
Most Significant Ethical Challenges Associated With AI Governance

Kiran Sharma Panchangam Nivarthi

Abstract

AI is radically transforming our economy and society in unprecedented ways and on an unprecedented scale. It may have an impact on humanity that's comparable to the industrial revolution and the internet. AI governance is crucial to ensure that AI development and growth remain in the best interest of humanity. However, the ever-evolving nature of AI, its definition, scope, and its developing impact on society and the economy make AI governance a challenging endeavor. The AI governance challenges can be divided into several categories and domains, including ethical challenges. In this article, we have identified five of the most significant ethical challenges associated with AI governance. It's not the full extent of ethical challenges because that range will evolve alongside AI, but it aims to serve as a comprehensive overview of the ethical challenges the world is facing today in the adoption and integration of AI with existing processes, along with some solutions and actionable insights. An overarching challenge with ethical AI governance is to ensure that it's not too restrictive to throttle progress or so permissive that there are no tangible protections against the negative ethical impact of AI.

Keywords:

AI governance;
Ethical challenges;
Ethical Challenges
Associated With AI
Governance;

Copyright © 2024 International Journals of Multidisciplinary
Research Academy. All rights reserved.

Author correspondence:

Kiran Sharma Panchangam Nivarthi,
31108 Algonquin Trail, Chisago City, MN, USA
Email: Kiransnandu@gmail.com

1. Introduction

Even though it's currently in its early stages of maturity, Artificial Intelligence (AI) is predicted to be as transformative for the global economy, modern societies, and the world as a whole as the internet has proven to be. Its impact on the fourth industrial revolution is predicted to be all-encompassing – revolutionizing everything in unprecedented ways and on an unprecedented scale [1]. One early estimate (made in 2018) of AI's impact on the global economy was that it would deliver additional economic output of about \$13 trillion by 2030 and would boost global GDP by about 1.2% every year [2].

However, this was before mainstream generative AI products like ChatGPT, Midjourney, and Dall-E were made available commercially, and there was limited information on their potential impact on different industries. There have been several studies on how generative AIs like Dall-E and Midjourney are reshaping the creative process and, more importantly, how they are revolutionizing the job market and workplaces built around creative outputs, and they have highlighted that job replacement is a viable possibility [3]. The fact that AI generative models need human artists' work to learn is often cited as a buffer against AI replacing jobs [4], but it's only a

matter of pace, and it will not be a hindrance when AI models are trained on AI-generated data. But if we broaden our spectrum to include a wide range of AI applications, their job replacement potential is enormous.

One study identified that about two-thirds of the jobs in the US alone would be exposed to some degree of AI automation, and about 25% to 50% of this exposed workload may be replaced, leading to diminished professional roles and lost jobs [5]. Out of all generative AIs, ChatGPT has been most extensively studied for its impact on creative output and the job market, and it has been identified that some industries and subsets of workers in those and other industries are more vulnerable than others [6].

Job replacement is just one of the major ethical and practical concerns associated with AI, and they deal with just one aspect of AI (generative) because generative AI products have garnered the largest user base. Several other challenges associated with AI research and their impact on various industries, the modern economy, and society at large have been identified. Generative AI tools like ChatGPT and Midjourney are already making significant waves in the field of education, and the implications of these tools in the hands of students and educators are forcing policymakers and a wide range of stakeholders to reevaluate everything from curriculum development and assessment [7]. AI is also transforming the field of management since it can be used in decision-making by replacing conventional tenets of wisdom and experience with data and algorithms, which have their own set of challenges [8]. The decision-making aspect of AI is also transforming the manufacturing industries, albeit at a different pace, as it's still struggling with the balance of human decision-making and AI-driven decision-making [9], but there is no denying that, like most other economic units, manufacturing industries are also pursuing AI solutions at an incredible pace [10].

Healthcare is another area where AI, alongside the Internet of Things (IoT), has triggered significant changes, and due to the highly sensitive nature of this industry, this is widely studied from a regulatory and policy standpoint. It has already shown promising results in the field of diagnostics, which can lead to better healthcare delivery [11], even though they are currently mainly relegated to a supporting role [12]. The field of diagnostics, especially radiography, is poised for rapid adoption of AI-based technologies that can be integrated into their current processes [13] despite challenges in specific domains like interventional radiology [14].

Algorithms, including the ones that prepare our social media feed, have a significant impact on how young people perceive and explore the world [15], which can have significant real-world implications, from purchase decisions and boycotts to voting. AI and machine learning (ML) algorithms can enhance it significantly and may have beneficial uses, such as positive social engineering [16] and detecting social engineering attacks [17], but also negative ones.

In conclusion, AI is altering our societies and the global economy in unprecedented ways and in several different dimensions at once. Its exponential growth is simultaneously creating a wealth of new opportunities, but the flip side is new risks and ethical challenges. It's in humanity's best interest to ensure that the growth and development of AI technologies and tools are governed and maintained in a way that allows us to leverage its benefits and mitigate the ethical risks they pose. This is where AI governance comes in.

2. AI Governance

There are several overlapping definitions of AI governance, and most of them lean towards a business's AI adoption and integrations. However, we will also focus on AI governance as it pertains to government regulations and regulatory oversight.

