Artificial Intelligence Governance:
A Comparative Analysis of China, the European Union, and the United States

MPP Professional Paper

In Partial Fulfillment of the Master of Public Policy Degree Requirements
The Hubert H. Humphrey School of Public Affairs
The University of Minnesota

Ren Bin Lee Dixon

May 7, 2022

Signature below of Paper Supervisor certifies successful completion of oral presentation ***and*** *completion of final written version:*

______________________________ ____________________ ____________________
Steve Kelley, Paper Supervisor Date, oral presentation Date, paper completion
Affiliate Faculty

____ ____________________
Professor Maria Gini, Second Committee Member Date
CSE Distinguished Professor
Signature of Second Committee Member, certifying successful completion of professional paper

____ ____________________
Elizabeth M. Adams, Third Committee Member Date
CEO and Chief AI Ethics & Culture Advisor - Leadership of Responsible AI™
Signature of Third Committee Member, certifying successful completion of professional paper

Executive Summary

Artificial Intelligence (AI) has become increasingly more ubiquitous and deployed across many sectors and industries. While the technology is expected to bring transformative changes to society, there has been a growing urgency to establish robust governance frameworks to mitigate the issues and risks attendant with its deployment. A representative governance initiative was selected from China, the European Union, and the United States — as the three leading global AI regimes at present — to conduct a comparative analysis on their approaches. Based on the analysis, this paper makes nine broad and amendable AI policy recommendations:

1. Implement a centralized AI governance framework to ensure that all AI principles are effectively incorporated throughout the development and regulation of AI.
2. Establish robust data protection regulations to uphold individual privacy and encourage safe and secure collection, storage, and use of data.
3. Employ transparency as a compliance mechanism in high-risk AI to foster greater public confidence in the technology.
4. Require testing to enforce safety and compliance in any AI that carries risks.
5. Collaborate with global alliances to support the common advancement of AI, and collaborate with local stakeholders to develop and enforce AI regulations.
6. Invest in AI research areas including studies on its long-term impact, potential cyber threats, and impact assessment on AI governance.

7. Implement distributive and redistributive policies, such as strong antitrust laws, progressive taxation, basic income, and negative income tax models to counter the concentration of wealth and power engendered by AI.
8. Integrate AI-related skills in education systems and increase AI literacy among the public to preserve self-agency in a democratic society.
9. Anticipate AI implications on the job market and provide upskilling and reskilling programs for workers who will be most affected by the job market shift.

These policy recommendations were developed to address fundamental AI principles that had been identified and distilled from a corpus of over ninety AI governance initiatives published by the academia, private and public sectors, and multistakeholder groups. AI principles were chosen as the standard for policy analysis in this paper because they have been established — in the field of AI governance — as well-researched guidelines that can be used as the foundation for developing AI governance frameworks. The eleven principles that grounded the policy analysis and recommendation in this paper are (see Table 1):

Table 1.

Topics and Keywords for AI Principles

Topics	Keywords
For Human	for human, beneficial, well-being, dignity, freedom, diversity
Fairness	fairness, justice, bias, discrimination, prejudice
Transparency	transparency, explainable, predictable, intelligible, audit, trace
Privacy	privacy, data protection, informed, control the data
Safety	safety, validation, verification, test, controllability, human control
Accountability	accountability, responsibility
Security	security, cybersecurity, cyberattack
Share	share, equal, equity, power, distributive
Collaboration	collaboration, partnership, cooperation, dialogue
Sustainability	sustainability, environment, Sustainable Development Goals
Long-Term AI	AGI, superintelligence, higher level AI

Note. AI principles topics and keywords based on Linking Artificial Intelligence Principles by (Zeng et al., 2018)

AI is a general-purpose technology that could potentially generate a monumental shift in society and humanity. Its impact has been compared to the harnessing of electricity and the industrial revolution (Lynch, 2017). To provide a deeper look at the benefits the technology brings, the paper presents examples of AI application in transportation, finance, and environmental sustainability. The risk of AI misuse is also explored in cases relating to data harvesting, social credit system, and algorithmic biases. AI risks can be broadly grouped into policy areas based on the level of urgency, the rate they're occurring, and their impact on people. For instance the short- to mid-term risks includes (Calo, 2017):

1. Justice and Equity – underrepresentation in training data has led to algorithmic biases and discriminatory outcomes in high-impact sectors.
2. Lethal Autonomous Weapons – countries are unable to agree on a preemptive ban, leaving ethical concerns, risks, and technological advancement unresolved.
3. Safety – lack of transparency and explainability in AI systems complicates the question of accountability in safety issues.
4. Privacy and Power – unrestricted access to data collection and usage have contributed to the concentration of wealth and power among a few large online entities.
5. Security – the unique nature of AI could attract novel security threats that could be more efficient, accurate, and on a larger scale.
6. Labor Displacement – rapidly emerging AI technologies are anticipated to disrupt the job market, subjecting certain tasks to higher risks of automation.

The longer-term risks in AI are primarily related to the expected arrival of powerful AI, that is Artificial General Intelligence and superintelligence (Bostrom, 2014). Against this backdrop, AI researchers Dafoe and Bostrom hypothesized and cautioned against several potential extreme risks that could arise from advanced AI (Bostrom et al., 2018; Dafoe, 2018).

1. Robust totalitarianism – AI can be used as a powerful tool for monitoring and manipulating large populations, concentrating power in the hands of a few elites.

2. Great power war – the pursuit of lethal autonomous weapons creates the availability of extreme military advantage that could engender more complex crisis dynamics that may escalate rapidly due to automation.
3. Value misalignment – the possibility of more advanced AI that is not aligned with human values could result in harmful outcomes.
4. Value erosion from competition – even when measures are put in place to avoid the previous scenarios, increasingly competitive AI environments could erode those values and instigate progressively harmful compromises in pursuit of more power and wealth.

In light of the rapid emergence of the benefits and risks involved in the deployment of AI, numerous countries have been working toward developing governance frameworks to regulate the deployment of AI. The context of AI governance along with its challenges are examined from a selection of academic literature that studied the various governance efforts, to identify trends and limitations that have emerged within this field.

While this paper attempts to present an overview of the state of the current AI governance efforts in China, the European Union, and the United States, the analysis and policy recommendations presented should not be considered exhaustive. AI is an emerging technology and its regulatory landscape is still developing and maturing at different rates. There are constantly new governance initiatives and analyses generated from different contexts and perspectives, contributing to the discourse of AI governance. It is essential that policymakers explore the most current and relevant governance landscape to identify the best practices that are suitable for their specific context and purposes.

Table of Contents

Introduction

In the past decade, Artificial Intelligence (AI) has become increasingly more ubiquitous and deployed in various sectors and in a wide range of applications. According to the AI Index Report of 2021 published by the Stanford Institute for Human-Centered Artificial Intelligence (HAI), the number of AI publications in the world doubled from 2010 to 2021, growing from 162,444 to 334,497 (Zhang et al., 2022). Meanwhile, global AI adoption rate in 2021 was at 56 percent, up 6 percent from 2020. As a result of its continued advancement, AI has become steadily more affordable (the cost of training an image classification system in 2021 decreased 63.6 percent compared to 2018) and operates at a higher accuracy in tasks such as recommendations, object detection, and language processing. While the latest AI innovations are widely covered in news media, fueling futuristic promises of humanoid robots, the failings of AI compounded by the fear of its rapid and potentially exponential growth have also raised concerns globally and on the matter of security (Bostrom, 2014; Brynjolfsson & McAfee, 2016; Wakefield, 2022). Its rapid growth has led to policy lags, however, governance regimes are slowly emerging (Taeihagh, 2021). According to the 2021 AI Index, the number of laws containing "artificial intelligence" grew from just one in 2016 to 18 in 2021 (Zhang et al., 2022).

AI, however, is not a recent invention, rather it traces its emergence back to the 1950s when a group of computer scientists gathered at Dartmouth College for a summer workshop to explore the possibility of building a machine that could simulate human intelligence (Crevier, 1993). Since then, AI has been through a tumultuous history from highly funded periods, largely from the United States Department of Defense (DoD) through The Defense Advanced Research Projects Agency (DARPA).

These periods were contrasted with low periods known as AI Winters when expectations far exceeded the performance and achievements of AI, leading to disappointments and withdrawal of research funding. Research and innovation in AI continued with less fanfare over the 1980s to the 1990s, but interest in the technology was rekindled when graphic processing units (GPUs) became increasingly more powerful (Morris et al., 2017). The increase in processing power led to a renaissance of machine learning models that were proposed in the 1950s, but were dismissed at the time due to computational limitations (Crevier, 1993). The convergence of machine learning and big data[1] resulted in several AI achievements in the late 1990s and early 2000s that caught the attention and imagination of the public and private sectors. For instance, IBM's Deep Blue defeated Chess Grandmaster Garry Kasparov in 1997, IBM Watson won Jeopardy! against former champions Brad Rutter and Ken Jennings in 2011, and in 2016 Google's AlphaGo defeated Go world champion Lee Sedol. The victory by AlphaGo was particularly significant because, unlike chess, the possible moves in Go are infinite. The AI in AlphaGo was programmed with deep learning and neural networks model instead of a search tree model, and thus was able to teach itself to play and create its own moves (Byford, 2016). The victory demonstrated that AlphaGo was able to learn from every game it played and come up with original and unconventional moves that stunned even seasoned Go players.

Today AI is used in nearly every thinkable application and sector, from the voice assistants in our homes and phones, to the AI imaging programs that are assisting computer-imaging diagnosis

[1] While the Oxford dictionary defines big data as sets of information that are too large or too complex to handle, analyze or use with standard methods; the more commonly used definition of big data in AI is the three V's – Volume, Velocity, and Variety. In short, big data is an immense amount of data containing a wide variety of information that is increasing rapidly over time (Barocas & Selbst, 2016).

(CAD) to detect diseases and risk factors in healthcare (Ting et al., 2018). The definition and understanding of what constitutes an AI have also evolved over time, though there is not yet a definitive definition. Instead, AI is widely understood as machines that can carry out tasks that normally require human cognition such as decision-making, recommendations, speech and mobility, and predictive analytics that are beyond human capabilities (National AI Initiative, n.d.; *The OECD Artificial Intelligence (AI) Principles*, n.d.).

While the technology is expected to bring about transformative changes in nearly every sector, issues attendant with the deployment of AI and the potential risks have given rise to a growing urgency to establish robust governance frameworks to mitigate these issues and risks. Thus, the purpose of this paper is to provide policymakers with recommendations to navigate a future permeated with AI, while preserving sustainability in society and the environment. The policy recommendations will be based upon a comparative analysis of the governance regime that is emerging from key global AI leaders such as the United States, China, and the European Union. Because these three regimes wield great influence over the future of AI innovation and the global governance of the technology, understanding their governance context, ambitions, and motivations can help policymakers better navigate the developing field of AI governance and apply it within their own domains. Brundage and Bryson wrote,

> "The key question is not whether AI will be governed, but how it is currently being governed, and how that governance might become more informed, integrated, effective, and anticipatory" (Brundage & Bryson, 2016, p. 2).

Methodology

To fulfill the purpose of this paper of proposing AI policy recommendations, this paper conducted a comparative analysis of key AI governance regimes from China, the European Union, and the United States to discern their distinct approaches in addressing AI principles and mitigating the risks associated with AI. First, a critical review was conducted to identify the key AI principles from a corpus of governance documents generated in the past few years. Next, China, the European Union, and the United States were selected as the governance regimes to be examined, and a systematized review was carried out to analyze the current context of their governance initiatives. Within this step, a systematic search was conducted to collect all the relevant AI governance documents from these three regimes. For the sake of clarity and conciseness, only the most recent and comprehensive governance initiative developed by the main governing bodies from each regime were selected for the comparative analysis. Initiatives that were developed by think tanks, academia, or the private sector for specific applications or sectors were excluded. For instance, regulations and policies on data protection and privacy were excluded from the analysis even though data plays a huge role in AI. The main purpose is to compare how effectively and comprehensively each governance initiative addresses the AI principles, as well as their potential tradeoffs and implications.

The following documents were selected for the comparative analysis in this paper (European Commission, 2021; The National New Generation Artificial Intelligence Governance Specialist Committee, 2021; Vought, 2020):

(1) *Ethical Norms for the New Generation Artificial Intelligence* (hereinafter "Ethical Norms") published by the People's Republic of China Ministry of Science and Technology (MOST) in 2021,

(2) *A Proposal for Regulations in Artificial Intelligence* (hereinafter "AI Act") by the European Commission in 2021, and

(3) the *Guidance for Regulation of Artificial Intelligence Applications* (hereinafter "Guidance") released by the Executive Office of the U.S. President through the Office of Management and Budget.

Since 2016, there have been numerous efforts to regulate AI through the development of guidelines, strategies, and standards. However, the AI Act is noteworthy for being the world's first bid at a comprehensive AI regulatory framework (Circiumaru, 2021; Fjeld et al., 2020; Zhang et al., 2021). The AI Act is over a hundred pages long and includes detailed paragraphs under each article articulating the regulatory implications. Due to its length, only the Explanatory Memorandum section of the AI Act was included in the analysis (European Commission, 2021, pp. 1–16). Comparatively, the translation of the Ethical Norms was only six pages and the Guidance reached sixteen pages, and were thus included in their entirety for the comparative analysis.

Once the representative governance initiative was selected from each governance regime, they were then compared against the key AI principles using a matrix table (see Table 3 in Appendix). This was then followed by a discussion of the effectiveness of their regulatory strategies to help identify strength and gaps among the examined frameworks. Finally, based on the matrix table and discussion, a series of broad and amendable policy recommendations were developed to meet the AI principles, mitigate risks associated with the technology, and secure a sustainable and thriving future for humanity.

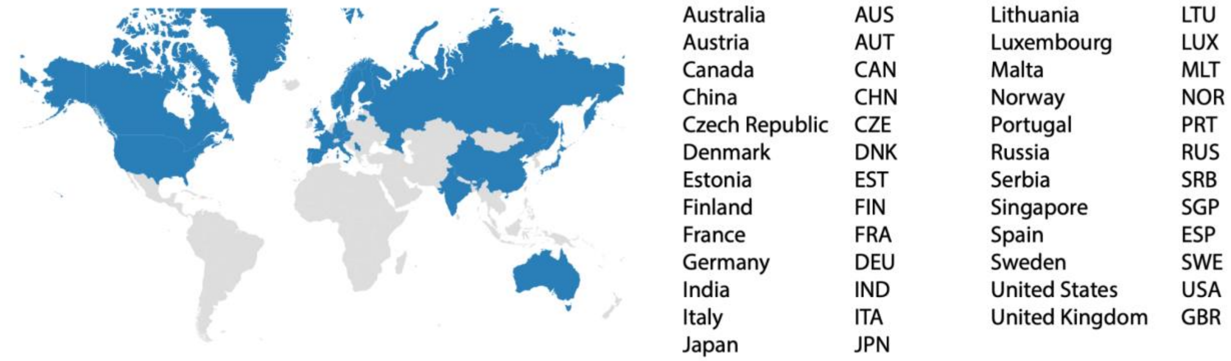
Limitations

There are limitations to focusing only on China, the European Union, and the United States, in the comparative analysis while omitting other countries and regions that have also been actively investing, developing, and deploying AI. The implications of examining only these three regimes are that the policy recommendations derived from this analysis will be to a degree skewed toward their unique realities, governance and cultural context. Despite their advances in AI as a whole, all three regimes are at varying stages of implementing an AI governance framework. Due to the diverse status of governance across these three regimes, there will be some policy gaps among them. Additionally, these documents are relatively recent and there has not been substantial data and evidence on their effectiveness and impact on societies and AI innovation.

The re-emergence of AI in the past decade implies that the national governance of the technology is still in its infancy. Van Berkel et al. identified only 25 countries with an existing national AI governance framework, which they defined as including national policies and strategies (see Figure 1).

Figure 1

Overview of The 25 Countries Identified with a National AI Policy or Strategy Frameworks



Australia	AUS	Lithuania	LTU
Austria	AUT	Luxembourg	LUX
Canada	CAN	Malta	MLT
China	CHN	Norway	NOR
Czech Republic	CZE	Portugal	PRT
Denmark	DNK	Russia	RUS
Estonia	EST	Serbia	SRB
Finland	FIN	Singapore	SGP
France	FRA	Spain	ESP
Germany	DEU	Sweden	SWE
India	IND	United States	USA
Italy	ITA	United Kingdom	GBR
Japan	JPN		

Note. From A Systematic Assessment of National Artificial Intelligence Policies: Perspectives from the Nordics and Beyond by van Berkel et al. (2020).

There is also an imbalance in the regions that AI governance research is coming from, with most literature originating from the global north that are also analyzing AI initiatives produced largely in the same region (van Berkel et al., 2020). This could have implications on the interpretation of and the weight given to the AI principles that are used as fundamental guidelines for policies and regulations. The meaning and significance behind each AI principle may not be universal and there may be nuances among different cultures. Subsequently, a similar perspective could also be applied to the weight and approaches implied in the policy recommendations developed in these documents.

