Governing Artificial Intelligence: An Analogical Approach to Cultivate Shared Knowledge and Generate Newfound Insight

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

William Greig

Fall 2023

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Kathryn A. Neeley, Associate Professor of STS, Department of Engineering and Society

Introduction

Artificial intelligence is a promising, powerful technological capability that has shown tremendous potential in various industries such as commerce, transportation, and education. However, as with any emerging technology, there is a need for governance. Supportive policies can be instrumental in facilitating technological development while detrimental regulations can halt technical progress altogether. The motivations for governing emerging technologies vary but typically address and weigh "private sectors interests, such as business disruptiveness and profits" as well as "public interests such as human dignity and identity" (Digital Regulation Platform, 2023, para. 4). As it stands, the United States has not passed many federal regulations about artificial intelligence; the two enacted federal bills are not strict regulations but rather outline how to regulate artificial intelligence in the future. Problematically, this lack of governance might yield costly results. With the global AI market expected to exceed $1.81 trillion in 2023, countries that readily adopt AI "could capture an additional 20 to 25 percent in economic benefits" compared to countries that do not choose to embrace AI (McKinsey Global Institute, 2018, pg. 3). Additionally, with the increasing utilization of artificial intelligence, public concerns regarding data privacy, algorithmic bias, and transparency are similarly increasing. Despite the importance and criticality of governing artificial intelligence, it is not a straightforward task. The main difficulty is a lack of shared understanding regarding artificial intelligence. Stakeholders–such as policymakers, researchers, and citizens–have varying levels of knowledge and unique perspectives surrounding artificial intelligence. Without a common ground of understanding, it is impossible to have an efficient and effective policymaking process.

In this paper, I argue that using analogies to understand artificial intelligence is both beneficial and imperative in cultivating a shared understanding of the nuances of the technology.

As an overview, this paper provides a general background of artificial intelligence regulation, introduces a fundamentally overlooked problem in regulating technologies, describes a flexible analogical framework, and uses this framework to explore technological analogies about social concerns surrounding artificial intelligence.

Artificial Intelligence Definition and Prevalence

Before delving into artificial intelligence research, it's important to understand exactly what artificial intelligence is and how it might be defined in various contexts. Formally, artificial intelligence (AI) is an overarching term used to describe a vast array of technologies and disciplines concerned with a machine's ability to emulate human problem-solving and decision-making abilities (Müller, 2020). The general term "AI" can encompass both fields of study–such as machine learning or natural language processing–and specific implementations such as ChatGPT or robotic assembly machines. To refine and clarify this colloquial definition of AI for the purposes of this paper, AI is better understood as a *technological capability* rather than a specific implementation. AI is not necessarily one device but a dynamic technical power enabling the development of new systems. This distinction is important as it highlights the breadth of AI's potential application as well as its transformative impact across various industries.

AI has already seen massive utilization in numerous sectors including autonomous vehicles for transportation, generative AI in commerce, and learning tools for education. Many companies and fields that utilize AI see massive improvements in effectiveness, efficiency, and performance. A statistical report by Gitnux indicated that autonomous vehicles are expected to save $234 billion in the US by 2030 due to reduced accidents. These self-driving cars are projected to prevent roughly 8.5 million car accidents annually (Castillo, 2023, para. 15).

According to Forbes, in industry, over 97% of surveyed commerce companies believe that ChatGPT will benefit their business (Haan, 2023, para. 18). Additionally, a Quizlet survey found that 73% of students who use AI technologies, such as ChatGPT, agree that they study faster and more efficiently and 42% of students reported that AI created a more equitable learning environment (Quizlet, 2023, paras. 2-3). Although these statistics are only an early representation, they seem promising for the potential benefits that AI could bring.