IBM defines AI governance as “*the ability to direct, manage, and monitor the AI activities of an organization,*” with compliance, trust, and efficiency as three core drivers behind AI governance trends in organizations [18]. A Deloitte report defined it as a spectrum of capabilities ensuring the responsible use of AI, which requires both conventional and differential governance constructs [19]. A paper dedicated to defining AI governance conceptualized it in the realm of corporate governance and, within it, IT governance and its overlap with data governance. It also identified

some of the challenges associated with AI governance, stemming from a fragmented understanding of the concept itself [20].

If we zoom out from private sector businesses and include the public sector in the mix as well, we identify that defining AI governance and its scope becomes more difficult. The public sector has access to far more data than the private sector, but it also has stricter data controls and regulatory restrictions [21]. Still, it has been a major concern since the early days of AI, when there was much less clarity on AI's impact on the world, and many research programs and studies focused on AI governance and its overlap with policy-making and developing new regulations. From a policy perspective, one research paper introduced a conceptual framework for policymakers to tackle the issue of AI governance, which identified one characteristic of a global AI governance model – flexibility to reconcile with different cultures and legal systems [22]. Another identified key considerations for AI governance to mitigate AI risk and described factors influencing AI governance (on a policy-making level), like problems a government is trying to solve, politics, policy formation, and adoption challenges [23].

Most definitions of AI governance boil down to this: Responsible use of AI in the given context. If it's a corporation, then AI governance would focus on ensuring that AI development, adoption, and use do not exceed regulatory or ethical boundaries and remain aligned with the business goals. For governments and governing bodies, internal AI governance will ensure that they do not exceed their mandate and that their AI-augmented operations are in the best interests of the public (as determined by legislation). When they define AI governance and its scope for the sake of policy, they should safeguard the interests of the people and uphold the legal and ethical principles without throttling AI's progress [24]. Job replacement, weaponization of AI, using AI to circumvent responsibility and accountability, algorithmic bias, and lack of transparency are some of the chief ethical concerns when it comes to defining AI governance and its scope in an organization or a state [25].

Even though the legal system has been relatively quicker in responding to the need for ethical AI governance, it's still catching up to AI penetration. However, there is a growing understanding of world leadership and global tech leadership regarding the critical importance of AI governance. But the current state of AI governance is fragmented, to say the least. AI governance has matured and has been refined and fine-tuned in industries like healthcare and education, where its practical applications and use have been growing at a rapid pace, but it's lacking in other industries.

3. Importance of AI Governance

The most overarching importance of AI governance is to ensure that the development of advanced AI technologies and tools is for the benefit of humanity in general [26]. AI has the potential to be a great disruptor, not just for individual industries and market segments [27] but the economy as a whole, alongside overlapping advances like automation [28]. It's beneficial and ethical when it's leveling the playing field [29] and offering equal opportunities to underdeveloped segments of the global economy, but despite its "Darkhorse" capabilities, most cutting-edge AI technologies will inevitably consolidate existing economic powerhouses, further exacerbating the power balance. This is something AI governance can mitigate if it is not neutralized.

Another crucial importance of AI governance is to mitigate or, ideally, prevent its weaponization (on a national scale) and worst uses. Parallels between the AI "arms race" and the nuclear arms race have already been drawn in literature, and one ominous prediction is that it's already too late to stop the AI arms race [30]. Even though, according to one paper, the theory of the AI arms race has been debunked [31], the possibility that AI can be weaponized in a number of ways still exists, and the right AI governance policies and framework can act as an effective deterrent. The danger here might be even more unique than that of nuclear arms because cutting-edge AI tech can be developed much more discretely.

AI is changing many disciplines in unprecedented ways, including healthcare and academia. In healthcare, the overlap with AI has resulted in a wealth of new challenges, including the accountability issue because if medical decisions are taken or influenced by AI, the responsibility and accountability will be spread out from medical professionals using these technologies all the way to developers. Other challenges include privacy concerns, job loss, etc. [32]. Similarly, tools like ChatGPT are already radically changing how people learn, how assignments are completed,

and even the assessment processes. Unless governed properly, with the right ethical values instilled into the AI development process and use, these tools may cause more harm than good [33].

Many early predictions about AI have been proven wrong. It was assumed that AI would not replace human creativity and that its impact on the creative market may be minimal, but generative AI arts have emerged as a massive market and are impacting artists all around the world [34]. On the other end of the spectrum, language models like ChatGPT-4 not only produce code [35] but also test it, which may significantly reduce the burden of manual coding and impact jobs. It should serve as a reminder that both positive and ominous predictions about AI should be approached with a degree of healthy skepticism. The influence we can have on AI's growth and its impact should also never be discounted, and AI governance is an important part of it.

AI is growing too rapidly for existing control measures, oversights, and ethical governance to keep pace with, and there is a need for an AI governance framework that covers most of the major ethical considerations and can evolve alongside AI. Ethical AI governance can significantly mitigate these challenges and positively curate the changes AI is triggering in the world. AI governance is important at every level, from organizations that simply use AI tools to governing bodies like the European Union so that its untethered and unsupervised growth doesn't harm the economy and societies they are created to govern and protect.

4. AI Governance Most Significant Ethical Challenges

The following ethical challenges are identified from existing literature and a variety of other sources. These ethical challenges encompass more than just the development of AI governance strategy and protocols; they also cover its implementation.

Ethical concerns may differ from industry to industry and among different stakeholders, but the following are some of the broadest ethical challenges that people responsible for AI governance have to contend with.