Therefore, the policy considerations that were based on the analyses of primarily developed nations may not necessarily be adequate for developing economies. For instance, AI has the ability to assist developing nations leapfrog their healthcare system through applications in telemedicine as a way to mitigate healthcare worker shortages and extend healthcare to rural populations (Yayboke & Carter, 2020). To realize such a vision, policymakers in developing economies may have different goals and tradeoffs to consider compared to developed economies. In regions that are plagued by inaccessible healthcare, vulnerable to climate change issues, and threatened by geopolitical conflicts, risk levels and tradeoffs may be assessed differently to prioritize better health access, AI solutions for climate change issues, or advancement in military AI. Given the circumstances, developing economies may have the opportunity to envision a more disruptive strategy and sustainable outcome through the deployment of AI. Hence, the policy recommendations made in this paper are intended as a broad conceptual framework for policymakers to use as reference or as a starting point for developing AI regulations and policies.

While this paper attempts to address the transformative potential that AI will bring and its associated risks, it will not be able to address the myriad of benefits and risks emerging from different sectors and different applications of AI. There is also not necessarily a universal approach to mitigate a similar group of risks in different cultural and application contexts. For example, people from individualistic cultures place a greater value on their privacy and are more resistant to data collection, while people from collectivist cultures are more likely to disclose their personal information for the benefit of the community (Li et al., 2017). Facial recognition as verification on a single-person device may be less harmful than the use of facial recognition identification in a public space for surveillance purposes.

Finally, given that AI is a general-purpose technology, its impact has such a vast reach that this paper is unable to provide a comprehensive account of its full implications and thus it does not presume such a position. This paper simply intends to offer policy recommendations that can help shape the outcome of AI so that the benefits are maximized to help humanity thrive in the long-term future, and to minimize the risks associated with the technology.

Benefits of AI

AI is widely regarded as a general-purpose technology[2] and is expected to bring transformative benefits and engender a transitional impact comparable to the harnessing of electricity and the industrial revolution (Lynch, 2017). Computer scientist Andrew Ng has said that AI, like electricity when it was first discovered, will change the way the world operates, disrupting transportation, manufacturing, agriculture, and healthcare. Similarly, AI is poised to have an impact on social welfare, healthcare, domestic security, transportation, military, education, finance, and climate change.

> "Just as electricity transformed almost everything 100 years ago, today I actually have a hard time thinking of an industry that I don't think AI will transform in the next several years," Andrew Ng (Lynch, 2017).

Presently, AI has been deployed in applications to help us make decisions faster based on vast amounts of data, from recommending lifestyle choices; to deciding outcomes in recidivism sentencing, financial loans, and healthcare services. Its ability to identify patterns otherwise indiscernible to humans enables applications in medical imaging, weather forecasting, and traffic route planning (Yu & Alì, 2019). To provide a deeper understanding of the technology's applications and benefits, a few examples are expanded further below.

[2] Elhanan Helpman defined general-purpose technologies as technologies that have "the potential to affect the entire economics system and can lead to far-reaching changes in such social factors…" (Helpman, 1998).

Transportation

AI can be implemented in the transportation systems in many different aspects (Dixon, 2021). Autonomous vehicles including cars, freight trucks, and buses, can operate on various AI applications, such as camera sensors, radar detection units, navigational systems and more. Traffic infrastructure built with AI abilities will enable smoother traffic flows reducing congestion and carbon emission, while AI-enabled traffic planning can optimize public transportation routes. Through enabling automation, providing greater efficiency, and a reduction in human error, AI has the potential to reduce overall carbon emissions from transportation (which is the sector with the highest amount of carbon emission (US EPA, 2015)), increase transportation safety, and accessibility and independence among the vulnerable population.

Finance

With its ability to process massive amounts of data in a very short time, AI applications in stock exchanges have grown and connected market trading activities directly with individual users. Robo-advisors "create personalized investment portfolios, obviating the need for stockbrokers and financial advisers" (Popper, 2016). AI is also used in fraud detection with its ability to discover abnormalities in vast financial data and detect fraudulent transactions earlier (Allen & West, 2018). Other financial services that involve credit screening of potential customers, such as mortgage lending, loans, and insurance, have also benefited from the efficiency and decision-making aspects of AI. However, these applications have been heavily scrutinized for the biased and discriminatory outcomes that have occurred (Bartlett et al., 2019).

Environmental sustainability

In a 2020 study that surveyed the role of AI in achieving the UN Sustainable Development Goals (SDGs), the authors found that AI could enable 93 percent of the targets identified under the environmentally related goals that are SDG 13 climate action, SDG 14 below water, and SDG 15 life on land (Vinuesa et al., 2020). In terms of climate action, AI's ability to connect and process data from a wide selection of databases can help produce more accurate climate change models to help researchers understand and predict their likely impacts. AI applications have also been proposed and deployed to detect environmental harms such as oil spill, deforestation, and desertification, enabling authorities to plan, manage, and mitigate these incidents more efficiently. Additionally, the use of AI as the foundation of smart cities with carbon-efficient infrastructure, will also reduce overall carbon emissions.

The examples described above are just a small sample of the numerous AI applications that are already or will be deployed in these sectors and more. Due to AI being a general-purpose technology, it would be challenging to find a sector where AI would not be able to serve and enhance.

Risks of AI

Despite its potential to bring transformative benefits to humanity, the widespread adoption of AI has led to unintended consequences (Helpman, 1998). This includes concerns of a potential existential risk that could be brought about by a sudden growth explosion in AI that exceeds human intelligence and control (Bostrom, 2014). More presently, AI applications in policing programs such as the use of facial recognition to identify law offenders, the predictive analytics used in crime prevention and for identifying high-risk youth, have exhibited highly problematic biases due to the algorithms and training data in these programs (O'Neil, 2016). A few cases of AI misuse are highlighted below to illustrate the extent of damages and risks associated with the technology.

Data harvesting by Cambridge Analytica

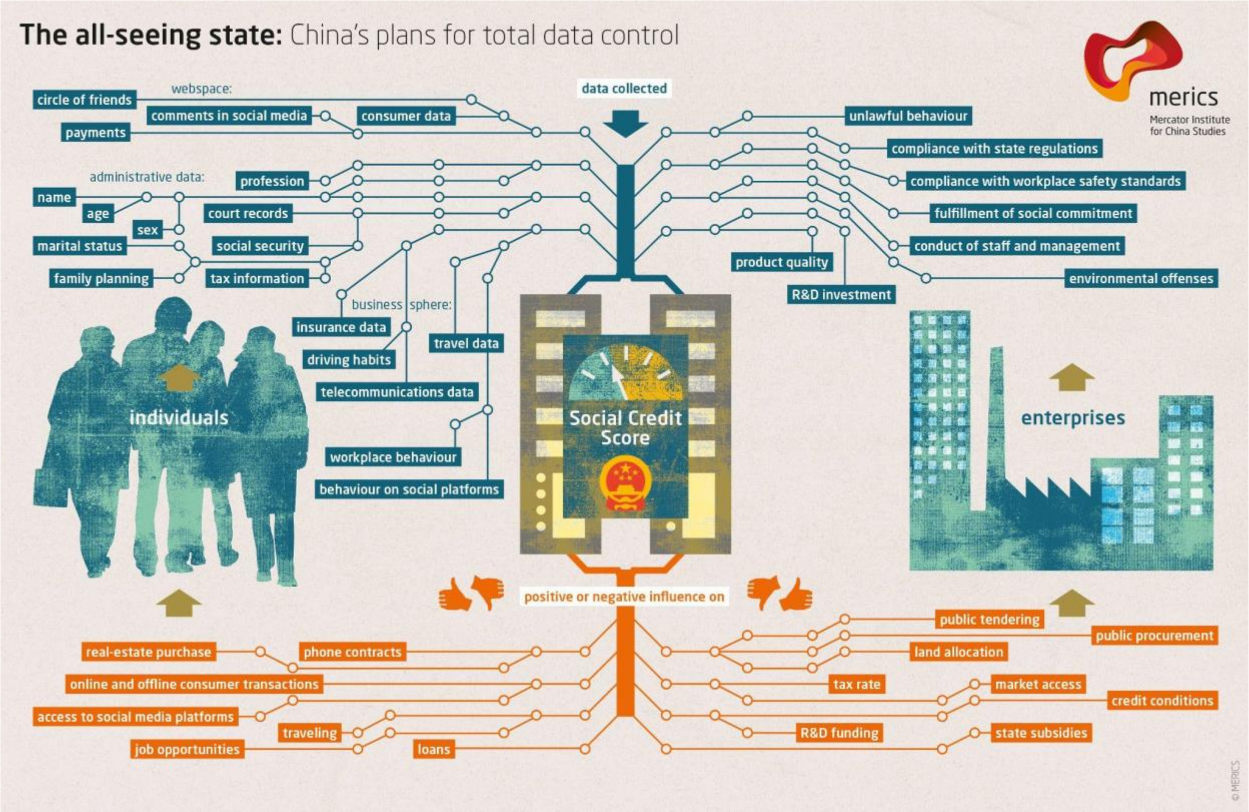
The 2016 US presidential election became a clear example of how unregulated AI usage can infringe upon democratic processes. Donald Trump's campaign team had hired Cambridge Analytica (CA), a now defunct British political consulting firm, as part of his 2016 presidential campaign strategy to influence voters through Facebook. CA developed an application that harvested Facebook user data and proceeded to create psychological profiles of the users using an algorithm that CA had developed. Based on these profiles, CA was able to create highly personalized ads to target vulnerable users and used fear tactics to influence them to vote for Trump instead of Hillary Clinton (Cadwalladr & Graham-Harrison, 2018). This incident underlined the importance of data governance in AI since the technology collects, generates, and thrives on big data (O'Leary, 2013). It also highlighted the absence of adequate data regulation in the United States that can adapt to the fast evolving nature of data and AI . Through the use of Facebook's platform, CA demonstrated how AI can be exploited to psychologically profile online users to target their vulnerability and use fear to influence their political decisions, effectively diminishing their civil liberties.

Social credit system in China

China's deployment of AI as an authoritarian governance tool has been critically scrutinized globally, especially by human rights defenders and civil society for violating human rights (Dragu & Lupu, 2021). Apart from having built an extensive network of surveillance cameras and using facial recognition to monitor its citizens, the Chinese government has rolled out a pilot program of social credit systems (see Figure 2).

Figure 2

Overview of China's data centralization strategy through the social credit system.



Note. Graphic designed by Mercator Institute of China Studies MERICS (Drinhausen & Brussee, 2021).

The program intends to track, regulate, and promote core socialist values[3] directly to citizens through their mobile devices (Gow, 2017; Liang et al., 2018). Critics argue this creates a highly advanced digitized panopticon[4] that can be extremely pervasive and heavily encroach on individual privacy and freedom, because the application can track and monitor every single transaction, interaction, and other behaviors a person performs on their mobile device (Chorzempa et al., 2018). Additionally, human rights defenders are concerned that these applications are being used to curb political speech and dissenting voices, which leaves activists and dissenters vulnerable to arbitrary persecution. The use of AI as a highly efficient and effective authoritarian tool has been cautioned by scholars, especially because of its prevalence in people's mobile devices (Ünver, 2018). This condition makes it easy and efficient for authoritarian governments to harvest big data on their people to map out their behaviors and preferences, enabling the government to directly monitor, influence, and potentially manipulate people's behavior.

Algorithmic biases in recidivism and healthcare assessments

Biases in AI systems have also led to discriminatory outcomes in applications used within high-stake sectors, such as criminal justice and healthcare. A ProPublica report in 2016 exposed algorithmic bias in a recidivism prediction application used in criminal justice (Angwin et al., 2016). The report revealed that the machine biased favorably toward white defendants, mislabeling whites as lower risk

[3] Core socialist values is a set of moral principles promoted by China's central authorities since 2012 that includes prosperity, democracy, civility, harmony, freedom, equality, justice, the rule of law, patriotism, dedication, integrity and friendliness.

[4] Contrary to the negative coverage in western media, the social credit system has received positive responses in Chinese social media, while other scholars have suggested that its capacity to create 'social trust' against the backdrop of the perceived moral decline in the country, could be a welcomed solution (Roberts et al., 2021).

more often than blacks. In contrast, black defendants were more likely to be mislabeled with higher recidivism at twice the rate as white defendants. The writers discovered that despite using the AI applications in their judicial decision-making, the Sentencing Commission did not conduct an impact assessment to evaluate the risk scores, which turned out to be only 20 percent accurate. Inaccuracy in AI applications used in high-stakes sectors can have an immense negative impact on people, and in this case, the outcome led to wrongful incarceration and infringement on individual freedom. The investigated cause of algorithmic bias has pointed to poor data quality in training datasets used in machine learning (Barocas & Selbst, 2016). Flawed methods employed by AI developers used for training these machines, such as data labeling and the use of proxies (Obermeyer et al., 2019). Data quality can be negatively affected by factors, including data collection methods and structural discrimination. In the case profiled in the 2016 ProPublica report, the collected data was inherently biased against blacks due to questions, such as *"Was one of your parents ever sent to jail or prison?"* and *"How many of your friends/acquaintances are taking drugs illegally?"* (Angwin et al., 2016). These questions may seem reasonable for evaluating recidivism rate, but in the context of the United States where the incarcerated population is disproportionately black, the data collected through this method were prejudiced against blacks. Thus, when machines were trained using prejudiced datasets they simply replicated the rule, drew prejudice inferences, and further systematically reinforced the discriminatory practices in society (Barocas & Selbst, 2016).

Equally important is the process of how machines are trained to develop the algorithms that would determine its intended outcomes. The choice of labeling data during machine learning can have a significant impact on the intended outcomes. In a 2019 study on racial biases in healthcare algorithms, researchers found that the AI system that was intended to identify high-risk patients with complex health needs for the purpose of providing them with adequate healthcare, was substantially biased against black patients (Obermeyer et al., 2019). Accordingly, healthcare cost was labeled as a predictor for health risk when training the algorithm. While healthcare cost and health needs are indeed highly correlated, the disparity in healthcare accessibility and discriminatory practices have resulted in a relatively lower medical cost for blacks. Thus, despite evidence showing blacks are more prone to health issues and at a greater severity – and therefore requiring greater healthcare needs – the algorithm produced biased predictions favoring whites for healthcare needs because historically their medical costs are relatively higher than blacks' (Obermeyer et al., 2019).

Progression of AI Capabilities

While the benefits and risks discussed above focused on existing AI applications, the discussion of risk and benefits in AI extends to the anticipated progression of AI capabilities in the future. The progression of AI can be categorized into three levels of abilities that are (Bostrom, 2014, p. 22; Pennachin & Goertzel, 2007, p. 1):

1. Artificial Narrow Intelligence (ANI) is sometimes known as Weak AI. Weak AI was defined as such for its ability to execute predefined tasks and make decisions within a narrowly defined scope, and is the AI currently deployed in operation.
2. Artificial General Intelligence (AGI) is AI that will be able to carry out tasks over a wide array of domains and achieve cognitive processes comparable to human abilities, and

3. Artificial Super Intelligence (ASI) also known as superintelligence, refers to AI that surpasses human intelligence in nearly all domains of interest and is likely to self-improve beyond the scope of human comprehension.

Bostrom et al. argued that when AI reaches superintelligence, it could potentially replace almost all human labor, including conducting scientific research and other inventive activities (Bostrom et al., 2018). Since most of the risks of advanced AI that will be discussed in this paper are largely based on work by Bostrom and Dafoe, the term superintelligence (which is the term they use) will be used to describe Artificial Super Intelligence.

AI Policy Areas

Inevitably, AI is a powerful tool that promises transformative benefits to humanity but there will also be certain risks involved in its usage, including those described above. These risks can be framed into policy areas based on the level of urgency, the rate they are occurring at, and their impact on people. The examples discussed in the previous section are some of the more urgent and high-impact risks that need to be addressed in the short- and mid-term. As AI abilities advances in the future there will be other risks involved, though these longer-term risks contain uncertainties in its level and magnitude.

Short- to mid-term

In order to better understand the foreseeable risks and concerns AI can bring in the short- to mid-term future, Calo proposed some key policy considerations in the following topics, though these should not be treated as a comprehensive list (Calo, 2017).

Justice and Equity. Due to underrepresentation in machine learning training dataset, algorithmic biases have led to discriminatory outcomes across numerous sectors including policing, finance, health, and even criminal justice (Allen & West, 2018). Policies should reduce these biases and ensure that the risks and benefits of AI are equitably distributed.

Lethal Autonomous Weapons (LAWs). While the definition of LAWs varies across countries, it is generally understood that they are fully autonomous and have the potential to kill human targets (Congressional Research Service, 2021). In 2019, the United Nations Convention on Certain Conventional Weapons (CCW), which has been one of the key international bodies examining the implications of LAWs, proposed eleven guiding principles on LAWs including the application of international humanitarian law to the development and use of LAWs. However, countries are still debating a preemptive LAWs ban leaving the question of ethical concerns, risks, and technological advancement unresolved (Congressional Research Service, 2021).