The Need for Governance

As with any emerging technology, the future of AI depends on how society–from citizens to organizations–and the government choose to control and utilize it. As environmental policy researchers Almgren and Skobelev (2020) explain: "the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development is essential to achieve longer-term objectives in [...] technological transformation" (pg. 10). Regulations can be imposed on technologies that inhibit development due to compliance costs, slow approval processes, and barriers to entry. The FDA is a great example of this whereby new pharmaceuticals require extensive testing and verification in order to be sold to the public; this process is incredibly expensive and can take years to finish (Digital Regulation Platform, 2023). Alternatively, regulations can be supportive of the development of technology if implemented correctly. Financial investments and educational programs help emerging technologies gain traction and attention which is necessary for any technology to grow and develop quickly. Standardization–which enforces a standard across the industry–can be effective in establishing sustainable development (Grant, 2021). Although these policies are promising and appealing on paper, in reality, it is much harder to develop and pass regulations without ample support.

Within the United States, little progress has been made–especially compared to China and Europe–in governing AI. To date, there have only been two major federal policies released by the US surrounding AI: *Blueprint for an AI Bill of Rights* and *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* (The White House, 2023). Both documents discuss and address certain social concerns such as data privacy, transparency, and algorithmic biases. Importantly, however, both documents simply outline future regulations–essentially describing concerns to be considered while designing and deciding future governing policies; the current attitude towards federal AI regulation in the US is self-governance, whereby companies are expected to uphold certain conditions of the *AI Bill of Rights* without strict enforcement (Kang, 2023). Notably, this lack of federal regulation is in part due to the United States' difficult legislative process. A study conducted by Stanford Human-Centred Artificial Intelligence (HAI) Institute found that among 130 AI-related bills proposed to Congress in 2021, only 3 were actually passed (Lynch, 2022, para. 10). Figure 1 demonstrates this discrepancy very clearly, illustrating the wide gap between proposed and passed bills between 2015 and 2021. According to an AI policy tracker for 2023, over 53 actionable AI legislations have been proposed but none have passed thus far (Artificial Intelligence Legislation Tracker, 2023).

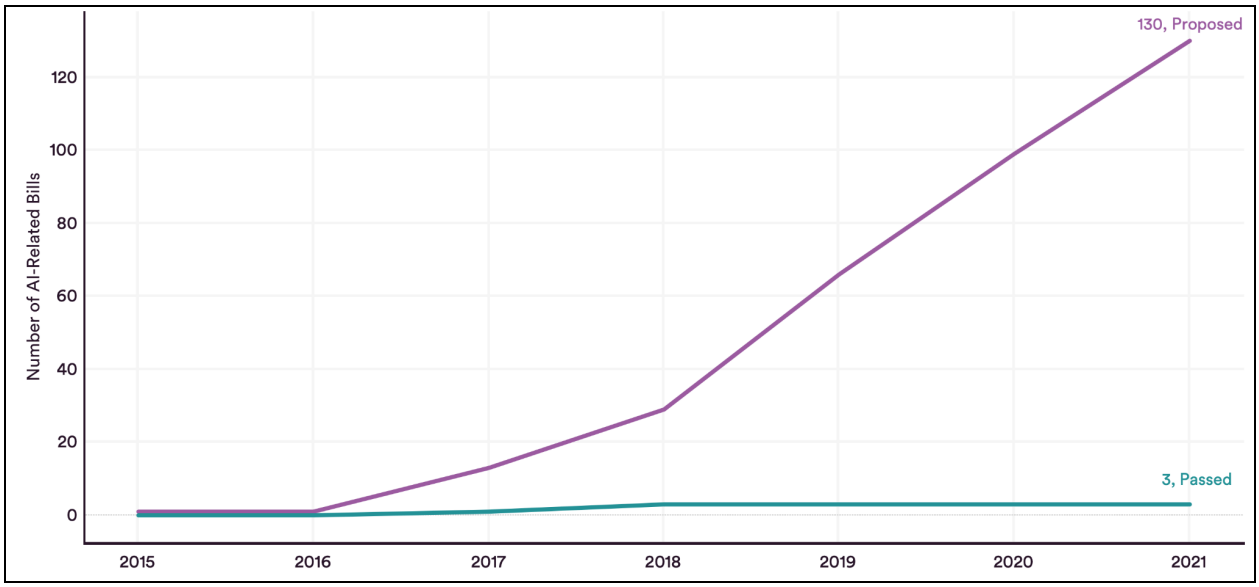

Figure 1 – Number of AI-Related Bills in the United States, 2015–21 Proposed vs. Passed (Lynch, 2022). Comparatively low number of passed bills contrasted by the growing total number of proposed bills.