4.1 Job Replacement

Ironically, AI replacing jobs is not a "new" ethical dilemma. It's something humanity has been going through since the Industrial Revolution and maybe before even that. Mass production of vehicles ended the horse and carriage industry, and printers took the jobs away from scribes. We are already observing the tangible impact of AI replacing jobs [36] in various industries, and even where AI is not replacing jobs, it's already having a negative impact on the morale and thinking process of a sizable segment of the workforce [37]. It can also lead to wage stagnation or decline by reducing the value of human labor and talent. This indicates the importance of AI governance interacting with or working in conjunction with regulatory bodies (like ministries of labor) that govern the job market in a country. More people losing their jobs can also strain the benefit system of a country.

This challenge is particularly difficult to navigate in AI governance because, from a profitability perspective, it's detrimental for organizations to not choose the most optimal resources for their business operations. However, for a country or any society, a massive rise in joblessness can lead to a wide range of social and security challenges. Reconciling the two different interests from a governance perspective may require each side to concede some ground.

4.1.1 Examples and Projections

We are already seeing examples of AI replacing jobs in various markets.

- An OECD report consolidated 96 case studies with 325 underlying interviews focused on AI's impact on job replacement. It covered eight countries and multiple sectors, most prominently finance and manufacturing. The report identified Customer Service Representatives (14 instances), Maintenance & Repair Workers (14 instances), and Electromechanical Equipment Assemblers (9 instances) as the occupations most affected by AI (and automation) implementation. It also identified various tasks that were automated (at least partially) with AI, reducing human intervention and the number of workers, like responding to and classifying emails and identifying spare parts [38].

- A PwC study commissioned by the UK Department for Business, Energy, and Industrial Strategy in 2021 estimated that about 7% of existing UK jobs will be displaced because of AI in the next five years, 18% in ten years, and over 30% in the next 20 years. The retail sector was identified as the most vulnerable to this effect. However, the report also identified several new job creation avenues, albeit requiring different skill sets [39].
- According to one tracker, US firms laid off 4,600 people between May 2023 and January 2024 because of AI [40]. A separate report identified 5,500 job cuts by US-based tech giants in just the first two weeks of 2024. For two of the tech giants, Google and Amazon, the lay-offs coincided with or came after significant investments in an AI startup [41].
- A report by Asana that was generated after interviewing 4,546 knowledge workers in the United States and the United Kingdom identified that 29% of their tasks were replaceable with AI [42].

4.1.2 Possible Solutions

While it's not a solution per se, the opposite of job displacement and replacement is also happening or expected to happen, thanks to AI, i.e., job creation. A study concentrated on the German job market identified that industrial robots that displace existing jobs in the logistics industry create enough new jobs to offset the number of jobs it displaced [43]. One problem with it, and most of the other studies focused on job displacement and replacement by AI and AI-augmented automation, is that job displacement and job creation take place in different labor domains. This is particularly true in industrial settings. However, for creative and a wide range of white-collar jobs where AI tools are replacing human workers or reducing workforces, researchers are proposing, and organizations are experimenting with "human-AI-cooperation" models. Six researchers presented a human-AI co-creation model at a conference that focused on human-AI collaboration practices and frameworks and the importance of dispelling the idea of directly competing with AIs [44].

Governments and regulatory bodies are addressing this ethical dilemma as well. A national collective agreement in Italy in 2017 included individual rights for workers, stating that they would be trained or moved to other tasks if their jobs were being replaced by AI or automation [45]. A World Bank brief evaluated the existing employee/job market support structures for AI-based job displacement like Employee Insurance, redundancy payments, and income support programs. Not all countries have all of these in place, but augmenting them for widespread AI job displacement in the free market might serve as an important line of defense. The brief also suggested that government or global initiatives should focus on reintroducing the displaced workers to the market after training or relocation assistance to job markets where they may still have a chance [46].

As for organizations, PwC has already invested \$1 billion to give its 75,000 employees training on their AI chatbot. It's an active measure to ensure the company doesn't have to cut any jobs because of AI [47]. Upskilling the existing workforce and providing training on advanced technologies is one of the most common preemptive alleviative measures against the ethical challenge of job displacement that AI poses.

4.2 Bias and Discrimination

Bias is one of the major ethical concerns associated with AI governance, as biased performance and suggestions have been observed in many ungoverned AI applications, especially in the healthcare sector [48]. Data has been identified as the primary root of these biases, and this has led to a significant overlap of AI governance with data governance and indicates the need for comprehensive governance protocols that are not limited to the development and deployment of AI systems [49]. The challenge becomes even more significant if you consider the inherent bias in data and strive to design a governance mechanism that leads to the development of AI systems that *counteract* this inherent bias [50]. Another governance challenge is of scale. Current frameworks developed to tackle the issue of biasness in the development and deployment of AI models introduce a significant amount of red tape [51] that is currently deemed necessary, but if these frameworks are enforced as governance models, they may push the AI model out of reach of smaller businesses.

AI governance frameworks and laws have been studied in the context of existing discrimination laws to produce conducive governance guidelines, but it has its own sets of challenges, particularly the subtle, unintuitive, and, in some cases, intangible discrimination of AI models [52]. The discrimination risk is prevalent in multiple facets of our economy and society and overlaps with other risks, such as the impact on human jobs if ungoverned or ineffectively governed AI models are integrated into the hiring process [53]. This broadens the scope of governance and areas of intersections exponentially.

While bias and discrimination are significant ethical challenges associated with AI, and ideally, they should be addressed at the governance level, they are already being tackled by many organizations and other entities just through adherence to good screening and development practices. Organizations that properly vet the AI tools they use to ensure that they match their internal policies against bias and discrimination, and if they are commissioning an AI tool or developing one internally, tackling this bias is typically part of the process.