Safety. The overarching issue in safe AI links to explainability. When an AI system reaches a decision or executes an action that cannot be explained due either to its highly complex architecture as a result of recursive self-improvement, or a matter of proprietorship – it creates a "black box" model (Yu & Alì, 2019). The lack of transparency and accountability can become dangerous in machine automation, autonomous vehicles, and prosthetics, when an erroneous outcome harms humans and is unexplainable. The consequence of which complicates the question of accountability and liability as well.

Privacy and Power. AI operates on big data and its data intensive nature engenders concerns in privacy issues and the parity of power distribution. As illustrated in the Facebook and Cambridge Analytica case, user data was collected to build psychological profiles to target their vulnerabilities. With AI applications embedded in our mobile devices, vehicles, and voice assistants, highly granular details about our everyday lives can be gathered to reveal intimate behavior patterns that are otherwise indiscernible by human observation (Yu & Alì, 2019). Data that contains such information is invaluable in the hands of government, politicians, and advertisers, who can then use AI to manipulate the public for their own agenda and profit. Additionally, these data are presently collected by concentrated within a few large entities, resulting in a power imbalance that doesn't necessarily benefit the public (Calo, 2017).

Security. A group of distinguished AI experts from diverse disciplines and organizations published a report in 2018 on the malicious use of AI (Brundage et al., 2018). Based on the unique capabilities of AI, they anticipated several changes in the realm of cybersecurity, such as the expansion of existing threats, emerging novel threats, and attacks becoming more sophisticated in that they will be more effective, targeted, and difficult to trace. These attacks would threaten digital, physical, and political security.

Labor displacement. Concerns about labor displacement due to the advent of machines have been around since the industrial revolution. However, because of its expected exponential growth and transformative impact, AI's disruption to the labor market is expected to happen faster and affect nearly every sector of the economy (Brynjolfsson & McAfee, 2016). For instance, predictive models based on the U.S. labor force have suggested that workers in transportation and logistics industries, office administration, and production are at the highest risk of replacement by automation (Frey & Osborne, 2013). Other research suggested certain jobs, such as mid-level skilled jobs, would have partial tasks automated but not necessarily completely replaced by automation (Autor, 2015). In these cases, the AI applications would complement and enhance the human workers in their job roles.

Long-term

In the longer term, uncertainties in the complexity of AI systems compounded by its capacity for exponential growth "pose tremendous opportunities and risks for humanity" (Dafoe, 2018). One of the most widely quoted risks is the existential risk of an AI singularity in which the technology gains superintelligence and is not fully aligned with human values (Bostrom, 2014; Dafoe, 2018). To be clear, superintelligence can potentially bring profound benefits to humanity as well. With the proper guardrails in place to shape its development and deployment, we can better ensure a generally beneficial outcome. While the timeline of the arrival of superintelligence is still far from certain, a survey among leading AI researchers revealed that the majority of them expect it to occur within this century (Bostrom, 2014). Against this backdrop, AI researchers Dafoe and Bostrom hypothesized and cautioned against several potential extreme risks that could arise from advanced AI (Bostrom et al., 2018; Dafoe, 2018).

Robust totalitarianism. The use of AI in facial recognition, emotion detection, and its ability to track and analyze our extensive digital footprints[5], could enable it to be used as a highly accurate and targeted tool for behavior monitoring and manipulation. A tool of this magnitude can effectively shift and concentrate power in the hands of the elites, leading to robust totalitarianism.

Great power war. Even now, several countries are invested in developing autonomous weapons to upgrade their military power to obtain a critical warfare advantage, and to use them as a strategic conflict deterrence (Blasko, 2011). The availability of such an extreme first-strike advantage could present a higher possibility that powerful actors could order a preventive strike as a peacetime deterrence. Additionally, the crisis dynamic with advanced AI could become more complex, and depending on the level of automation, it could also lead to more rapid escalation, risking unintentional and unmanageable great power war.

Value misalignment. There have been instances where in an attempt to achieve its given tasks, AI inadvertently generated unintended consequences, such as chatbots becoming abusive and robots developing their own language.[6] While these scenarios are harmless in weak AI, an advanced AI that is not built with or aligned with human values could lead to significantly harmful outcomes and in extreme cases existential risk.

[5] Most Chinese citizens conduct nearly all their daily transactions and interactions through mobile apps that are validated through “real-name registration”. These apps are necessary for navigating their daily lives, whether it’s hailing a cab, scheduling a doctor's appointment, or ordering food at a restaurant. Therefore the user’s online and offline behavior are linked to their personal information. As a result, a comprehensive and detailed profile of each and every user (citizens in this case) is created and provides the government with an accurate profile of their citizens (Roberts et al., 2021).

[6] Microsoft developed a chatbot that turned abusive after 24 hours interacting with humans on Twitter. Facebook AI Research shut down an experiment after two AI agents that were supposed to simulate human dialog began communicating in their own language that was indiscernible to humans.

Value erosion from competition. While measures could be put in place to avoid the previous scenarios, there is still risk of value erosion as AI actors[7] increasingly prioritize attaining a competitive advantage against others. An AI race could instigate progressively harmful tradeoffs in the pursuit of gaining more power and wealth.

AI Governance

AI has many benefits and is expected to bring transformative changes across all sectors and all levels of society. It is therefore imperative that the uncertainties surrounding its development are not feared, but rather studied and understood, and that its risks are not underemphasized but adequately addressed. AI governance will be crucial in helping humanity navigate a future permeated by AI. Indeed, the purpose of AI governance is to maximize the power of AI to sustain a thriving global community in which resources and benefits are equitably distributed, while ensuring resilience against security threats. The arrival of a rapidly advancing technology has left many regulatory frameworks lagging behind AI development, although there has been an increasing response from researchers, governments, and international agencies to address the issues arising from AI deployment and to shape its future outcomes. The context of AI governance along with its challenges were examined from a selection of academic literature that studied various governance efforts to identify trends and limitations that have emerged within this field.

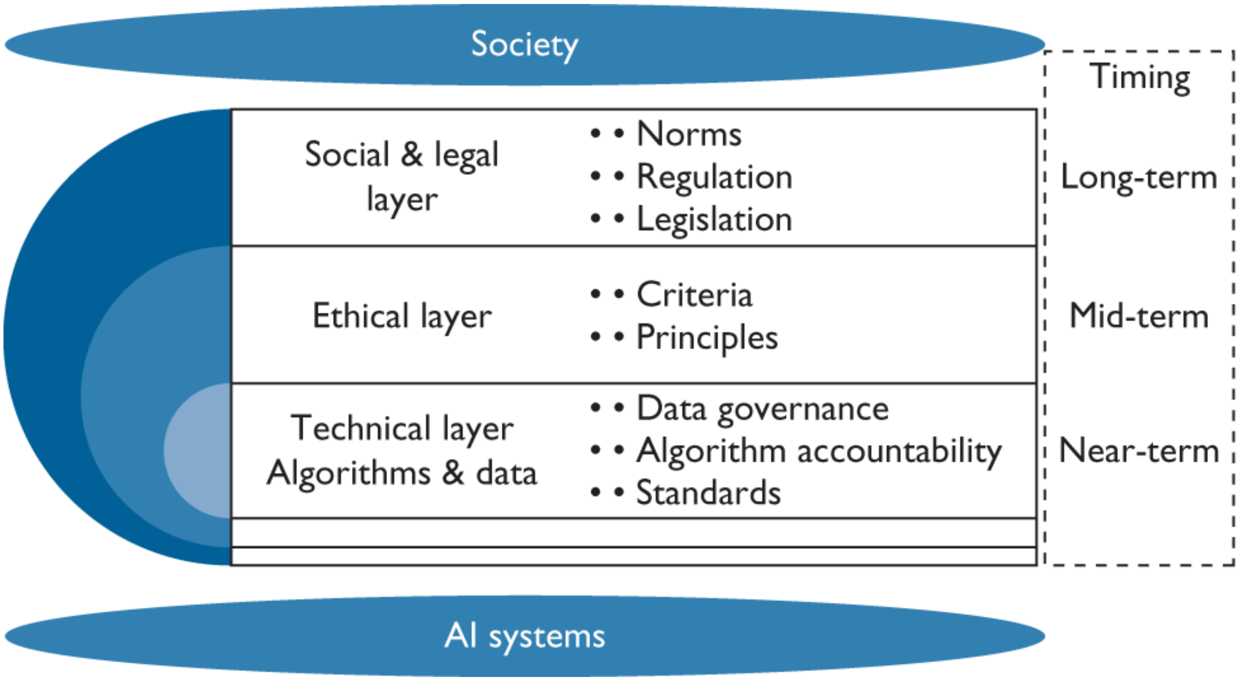
[7] UNESCO defines AI actors as "any actor involved in at least one stage of the AI system life cycle, and can refer both to natural and legal persons, such as researchers, programmers, engineers, data scientists, end-users, business enterprises, universities and public and private entities, among others" (UNESCO, 2021).

Existing regulatory frameworks, such as internet governance, space law, aviation safety, and the Chemical Weapons Convention, have been used and proposed as reference sources for developing and proposing global AI governance (Butcher & Beridze, 2019). For instance, McGregor et al. (2019) suggested international human rights law as an approach to address the gaps in algorithmic accountability proposals intended to mitigate the infringement of human rights caused by biased algorithms in decision-making AI (McGregor et al., 2019). The UN Convention of Certain Conventional Weapons has also adopted international human rights law as guiding principles for regulating Lethal Autonomous Weapons (LAWs) (Convention of Certain Conventional Weapons, 2019). International human rights laws consist of internationally protected rights that can be used as a guiding framework for AI actors to identify potential factors that could lead to harm. Policymakers can also apply this framework to develop regulatory requirements across the lifecycle of AI to keep the system compliant with protected human rights (McGregor et al., 2019).

To capture the complex nature of AI governance, Gasser and Almeida (2017) referred to the layered models used in internet governance to propose an AI governance model with interacting regulatory layers including social, legal, ethical, and technical foundations (Gasser & Almeida, 2017). The authors proposed situating the layered model between society and AI systems, whereby corresponding governance tools for each layer can be developed at different points of time. For instance, technical governance proposals such as standards setting and algorithm accountability principles could be developed in the near-term. While specific regulations for mature AI applications can be developed at the mid- and long-term (see Figure 3).

Figure 3

Layered AI Governance Model



Note. From Gasser, U., & Almeida, V. (2017). A Layered Model for AI Governance. IEEE Internet Computing, 21(6), 58-62.

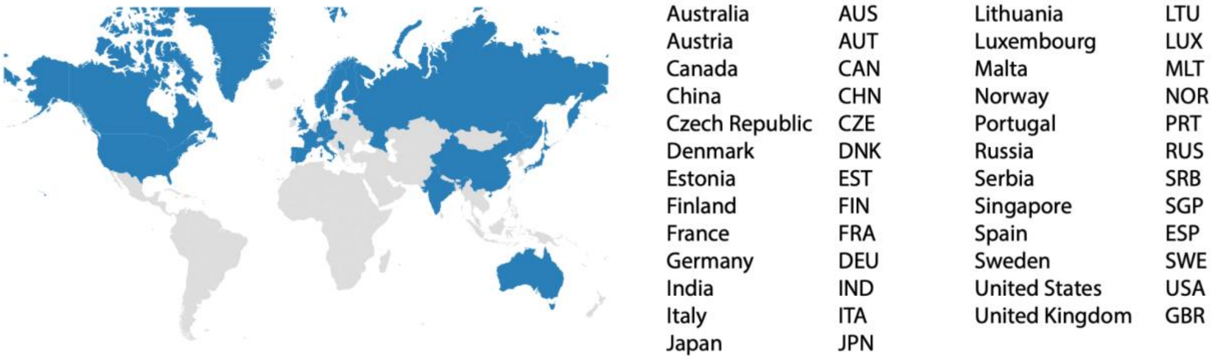
Nonetheless, layering policy frameworks have been criticized by Yoo (2013) who argued that modularizing clusters of tasks can reduce functionality and efficiency (Yoo, 2013). Instead, Yoo cautioned policymakers to adopt a more dynamic perspective that allows layered structures to change over time. As AI is further deployed in the public domain, it is possible that future governance structures may evolve to take advantage of the efficiency AI provides, which will also affect regulatory efforts across other sectors. Whether the most effective governance structure would take the form of a layered model, a top-down model as espoused by China, a centralized model similar to the European

Union, a sectoral approach such as the United States, or a combination of these options – warrants investigating in future research.

Despite the vast amount of studies on the topic of AI governance, an OECD report from 2019 identified only fifty countries—including the European Union—that have either developed or are in the process of developing a national AI strategy (Berryhill et al., 2019). The following year in 2020, a conference paper identified 25 countries that had successfully established a national AI policy or strategy framework[8] (see Figure 4) (van Berkel et al., 2020).

Figure 4

Overview of the 25 countries identified with national AI policy in 2020.



Australia	AUS	Lithuania	LTU
Austria	AUT	Luxembourg	LUX
Canada	CAN	Malta	MLT
China	CHN	Norway	NOR
Czech Republic	CZE	Portugal	PRT
Denmark	DNK	Russia	RUS
Estonia	EST	Serbia	SRB
Finland	FIN	Singapore	SGP
France	FRA	Spain	ESP
Germany	DEU	Sweden	SWE
India	IND	United States	USA
Italy	ITA	United Kingdom	GBR
Japan	JPN		

Note. A Systematic Assessment of National Artificial Intelligence Policies: Perspectives from the Nordics and Beyond (van Berkel et al., 2020)

Although these two papers show that countries are steadily making progress in developing AI governance within their territory, the number of countries that are actively engaged in this field remains relatively low.

[8] Defined by the authors as an official document issued by the national government detailing a national AI policy or strategy framework that applies to the entirety of AI and not just a specific application, such as autonomous vehicles.

The lack of an established AI governance framework, especially in countries that are heavily invested in its innovation, should be addressed swiftly to avoid policy lags that can lead to significant ramifications on societies. Such ramifications have been observed in algorithmic biases and the mishandling of user data, which demonstrated the severity of the adverse impact AI can have on human rights[9] and election integrity[10] (Angwin et al., 2016; Cadwalladr & Graham-Harrison, 2018). Indeed, it is urgent that the principles and mechanisms necessary to ensure a desirable outcome are promptly embedded while stakes are still relatively low, as compared to a future when Artificial General Intelligence is achieved and AI becomes even more intricately woven into society (Dafoe, 2018).

Given its far reaching impact and its deployment of diverse applications, the consensus among AI governance research is that framework development should involve a "holistic, multi-disciplinary, and multi-stakeholder" approach (Rossi, 2018). AI developers, users, policymakers, and advocacy groups in public-private-academic sectors are urged to promote robust collaborations when developing regulatory frameworks. Collaborative governance should also expand beyond borders because many AI-related issues overlap extensively and are not limited within national boundaries, some AI infrastructure are internationally connected, and global co-operations will be crucial to prevent a global AI race (Cave & ÓhÉigeartaigh, 2018). Establishing a global AI governance framework will require careful consideration for the plurality of stakeholders, legal systems, and

[9] COMPAS developed an algorithm for recidivism scoring that allegedly produced racially biased results (Angwin et al., 2016).

[10] Cambridge Analytica harvested user data from Facebook to create highly personalized election ads that targeted vulnerable users using fear tactics to sway votes for Donald Trump against Hillary Clinton (Cadwalladr & Graham-Harrison, 2018).

cultures to avoid introducing coercive norms and regulations. Gasser and Almeida (2017) advocated for a global AI governance system that "must be flexible enough to accommodate cultural differences and bridge gaps across different national legal systems" (Gasser & Almeida, 2017). Thus, the input from a variety of stakeholders will be critical for building a sustainable and effective AI governance framework.

Challenges in AI Governance

The complexities of the technology and the uncertainties that surround its future make it challenging for policymakers to design and implement effective governance. The abundance of AI principles, voluntary standards, ethical guidelines, and strategies easily eclipse the number of initiatives with enforceable mechanisms (Taeihagh, 2021). While self-regulatory "soft laws" are commended for their adaptability to the rapid development of AI, this approach essentially contains 'nonbinding norms and techniques' and are thus not enforceable (Larsson, 2020; Taeihagh, 2021).

The complexity of the technology has also translated into information asymmetries among different AI stakeholders (government, private, and individuals), resulting in challenges in complying with regulations and laws that are too vague, as well as insufficient technical literacy preventing users from exercising self-agency (Gasser & Almeida, 2017). The re-emergence of the technology in the last decade has been disruptive and its rapid progress has created a policy lag (that is likely to continue for the same reasons).[11] Existing governance frameworks have been inadequate for remedying the societal

[11] The lack of awareness and understanding of the challenges posed by disruptive technologies is a problem for regulators. For example, large technology companies such as Alphabet, Meta, and Amazon have amassed significant information and

problems that arose and regulators are further constrained by their limited knowledge on the technology (Taeihagh, 2021).