Governing AI is important for three primary reasons: economic opportunities, technological progress, and societal concerns. A report from Grand View Research (2023) estimated that the global AI market will reach $1.81 trillion in 2030, up from $136 billion in 2022. Countries with higher adoption rates and supportive AI regulations will contribute more to the global AI market and is imperative for countries to take regulatory action quickly to remain competitive. As explained in a report by McKinsey Global Institute (2018), "the impact of AI depends on the level of its adoption by corporations and government entities" (pg. 9). According to their financial models, early AI adopters and supporters "could capture an additional 20 to 25 percent in economic benefits" compared to countries that do not choose to embrace AI (McKinsey Global Institute, 2018, pg. 3). Beyond economics, AI facilitates the development of revolutionary technologies that can directly address previously unsolvable problems. As mentioned, self-driving cars have been shown to be statistically safer than human-operated

vehicles and could possibly save millions of lives in practice (Castillo, 2023). Governance towards these advantageous technologies is beneficial to both the future of AI and the sectors that adopt it. Additionally, regulations can discretely address the societal concerns and rights of citizens whom an emerging technology might affect. For AI, issues surrounding data privacy, algorithmic bias, and explainability are at the forefront of concern–highlighted by both *Blueprint for an AI Bill of Rights* and *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* (The White House, 2023).

The Problem in Governance

Action–or lack thereof–is not the only limiting factor with regard to governing AI. The main difficulty in governing AI is rather straightforward but often overlooked: shared understanding. Theresa Kushner, an AI and data consultant at NTT Data, explains that "understanding [AI] is as important in today's world as understanding how to read" (Brunelle, 2022, para. 29). She argues that AI literacy–the ability to reason "what an AI project can do, how it is managed, how the data is selected, how the output is reviewed and monitored"–is crucial for "moving to the next level of [AI] maturity" in corporations (Brunelle, 2022, para. 30). Theresa continues, indicating that "few people in organizations actually understand what AI is and is not" (Prokopeak, 2022, para. 5). However, this idea is not unique to organizations; it's extendable to the general population. Pew Research reported that 80% of individuals surveyed about AI and data privacy said they are concerned about how much personal data is collected by companies and the government. At the same time, 59% of respondents indicated that they are not sure how their data is actually being used by companies (Atske, 2020, para. 8). This report demonstrates that despite people's concern regarding personal data collection, a vast majority of individuals were uncertain as to how their data was being used and operated on.

This paper aims to address this discrepancy by introducing a robust, flexible framework for cultivating a shared understanding of emerging technologies and offering a concrete analysis of certain AI issues and capabilities using analogies.

The Analogical Framework

Schwarz-Plaschg addresses one understated obstacle in governing emerging technologies: shared understanding. There is an innate disconnect between emerging technology and its stakeholders. Different stakeholders–such as scientists, policymakers, or citizens–have various levels of understanding of technological capabilities and limitations. Achieving common ground is of utmost importance to ensure each stakeholder can effectively address their concerns related to an emerging technology. Schwarz-Plaschg understands this disconnect but argues that regardless of a stakeholders' understanding, "they are very much analogical animals in the sense that they rely on analogies when thinking and debating about emerging technological developments" (pg. 2). This is an interesting observation and sets the basis of her work, arguing that analogies can be used to not only establish common ground but also expand our understanding of the technology. Schwarz-Plaschg also discusses how analogies can be used to explore future implications of technology as well as the overlooked rhetorical power analogies hold in debate and policymaking decisions. Ultimately, Schwarz-Plaschg provides a convincing argument for the utilization of analogies in policymaking discussions for emerging technologies due to its ability to cultivate shared understanding, expand current knowledge, and explore future possibilities and limitations through imaginative analogies.

As mentioned, Schwarz-Plaschg provides a flexible framework for understanding emerging technologies through imaginative analogies. These analogies function as comparisons to previous or alternative technologies to provide context and a foundation for argumentation.