4.2.1 Examples

AI tools and technologies have been deployed at various organizational levels, and different scales, and many of them have displayed examples of bias and discrimination.

- Amazon developed an algorithm in 2014 to screen resumes, and they trained that algorithm on their own data set – ten years of resumes they had access to and the profiles of the company employees. However, since the bulk of the workforce was white and male, that's what the algorithm developed a preference for, leading to sexist and racist screening practices. It was one of the earliest examples of how, based on flawed data, AI models can be trained to have inherent biases and, as a result, discriminatory outputs [54].
- AI-based facial recognition systems tend to be biased towards some racial groups and skin tones. One common theme across multiple such systems is they are more accurate with lighter skin tones compared to darker ones. Since facial recognition has become an important part of our daily lives, from unlocking phones to surveillance systems, this bias can lead to discriminatory profiling, bad consumer experience, and legal injustice in certain cases [55].
- AI models can be gender and racially-biased because of their data sets and wrongly defined parameters. There are several examples of biased practices in the healthcare system being transferred to AI models being trained, like the fact that only 17% of cardiologists identify that women may have a higher risk of heart diseases compared to men while the majority of the population leans towards a higher prevalence of heart diseases in men. AI models that are trained with biased diagnostic data may underdiagnose or misdiagnose women [56].
- AI algorithms are also being deployed/integrated into the justice system for a wide variety of use cases, including identifying the hot spots for heavy police concentration and surveillance to short-listing inmates for an early release. These algorithms might show bias towards certain neighborhoods/areas and certain population segments, perpetuating the existing hurtful stereotypes and systematic biases [57].

It's important to note that AI bias and discrimination are complex challenges to navigate because they manifest differently in different industries and use-case scenarios, and they are also highly reliant on the training data. AI models trained on massive amounts of online data that either can't or haven't been screened for bias and discrimination are more vulnerable to this. They may require the use of stringent measures against bias and discrimination placed at training or implementation levels.

4.2.2 Possible Solutions

A wide range of solutions have been adopted to remove biases and discrimination from AI models deployed in various industries. One solution that permeates almost all domains is managing data and developing training algorithms and processes with the shortcomings of data in mind. For example, if some racial, ethnic, or gender groups are over or underrepresented in the training data set, the training data is readjusted, or the training parameters have to be adjusted for discrepancies

in the data. Strategies like data curation, algorithmic debiasing, and inclusive data collection can be instrumental in mitigating this issue. Human oversight or human-in-the-loop in training or decision-making processes can also reduce bias in AI model training and implementation [58] [59].

4.3 Trust

Most users of technology are typically unaware of how it works and how all the processes and physical elements come together to make it function. However, if they truly wish to develop an understanding of technology, there are ample resources available. However, many AI models are black boxes and uninterpretable, which creates a state of distrust because it's difficult to trust an AI model's decisions and suggestions without comprehending how it reached them [60]. Even though the right AI governance can go a long way towards establishing trust between AI systems and models [61], there are several challenges in developing this element of AI governance, including the coverage of all the nuances and complexities associated with trust between humans [62].

4.3.1 Examples

Trust in AI and AI-powered technologies varies greatly based on the level of interaction, dependency, and situations.

- Autonomous vehicles that rely upon a wide range of technologies to navigate traffic and other obstacles on the road and are tasked with important decisions regarding passenger and pedestrian safety have long been a concern. The famous “baby or grandma” ethical dilemma is a prime example of this. A study that included 18 participants and Autonomous Vehicle (AV) simulations of four different risk levels found that trust in AI increased when the risk was lower but low when the risk increased [63].
- A study including 114 participants tested the AI-human trust hypothesis when it comes to a particular decision-making scenario, i.e., buying the stock of a company. The study revealed that humans tended to trust the AI model more when it gave information in a precise format [64]. This shows that not just an AI's development but how it interacts with the users can have a significant influence on trust.

4.3.2 Possible Solutions

The possible solutions that focus on developing trust between AI and humans have to cover different types of trust – like cognitive and emotional trust and rely upon factors like how an AI is presented and how it interacts with humans [65]. Other solutions delve deeper into the nature of AI and human trust and show that defining the trust between AI and humans, limiting its scope, and differentiating between AI trust and anthropomorphic trust that so far has been reserved for humans may benefit in drawing realistic trust dimensions [66]. Other significant steps taken in this direction are related to transparency – making AI models more open and implementing explainable AI techniques for the development of AI tools instead of black-box learning.

4.4 Privacy

Privacy is a complex ethical challenge to address through AI governance because it thrives on data, including personal data, and it would be nearly impossible to develop good AI models without a healthy, comprehensive, and extensive data set, which often leads to complex privacy issues. Take healthcare as an example. If an AI model is being trained to aid in diagnostics, identifying the most vulnerable individuals or population segments, or simply studying disease patterns, it needs identifiable data like age, gender, weight, etc. The largest health institutions in any country may have access to enough data to train models, but even those extensive data sets are localized and do not often fully represent the underlying regional or national population, so data sharing between healthcare facilities to train models is a viable practice. If AI governance forbids it, the risk is training biased models. If it doesn't, patient data may be shared with a number of AI models without their knowledge [67]. The challenges become more significant if AI development takes place outside the industry fold, with conflicting interests and practices.

4.4.1 Example

Unless it's designed intentionally for this purpose, most AI privacy flaws or issues are unintentional. However, they are concerns that should be addressed at the development/governance level.

