of AI and its impact, we build bridges between relevant constituents, and we build innovative solutions to help communities and people all over the world enjoy the benefits of AI and avoid its risks. Incorporated since 2016 and funded through a diverse community of donors, we provide an independent and nuanced perspective on the governance of AI. For more about our work, please visit our website.
Expertise	As one of the most prominent applications of AI, facial recognition systems depend on machine learning (ML) models that are often opaque "black boxes" and may be trained on real-world data that can reflect discriminatory biases. It is critical to understand how different approaches to bias and fairness in AI systems such as facial recognition technology can inform law and regulation surrounding AI. To learn more about current research and what policymakers can do about the social and ethical implications of AI, please see our fact sheet.	As one of the most powerful general purpose technologies of the 21st century, AI is a key strategic factor in economic growth and innovation policy. Support for public and private sector R&D, STEM education, and SMEs all contribute to AI competitiveness. The United States currently leads in private sector AI R&D, venture capital, and talent, while China has more AI research publications and supercomputers and generates more data overall. To learn more about current research and what policymakers can do about the economic and technological leadership implications of AI, please see our fact sheet.
Narrative	When Robert Julian-Borchak Williams went to work in his office at an automotive supply company in Detroit, he had no idea he would be handcuffed and arrested later that day in front of his wife and two young daughters. That day, Robert became one of the first Americans wrongfully arrested because of a false match of a facial recognition algorithm, an example of how faulty or misused AI algorithms can go awry. To hear more about how AI went wrong in this case and what policymakers can do about the social and ethical implications of AI, please read more about Robert's story.	Few could have predicted that when Google's AI program AlphaGo defeated Chinese Go champion Ke Jie in 2017, it would spur a Sputnik-like technological competition between the United States and China. Soon after, Chinese President Xi Jinping stated that China must "ensure that critical and core AI technologies are firmly grasped in our own hands." In response, U.S. President Joe Biden has said that "we're in competition with China and other countries to win the 21st Century." To hear more about the AI race and what policymakers can do about the economic and technological leadership implications of AI, please read more about the story behind the AI race.

Fact sheets and stories used in treatments are available at the links below or by request:

- **Expertise + Ethics**: https://thefuturesociety.org/wp-content/uploads/2021/10/legislator-fact-sheet-about-ai-ethics.pdf
- **Narrative + Ethics**: https://thefuturesociety.org/wp-content/uploads/2021/10/facial-recognition-gone-awry.pdf
- **Expertise + Competition**: https://thefuturesociety.org/wp-content/uploads/2021/10/legislator-fact-sheet-about-ai-innovation.pdf
- **Narrative + Competition**: https://thefuturesociety.org/wp-content/uploads/2021/10/race-for-ai-leadership.pdf

C.6 Additional References for Appendix C

Anderson, S. E., DeLeo, R. A., & Taylor, K. (2020). Policy entrepreneurs, legislators, and agenda setting: Information and influence. *Policy Studies Journal*, *48*(3), 587–611. https://doi.org/10.1111/psj.12331

Butler, D. M., Karpowitz, C. F., & Pope, J. C. (2012). A field experiment on legislators' home styles: Service versus policy. *The Journal of Politics*, *74*(2), 474–486. https://doi.org/10.1017/s0022381611001708

Fisher, S. H., & Herrick, R. (2013). Old versus new: The comparative efficiency of mail and internet surveys of state legislators. *State Politics & Policy Quarterly*, *13*(2), 147–163. https://doi.org/10.1177/1532440012456540

Squire, P. (1992). Legislative professionalization and membership diversity in state legislatures. *Legislative Studies Quarterly*, *17*(1), 69–79. https://doi.org/10.2307/440081