Naturally, these analogies help draw parallels to the technical capabilities as well as possibly social or ethical concerns between the comparative technologies. If constructed well, Schwarz-Plaschg argues that analogies can "[work] out new or less obvious similarities [that can] draw attention to areas and issues that have been hitherto overlooked", a characteristic that is highly valued for a multistakeholder process such as policymaking (pg. 6). Beyond shared understanding, the broader utilization of analogies is for exploring future implications. Comparing past technologies and their outcomes can certainly inform, although not outright predict, the outcomes of similar technology today.

One powerful aspect of imaginative analogies is that differences between technologies are seen as positive. As stated by Schwarz-Plaschg, inconsistencies in technologies are treated as "an expression of the complexity of the issue and its different dimensions. Building multiple—also contradicting—analogies then does not signify inability but imaginative ability." This is a profound statement as it suggests that technological differences are to be highlighted and scrutinized carefully. It also reinforces the fact that this framework is incredibly flexible; a piece of technology and its issues can be addressed from numerous perspectives and backgrounds. As an example, concerns surrounding AI job displacement can be compared to both industrialization caused by the steam engine and the degradation of writing industries from the creation of the printing press. Both a serviceable analogies and can offer different perspectives to explain AI's potential impact on job displacement.

Research Methodology

To understand this research methodology in this paper, refer to Figure 2 below. This diagram starts by identifying various issues or technical capabilities pertaining to artificial intelligence. For each point, find a source, line of discussion, or argument that compares AI–or

any component of it–to another technology. Finally, assess the validity and benefit of the comparison by highlighting any newfound knowledge or insight gained by the analogy.

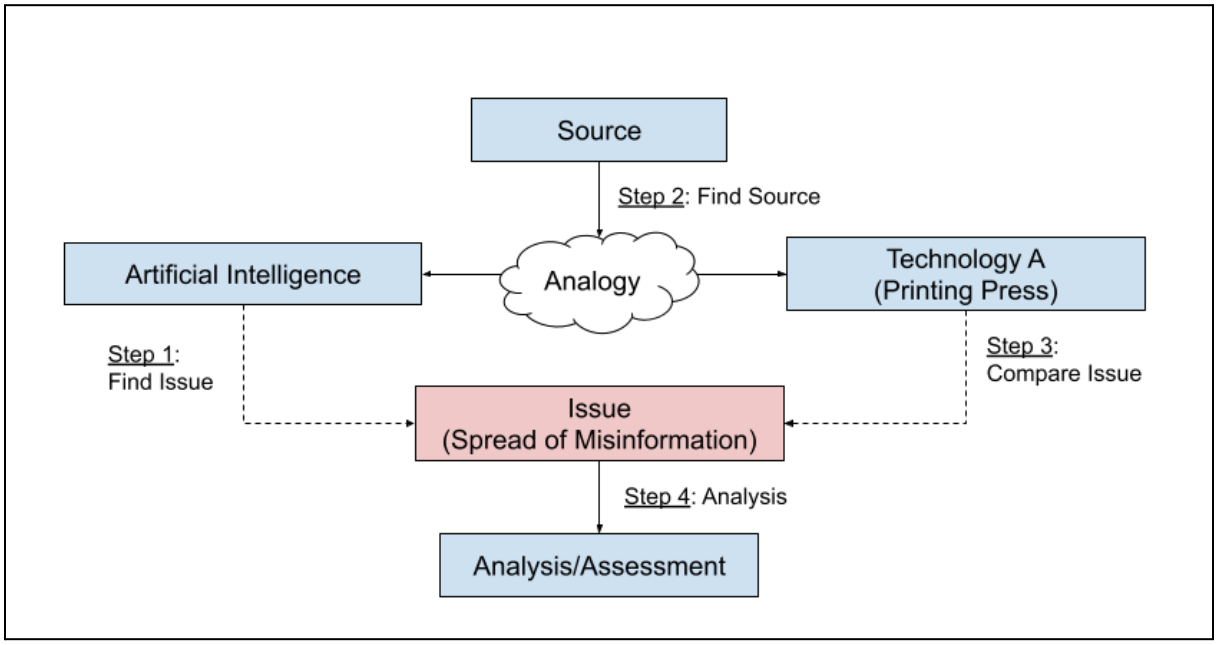

Figure 2 – Example Analogy using Analytical Framework (Created by Author). The methodology applied to Artificial Intelligence and the Printing Press to highlight similar issues regarding the spread of misinformation.