- Smart meters can be designed to extract data and infer insights from that data regarding appliances in the homes, peak energy use times, and other data parameters that, when combined with other data sets, can reveal things like preferences, routines, and the economic status of households [68].

4.4.2 Possible Solutions

These privacy concerns can be addressed by adopting the right development and AI model training practices, which in turn can be ingrained in AI governance. That can also be achieved by expanding its dimension to interact with data privacy laws. Another step in the right direction when it comes to AI privacy concerns is anonymizing training data.

4.5 Liability and Accountability Challenges

Can an AI model or system be held liable for its harmful decisions, suggestions, or actions, or does the liability pass through to the developers of the model or, even farther behind, to the individuals or teams responsible for collecting the data for training the data model? This is just one of the many difficult questions that individuals and entities responsible for AI governance need to answer when addressing the elements of liability and accountability in AI policies.

Regulations have already been prepared and proposed to this end, particularly the EU's AI Liability Directive (AILD), but its positive impact and potency are still in question [69]. One governance challenge identified in this domain is the liability gaps that exist when existing liability laws and regulations are applied to AI models and systems [70]. For many autonomous AI systems with unpredictable nature, it's difficult to assign fault to the producer or user. It's also challenging to trace causality and prove that whether the harm stemmed from negligence in development because even an ethically developed AI may cause harm in certain circumstances [71]. The lack of precedent for the legal system is another challenge in governing liability in AI models [72].

We can extend the legal challenges of liability to a broader dimension of ethical and social accountability. There are several challenges associated with integrating accountability in AI systems, which have to be considered from an AI governance perspective, including addressing social dilemmas as a tool. Can it really be considered inherently ethical or unethical [73]? Because if not, then how can it be held accountable? The question of AI accountability and, consequently, its governance becomes especially prominent in the public sector, where it has a direct impact on the lives of citizens [74] [75]. Another AI governance challenge is how AI tools can shift, modify, and even undermine existing accountability measures in place in certain industries and operations [76].

4.4.1 Examples

- In 2019, a Tesla Model S crashed into a car and killed the two passengers inside. That was one of the prime examples of finding the accountable party/parties when an autonomous vehicle causes an accident and kills someone.
- An image-based diagnoses algorithm mistook surgical marks for malignant melanoma and increased the rate of false positives by 40%. This raised concerns over accountability, whether it rested on developers or people supposed to provide human oversight to these diagnoses [77].

4.4.2 Possible Solutions

The accountability problems are an ethical dimension of AI governance that requires comprehensive, multi-pronged solutions. The facets of these solutions may include devising proper and clear ethical guidelines for AI development and implementation, regulatory frameworks enforcing these guidelines, algorithmic impact assessments before deployment, and transparency practices.

5. Limitations

These are just some of the most significant ethical challenges associated with AI governance right now. It's difficult to produce an evergreen list of challenges because of the fractured AI governance landscape and different rates of AI growth in different market segments and regions. Many of the challenges in the most evolved areas of AI governance (like healthcare) may not be relevant to budding areas of governance, like AI's intersection with IoT. Hence, any comprehensive list of ethical challenges of AI governance will essentially be ever-growing.

Another problem is that only a relatively limited number of AI tools are available to the general market, and many tools and AI models are working behind the curtain and are, therefore, understudied. The ethical challenges associated with such AI models may remain limited to specific industries or regulatory segments, so there might be a whole range of ethical challenges in AI governance that are not accessible to academia.

The overlap of AI governance, data governance, and other relevant governance may lead to a broader range of ethical challenges that we might predict but not define in their full scope. For now, we have observed only a limited impact of AI technologies on our society, and it's heavily concentrated towards generative AIs. When other AI facets like predictive AIs and discriminative AIs (and their tools) become more commonplace and a more impactful part of our societies and economies, we may encounter a completely new set of ethical challenges.

6. Conclusion

Identifying even some of the most prominent ethical challenges associated with AI governance can give corporations, regulatory bodies, governments, and international bodies an adequate amount of useful information when developing their AI governance policies and standards. Appreciating and recognizing these challenges is the first step towards developing policies that can circumvent or at least mitigate these challenges. However, significant research is needed to identify the prevalence, permeance, and impact these individual challenges can have on the overall AI landscape, research, and the development of AI models.

The primary goal of AI governance is to ensure beneficial AI development and growth. Balancing optimal AI growth with minimal negative impact in isolated areas (individual businesses or niche markets), broad industries (healthcare, legal, etc.), and society as a whole might be the most significant overarching challenge of AI governance. However, it's important to note that precedent exists for AI governance stakeholders to draw insights from, like how corporations are aligning their AI development with their ethical values and how various governments are ensuring the safety and privacy of their citizens by controlling AI development and usage. Many of these things have worked well while others have failed, and they can help guide the people responsible for AI governance policies. AI can also trigger a new era of collaboration among different industries, corporate stakeholders, governments, and researchers from different sectors and countries. Technologies like blockchain and Web 3 that lean towards a more decentralized internet might be naturally conducive to global collaboration, and AI governance stakeholders can adopt and leverage these technologies to tackle AI governance ethical challenges as a collective.