Tech companies have also been observed to have expanded their influence and power through their involvement in proposing AI principles that contribute to AI governance frameworks. These conglomerates have been lobbying their framing of AI issues and policies through participation in multistakeholder AI expert groups commissioned by governments (Cath, 2018). Extending corporate interest into the regulatory domain can create an imbalance of interest in regulatory frameworks that favor the interest of technology companies. This is cause for concern as the structural nature of AI has been concentrating resources, wealth, and power in the hands of technology companies more than ever before. Thus, a regulatory framework that is dominated by corporate interests is less likely to distribute the benefits and resources generated by AI equitably.

There are also structural challenges that the technology poses to governance, specifically in the area of spatial jurisdiction and sectoral constraints. AI has a vast reach across sectors and borders and its diversified applications have cross-cutting regulatory implications (Gasser, 2017). For example, the European Union General Data Protection Regulation (GDPR) claimed Google Analytics activities in its region violated its conditions because Google Analytics communicates with US based servers (Stupp, 2022). The GDPR approach for data localization rules will have implications for other European Union companies using US based cloud services (Lomas, 2022). Meanwhile, technology has expanded traditional applications such as phones to smartphones that now encompass myriads of

resources that placed them in an advantageous position in regulating AI over governments, who used to hold the traditional role of distributing and controlling resources in society (Guihot et al., 2017).

applications, and these technologies are likely to evolve even further with AI. The emergence of new AI applications will likely push the definition of application-specific laws even further, as demonstrated in a 2018 case when the U.S. Federal Bureau of Investigation seized the cell-site location information (CSLI) of an individual's cell phone through the provisions of the Stored Communications Act, which was an act that was enacted in a time when phone data was simply communication data. Consequently, the CSLI enabled the FBI to track the person's whereabouts over a long period of time, leading to debates on whether such practice violated a reasonable expectation of privacy (*Carpenter v. United States*, 2018).

It would be interesting to explore what a future governance regime would look like – ideally one that can provide the guardrails needed for AI to advance in a way that complies with AI principles, while providing a conducive environment for sustainable progress.

Why AI Principles?

While AI governance initiatives with enforceable mechanisms such as regulations and policies have only begun to emerge in the last couple of years, there has been a wealth of AI principles and ethical frameworks generated in the past five years by various stakeholders from academia, and the private and public sectors (Cath, 2018; Fjeld et al., 2020; Jobin et al., 2019). Though non-binding in nature, these principles and ethical frameworks proffer an abundance of well-researched guidelines that can be used as the foundation in the development of an AI governance framework. This approach was echoed in the findings by van Berkel et al. (2020), whereby they found a strong overlap in the frequency of ethical principles being discussed in national governance documents, suggesting a movement of AI governance framework that is being built upon a foundation of AI principles (van Berkel et al., 2020). Indeed, Raji et al. (2020) proposed situating AI principles as the standard for evaluating the development of AI lifecycle[12] and in internal audits when formalized guidelines are not available (Raji et al., 2020). By codifying compliance with AI principles into a risk analysis framework, this proposed method essentially implements the principles into practice.

Several papers have analyzed and distilled numerous AI principles documents into a few key topics, with some variations in the chosen representative words (for example, non-maleficence, humanity, beneficial, and freedom represent the principle that AI should have a positive impact on humanity). The literature considered for determining the scope of AI principles in this paper, included Zeng et. al (2022) and Fjeld et al. (2020), Floridi & Cowls (2019), Greene et al. (2019), and Jobin et al. (2019). In the end, the website *Linking AI Principles* (https://www.linking-ai-principles.org/) created by Zeng et al. based on their 2019 paper was selected as the main source of reference as it was the most comprehensive and up to date. On the website, Zeng et al. collected and analyzed a corpus of AI principle documents from 2016 onwards. First, they manually selected the core terms, and then using a natural language processing algorithm they identified and distilled down the keywords for each core term. The result is an overview of the key topics and their related keywords as shown in Table 2 (Zeng et al., 2022).

[12] AI system lifecycle phases involve: i) 'design, data and models'; which encompass planning and design, data collection and processing, as well as model building; ii) 'verification and validation'; iii) 'deployment'; and iv) 'operation and monitoring'. These phases often take place in an iterative manner and are not necessarily sequential (OECD AI Policy Observatory, 2019).

Table 2.

Topics and Keywords for AI Principles

Topics	Keywords
For Human	for human, beneficial, well-being, dignity, freedom, diversity
Fairness	fairness, justice, bias, discrimination, prejudice
Transparency	transparency, explainable, predictable, intelligible, audit, trace
Privacy	privacy, data protection, informed, control the data
Safety	safety, validation, verification, test, controllability, human control
Accountability	accountability, responsibility
Security	security, cybersecurity, cyberattack
Share	share, equal, equity, power, distributive
Collaboration	collaboration, partnership, cooperation, dialogue
Sustainability	sustainability, environment, Sustainable Development Goals
Long Term AI	AGI, superintelligence, higher level AI

Note. AI principles topics and keywords based on Linking Artificial Intelligence Principles by (Zeng et al., 2018) with several additional keywords including diversity, environment, planet, Sustainable Development Goals, equity, power, distributive, and data protection. Keywords that were removed were education (under "for human") and confidential (under "privacy").

In this paper, the topics in Table 2 have been listed in the order from most mentioned to least. For example, the topic of "for humans" was most mentioned at 414 times in the documents surveyed by Zeng et al., while the topic of "long-term AI" was lowest with only 31 mentions over seven documents from a list of ninety (see Table 3 in Appendix). Moreover, some phrases were added to expand the keywords section to reflect the topics that are relevant in this paper. For example, diversity was included to cover the diversity of legislative parameters, culture, and context in human interactions. Distributive was included under share to highlight the need to ensure that benefits and risks from AI are equitably distributed across societies. Environment was included under sustainability to underline the importance of ensuring environmental sustainability throughout the development and usage of AI systems. The United Nations' Sustainable Development Goals (SDGs) were also included under sustainability to capture the globally endorsed mission to achieve "peace and prosperity for people and the planet, now and into the future" (United Nations General Assembly, 2015). A couple of keywords from the original table were removed because they were considered more of a policy area (education) or because its definition was already covered by other keywords (confidential). The following is a summary of each AI Principle.

For human. It could be argued that the one overarching topic in all AI research and discourse is how the technology will impact humanity. Among all the other listed AI principles, "for humanity" is perhaps the most prominent principle having the highest number of mentions at 414 counts (see Table 3 in Appendix). This principle urged for AI systems to be compatible with human values, be beneficial to humanity, and uphold human rights and diversity in humanity (Zeng et al., 2022). Diversity was added under this principle to highlight the diversity of culture, knowledge, and the rich context of human interaction that should be preserved in our interactions with AI. In developing a technology that aims to simulate human cognition and interaction such as AI, it is important to be aware of how AI will affect our humanness. According to Dick, humanness is an evolving concept that is relatively defined by our interactions, and in this case in relation to AI (Dick, 2021). Special attention should be given when developing AI systems especially in understanding the implications of where and how training data is acquired, and the algorithms and rules that are developed based on these data, which will determine the outcomes of AI systems. For example, algorithms developed within the context of the Chinese authoritarian governance structure may not benefit a democratic context. On the other hand, AI systems trained with WEIRD (Western, Educated, Industrialized, Rich, and Democratic) data may not yield optimal results in a developing economy in the global south. Thus, AI actors should be cognizant of their interpretation when defining the principle of "for human" and to avoid imposing a universal definition across humanity, because as Katz claimed "Like whiteness, AI aspired to be totalizing to say something definitive about the limits and potential of human life based on racialized and gendered models of the self that are falsely presented as universal" (Katz, 2020, p. 10).

Fairness. The next most cited AI principle is fairness with 374 mentions across the surveyed documents (see Table 3 in Appendix). Fundamentally, AI systems should prevent discriminatory outcomes, promote social justice and fair competition, and ensure inclusive access to the technology. Fairness came into focus as AI applications are increasingly being used to assist decision-making processes related to recidivism, recruitment, finance, insurance, and medical care (Angwin et al., 2016; Bartlett et al., 2019; Dastin, 2018; Obermeyer et al., 2019; Prince & Schwarcz, 2019). Considerations should be given when defining fairness in AI systems that are built upon algorithms devoid of context and culture, since fairness is a complex and multifaceted concept that depends on context and culture (Bennett & Keyes, 2020).

Transparency. Transparency in AI systems covers two major themes that can be understood as explainability and the right-to-know. Most of the documents surveyed insist that AI should avoid "black box" scenarios, which are AI models whereby the decision process cannot be explained or traced, either due to proprietary concerns or the design of the system (Yu & Alì, 2019). Transparency in this context has legal implications in cases where decisions that are reached by an AI system have discriminatory outcomes or result in safety issues (autonomous vehicles). With transparency, humans can maintain oversight and control of the systems. People should also have the right-to-know when and how AI is being used, such as AI applications in government settings or when one is interacting with AI. The right-to-know also extends to users being informed on how their data is being collected, stored, and used.

Privacy. The privacy of human users should be protected in the application of AI. AI's nature to operate on big data has pushed the need for better privacy oversight to uphold international human rights as enshrined in various international standards (OHCHR, n.d.). The main focus in data privacy has been on the handling of user data through "consent" and "control". The European Union General Data Protection Regulation (GDPR) and China's Personal Information Protection Law (PIPL) are examples of comprehensive data protection regulations; however, these initiatives will not be discussed rigorously within the scope of this paper. Equally important is the use of AI as surveillance tools that can invade individual privacy and violate international human rights law. Apart from the Australian government, the European Commission, UNESCO, Dubai, Google, and University of Montreal, this particular aspect was not widely discussed among the surveyed documents on *Linking AI Principles* website perhaps because many (176) countries are actively deploying AI for surveillance purposes (Feldstein, 2019).

Safety. Safety in AI systems should be safeguarded particularly in the application of safety components, robotics, autonomous vehicles, and prosthetics. The white paper on AI principles by Fjeld et al. succinctly defined safety as the proper internal functioning of an AI system and delivering its intended outcomes to avoid harm (Fjeld et al., 2020). Ensuring safety in these applications prevents harm from occurring toward humans.

Accountability. Mechanisms to ensure responsibility and accountability in AI systems should be in place before and after the deployment of AI. Entities and individuals that are accountable for an AI system should be identified where necessary, while venues for addressing redress when a person or society has experienced adverse impact should be made accessible. The principle of accountability is especially significant for its role in supporting other principles by ensuring that they are adequately implemented in AI systems. Clear legal responsibility and a regulatory framework that can adapt to the evolving capabilities of AI will also be necessary.

Security. Though security and safety in AI are often discussed together, the security aspect of AI focuses on external threats toward AI systems, such as cyberattacks and data breaches. Risk of a security breach in an AI system can be of great concern in applications where the system is autonomous or handles large amounts of sensitive user information or is connected to safety components that when malfunctioning can harm human users. AI is essentially a highly efficient and scalable tool that is also extending beyond human capabilities, which produces discrete cybersecurity implications (Brundage et al., 2018). When used as an efficient and effective tool in cyberattacks, AI can enable fewer bad actors to carry out targeted attacks at a higher rate that are also difficult to trace. The new capabilities in AI (such as hijacking delivery drones for attacks) and its specific vulnerabilities (such as weaknesses in its goal definition) could also be exploited giving rise to novel threats requiring distinct security measures (Brundage et al., 2018).

Share. AI should be developed in a way that ensures equity is observed throughout its lifecycle, and that benefits and risks are equitably distributed. With its potential to convey resources (such as data to knowledge) to power with minimal effort, AI can easily concentrate wealth to a few large entities —especially to actors who deploy and operate the AI system (Calo, 2017). Therefore, the need to ensure that such power is equitably distributed will become even more important. In the longer term, when AI reaches superintelligence, a small fraction of a nation's gross domestic product could potentially translate into enormous economic growth, and at that point, considerations for magnanimous policy would become relevant (Bostrom et al., 2018).

Collaboration. Global collaboration has been promoted as one of the key AI principles to foster an all-benefiting and conducive environment for building a consistent global governance that can accelerate innovation based on shared priorities. More importantly, the push for prioritizing a collaborative global AI ecosystem can also counter the threat of an AI race. An AI race could risk overlooking or choosing tradeoffs over AI safety precautions, which can have a profoundly adverse impact on societies since AI is prevalent in nearly all aspects of our lives and negative impacts can be amplified (Armstrong et al., 2016). Cave and ÓhÉigeartaigh believe that an AI race could also increase the risk of real conflicts leading to military arms races involving Lethal Autonomous Weapons (LAWs)

(Cave & ÓhÉigeartaigh, 2018). Cooperation and partnership among public-private-academic is also encouraged to foster a stronger adoption of AI principles throughout the lifecycle of AI, architecture, and ecosystem, and to accelerate the progress of AI innovation. For instance, collaboration in standard setting, developing ethical frameworks, and developing solutions for global challenges such as climate change.

Sustainability. The Sustainable Development Goals (SDGs) was added to this principle because it brings a more comprehensive approach to the principle by encompassing sustainability. The Sustainable Development Goals promote thriving economies, advancing health and education, reducing inequality, tackling climate change, and helping to preserve the environment (United Nations General Assembly, 2015). AI is situated as an efficient and powerful tool that can produce solutions to address these goals, however, it is equally important to note the difference between developing AI solutions that enable sustainability goals, and developing and deploying AI in a sustainable way (van Wynsberghe, 2021). For instance, Deep Learning models for natural language processing have been shown to require high energy demand, which could generate higher environmental cost as long as renewable energy sources are not readily available (Strubell et al., 2019). Such environmental costs should therefore be carefully evaluated along with the advancement of AI. On top of the Sustainable Development Goals, the Doughnut economy theory proposed by Kate Raworth — which illustrates certain planetary resource "boundaries" — could also be used as a potential guideline for determining sustainable development in AI (Raworth, 2017). Several cities (Sydney, Melbourne, Berlin, Brussels, and Amsterdam) have begun implementing the Doughnut framework to transform their economies into more sustainable models. Thus, when deploying AI in social infrastructure and public services within cities that are adopting the Doughnut model, the economic model can be incorporated as part of the guiding framework for ensuring sustainable design in AI. Lastly, since the foremost AI principle is explicitly centered on humans, policymakers should take extra consideration on ensuring the narrowly defined anthropocentric principle balances the ideology that humans and nature are in fact intrinsically connected (Jackson & Palmer, 2015).

Long-term AI. The consideration for long-term AI impact had the lowest presence among all the surveyed documents by Zeng et. al (see Table 3 in the Appendix) with only 31 mentions over seven documents from a total of ninety. Under this principle, AI actors should take into account the immense impact Artificial General Intelligence (AGI) and superintelligence could have on the future of humanity when developing the technology and its governance framework. For instance, policymakers and AI actors should ensure that powerful AI in the future is aligned with human values instead of simply programmed to achieve predefined goals (Russell, 2020). Past evidence has demonstrated the risk of AI diverging from human values in the process of accomplishing its tasks. Examples include chatbots turning abusive after interacting freely with humans, and "reward hacking" whereby AI systems discover a different method for reaching their goals with unintended consequences (Kim et al., 2021). Understandably, the more prevalent AI principles target the most urgent issues burgeoning from the deployment and advancement of the technology today. However, AI with a higher level of capabilities as those described in AGI and beyond will have a profound impact on humanity comparable to the agricultural and industrial revolution, thus guardrails should be instilled now to embed some level of control over its outcomes while the "stakes are still relatively low" (Dafoe, 2018; Karnosfsky, 2016).

Many of the AI principles are interconnected, dependent on other principles to succeed, as well as enable other principles to be realized. For instance, transparency in a prison sentencing AI system will enable it to be audited to ensure that its decision-making processes are fair. Hence, when developing AI policies and regulations these principles should form a holistic foundation and not be siloed into individual criteria to be met (Gasser, 2017).

Why China, the European Union, and the United States?

China, the European Union and the United States were identified as the current leading AI powers in the world, surpassing their peers in a series of progress indicators including number of AI talents, amount of research conducted, number of AI companies, adoption rate of AI systems, amount of data, and computing power (Castro & McLaughlin, 2021). Based on their position as AI leaders, the implicit assumption is that their status would also imply a relatively higher level of AI innovation, a wider extent of deployment, and a longer history of impacts brought about by the adoption of the technology. Based on this assumption, these three regimes would also have relatively stronger research in AI impact and in developing the governance framework to address it. Additionally, their position as leading global AI powers make them influential actors in the global arena of governance setting (Bal & Gill, 2020; Daly et al., 2020; S.1260 - 117th Congress (2021-2022), 2021). Consequently, their approach to AI governance will have a significant influence over the emergence of AI governance globally.