Initial Artificial Intelligence Analogies

Following the research methodology described above, there were numerous common analogies–such as the steam engine or CCTV–as well as uncommon analogies–like genetic engineering and space exploration–made to AI. Each comparison served to Figure 3, created by the author, enumerates a select few analogies found while researching discussions regarding various social issues of AI. The left column is the analogy (technology) with which AI is being compared, the middle column is the issue that is being addressed, and the line of argumentation is in the right column.

Figure 3 – Analogies of AI in Public Discourse (Created by Author). Table organizing various issues surrounding AI and analogies made with other technologies in public discourse.

Analogy With	Issue	Line of discussion and argumentation
CCTV	Privacy Concerns	Comparison of AI privacy concerns to surveillance systems which capture data without explicit permission
Steam Engine	Job Displacement	Influence of AI in manufacturing and transportation will reshape the industry
Human Bias	Algorithmic Bias	A common comparison to understand why algorithms perpetuate bias
Printing Press	Spread of Misinformation	Comparison to the printing press whereby technology could be used to spread propaganda and misinformation quickly
Genetic Engineering	Self-Regulation	Comparison of self-regulated restrictions and principles for both gene editing and AI
Nuclear Power	Responsibility	Comparison to nuclear power in generating and following national guidelines of operation
Space Exploration	International Cooperation	Arguing AI development at a similar stage and the importance of protecting the world
3D Printing	Hype and Anticipation	The hype of 3D printing is similar to AI but has yet to transform the manufacturing industry as expected

From this visualization alone, it's clear how analogies can clearly be useful in establishing shared understanding as, even without prior knowledge of AI, we can generally understand how AI relates to known technologies such as CCTV or 3D printing. However, the real power of analogical frameworks is generated via the careful analytical observations of each analogy and any resulting newfound insights.

Analogous Flexibility

Demonstrating the flexibility of the analogical framework, Yuying Chen-Wynn (2023), who serves as the Head of AI at PEAK6, makes three distinct comparisons to AI: the ballpoint pen, the printing press, and the internet. Each of these analogies lends itself to understanding a particular capability and issue surrounding AI. Focusing on the printing press and AI, the similarity is rather nuanced. As Chen-Wynn argues, the printing press "expanded the access to information and ideas" which is analogous to AI digitalization whereby technologies such as optical character recognition have helped convert physical data into digital spaces (para. 8). However, the main takeaway was not in the positive similarities between the printing press and AI but rather the consequences of the printing press: "[the printing press] caused the decline of scribes, illuminators and copyists [...] it was essentially the first automation technology that replaced manual labor." (para. 12). This idea–job displacement–was not readily apparent when first making the analogy between AI and the printing press. More often than not, the more common and straightforward comparison is that of the steam engine in the Industrial Revolution. Regardless, with the advent of AI, the possibility of losing manual labor jobs due to automation is rather likely. This comparison reinforces a terrifying yet entirely real concept of entire industries vanishing due to emerging technologies (Chen-Wynn, 2023).

In a separate article written by Jonny Thomson (2023), demonstrating the flexibility of the analogical framework, the same comparison–AI and the printing press–was made but focused on another aspect of their comparison: misinformation. Thomson indicates that "during the 1500s, people worried about the easy dissemination of false information by mass-produced pamphlets" (para. 10). Today, people are similarly concerned with ChatGPT and its spread of misinformation through "plausible sounding but incorrect responses." (para. 10). This focus is

interesting as it highlights an important phenomenon: even the most advanced technologies have the same underlying problems that historical technologies have. Here, it's evident that the technology itself is not the issue; the printing press alone will not spread information and, as mentioned before, AI systems will only perpetuate *human* bias. The insight isn't readily apparent but it is certainly a valuable understanding: communicative technology is a mechanism to spread false information but is not the root cause of misinformation (Thomson, 2023).