References

- [1] G. Velarde, "Artificial Intelligence And Its Impact On The Fourth Industrial Revolution: A Review," *International Journal of Artificial Intelligence & Applications (IJAIA)*, vol. 10, no. 6, November 2019.
- [2] J. Bughin, J. Seong, J. Manyika, M. Chui, and R. Joshi, "Notes From The AI Frontier Modeling The Impact Of AI on The World Economy," Mckinsey Global Institute, 2018.
- [3] A. M. Radhakrishnan, "Is Midjourney-Ai A New Anti-Hero Of Architectural Imagery And Creativity?: An Atypical Era Of Ai-Based Representation & Its Effect On Creativity In The Architectural Design Process.," *Global Scientific Journals*, vol. 11, no. 1, 2023.
- [4] S. Shuhan, Y. Chen, M. Hua, and M. Ye, "Measuring designer's use of Midjourney on the technology acceptance model," in *IASDR 2023 Life-Changing Design*, Milan, Italy, 2023.
- [5] A. P. D. Castillo, "Generative AI, generating precariousness for workers?" *AI & Society*, 2023.
- [6] E. Felten, M. Raj, and R. Seamans, "How will Language Modelers like ChatGPT Affect Occupations and Industries?" *arXiv*, 2023.
- [7] T. K. F. Chiu, "The impact of Generative AI (GenAI) on practices, policies, and research direction in education: a case of ChatGPT and Midjourney," *Interactive Learning Environments*, 2023.
- [8] F. H. Jordi Canals, *The Future of Management in an AI World: Redefining Purpose and Strategy in the Fourth Industrial Revolution*, Palgrave Macmillan Cham.
- [9] L. Krishnan, "Impact of Using AI in Manufacturing Industries," *Journal of the International Academy for Case Studies*, vol. 28, no. Special Issue 4, 2022.
- [10] S. Shao, Z. Shi, and Y. Shi, "Impact of AI on employment in the manufacturing industry," *International Journal of Financial Engineering*, vol. 9, no. 3, 2022.
- [11] M. Ali, "A Comprehensive Review of AI's Impact on Healthcare: Revolutionizing Diagnostics and Patient Care," *BULLET: Jurnal Multidisiplin Ilmu (INPRESS)*, vol. 2, no. 4, 2023.
- [12] M. H. R. Mehrizi, P. v. Ooijen and M. Homan, "Applications of artificial intelligence (AI) in diagnostic radiology: a technography study," *European Radiology*, vol. 31, p. 1805–1811, 2020.
- [13] M. Hardy and H. Harvey, "Artificial intelligence in diagnostic imaging: impact on the radiography profession," *The British Journal of Radiology*, vol. 93, no. 1108, 2019.
- [14] T. Boeken, J. Feydy, A. Lecler, P. Soyer, A. Feydy, M. Barat and L. Duron, "Artificial intelligence in diagnostic and interventional radiology: Where are we now?," *Diagnostic and Interventional Imaging*, vol. 104, no. 1, pp. 1-5, 2023.
- [15] J. Swart, "Experiencing Algorithms: How Young People Understand, Feel About, and Engage With Algorithmic News Selection on Social Media," *Social Media + society*, 2021.
- [16] N. C. J. H. F. e. a. Tomašev, "AI for social good: unlocking the opportunity for positive impact," *Nature Communications*, vol. 11, 2020.
- [17] M. Lansley, N. Polatidis and S. Kapetanakis, "SEADer: A Social Engineering Attack Detection Method Based on Natural Language Processing and Artificial Neural Networks," in *International Conference on Computational Collective Intelligence*, 2019.
- [18] IBM, *What is AI governance?* <https://www.ibm.com/analytics/common/smartpapers/ai-governance-smartpaper/#ai-governance-delivers>.
- [19] Deloitte, *AI governance for a responsible, safe AI-driven future*, <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/risk/us-ai-governance-for-a-responsible-safe-ai-driven-future-final.pdf>, 2021.
- [20] M. Mäntymäki, M. Minkkinen, T. Birkstedt, and M. Viljanen, "Defining organizational AI governance," *AI and Ethics*, vol. 2, p. pages603–609, 2022.
- [21] M. Kuziemski and G. Misuraca, "AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings," *Telecommunications Policy*, vol. 44, no. 6, 2020.
- [22] U. Gasser and V. A. Almeida, "A Layered Model for AI Governance," *Internet Governance*, 2017.
- [23] B. Perry and R. Uuk, "AI Governance and the Policy-making Process: Key Considerations for Reducing AI Risk," *Big Data and Cognitive Computing*, vol. 3, no. 2, 2019.
- [24] A. Taeihagh, "Governance of artificial intelligence," *Policy and Society*, vol. 40, no. 2, p. 137–157, 2021.