AI Governance: A Comparative Analysis Across China, the European Union, and the United States

When analyzing the three governance initiatives from China (Ethical Norms), the European Union (AI Act), and the United States (Guidance), the AI principles that were most widely covered were For Human, Privacy, Transparency, and Safety. Additionally, Safety was the most extensively covered by both the European Union and the United States (Table 3 in Appendix). China expectedly had some of the briefest guidance in its description due to the nature of the document (nonregulatory) and its length in general. However, in the case of Accountability, Security, and Sustainability, China either matched with their Western counterparts or exceeded in their considerations and guidance for these principles. The principle of Long-Term AI was not covered in any of the documents, suggesting the focus of current regimes are on the most urgent issues rather than the future risks of Artificial General Intelligence and superintelligence.

There is currently no standard definition of AI and the technology is defined in similar variations between the European Union and the United States. China, on the other hand, does not seem to have an official definition for AI, as a search for it revealed no results. Whether there is a strategic ambiguity in China's non-definition of AI is unclear. The European Union defines AI as an "Artificial Intelligence system" (AI system), which means software that is developed with one or more of the techniques and approaches listed in Annex I[13] and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with" (European Commission, 2021). The United States has a similar understanding of AI and defined it in the National Artificial Intelligence Act of 2020, 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. Artificial intelligence systems use machine and human-based inputs to – (A) perceive real and virtual environments; (B) abstract such perceptions into models through analysis in an automated manner; and (C) use model inference to formulate options for information or action" (Johnson, 2020). The similarities between the European Union and the United States definition of AI underpin their partnership in global AI alliances such as in the OECD and the Global Partnership for Artificial Intelligence (GPAI). However, the similarity between these two regimes diverges when it comes to their approaches to regulating and developing AI. How this discrepancy will affect their global partnership remains to be seen as the alliances are still relatively recent and in its initial stages.

Regardless, understanding how AI is defined and not defined by these different regimes lends an important lens to analyzing their reasoning and motivation in their governance approaches, the scope of their regulatory effort, and the potential implications from the strategies they chose.

[13] The AI Act intentionally made the list of techniques and approaches under Annex I amendable to accommodate the evolving technology. The list currently consists of: (1) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods, including deep learning; (2) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems; (3) Statistical approaches, Bayesian estimation, search and optimization methods.

For Human

The AI Act centered the European Union Charter of Fundamental Rights (mentioned 33 times) as one of their grounding principles in regulating AI. "It is in the Union interest… to ensure that Europeans can benefit from new technologies developed and functioning according to Union values, fundamental rights and principles" (AI Act, p. 1). With a strong focus on fundamental rights, the European Union declared that its AI regulations are human-centric, thereby framing its regulatory framework largely on AI's impact on humans. This had clear implications on its risk-based approach that squarely focused on the technology's harm to humans. For instance, practices of AI systems that violate fundamental rights were categorized as unacceptable risks and thus prohibited, with effective redress for affected persons proposed if infringements still occur.

The Guidance stated that AI regulations should promote American values, including freedom, human rights, and human dignity. The Guidance urged agencies to select approaches that would maximize net benefits and to avoid a precautionary approach that could prevent society from enjoying the benefits of AI. This contrasts with the European Union's scientific and technological policy approach that is founded upon the precautionary principle when dealing with complex uncertainties (*The Precautionary Principle: Decision-Making under Uncertainty*, 2017). The Guidance, in congruence with the United States regulatory tradition, recommended employing cost-benefit analysis when considering regulations. The Guidance referred to Executive Order 12866 that defined benefits as "potential economic, environmental, public health and safety, and other advantages; distributive impacts; and equity" (Guidance, p. 5).

While the European Union framed its regulatory framework as human-centric, China recommended that all AI activities should be people-centered, suggesting an approach that is human-centered by design（以人为本）. A unique perspective that the Ethical Norms offered was the promotion of human-computer harmony（促进人机和谐友好）to achieve sustainable advancement of humanity and the natural environment. The European Union and the United States both perceived AI as a technological tool to be harnessed for maximum benefits, China on the other hand seemed to envision a more integrated dynamic between AI, humans, and the environment.

Fairness

When considering regulations or nonregulatory approaches, the Guidance recommended considering the issues of Fairness and nondiscrimination in the outcomes and decisions produced by AI applications, as well as whether the application may reduce levels of unlawful, unfair, or otherwise unintended discrimination as compared to existing processes. It also recommended agencies to be transparent regarding the impacts that AI applications may have on discrimination, such as clearly articulating the potential impact and how specific regulatory efforts will mitigate biases and risks.

The Ethical Norms urged the promotion of Fairness and Justice in AI activities, including upholding inclusivity, lawful rights, equal opportunity, and fair sharing of AI benefits. To achieve that, the Ethical Norms prescribed incorporating these ethics throughout the lifecycle of AI. Further elaboration on how these ethics will be incorporated at each stage of the AI lifecycle, however, was not indicated. Additionally, the Ethical Norms recommended rigorous ethical review when handling data and developing algorithms for AI systems to avoid biases and discrimination.

The AI Act drew on the European Union Charter of Fundamental Rights and existing European Union legislation on data protection, consumer protection, and non-discrimination and gender equality when addressing the issue of fairness and non-discrimination. The AI Act also offered specific recommendations for mitigating biased AI outcomes, such as testing data prior to deployment of AI systems, risk management, and incorporating human oversight throughout the lifecycle of AI. Additionally, AI-based social scoring conducted by public authorities for general purposes is prohibited, as it may lead to discriminatory outcomes.

Transparency

Transparency was linked with Security in the Ethical Norms, suggesting its potential value as a mechanism for enhancing security within AI systems. China recommended Transparency throughout the algorithm lifecycle to enable audits and verifications, and avoid black box situations where the outcome of an AI system cannot be traced or explained either because the system has recursively self-improved to a level of complexity that even their programmers cannot comprehend, or is a proprietary issue. Transparency is also encouraged in the form of user-awareness whereby users have the right to always be clearly informed when interfacing with AI, and the right to continue or discontinue such interactions.

The European Union placed different expectations on Transparency based on the risk levels determined in the AI Act. Strict requirements for transparency and traceability were applied to high-risk AI systems to mitigate the risks of harm to fundamental rights and safety. Non-high-risk AI systems were subjected to limited obligations similar to those imposed by China — that is to inform users when they're interacting with AI or when their personhood is being analyzed, in order to allow people to make informed choices in these situations. Transparency was also situated as a mechanism of enforcement for the AI Act, and for enabling effective redress for persons affected by AI outcomes through ensuring traceability.

The United States covered the concept of Transparency extensively in the Guidance, however, the recommendations for Transparency were mainly directed at practices by and communications from federal agencies rather than technological transparency in AI systems. Agencies are expected to transparently articulate potential impacts of AI applications, and to use transparency as a tool to increase public trust and understanding in the technology. Whether this approach was strategically chosen to protect innovation and intellectual property rights is uncertain,[14] however, it would correspond with the country's intention of boosting innovation and their competitiveness in the AI market. For instance, the Guidance explicitly recommended that "While narrowly tailored and evidence-based regulations that address specific and identifiable risks could provide an enabling environment for U.S. companies to maintain global competitiveness, agencies must avoid a precautionary approach that holds AI systems to an impossibly high standard such that society cannot enjoy their benefits and that could undermine America's position as the global leader in AI innovation" (Guidance, p. 2).

[14] In general, there is a call for greater transparency in AI systems especially in outcomes that have significant impacts on humans such as in recidivism assessment applications. However, these claims are usually "confronted with the observation that algorithms have proprietary nature and are protected under trade secret law" in the United States (Pedreschi et al., 2019).

Privacy

The principle of Privacy is predominantly framed through the lens of data protection across all three regimes. Both China and the European Union explicitly underlined the need to uphold data quality to ensure accuracy and robustness in AI systems to mitigate risks to privacy. The European Union pointed to other existing legislation that supports data governance and protection in the regime, including the EU General Data Protection Regulation (GDPR), which has become one of the key authorities in data protection (Bradford, 2020). The China Personal Information Protection Law carries a similar authority to the GDPR but perhaps it only came into effect in October 2021 – a month after the Ethical Norms was released – and was thus not referred to in the document.

The Guidance too recommended protecting reasonable expectations of privacy through regulatory or nonregulatory responses, depending on the risk, and drew attention to the many existing governance frameworks for privacy considerations. Additionally, because of the nature of AI that is data-dependent, the U.S. guidance discussed increasing access to government data and even suggested providing more granular data rather than aggregate data, to support the advancement of AI. While the Guidance maintains that any data handling must be consistent with current legal standards and policies, the implications of providing granular data in AI (machine learning) data training means providing more specific and potentially revealing information on users, which could increase the risk of privacy infringement.

Safety

The European Union risk-based approach was anchored in assessing AI risks to safety, health, and fundamental rights. By framing their regulatory efforts through the lens of risk management, the AI Act was fundamentally built upon the principle of Safety. TITLE III in the AI Act elaborated on the risk-based approach the European Union has taken that involved classification of AI risks, legal requirements, and governance and assessment. Special attention was given to high-risk AI systems by imposing stringent and mandatory requirements, such as high-quality data, documentation and traceability, transparency, human oversight, accuracy and robustness, pre-deployment conformity assessments, and registration of stand-alone high-risk AI systems to minimize risks. Observed in this approach is the use of the Transparency principle as a mechanism for ensuring safety in AI.

A risk-based approach was also recommended by the United States in its Guidance for determining regulatory and nonregulatory efforts in mitigating risk to safety. Instead of proposing a standard risk framework such as the European Union, the United States delegated the definition of risk to federal agencies and advised them to "determine which risks are acceptable and which risks present the possibility of unacceptable harm, or harm that has expected costs greater than expected benefits," which is consistent with their federal governance structure and reliance on cost-benefit analysis (Vought, 2020, p. 4). The Guidance also advocated for upholding safety throughout the AI lifecycle to promote public trust in AI. In contrast with the European Union's comprehensive effort to regulate (high-risk) AI, the Guidance appeared more restrained and cautioned against a precautionary approach as it could prevent society from enjoying the full benefits of AI and undermine the country's position as a global AI leader. Within the Guidance, the United States recommended assessing the magnitude and nature of potential impact should an AI fail or succeed, as a way to determine the appropriate regulatory or nonregulatory response, which echoed their fundamental regulatory principle that all activities involve tradeoffs (The Guidance, p.4).

The risk of malicious deployment and use of AI was cautioned too in the Ethical Norms, which categorically prohibits any AI that could endanger public safety whether intentionally or unintentionally due to failure to comply with laws and regulations, ethics, standards, and norms. Though not nearly as comprehensive as the recommendations proposed by the European Union and the United States, China urged strengthening risk prevention through greater research and launching a monitoring and assessment mechanism that could warn of potential risks. Additionally, as an overarching safety measure, the Ethical Norms explicitly outlined that humans should always have full autonomous decision-making rights in their interaction with AI, the operation of AI systems, and that AI is always under human control. The debate between human control and full autonomy is a focal point in the field of lethal autonomous weapons (LAWs), and this statement appears to reinforce China's position that LAWs should be under human control and not fully autonomous (Kania, 2020).[15]

Accountability

In the Ethical Norms, China insisted that humans are ultimately the responsible entities in the realm of AI. They called for clearly defined responsibilities throughout the AI lifecycle and to establish

[15] China's position against fully autonomous AI may be only limited to usage and not the development or production of autonomous AI systems, including lethal weapon systems (Roberts et al., 2021).

accountability mechanisms but did not offer any examples of such mechanisms. Meanwhile, the AI Act dedicated a subsection on Governance and Implementation that delineated monitoring and reporting obligations for both providers of AI systems, and the facilitation of an EU-wide database for registering stand-alone high-risk AI systems to keep them accountable.

The United States approach to Accountability involved transparency and public participation, which is in accordance with Executive Order 13563, "Improving Regulation and Regulatory Review," whereby regulations "shall be adopted through a process that involves public participation" (“Executive Order 13563 -- Improving Regulation and Regulatory Review,” 2011). Public participation was encouraged in the rulemaking process related to AI, and when AI uses information about individuals.

While all three regimes recommended approaches to instilling accountability in AI, their proposals were minimalistic and broad without specific recommendations on how individual AI actors will be held accountable to the design, deployment, and operation of the technology.

Security

One of the key elements considered in the Guidance was ensuring Security in the deployment and use of AI, and this principle was further extended to protecting national security. The Guidance recommended incorporating security throughout the AI lifecycle, including consideration for methods to provide systemic resilience against cybersecurity threats, and to prevent adversarial use by bad actors. Accordingly, regulatory and nonregulatory efforts should examine unique AI vulnerabilities, and take appropriate measures to protect national security. While there are existing voluntary frameworks specific to cybersecurity applicable to AI, the Guidance urged establishing standards that could further bolster the technical aspects of security.

Similarly, China espoused incorporating Security throughout the lifecycle of AI and to enhance its ability to resist interference. One of the methods suggested in the Ethical Norms for protecting system security was to facilitate an active feedback mechanism for immediate reports on security vulnerabilities, regulatory vacuums, and policy lags discovered during the use of AI. This feedback mechanism reflects the United States support for public participation, though the self-initiated feedback response proposed by China could facilitate greater monitoring and regulating to achieve a comprehensive governance framework.

Because the scope of the AI Act analyzed in this paper was limited to its explanatory memorandum section, there weren't any significant mentions of security besides a reference to Chapter 2 that resides in the full proposal. In that chapter, the legal requirements for high-risk AI systems are described, including recommendations for addressing accuracy, robustness, and cybersecurity.

Share

The only mention of Equality, which is a keyword in the principle of Share, in the AI Act was when it denoted that the fundamental rights contained within the European Union Charter must be protected in the deployment of and usage of AI, including equality between men and women.

The tradition of practicing cost-benefit analysis in the history of U.S. regulation provided an advantage to their recommendations for addressing the principle of share. The Guidance was the only document among the three surveyed that included clear guidance for considering the distributional effect from the benefits and risks generated by the deployment and usage of AI. Further, the Guidance cautioned against anticompetitive effects from the application of AI that can reinforce the power of a few market leaders, preventing market entrants from entering and thriving in the AI market.

Similarly, China too advocated for regulating both the market and equal sharing of AI benefits. China described in their Ethical Norms that AI benefits should be shared equally by all and to provide appropriate alternative AI products and services with respect to vulnerable groups and special groups to ensure that there are no barriers to the equal use and enjoyment of AI. When it comes to regulating the market, China chose a strong message that demanded strict compliance with and respect for regulations governing market entry, competition, and other market activities. Entities are prohibited from disrupting market order through data or platform monopolies.

Collaboration

All three regimes encouraged global collaboration and dialogue when developing AI governance frameworks with the objective of furthering one's own agenda, such as having a role in shaping the global norms and standards in AI, and promoting their own brand of AI innovation. Representative statements include the following:

> "Promote the formation of AI governance frameworks and standards that have a far-reaching consensus" (Ethical Norms, p. 4).
>
> "The proposal also strengthens significantly the Union's role to help shape global norms and standards and promote trustworthy AI that is consistent with Union values and interests" (AI Act, p. 5).

“Accordingly, agencies should engage in dialogue to promote compatible regulatory approaches to AI and to promote American AI innovation” (Guidance, p. 11).

The European Union structure is fundamentally based upon a regional collaboration, which subsequently informs its governance framework. Therefore, the European Union sees cooperation as crucial for ensuring a frictionless and successful implementation of the AI Act in the region and cautioned against countries developing their own national rules. “An emerging patchwork of potentially divergent national rules will hamper the seamless circulation of products and services related to AI systems across the European Union and will be ineffective in ensuring the safety and protection of fundamental rights and Union values across the different Member States” (AI Act, p. 6).

In the Guidance, the US quoted Executive Order 13609, "Promoting International Regulatory Cooperation " that urged collaboration with international regulatory cooperation when developing regulatory frameworks. This could point to the country’s participation in various multistakeholder initiatives, including the AI alliance in the Organization for Economic Cooperation and Development (OECD), the Global Partnership on AI (GPAI), G7 and G20 discussions, bilateral partnerships on AI research and development, and AI collaborations for defense.

Sustainability

China sees AI as a tool to help the country and its people achieve sustainable development both socially and environmentally. In the Ethical Norms, the Chinese government advised against pursuing quick successes and short-term benefits in the development of AI rather, to focus on healthy and sustainable development of the technology.

The right to a high level of environmental protection and improvement in the quality of the environment is included in the European Union Charter of Fundamental Rights. The European Union sees AI as a tool for protecting these rights by achieving environmental sustainability through the use of AI applications. The European Union cited AI's ability to improve predictions, and optimize operations and resource allocations, as AI's potential to mitigate environmental issues and deliver beneficial outcomes in high-impact areas in climate change.

The Guidance quoted the Executive Order 12866 recommendation that regulatory approaches should maximize net benefits including environmental benefits.

Long-Term AI

Considerations for long-term AI such as the potential impacts from Artificial General Intelligence (AGI) and superintelligence were not discussed in any of the documents surveyed.