Ethical Analogies

A report published by Google entitled *Perspectives on Issues in AI Governance* introduced certain ethical dimensions of AI and compared these issues to other transformative technologies. The first comparison was to genetic engineering particularly concerning self-regulation. In the 1970s, genetic researchers reached a "voluntary agreement [...] to impose self-regulatory restrictions" (pg. 5). In 2017, AI researchers were "inspired" and led to the self-adoption of the newfound Asilomar AI principles–a collection of core values to guide safe, secure, and fair development AI. These principles enforce the idea that self-governance is at an individual level and suggest developers of AI technologies have a responsibility to adhere to these guidelines despite a lack of enforced regulation. This idea of responsibility was also reflected in a comparison of AI to nuclear technology: "Nuclear reactions can also be used to produce highly destructive weapons. While new nations continue to develop nuclear technology, national guidelines, and international non-proliferation agreements have proven a strong framework for setting and maintaining expectations of responsible behavior" (pg. 5). In essence, this comparison serves two main ideas: developers of AI must ensure its power and potential aren't misused for malintent and also to utilize national frameworks as a guideline for responsibility.

The last comparison made in this report was AI and space exploration. This comparison was rather abstract, arguing that AI is at a similar developmental stage as "space exploration was in 1985 when the UN formed its committee for the peaceful exploration of space". The report argues further that the treaty was "instrumental in providing the [...] principles to underpin national guidelines and legislations [...] in developing their own space programs" (pg. 5). In essence, this analogy extends that AI research is in a catalytic state whereby huge advancements in the field could be made with proper national guidelines. The analogy emphasizes the potential impact of creating regulatory guidelines for AI that could ultimately promote rapid technological development. (Perspectives on Issues in AI Governance, n.d).

Non-Technical Analogy

The most common comparison to explain algorithmic bias is a direct comparison to human discrimination. From a basic perspective, algorithmic bias occurs due to an unrepresentative dataset used to train and develop AI models. Pethig and Kroenung, in a research paper concerning gender biases in humans and algorithms entitled "Biased Humans, (Un)Biased Algorithms?", explain this idea clearly: "many algorithms learn automatically from historical data, [...] introducing additional sources of bias incorporated in the underlying training data" (Pethig and Kroenung, 2023). Importantly, the AI systems themselves are not the issue–humans are. "Ethical Implications and Accountability of Algorithms", by Kirsten Martin, expands on this idea, indicating that "sentencing algorithms in sentencing illustrate a similar problem with unjust biases perpetuating human discrimination. Similarly, an algorithm for university admittance could be as discriminatory by design or the algorithm could be trained on data with historical biases." (Martin, 2023) This analogy is rather insightful and important for an understanding of intelligent systems. The main takeaway is that a discriminatory facial recognition system is not

faulty nor is it correct. AI is simultaneously an extension of technological ability and a reflection of humans' biases and discriminatory patterns. From a compliance perspective, this insight is important in assigning responsibility and liability of AI system developers as their implicit biases–unintentional or not–are culpable for biased software.

Newfound Insights

For each analyzed analogy, some amount of newfound information or understanding about AI was discovered. Summarized in Figure 4, this table shows the specific knowledge or observation that was generated from analyzing the analogy and teasing out the similarities or differences in technologies. In retrospect, these insights or knowledge seem abundantly obvious; however, without the analysis technique of using analogies to compare ethical concerns and societal impacts between similar technologies, these insights would not have been garnered easily.

Figure 4 – Insight Gained from Analyzed Analogies (Created by Author). Table organizing various issues surrounding AI and analogies made with other technologies in public discourse.

Analogy With	Issue	Insight or Knowledge Attained
Human Bias	Algorithmic Bias	Algorithmic bias is a reflection of human discrimination and bias
Printing Press	Job Displacement	AI could possibly destroy entire industries or workforce populations
Printing Press	Spread of Misinformation	AI is a mechanism for spreading misinformation, not the cause
Genetic Engineering	Self-Regulation	Prior policies can be examined for self-regulation guidance
Nuclear Power	Responsibility	Ensure AI isn't misused for malicious intention
Space Exploration	International Cooperation	AI research is in catalytic state, on the frontier of profound progress