- [25] J. Butcher and I. Beridze, "What is the State of Artificial Intelligence Governance Globally?" *The RUSI Journal*, vol. 164, no. 5-6, 2019.
- [26] A. Dafoe, "AI governance: a research agenda.," *Governance of AI Program, Future of Humanity Institute, University of Oxford: Oxford, UK*, vol. 1442, 2018.
- [27] A. Mer and A. S. Viridi, "Artificial Intelligence Disruption on the Brink of Revolutionizing HR and Marketing Functions," in *Impact of Artificial Intelligence on Organizational Transformation*, Scrivener Publishing LLC, 2022.
- [28] A. Pavaloiu, "The Impact of Artificial Intelligence on Global Trends," *Journal of Multidisciplinary Developments.*, vol. 1, no. 1, pp. 21-37, 2016.
- [29] D. McDuaie-Ra and K. Gulson, "The backroads of AI: The uneven geographies of artificial intelligence and development," *Area*, vol. 52, no. 3, pp. 461-662, 2020.
- [30] E. M. Geist, "It's already too late to stop the AI arms race—We must manage it instead," *Bulletin of the Atomic Scientists*, vol. 72, no. 5, pp. 318-321, 2016.
- [31] P. Scharre, "Debunking the AI arms race theory," *Texas National Security Review*, 2021.
- [32] D. Lee and S. N. Yoon, "Application of Artificial Intelligence-Based Technologies in the Healthcare Industry: Opportunities and Challenges," *International Journal of Environmental Research and Public Health*, vol. 18, no. 1, 2021.
- [33] J. G. Meyer, R. J. Urbanowicz, P. C. N. Martin, K. O'Connor, R. Li, P.-C. Peng, T. J. Bright, N. Tatonetti, K. J. Won, G. Gonzalez-Hernandez and J. H. Moore, "ChatGPT and large language models in academia: opportunities and challenges," *BioData Mining*, 2023.
- [34] H. H. Jiang, L. Brown, J. Cheng, M. Khan, A. Gupta, D. Workman, A. Hanna, J. Flowers, and T. Gebru, "AI Art and its Impact on Artists," in *AIES '23: Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 2023.
- [35] Y. Feng, S. Vanam, M. Cherukupally, W. Zheng, M. Qiu, and H. Chen, "Investigating Code Generation Performance of ChatGPT with Crowdsourcing Social Data," in *Proceedings of the 47th IEEE Computer Software and Applications Conference*, 2023.
- [36] D. H. Kim, "A Study on the Job Replacement Impact of ChatGPT and Education Method," *European Journal of Science, Innovation and Technology*, vol. 3, no. 4, 2023.
- [37] D. Vorobevea, Y. E. Fassi, D. C. Pinto, D. Hildebrand, M. M. Herter, and A. S. Mattila, "Thinking Skills Don't Protect Service Workers from Replacement by Artificial Intelligence," *Journal of Service Research*, vol. 25, no. 4, 2022.
- [38] A. Milanez, "The impact of AI on the workplace: Evidence from OECD case studies of AI implementation (OECD Social, Employment and Migration Working Papers No. 289)," OECD, 2023.
- [39] PwC, "The Potential Impact of Artificial Intelligence on UK Employment and the Demand for Skills - A PwC Report for the Department BEIS," UK Govt. , 2021.
- [40] J. Constantz, "AI Is Driving More Lay-offs Than Companies Want to Admit," Bloomberg, 2024. [Online]. Available: <https://www.bloomberg.com/news/articles/2024-02-08/ai-is-driving-more-layoffs-than-companies-want-to-admit>.
- [41] C. Catherine Thorbecke, "The tech sector is pouring billions of dollars into AI. But it keeps laying off humans," CNN, 2024. [Online]. Available: <https://edition.cnn.com/2024/01/13/tech/tech-layoffs-ai-investment/index.html>.
- [42] T. W. I. L. (. Asana), "The State of AI at Work," Asana, 2023.
- [43] J. Furman and R. Seamans, "AI and the Economy," *Innovation Policy and the Economy*, vol. 19, 2019.
- [44] Z. Wu, D. Ji, K. Yu, X. Zeng, D. Wu, and M. Shidujaman, "AI Creativity and the Human-AI Co-creation Model," in *HCI 2021: Human-Computer Interaction. Theory, Methods, and Tools*.
- [45] A. Deshpande, N. Picken, L. Kunertova, A. D. Silva, G. Lanfredi and J. Hofman, "Improving working conditions using Artificial Intelligence," *Policy Department for Economic, Scientific, and Quality of Life Policies*, 2021.
- [46] A. D. Schmillen, "Causes and Impacts of Job Displacements and Public Policy Responses," Research & Policy Briefs From the World Bank Malaysia Hub, 2020. [Online]. Available: <https://documents1.worldbank.org/curated/en/621601588770464452/pdf/Causes-and-Impacts-of-Job-Displacements-and-Public-Policy-Responses.pdf>.
- [47] J. McGregor, "PwC's \$1B Investment Will Give Every Worker AI Training—And Staffers Access To Chatbot Assistants," Forbes, 2023. [Online]. Available:

<https://www.forbes.com/sites/jenamcgregor/2023/11/08/pwcs-1b-investment-will-give-every-worker-ai-training-and-staffers-access-to-chatbot-assistants/?sh=4a636cc55b8c>.