Discussion: The Context of Governance in China, the European Union, and the United States

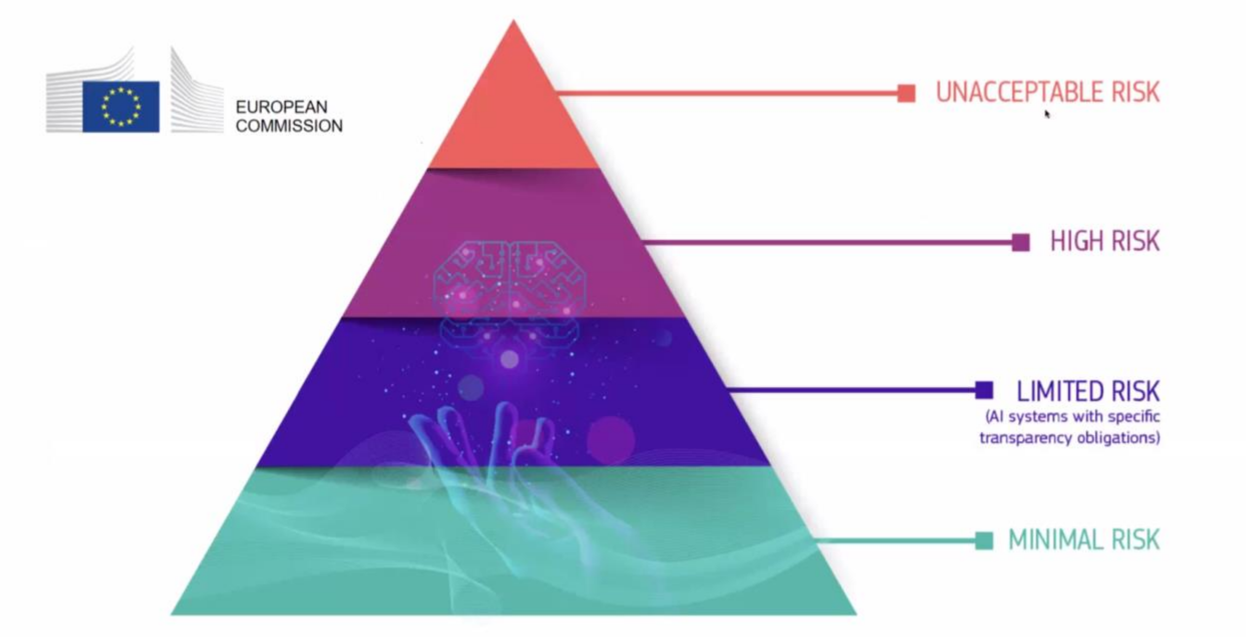
All three regimes have established numerous governance documents on AI that are application or sector specific that supplement the three documents examined in this paper. Notably, these include the EU General Data Protection Regulation (GDPR) that governs data protection and privacy that came into effect in 2018, the U.S. National AI Initiative that contains an extensive list of AI strategy documents from national and federal agencies, and China's recently released three-year road map for governing internet algorithms in 2021. Thus when surveying the comparative analysis of these three selected documents, it is important not to consider them in isolation but to understand that they are supported by an expanding governance framework. The distinct governance structures of the three regimes and their cultural context also informed their approaches to governing AI.

The European Union

The crux of the AI Act is its adoption of a risk-based approach that categorized AI systems based on their risk-level (see Figure 5):

Figure 5

The Proposed Artificial Intelligence Act of the European Commission (AIA)



Note. Image from the European Commission 2021.

1. Minimal risk – such as AI-enabled video games or spam filters are permitted for use with no restrictions.
2. Limited risk – such as impersonation bots, chatbots, and deepfakes are permitted for use with specific transparency obligations.

3. High risk – includes AI applications used in safety components, biometric ID, education, employment, access to private and public services, law enforcement, border control, justice systems, and democratic processes. The AI Act outlined comprehensive regulatory requirements for these AI systems.
4. Unacceptable risk – the use of AI in social credit scoring and behavior manipulation applications (including toys) are prohibited.

While the AI Act appears to determine AI risk-levels based on its potential harm toward the safety, livelihood, and rights of people, this risk assessment does not explicitly address harm against the environment and the scale of impact (European Commission, n.d.). For instance, deepfakes of authoritative figures can easily reach millions of users through social network platforms, resulting in a new and more convincing form of fake news. Based on these risk levels, the European Union proposed "proportionate measures" to mitigate the corresponding risks, including prohibiting practices of AI systems with unacceptable risk, imposing greater regulatory burden on high-risk AI systems, and minimal regulation on low-risk AI systems. As a consequence, the European Union admitted that the AI Act could potentially restrict certain freedom in businesses, art, and science, "to ensure compliance with overriding reasons of public interest such as health, safety, consumer protection and the protection of other fundamental rights ('responsible innovation') when high-risk AI technology is developed and used" (European Commission, 2021, p. 11). Stifling innovation as a tradeoff for adopting a precautionary approach has been critiqued as an outcome from the European Union's general attitude toward uncertainties in science and technology policies (Brattberg et al., 2020). Despite efforts to balance its precautionary approach through the use of regulatory sandboxes to test out low-risk AI innovation, experts are still worried that the strict regulations (such as the GDPR limiting data access, which is crucial for advancing AI) could continue to hold the regime back from achieving its ambition of becoming an AI innovation leader (Brattberg et al., 2020). Nevertheless, the AI Act is still a step in the right direction for the European Union to establish a coordinated regulatory effort across the region that has been otherwise plagued with fragmented attempts by member states. Additionally, the AI Act coupled with their influential GDPR sets the European Union apart from other AI regimes as a forefront actor in establishing "best practices, global standards and norms" that will help guide the future development of AI (Brattberg et al., 2020). Indeed, the European Union is determined to establish their own brand of AI that is "made in Europe" and distinguish itself from other global AI powers such as China and the United States by promoting strong ethics, human-centric design, and an AI ecosystem that's underpinned by the EU Charter of Fundamental Rights (Brattberg et al., 2020).

The United States

The United States on the other hand — though it provided similar recommendations to the European Union such as "proportionality" and "risk-based approach" — explicitly advised against precautionary measures at the cost of innovation and growth. The Guidance proposed reducing regulatory barriers that may harm the United States competitiveness, stating that "agencies must avoid a precautionary approach that holds AI systems to an impossibly high standard such that society cannot enjoy their benefits and could undermine America's position as the global leader in AI innovation" (Vought, 2020, p. 2). Against that backdrop, the United States continues to promote self-regulation and voluntary compliance in the governance of AI.

The United States has a long tradition of positioning cost-benefit analysis as a decisive tool when assessing the need for regulatory efforts (Sunstein, 2018). This method of monetizing benefits and determining the cost of risk to specify the corresponding regulation contrasts with the European Union approach of applying precaution on uncertainties. Whether this approach is effective in promoting greater AI innovation and deployment remains to be seen. However, it is worth noting that the United States currently boast the world's largest number of AI startups[16] and an AI adoption rate of 55 percent (in North America), which is slightly higher than Europe's 51 percent in 2021 (Zhang et al., 2022).

The United States governance structure is also distinct from the European Union and China in that the former regulates through a federal governance structure, while the latter two adopts a more centralized form of governance. The Guidance is literally a guiding document proposing recommendations for U.S. federal agencies on how to develop their own AI regulations, which implies that each agency will develop their own sector-specific regulations. Whereas the European Union and China expect the entirety of their regimes to follow the governance initiatives published by the central governing body. Nevertheless, recent directives from the United States government, such as Executive Order 13859 Maintaining American Leadership in Artificial Intelligence and the U.S. Innovation and Competition Act (USICA), clearly signaled the United States' intent on bolstering its global hegemony in AI.

[16] According to the AI Index Report 2022, the United States led with 299 newly funded AI companies in 2021, followed by China with 119, and the European Union with 96 (Zhang et al., 2022).

While ambitious, the manner in which the United States articulated its competitive intentions carried unmistakable "AI race" rhetoric. In the National Security Commission on Artificial Intelligence (NSCAI) final report released in 2021, the NSCAI urged the U.S. President and its Congress in "Defending America in the AI Era" and "Winning the Technology Competition" (National Security Commission on Artificial Intelligence, 2021). This report went on to inform the Intelligence Authorization Act, the U.S. Innovation and Competition Act, and other non-defense science and technology legislation (NDIA, 2021). Using an "AI race" framing was strongly cautioned against by Cave and ÓhÉigeartaigh, as they argued that the rhetoric alone could instigate an actual AI race, which could risk safety in AI development and escalate an AI arms race that can cause military conflicts (Cave & ÓhÉigeartaigh, 2018).

China

When China released the Ethical Norms, Sheehan suggested that because the document leans toward self-regulation, it diverges from the country's usual hands-on approach to regulatory initiatives (Sheehan, 2022). This however, does not necessarily imply that its nonregulatory nature wouldn't hold control and influence over AI actors developing, deploying and operating AI in China. Indeed, the Ethical Norms did not contain explicit enforcement mechanisms. The wording in the document (forbids 禁止, and strictly comply 严格遵守）however, suggested prescriptive directives rather than recommendations, which is consistent with China's paternalistic leadership and top-down governance (Fairbrother, 2013). China's emphasis on social harmony in its Ethical Norms ties back to their core socialist principles, and their intention to use AI as a normative tool for promoting social and moral governance (Roberts et al., 2021). The country has been struggling to patch up a "moral vacuum" left behind by the cultural revolution from the Maoist period and it has been the goal of the government to boost the country's moral integrity (Roberts et al., 2021; Yan, 2009). The Social Credit System is one such attempt by the government to regulate the private sector's activities and the social behavior of citizens.

Nonetheless, there are concerns that when a strong centralized governance structure enhances its influence through a powerful technology such as AI, it could result in a technological deterministic society or what some researchers have called "digital authoritarianism" (Polyakova & Meserole, 2019; Ünver, 2018). Although the Chinese government has outlined clear regulatory guidance, accordingly these legislations may be weak in implementation because of "the many loopholes, a ruling Party with legislative supremacy, and an influential government power that is not held accountable through democratic mechanism" (Roberts et al., 2021, p. 69). The use of AI-enabled governance over a population that generates an enormous amount of digital footprint, provides the government with an immense power to surveil, control, and potentially even manipulate the behavior of its citizens. Certainly, the Chinese government has been accused of genocide against the Uyghur population in the autonomous region of Xinjiang, and for using internal surveillance and tracking applications in mobile phones to monitor and oppress the Uyghurs (Human Rights Watch, 2019).

Preventing anti-competitive markets

The effort to curb the expanding power amassed by large tech companies was also observed in both the Ethical Norms and the Guidance. Despite still being at the initial stage of AI progress, tech behemoths have grown beyond governmental oversight at the cost of user safety and privacy (Calo, 2017; Taeihagh, 2021). Within the Ethical Norms, the Chinese government explicitly forbids AI to facilitate a market that enables monopolies, which the country has witnessed in the growth of companies such as Tencent and Alibaba. Perhaps as a response to curb tech companies' growing power, China released in 2021 a three-year road map for governing internet algorithms titled "Internet Information Service Algorithmic Recommendation Management Provisions," which is expected to have major effect on limiting the power of internet platforms, regulating algorithmic recommenders, and reinforcing consumer power (Toner et al., 2021). By targeting algorithmic recommenders and empowering users with greater control over how their personal data can and cannot be used, the Chinese government aims to restore market order and achieve a harmonious society (The National New Generation Artificial Intelligence Governance Specialist Committee, 2021; Toner et al., 2021). The United States shares similar sentiments and has also outlined recommendations in the Guidance to prevent anticompetitive practices and barriers that hinder new market entrants. While the algorithmic recommender regulation proposed by China may be considered restrictive and imposing in the context of the European Union and the United States, it still warrants studying and monitoring China's progress as the first country to regulate algorithmic recommenders (Toner et al., 2021).

Shaping global AI governance

The comprehensive AI Act reflects the regime's ambition to "help shape global norms and standards" in AI that is consistent with their values and interests (European Commission, 2021). The power to influence global AI governance is also an ambition shared by China and the United States. The China Electronic Standardization Institute has been playing an increasingly prominent role as a member of the subcommittee of the International Organization for Standardization, which develops international standards for the AI Industry (Bal & Gill, 2020). However, China's participation has raised concerns among the national security community in the United States, and in 2021 the U.S. Congress passed the Innovation and Competition Act to directly counter China's progress in AI and its growing influence as a global AI authority (S.1260 - 117th Congress (2021-2022), 2021). In the past, the European Union has spearheaded regulatory efforts in data governance with the introduction of the EU General Data Protection Regulation (GDPR) that has since become the benchmark for data protection in nearly 120 countries (Bradford, 2020). Whether or not the AI Act would achieve a similar influence as the GDPR remains to be seen. Although as one of the first comprehensive AI regulatory frameworks in the world where so few countries and territories have even begun to establish an AI governance framework, the AI Act may benefit from the first-mover advantage (van Berkel et al., 2020). The European Union has the potential to establish its regulations as the global standard, giving itself a strategic advantage to collaborate more easily with countries and regions that adopt compatible protections, and effectively situating the region as a destination for AI innovation and businesses that are invested in safety and human rights.

Overall, all three regimes are positioning themselves as leading global AI powers through their own distinct strategy. The European Union espoused safety and human rights as key elements to the AI brand they're building, while the United States has placed a strong emphasis on innovation and optimization of AI benefits. On the other hand, China's goal to build a harmonious society by harnessing the benefits of AI tends to be overshadowed by its distinctively paternalisti c approach to control and regulate. Nevertheless, China plays a key role in the international standard setting and development of AI governance. Thus, it is important to "understand their (China) needs, ambitions, and motivations – and not just from a Western perspective" (Roberts et al., 2021).

Policy Recommendations

The following policy recommendations were developed based on the comparative analysis above. By borrowing from the governance initiatives produced by the three regimes, the most effective and complete proposals were selected to build the policy recommendations below. The chosen approach for this paper was to keep the policy recommendations broad and amendable. Because AI is going to affect nearly all sectors and all walks of life, there is a need to consider the "big picture of what this will mean for ethics, governance, and societal impact" (Allen & West, 2018). Amendable recommendations will also be more accommodating toward the evolving nature of AI, the differences in culture, legal systems, governance structure, and progress in AI deployment. Thus, specific recommendations targeting each stage of the AI lifecycle or specific sector and application, for instance, will not be discussed as such recommendations may differ depending on the different factors mentioned above.

Implement a centralized AI governance framework

Given the transformative impact AI has and will continue to have on society, a key condition for ensuring that all AI principles are effectively incorporated throughout the development and regulation of AI is to lay the foundation for a unified and comprehensive governance framework. Policymakers must implement an overarching, centralized governance framework that establishes key definitions and regulatory approaches, similar to the AI Act. A centralized governance framework is crucial for enabling interconnected AI systems to operate smoothly, and to better facilitate AI innovation across platforms. AI, like electricity, will be prevalent in nearly every sector and applications (Lynch, 2017). Thus, having an overarching governance framework that establishes a consistent standard across sectors, local borders, and agencies will be highly advantageous for accelerating and regulating the deployment and advancement of the technology.

The central governance framework should adopt a risk-based approach that is founded on harm against humans and the environment including the scale of impact, to categorize AI systems accordingly and enforce proportionate measures for regulating different levels of AI risks. By establishing clear risk levels, AI can be regulated proportionately according to its risk of harm without stifling innovation—lower risk AI can enjoy more innovation and application freedom, while higher risk AI will be more highly regulated to avoid producing harm. The method for determining risk levels should be treated with great care for AI applications with dual-uses—such as facial recognition for identification and verification—to monitor the evolving and emerging application of AI, and to avoid over- or under-regulating.

This governance framework will determine the risk assessment parameters for labeling AI risk levels that can then be used by other agencies and business entities for developing their own specific AI regulations. The centralized framework can be understood as a minimal requirement for regulating AI, and by providing a clearly defined risk-based model it can foster greater regulatory consistency across sectors. Additionally, the uncertainties regarding longer-term risks and benefits must be addressed using a precautionary approach because AI has a tendency to amplify its impact in intensity and scale, which could cause great harm to humanity (Armstrong et al., 2016).

Finally, the governance framework must be human-centered, as well as environmentally sustainable. While the Ethical Norms proposed human-centered AI design and the AI Act was framed to regulate AI harm against humans, a centralized AI governance framework must explicitly express that both regulatory efforts and AI development must be human-centered, as well as environmentally sustainable. This implies that regulations must incorporate or be founded upon relevant human rights laws (such as international human rights laws), Sustainable Development Goals, or the Doughnut economic framework (McGregor et al., 2019; Raworth, 2017; United Nations General Assembly, 2015). These initiatives can be used as compliance frameworks when testing AI systems prior to deployment and during operations. Furthermore, human-centered AI design must preserve the rich context of human interactions and AI actors must be cognizant of how AI is redefining humanness (Dick, 2021). To achieve this, the design and development stages of AI must include cross-disciplinary expertise from social sciences and humanities.