Conclusion

This paper explored a practical implementation and analysis of an analogical framework for artificial intelligence as a technical capability. The application of this methodology can be further explored in more analogies and more detail for a better, more well-rounded understanding of artificial intelligence. The results generated from this paper are significant in two aspects; one, these insights and observations can facilitate policymaking by ensuring stakeholders have a foundational level of understanding of the emerging technology. This can reduce stakeholders overstating or underemphasizing the potential capabilities, limitations, and problems of an emerging technology. Additionally, the results demonstrate the potential rhetorical power that analogical frameworks hold in policymaking spheres by enabling unique perspectives to arise from simple analogies. These analogically generated perspectives add robustness and completeness to policy creation by highlighting uncommon or overlooked dimensions. Nonetheless, there are still limitations to this framework and methodology. As defined earlier, artificial intelligence is a dynamic technical capability that is more complex and nuanced than a simple analogy can capture. Although analogies can capture particular aspects and issues concerning artificial intelligence, the nuances and unique differences of artificial intelligence, or any other emerging technology, might not be able to be captured neatly. As such, this observation necessitates that these governing strategies are revisited and revised according to how artificial intelligence–as a technical capability–has progressed and changed from its initial drafting.

References

Almgren, R., & Skobelev, D. (2020). *Evolution of technology and technology governance*. Journal of Open Innovation: Technology, Market, and Complexity, 6(2), 22.

Atske, S. (2020, August 17). *Americans and privacy: concerned, confused and feeling lack of control over their personal information*. Pew Research Center: Internet, Science & Tech. [company website]

Brennan Center for Justice. (2023, August 7). *Artificial Intelligence Legislation Tracker*. [company website]

Brunelle, P. (2022, April 16). *Data-Driven Work Cultures: NTT DATA's Theresa Kushner On How To Effectively Leverage Data To Take Your Company To The Next Level*. Medium.

Castillo, L. (2023, October 31). *Must-Know Self-Driving Cars Statistics [Current Data]*. Gitnux. [company website]

Chen-Wynn, Y. (2023, March 15). *Is AI the ballpoint pen, the printing press or the internet?* LinkedIn. [web article]

Digital Regulation Platform. (2023, August 9). *Transformative technologies (AI) challenges and principles of regulation*. World Bank.

Eggers, W., Kishnani, P., & Turley, M. (2019, June 18). *The future of regulation.* Deloitte Insights. [company report]

Grant, M. (2021, February 14). *Standardization*. Investopedia. [web article]

Google. (n.d.). *Perspectives on Issues in AI Governance*. [company report]

Haan, K. (2023, April 25). *24 Top AI Statistics And Trends In 2023*. Forbes Advisor.

Kang, C. (2023, July 21). *A.I. regulation is in its Early days.* The New York Times.

Lynch, S. (2022). *The state of AI in 9 charts*. Stanford HAI. [web article]

Martin, K. (2019). *Ethical Implications and Accountability of Algorithms.* Journal of Business Ethics 160, 835–850

McKinsey Global Institute. (2018, September). *Notes from the frontier: Modeling the impact of AI on the world economy*. McKinsey & Company.

Müller, Vincent C. (2020). *Ethics of Artificial Intelligence and Robotics.* The Stanford Encyclopedia of Philosophy (Fall 2023 Edition).

Pethig, F., & Kroenung, J. (2023). *Biased Humans, (Un)Biased Algorithms?*. Journal of Business Ethics, 183, 637–652.

Prokopeak, M. (2022). *The Many Challenges of AI Governance — And Why It Matters.* Reworked [web article]

Quizlet. (2023, July 24). *Quizlet's State of AI in Education Survey Reveals Teachers Are Surprise AI Champions*. PR Newswire. [web article]

Schwarz-Plaschg, C. (2018). *The Power of Analogies for Imagining and Governing Emerging Technologies*. Nanoethics, 12, 139–153.

Sheehan, M. (2023, July 10). *China's AI regulations and how they get made.* Carnegie Endowment for International Peace.

The China Electronics Standardization Institute. (2021, October 21). *Artificial Intelligence Standardization White Paper (2021 Edition)*. Center for Security and Emerging Technology.

The White House. (2023, March 16). *Blueprint for an AI Bill of Rights. Washington, D.C. Office of Science and Technology Policy*. [federal report]

Thomson, J. (2023, April 6). *People destroyed printing presses out of fear. What will we do to AI?*. Big Think. [web article]