- [48] G. S. Nelson, "Bias in Artificial Intelligence," *North Carolina Medical Journal*, vol. 80, no. 4, pp. 220-222, 2019.
- [49] D. Coates and A. Martin, "An instrument to evaluate the maturity of bias governance capability in Artificial Intelligence projects," *IBM Journal of Research and Development*, vol. 63, no. 4/5, 2019.
- [50] S. Reddy, S. Allan, S. Coghlan, and P. Cooper, "A governance model for the application of AI in health care," *Journal of the American Medical Informatics Association*, 2019.
- [51] R. Schwartz, A. Vassilev, K. Greene, L. Perine, A. Burt, and P. Hall, "Towards a standard for identifying and managing bias in artificial intelligence.," *NIST special publication*, 2022.
- [52] S. Wachter, B. Mittelstadt, and C. Russell, "Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI," *Computer Law & Security Review*, vol. 41, 2021.
- [53] A. Todolí-Signes, "Algorithms, artificial intelligence and automated decisions concerning workers and the risks of discrimination: the necessary collective governance of data protection," *Transfer: European Review of Labour and Research*, vol. 25, no. 4, 2019.
- [54] A. C. C. C. Max Langenkamp, "Hiring Fairly in the Age of Algorithms," *arXiv:2004.07132*, 2020.
- [55] R. Singh, P. Majumdar, S. Mittal, and M. Vatsa, "Anatomizing Bias in Facial Analysis," in *The Thirty-Sixth AAAI Conference on Artificial Intelligence (AAAI-22)*, 2022.
- [56] D. Cirillo, S. Catuara-Solarz, C. Morey, E. Guney, L. Subirats, S. Mellino, A. Gigante, A. Valencia, M. J. Rementeria, A. S. Chadha and N. Mavridis, "Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare," *NPJ digital medicine*, vol. 3, 2020.
- [57] M. Bagaric, J. Svilar, M. Bull, D. Hunter, and N. Stobbs, "The Solution To The Pervasive Bias And Discrimination In The Criminal Justice System: Transparent And Fair Artificial Intelligence," *American Criminal Law Review*, vol. 59:95, 2022.
- [58] R. Confalonieri, F. Lucchesi, G. Maffei, and S. Catuara-Solarz, "Chapter 9 - A unified framework for managing sex and gender bias in AI models for healthcare," in *Sex and Gender Bias in Technology and Artificial Intelligence Biomedicine and Healthcare Applications*, 2022, pp. 179-204.
- [59] E. Ntoutsis, P. Fafalios, U. Gadiraju, V. Iosifidis, W. Nejdil, M.-E. Vidal, S. Ruggieri, F. Turini, S. Papadopoulos, E. Krasanakis, I. Kompatsiaris, K. Kinder-Kurlanda, C. Wagner and e. al, "Bias in data-driven artificial intelligence systems—An introductory survey," *Wires - Data Mining and Knowledge Discovery*, vol. 10, no. 3, 2020.
- [60] W. J. v. Eschenbach, "Transparency and the Black Box Problem: Why We Do Not Trust AI," *Philosophy & Technology*, vol. 34, p. 1607–1622, 2021.
- [61] J. Roski, E. J. Maier, K. Vigilante, E. A. Kane, and M. E. Matheny, "Enhancing trust in AI through industry self-governance," *Journal of the American Medical Informatics Association*, pp. 1-9, 2021.
- [62] A. Jacovi, A. Marasović, T. Miller and Y. Goldberg, "Formalizing Trust in Artificial Intelligence: Prerequisites, Causes and Goals of Human Trust in AI," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 2021.
- [63] I. B. Ajenaghughrure, S. C. d. C. Sousa and D. Lamas, "Risk and Trust in artificial intelligence technologies: A case study of Autonomous Vehicles," in *2020 13th International Conference on Human System Interaction (HSI)*, 2020.
- [64] M. G. J. C. L. Jungkeun Kim, "When do you trust AI? The effect of number presentation detail on consumer trust and acceptance of AI recommendations," *Psychology & Marketing*, vol. 38, no. 7, pp. 1140-1155, 2021.
- [65] E. Glikson and A. W. Woolley, "Human Trust In Artificial Intelligence: Review Of Empirical Research," *Academy of Management Annals*, vol. 14, no. 2, p. 627–660, 2020.
- [66] M. Ryan, "In AI We Trust: Ethics, Artificial Intelligence, and Reliability," *Science and Engineering Ethics*, vol. 26, p. 2749–2767, 2022.
- [67] J. S. Winter and E. Davidson, "Governance of artificial intelligence and personal health information," *Digital Policy, Regulation and Governance*, 2019.
- [68] S. S. G. V. D. V. Jillian Carmody, "AI and privacy concerns: a smart meter case study," *Emerald Insights*, 2021.
- [69] P. Hacker, "The European AI liability directives – Critique of a half-hearted approach and lessons for the future," *Computer Law & Security Review*, vol. 51, 2023.

- [70] S. D. Conca, "Bridging the Liability Gaps: Why AI Challenges the Existing Rules on Liability and How to Design Human-empowering Solutions," in *Law and Artificial Intelligence*, The Hague, T.M.C. Asser Press, 2022, pp. 239-258.
- [71] M. Buiten, A. d. Streel and M. Peitz, "The law and economics of AI liability," *Computer Law & Security Review*, vol. 48, 2023.
- [72] O. J. Erdélyi and G. Erdélyi, "The AI Liability Puzzle and a Fund-Based Work-Around," in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 2020.
- [73] V. Baldi and L. Oliveira, "Challenges to incorporate accountability into artificial intelligence," *Procedia Computer Science*, vol. 204, pp. 519-523, 2022.
- [74] M. Loi and M. Spielkamp, "Towards Accountability in the Use of Artificial Intelligence for," in *ES '21: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 2021.
- [75] M. Busuioc, "Accountable Artificial Intelligence: Holding Algorithms to Account," *Public Administration Review*, vol. 81, no. 5, pp. 817-989, 2021.
- [76] A. Volkova, A. D. Patil, S. A. Javadi and H. d. Meer, "Accountability challenges of AI in smart grid services," in *e-Energy '22: Proceedings of the Thirteenth ACM International Conference on Future Energy Systems*, 2022.
- [77] S. Kennedy, "AI May Be More Prone to Errors in Image-Based Diagnoses Than Clinicians," health IT analytics, 2022. [Online]. Available: <https://healthitanalytics.com/news/ai-may-be-more-prone-to-errors-in-image-based-diagnoses-than-clinicians>.