Establish robust data protection regulations

To supplement a centralized AI regulatory framework, policymakers must establish a comprehensive and unified data protection governance regime to uphold individual privacy and encourage safe and secure collection, storage, and use of data. The China Personal Information Protection Law (PIPL)[17] and the EU General Data Protection Regulation (GDPR), for instance, empower users by giving them the right to consent and providing transparency on how their data is

[17] Even though China's PIPL has clear provisions regarding how individuals and organizations handle the means of data processing, it is unclear how the specific provisions for the Chinese state government will impact user data protection (Lee et al., 2021).

collected, processed, and used. The United States, on the other hand, has data protection laws that are sector- and state-specific, such as the Health Insurance Portability and Accountability Act, California Consumer Privacy Act, and Children's Online Privacy Protection Act. The segmented approach, however, could potentially be disadvantageous for protecting data in an age percolated with AI and big data. As each law is limited to its own domain, gaps and inconsistencies can appear across sectors and applications (McGeveran, 2016, p. 549). Additionally, since upholding individual privacy is one of the key objectives of data protection, extra attention must be given to accommodate the evolving concept of privacy as it responds to emerging technologies, and the meaning of privacy in different contexts and cultures (Li et al., 2017).

Big data as a key component in AI will be generated at great velocity and in great volume, from various sources and in various forms. Therefore, a consistent data governance standard will provide a safe and secure environment for supporting the lifecycle[18] of big data across sectors, borders, and applications, which will also contribute to the acceleration of AI deployment and innovation. Data quality must also be a focal point in data protection, as it plays an important role in machine learning training datasets. The quality of machine learning training datasets determines the robustness of an AI system and its compliance with AI principles. To ensure data quality, AI developers and researchers must implement measures to test for biases or inaccuracy that can lead to discriminatory outcomes in AI applications.

[18] Parallel to data lifecycle that involves collecting, storing, using, and distributing.

The availability of a robust data protection regulatory framework can also facilitate the creation of a "digital common" as proposed by UNESCO (UNESCO, 2021, p. 29). A "digital common" can provide a secure space for private and public sectors to share the data they have collected with stakeholders, both for research and to further advance AI innovation.

Employ transparency as a compliance mechanism

The principle of transparency was widely proposed as a compliance enforcement tool across all three governance initiatives from China, the European Union, and the United States. Policymakers must adopt this approach and require transparency practices involving high-risk AI and AI that has the potential to manipulate humans to uphold safety and compliance to AI principles:

- AI operators must be transparent in their testing processes and required to produce clear reports on test outcomes so that users and stakeholders can be informed of the benefits and risks that may occur from using the application.
- Audit trails must be included in AI designs to enable traceability and collect information for retrospective analysis when failures occur (Shneiderman, 2020).
- Design explanatory and exploratory user interfaces (such as a mortgage loan applications where users can use sliding bars for adjusting income, assets, and loan amount) where relevant, to allow users to modify their inputs in AI applications and understand how different variables contribute to different outcomes (Shneiderman, 2020).

- People must be provided with easy-to-understand statements explaining outcomes generated by AI that have consequential impacts on them. When necessary, these statements can be used for effective redress.
- Users must be informed when interfacing with AI that has the ability to manipulate people, such as AI that interacts with humans, that can detect emotions or reveal social categories, or manipulated content such as deepfakes.
- Where relevant, AI actors must explicitly implement AI principles throughout the lifecycle of AI and provide transparent reports on which principles are enabled and inhibited.

Transparency, when executed appropriately, can potentially mitigate the information asymmetry caused by rapidly advancing AI technologies. Therefore, transparency can also foster greater public confidence in AI (as proposed by the Guidance) that will, in turn, increase the widespread adoption of the technology.

Require testing to enforce safety and compliance

While the AI Act proposed pre-deployment testing for high-risk AI, this recommendation should be extended to all AI systems that carry risks. AI systems with low- to high-risk levels must be tested prior to deployment, continuously monitored, and periodically tested throughout operation to ensure that the systems consistently comply with AI principles. Both low- and high-risk AI systems have shown evidence of adverse impacts on people, from algorithm biases in search engines to discriminatory outcomes in recidivism assessment applications (Angwin et al., 2016; Noble, 2018). To uphold the standards of these tests, external oversight entities should be established to prevent any conflict of interest. Regulatory sandboxes, such as those proposed by the AI Act, can also be used to test out new applications in a controlled and time-limited environment, without compromising on human and environmental costs.

Validation testing prior to deployment should assess AI's compliance with regulatory checks that include elements from the Sustainable Development Goals and international human rights, in correspondence to the AI principles. Presently, there are various options and methods available to test for fairness in AI systems, including toolkits for detecting and mitigating algorithmic bias, fairness-enhancing interventions, and by building a collection of test cases to identify bias incidents (Shneiderman, 2020). But tools for testing AI's impact on Sustainable Development Goals and international human rights laws specifically still need to be developed to address the need in validation testing. Similarly, verification testing must also be conducted on AI systems prior to deployment and over periodic intervals during operations, to ensure that outcomes remain consistent and as expected. A consistent outcome in AI is important as it also indicates that the system maintains its value alignment, supporting the long-term AI principle that powerful AI should be aligned with human values. Especially with AI's ability to self-improve and enhance its algorithms, continuous monitoring will be crucial to maintain consistently safe outcomes and detect any value misalignment as early as possible.

Clear and transparent reports on the outcomes of these tests must be made available to AI actors, users, and stakeholders. These reports could help AI developers identify which Sustainable Development Goals or human rights the AI system enables or inhibits, and make the necessary adjustments to enhance the system's compliance with AI principles and regulatory requirements. Furthermore, continuous interval testing could help reduce the occurrence of failures and safety issues.

Collaborate with global alliances and local stakeholders

Countries should participate in global AI alliances to support the common advancement of AI, establish global governance, and attenuate the risk of an AI race. The AI principle of collaboration was recommended across all three governance initiatives examined in this paper, with the objectives of promoting regulatory influence and technology advancement. By bringing diverse expertise together, global alliances can help accelerate AI innovation to address global challenges such as climate change, affordable clean energy, and greater access to quality education. Participation in global standard setting for AI will be essential to facilitate a global AI ecosystem that encourages compatibility across borders and deter the development of malicious AI. Global alliances must implement measures to discourage rhetoric or intentions of an AI race and foster a strong common goal for advancing AI to benefit all of humanity.

Collaboration with local stakeholders will help policymakers take into consideration the many possible impacts AI can have on people. Policymakers should engage local stakeholders when developing AI policies and in enforcing regulatory requirements, such as reporting AI failures. Easily accessible platforms that encourage public engagement must be established to increase the public's enthusiasm to participate in these processes. For instance, reporting channels for AI failures must be clear and easily accessible by various groups of people and needs. Actions and redress pertaining to failures should also be transparent to keep AI actors accountable and incentivize people to report on failures consistently. These measures can foster greater public trust, uphold fairness, and provide equitable enjoyment of the benefits produced by AI.

Invest in AI research

Governments must provide continuous funding in AI research within the areas of its long-term impact on people and the environment, potential cyber threats and malicious use of AI, and impact assessment of AI governance on emerging technologies.

The AI principle of long-term advancement and impact on people and the environment was not addressed in any of the governance initiatives examined, perhaps because it involves many uncertainties. While the exact timeline for the long-term advancement of AI is less certain, AI is anticipated to have an increasingly profound impact on societies that could be both advantageous and adversarial (Bostrom, 2014; Bostrom et al., 2018). By investing in research to understand its future potential, policymakers can take advantage of the coming opportunities to maximize benefits, while reducing policy lags that could cause substantial and irreversible harm. Additionally, AI governance is an emerging field with numerous proposed governance models that must be continuously examined to ensure it is effectively upholding AI principles, not unnecessarily impeding innovation, and able to keep pace with the rapidly growing application of AI (Taeihagh, 2021).

Even though all three governance regimes recommended and possess cybersecurity governance, the unique characteristics of AI warrant investigation into novel security threats and malicious uses associated with AI. Robust security measures must be determined and implemented to prevent these security breaches. The AI principles urges strengthening security in AI applications, as it is paramount for upholding safety in the deployment of AI, especially in safety components such as those in autonomous vehicles, machinery, and medical devices. Furthermore, as AI is increasingly adopted in public administration, the prevalence of cybersecurity as a national security threat will intensify correspondingly.

Implement distributive and redistributive policies

To counter the concentration of wealth and power produced by AI and uphold the principle of share, policymakers must implement distributive and redistributive policies to rebalance the scale and ensure that the benefits and risks generated by AI are distributed equitably.

Stronger regulations must be enacted in antitrust laws to prevent mergers and acquisitions that prohibit a competitive market. The nature of emerging technology as an evolving market element, combined with the limitless potential in AI, will require specific scrutiny on mergers and acquisitions in tech-related (e.g. fintech, biotech, medtech, social media, etc.) companies. Specific investigation should be carried out to identify opportunities in emerging technologies that could lead to anticompetitive markets. For instance, Facebook's acquisitions of Instagram and WhatsApp were permitted at the time as the two smaller platforms were determined not in direct competition with Facebook. Presently, Facebook is being sued by the U.S. government for having become a social media monopoly and the government is seeking to require the sale of WhatsApp and Instagram (Hamilton, 2022). This case suggests that antitrust investigators must look beyond revenue and operational growth, and scrutinize the acquisition of data and users involved, and consider them as resource and profit. The accumulation of massive amounts of data and users — which are positioned as two key resources in AI advancement — could contribute to a power monopoly through data monopolization. Indeed, when Facebook proceeded to integrate all three platforms into a unified structure, the move should have prompted more scrutiny as a potential anticompetitive practice.

The growing wealth inequality within and across countries is projected to be exacerbated by the implications of AI on the labor market (Brynjolfsson & McAfee, 2016; Frey & Osborne, 2017). To slow growing wealth inequality, governments must implement progressive taxation on tech behemoths, such as taxation on monopoly rents and negative externalities created by AI deployment (Korinek & Stiglitz, 2021). In the long run, economists expect societies to become wealthier; Bostrom et al. envisioned a future when AI reaches a level of productivity that requires minimal gross domestic product (Bostrom et al., 2018). In this regard, various basic income and negative income tax models should be explored, and governments must determine which models will be best suited for their economies, societies, and cultures. For instance, Korinek and Stiglitz recommended pre-distributive policies in developing economies instead of redistribution, because the "capacity to tax" in those economies will be low (Korinek & Stiglitz, 2021). Additionally, a basic income strategy should be progressive and adaptable to the growing wealth generated by emerging technologies.

Non-AI Principles Policy Recommendations

Given its potential of having a transformative impact on societies, there will be other outcomes and implications that could arise from the deployment and use of AI. One of the implications of advancing toward a future permeated with AI is the need to improve AI literacy in the public so that individuals can continue to embrace their civic responsibilities and preserve their self-agency in a democratic society. Moreover, AI is also expected to transform the labor market as it offers greater efficiency, accuracy, and automation. Both matters can lead to fundamental shifts in societies that if left without any intervention could lead to harm against humanity. Therefore, this paper includes policy recommendations targeting these two areas in addition to addressing AI Principles.

Increase AI literacy and education

Education on AI-related skills and knowledge must be incorporated into education systems to prepare future generations that are capable of navigating a world permeated with AI. Governments must monitor market trends to identify necessary skills to promote AI talents that will advance AI innovation and ensure compliance to AI principles. With its potential for a transformative impact on society, AI actors should be trained in multidisciplinary fields on top of AI skills, including social sciences, ethics, and humanistic studies (Noble, 2018). A multidisciplinary training would better inform AI actors on the multifaceted impact their work will have on society and how to reduce the risk of harm on users and stakeholders.

AI literacy must be increased to empower the public and reduce the information asymmetry between AI actors and stakeholders. Greater AI literacy in understanding how AI decision-making can impact them could enable people to embrace their civic responsibilities and exercise their self-agency in a democratic society (Ferrer et al., 2021). Many of the AI issues that have occurred, such as algorithm biases in recidivism, finance, and healthcare, might have been addressed sooner if there had been greater AI literacy among stakeholders.

Anticipate future job demands

Governments must continue monitoring the implications of AI on labor market trends to identify the types of jobs that are most likely to become redundant, those that are more resilient, and new roles that are appearing in tandem with emerging technologies. While there are various theories and suggestions on how AI will affect the job market (replacement by automation, reinstatement effects counterbalancing displacement effect, task-based analysis), certain key themes can be derived from these discussions (Agrawal et al., 2019; Frey & Osborne, 2017). Low-skill tasks will most likely be automated, while tasks that require creative and social intelligence would be less susceptible to computerization (Frey & Osborne, 2017, p. 48). Based on these findings, policymakers should implement upskilling and reskilling programs for workers who will be most affected by automation and promote programs that offer creative and social skills.

Additionally, in the long-term future when advanced AI reaches a level of productivity that involves minimal gross domestic product as suggested by Bostrom et. al, the meaning of jobs and their purpose will likely change as well (Bostrom et al., 2018). By then a larger scale of redistribution of wealth and resources would be called for in the form of more magnanimous policies. Such a shift could prompt a fundamental review on the meaning of jobs, and a stronger emphasis on achieving a sustainable economy and a thriving humanity (Raworth, 2017).

Conclusion

AI has become ubiquitous in our societies and will continue to permeate extensively becoming an integral part of humanity in the future. Its impact on humanity is anticipated to be profound. It is crucial to set forth regulatory frameworks that can uphold AI Principles continuously and encourage a sustainable and beneficial development of AI innovations. Effective regulatory frameworks should maximize the benefits and diminish the risks of AI, and at the same time ensure that both benefits and risks are distributed equitably.

Admittedly, AI governance is at present a developing domain and best practices for different realities and contexts are yet to be determined in certainty. Nevertheless, the governance initiatives discussed in this paper outlined a number of reasonable approaches based on existing AI governance regimes, such as China's requirements for human oversight throughout the AI lifecycle, the European Union's risk-based approach, and the United States' wide-reaching recommendation for transparency in regulating and innovating AI.

A strong and comprehensive AI governance framework will be crucial for upholding AI principles and encouraging a safe and sustainable environment for advancing AI. Local and global collaboration in governance and innovation efforts will be immensely valuable for addressing the many risks and uncertainties that AI will bring. At the local level, AI stakeholders should have a role in shaping governance efforts that affect the outcomes of AI, empowering them to control the impact the technology will have upon them. Involving stakeholders in these processes reduces the risk of AI from overwriting their basic human rights. At a global scale, multilateral collaborations can help accelerate AI innovation to tackle global challenges, narrow the economic disparity among countries that could emerge from diverse AI deployment ability, and prevent the risk of an AI race.

Dafoe urged us to incorporate long-term measures for governing AI today while "the stakes are relatively low," another perspective to this framing is that we are still at the starting line of a future that is rapidly moving towards an AI age (Dafoe, 2018). Hence, we should seize the opportunity now to shape the future into a space and time where we want to be; a space and time where humanity and the natural environment can thrive and progress sustainably. Equally important to explore today is how we can embed human values into AI to ensure that our interactions with the technology and its impact on us will not detract our humanness. Hence, the decisions we make today will define not just the future of AI, but also of humanity.

References

Agrawal, A., Gans, J., Goldfarb, A., & Walter de Gruyter & Co. (2019). *The economics of artificial intelligence: An agenda*. University of Chicago Press.

Allen, J. R., & West, D. M. (2018, April 24). How artificial intelligence is transforming the world. *Brookings*. https://www.brookings.edu/research/how-artificial-intelligence-is-transforming-the-world/

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). *Machine Bias*. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Armstrong, S., Bostrom, N., & Shulman, C. (2016). Racing to the precipice: A model of artificial intelligence development. *AI & SOCIETY*, *31*(2), 201–206. https://doi.org/10.1007/s00146-015-0590-y

Autor, D. H. (2015). Why Are There Still So Many Jobs? The History and Future of Workplace Automation. *Journal of Economic Perspectives*, *29*(3), 3–30. https://doi.org/10.1257/jep.29.3.3

Bal, R., & Gill, I. S. (2020). *Policy Approaches to Artificial Intelligence Based Technologies in China, European Union and the United States* (SSRN Scholarly Paper ID 3699640). Social Science Research Network. https://doi.org/10.2139/ssrn.3699640

Barocas, S., & Selbst, A. D. (2016). *Big Data's Disparate Impact* (SSRN Scholarly Paper ID 2477899). Social Science Research Network. https://doi.org/10.2139/ssrn.2477899

Bartlett, R., Morse, A., Stanton, R., & Wallace, N. (2019). *Consumer-Lending Discrimination in the FinTech Era*. 51.

Bennett, C. L., & Keyes, O. (2020). What is the point of fairness? Disability, AI and the complexity of justice. *ACM SIGACCESS Accessibility and Computing*, *125*, 5:1. https://doi.org/10.1145/3386296.3386301

Berryhill, J., Heang, K. K., Clogher, R., & McBride, K. (2019). *Hello, World: Artificial intelligence and its use in the public sector*. OECD. https://doi.org/10.1787/726fd39d-en

Blasko, D. J. (2011). 'Technology Determines Tactics': The Relationship between Technology and Doctrine in Chinese Military Thinking. *Journal of Strategic Studies*, *34*(3), 355–381. https://doi.org/10.1080/01402390.2011.574979

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

https://books.google.com/books?id=7_H8AwAAQBAJ

Bostrom, N., Dafoe, A., & Flynn, C. (2018). *Public Policy and Superintelligent AI: A Vector Field Approach*. 28.

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

Brattberg, E., Raluca, C., & Rugova, V. (2020). *Europe and AI: Leading, Lagging Behind, or Carving Its Own Way?* https://carnegieendowment.org/2020/07/09/europe-and-ai-leading-lagging-behind-or-carving-its-own-way-pub-82236

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Allen, G. C., Steinhardt, J., Flynn, C., hÉigeartaigh, S. Ó., Beard, S., Belfield, H., Farquhar, S., … Amodei, D. (2018). The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. *ArXiv:1802.07228 [Cs]*. http://arxiv.org/abs/1802.07228

Brundage, M., & Bryson, J. (2016). Smart Policies for Artificial Intelligence. *ArXiv:1608.08196 [Cs]*. http://arxiv.org/abs/1608.08196

Brynjolfsson, E., & McAfee, A. (2016). *The Second Machine Age* (1st ed.). Norton Paperback.

Butcher, J., & Beridze, I. (2019). What is the State of Artificial Intelligence Governance Globally? *The RUSI Journal*, *164*(5–6), 88–96. https://doi.org/10.1080/03071847.2019.1694260

Byford, S. (2016, March 9). *Why Google's Go win is such a big deal*. The Verge. https://www.theverge.com/2016/3/9/11185030/google-deepmind-alphago-go-artificial-intelligence-impact

Cadwalladr, C., & Graham-Harrison, E. (2018, March 17). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian*. https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

Calo, R. (2017). *Artificial Intelligence Policy: A Primer and Roadmap* (SSRN Scholarly Paper ID 3015350). Social Science Research Network. https://doi.org/10.2139/ssrn.3015350

Carpenter v. United States, U.S. 119 (Supreme Court of United States 2018).

https://www.supremecourt.gov/opinions/17pdf/16-402_h315.pdf

Castro, D., & McLaughlin, M. (2021). *Who Is Winning the AI Race: China, The EU, or the United States? —2021 Update* (p. 49). Center for Data Innovation.

Cath, C. (2018). Governing artificial intelligence: Ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2133), 20180080. https://doi.org/10.1098/rsta.2018.0080

Cave, S., & ÓhÉigeartaigh, S. S. (2018). An AI Race for Strategic Advantage: Rhetoric and Risks. *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 36–40. https://doi.org/10.1145/3278721.3278780

Chorzempa, M., Triolo, P., & Sacks, S. (2018). Policy Brief 18-14: China's Social Credit System: A Mark of Progress or a Threat to Privacy? *Peterson Institute for International Economics*, 11.

Circiumaru, A. (2021, December 13). *Three proposals to strengthen the EU Artificial Intelligence Act*. Ada Lovelace Institute. https://www.adalovelaceinstitute.org/blog/three-proposals-strengthen-eu-artificial-intelligence-act/

Congressional Research Service. (2021). *International Discussions Concerning Lethal Autonomous Weapon Systems*. Congressional Research Service. https://sgp.fas.org/crs/weapons/IF11294.pdf

Convention of Certain Conventional Weapons. (2019). *Lethal autonomous weapon systems (LAWS), or weapons designed to independently select and engage targets without the need for manual human control, could enable military operations in communications-degraded or -denied environments where traditional systems may not be able to operate. LAWS are not yet in widespread development. However, as technology advances—Particularly artificial intelligence (AI)—A larger number of countries may consider developing and operating LAWS. This could hold potential implications for congressional oversight, defense investments, military concepts of operations, treaty-making, and the future of warfare.* United Nations Convention of Certain Conventional Weapons. https://undocs.org/pdf?symbol=en/CCW/MSP/2019/9

Crevier, D. (1993). *AI: The tumultuous history of the search for artificial intelligence*. Basic Books.

Dafoe, A. (2018). AI governance: A research agenda. *Governance of AI Program, Future of Humanity Institute,* *University of Oxford: Oxford, UK*, *1442*, 1443.

Daly, A., Hagendorff, T., Li, H., Mann, M., Marda, V., Wagner, B., & Wang, W. W. (2020). *AI, Governance and Ethics: Global Perspectives* (SSRN Scholarly Paper ID 3684406). Social Science Research Network. https://doi.org/10.2139/ssrn.3684406

Dastin, J. (2018, October 10). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

Dick, S. A. (2021, May 3). Making Up Minds. *Thinking Machines. History, Present and Future of AI*. https://thinking-machines.online/dick/

Dixon, R. B. L. (2021). *Decarbonizing Road Transportation: Re-envisioning Road Transportation with Artificial Intelligence* (p. 24). Humphrey School of Public Affairs.

Dragu, T., & Lupu, Y. (2021). Digital Authoritarianism and the Future of Human Rights. *International Organization*, 1–27. https://doi.org/10.1017/S0020818320000624

Drinhausen, K., & Brussee, V. (2021, March 3). *China's Social Credit System in 2021: From fragmentation towards integration | Merics*. MERICS. https://merics.org/en/report/chinas-social-credit-system-2021-fragmentation-towards-integration

European Commission. (n.d.). *A European approach to artificial intelligence | Shaping Europe's digital future*. Retrieved March 1, 2022, from https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

European Commission. (2021). *Proposal For A Regulation Of The European Parliament And Of The Council Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts*. https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206

Executive Order 13563—Improving Regulation and Regulatory Review. (2011). *Whitehouse.Gov*. https://obamawhitehouse.archives.gov/the-press-office/2011/01/18/executive-order-13563-improving-regulation-and-regulatory-review

Fairbrother, G. P. (2013). The Chinese Paternalistic State and Moral Education. In *Citizenship Education in China* (1st ed., pp. 11–26). Routledge.

Feldstein, S. (2019). *The Global Expansion of AI Surveillance*. https://carnegieendowment.org/2019/09/17/global-expansion-of-ai-surveillance-pub-79847

Ferrer, X., Nuenen, T. van, Such, J. M., Coté, M., & Criado, N. (2021). Bias and Discrimination in AI: A Cross-Disciplinary Perspective. *IEEE Technology and Society Magazine*, *40*(2), 72–80. https://doi.org/10.1109/MTS.2021.3056293

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 Electronic Journal*. https://doi.org/10.2139/ssrn.3518482

Floridi, L., & Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*, **1**(1). https://doi.org/10.1162/99608f92.8cd550d1

Frey, C. B., & Osborne, M. A. (2013). 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

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

Gasser, U. (2017, June 26). AI and the Law: Setting the Stage. *Berkman Klein Center Collection*. https://medium.com/berkman-klein-center/ai-and-the-law-setting-the-stage-48516fda1b11

Gasser, U., & Almeida, V. A. F. (2017). A Layered Model for AI Governance. *IEEE Internet Computing*, *21*(6), 58–62. https://doi.org/10.1109/MIC.2017.4180835

Gow, M. (2017). The Core Socialist Values of the Chinese Dream: Towards a Chinese integral state. *Critical Asian Studies*, *49*(1), 92–116. https://doi.org/10.1080/14672715.2016.1263803

Greene, D., Hoffmann, A. L., & Stark, L. (2019). Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning. *Proceedings of the 52nd Hawaii*

International Conference on System Sciences, 10.

Guihot, M., Matthew, A. F., & Suzor, N. (2017). *Nudging Robots: Innovative Solutions to Regulate Artificial Intelligence* (SSRN Scholarly Paper ID 3017004). Social Science Research Network. https://papers.ssrn.com/abstract=3017004

Hamilton, I. A. (2022, January 12). *The FTC can move forward with its bid to make Meta sell Instagram and WhatsApp, judge rules*. Business Insider. https://www.businessinsider.com/ruling-ftc-meta-facebook-lawsuit-instagram-whatsapp-can-proceed-2022-1

Helpman, E. (1998). *General Purpose Technologies and Economic Growth*. MIT Press. https://books.google.com/books?id=TSSePifW9Y4C

Human Rights Watch. (2019). *China's Algorithms of Repression: Reverse Engineering a Xinjiang Police Mass Surveillance App*. Human Rights Watch. https://www.hrw.org/report/2019/05/01/chinas-algorithms-repression/reverse-engineering-xinjiang-police-mass

Jackson, S., & Palmer, L. R. (2015). Reconceptualizing ecosystem services: Possibilities for cultivating and valuing the ethics and practices of care. *Progress in Human Geography*, *39*(2), 122–145. https://doi.org/10.1177/0309132514540016

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, E. B. (2020, March 12). *H.R.6216 - 116th Congress (2019-2020): National Artificial Intelligence Initiative Act of 2020* (2019/2020). https://www.congress.gov/bill/116th-congress/house-bill/6216

Kania, E. B. (2020). "AI weapons" in Chinese military innovation. *GLOBAL CHINA*, 23.

Karnosfsky, H. (2016, May 6). *Potential Risks from Advanced Artificial Intelligence: The Philanthropic Opportunity*. Open Philanthropy. http://www.openphilanthropy.org/blog/potential-risks-advanced-artificial-intelligence-philanthropic-opportunity

Katz, Y. (2020). *Artificial whiteness: Politics and ideology in artificial intelligence*. New York : Columbia University Press.

Kim, T. W., Hooker, J., & Donaldson, T. (2021). Taking principles seriously: A hybrid approach to value alignment in artificial intelligence. *The Journal of Artificial Intelligence Research*, *70*, 871–890. https://doi.org/10.1613/JAIR.1.12481

Korinek, A., & Stiglitz, J. E. (2021). *Artificial Intelligence, Globalization, and Strategies for Economic Development* (Working Paper No. 28453; Working Paper Series). National Bureau of Economic Research. https://doi.org/10.3386/w28453

Larsson, S. (2020). On the Governance of Artificial Intelligence through Ethics Guidelines. *Asian Journal of Law and Society*, 7(3), 437–451. https://doi.org/10.1017/als.2020.19

Lee, A., Shi, M., Chen, Q., Horsley, J. P., Schaefer, K., Creemers, R., & Webster Graham. (2021, September 15). Seven Major Changes in China's Finalized Personal Information Protection Law. *DigiChina*. https://digichina.stanford.edu/work/seven-major-changes-in-chinas-finalized-personal-information-protection-law/

Li, Y., Kobsa, A., Knijnenburg, B. P., & Carolyn Nguyen, M.-H. (2017). Cross-Cultural Privacy Prediction. *Proceedings on Privacy Enhancing Technologies*, *2017*(2), 113–132. https://doi.org/10.1515/popets-2017-0019

Liang, F., Das, V., Kostyuk, N., & Hussain, M. M. (2018). Constructing a Data-Driven Society: China's Social Credit System as a State Surveillance Infrastructure. *Policy & Internet*, *10*(4), 415–453. https://doi.org/10.1002/poi3.183

Lomas, N. (2022, January 13). Austrian website's use of Google Analytics breached GDPR. *TechCrunch*. https://social.techcrunch.com/2022/01/12/austrian-dpa-schrems-ii/

Lynch, S. (2017, March 11). *Andrew Ng: Why AI Is the New Electricity*. Stanford Graduate School of Business. https://www.gsb.stanford.edu/insights/andrew-ng-why-ai-new-electricity

McGeveran, W. (2016). *Privacy and data protection law*. Foundation Press.

McGregor, L., Murray, D., & Ng, V. (2019). INTERNATIONAL HUMAN RIGHTS LAW AS A FRAMEWORK FOR ALGORITHMIC ACCOUNTABILITY. *International & Comparative Law Quarterly*, *68*(2), 309–343. https://doi.org/10.1017/S0020589319000046

Morris, K. C., Schlenoff, C., & Srinivasan, V. (2017). Guest Editorial A Remarkable Resurgence of Artificial

Intelligence and Its Impact on Automation and Autonomy. *IEEE Transactions on Automation Science and Engineering*, *14*(2), 407–409. https://doi.org/10.1109/TASE.2016.2640778

National AI Initiative. (n.d.). *About Artificial Intelligence*. National Artificial Intelligence Initiative. Retrieved February 8, 2022, from https://www.ai.gov/about/

National Science and Technology Council. (2019). *The National Artificial Intelligence Research and Development Strategic Plan: 2019 Update* (p. 50). National Science and Technology Council. https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf

National Security Commission on Artificial Intelligence. (2021). *The National Security on Artificial Intelligence*. National Security Commission on Artificial Intelligence.

NDIA. (2021, October 2). *NDIA commends work of NSCAI as it ends service*. https://www.ndia.org/about/media/press-releases/2021/10/1/nscai

Noble, S. U. (2018). *Algorithms of Oppression*. NYU Press. https://doi.org/10.2307/j.ctt1pwt9w5.1

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

OECD AI Policy Observatory. (n.d.-a). *Policy initiatives for China*. OECD AI Policy Observatory. Retrieved February 28, 2022, from https://oecd.ai/en/dashboards/policy-initiatives?conceptUris=http:%2F%2Fkim.oecd.org%2FTaxonomy%2FGeographicalAreas%23China

OECD AI Policy Observatory. (n.d.-b). *Policy initiatives of the EU*. OECD AI Policy Observatory. Retrieved March 1, 2022, from https://oecd.ai/en/dashboards/policy-initiatives?conceptUris=http:%2F%2Fkim.oecd.org%2FTaxonomy%2FOrganisations%23EuropeanUnion

OECD AI Policy Observatory. (n.d.-c). *Policy initiatives of United States*. OECD AI Policy Observatory. Retrieved March 1, 2022, from https://oecd.ai/en/dashboards/policy-initiatives?conceptUris=http:%2F%2Fkim.oecd.org%2FTaxonomy%2FGeographicalAreas%23Unite dStates

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

Office of Science and Technology Policy. (2020). *American Artificial Intelligence Initiative: Year One Annual Report* (p. 36). The White House.

OHCHR. (n.d.). *International standards*. Retrieved February 21, 2022, from https://www.ohchr.org/EN/Issues/Privacy/SR/Pages/Internationalstandards.aspx

O'Leary, D. E. (2013). Artificial Intelligence and Big Data. *IEEE Intelligent Systems*, *28*(2), 96–99. https://doi.org/10.1109/MIS.2013.39

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown; eBook Collection (EBSCOhost). http://login.ezproxy.lib.umn.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=nlebk&AN=1109940&site=ehost-live

Pedreschi, D., Giannotti, F., Guidotti, R., Monreale, A., Ruggieri, S., & Turini, F. (2019). Meaningful Explanations of Black Box AI Decision Systems. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*(01), 9780–9784. https://doi.org/10.1609/aaai.v33i01.33019780

Pennachin, C., & Goertzel, B. (2007). Contemporary Approaches to Artificial General Intelligence. In B. Goertzel & C. Pennachin (Eds.), *Artificial General Intelligence* (pp. 1–30). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-68677-4_1

Polyakova, A., & Meserole, C. (2019). *Exporting digital authoritarianism: The Russian and Chinese models* (Foreign Policy, p. 22). Brookings Institute.

Popper, N. (2016, February 25). The Robots Are Coming for Wall Street. *The New York Times*. https://www.nytimes.com/2016/02/28/magazine/the-robots-are-coming-for-wall-street.html

Prince, A. E. R., & Schwarcz, D. (2019). Proxy Discrimination in the Age of Artificial Intelligence and Big Data. *Iowa Law Review*, *105*(3), 1257–1318.

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., &

Barnes, P. (2020). *Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing*. 12.

Raworth, K. (2017). *Doughnut economics: Seven ways to think like a 21st century economist*. Chelsea Green Publishing.

Roberts, H., Cowls, J., Morley, J., Taddeo, M., Wang, V., & Floridi, L. (2021). The Chinese approach to artificial intelligence: An analysis of policy, ethics, and regulation. *AI & SOCIETY*, *36*(1), 59–77. https://doi.org/10.1007/s00146-020-00992-2

Rossi, F. (2018). BUILDING TRUST IN ARTIFICIAL INTELLIGENCE. *Journal of International Affairs*, *72*(1), 127–134.

Russell, S. (2020). *Human Compatible: Artificial Intelligence and the Problem of Control*. Penguin Books.

S.1260 - 117th Congress (2021-2022): United States Innovation and Competition Act of 2021, S.1260, 117th (2021). https://www.congress.gov/bill/117th-congress/senate-bill/1260

Sheehan, M. (2022, January 4). *China's New AI Governance Initiatives Shouldn't Be Ignored*. Carnegie Endowment for International Peace. https://carnegieendowment.org/2022/01/04/china-s-new-ai-governance-initiatives-shouldn-t-be-ignored-pub-86127

Shneiderman, B. (2020). Bridging the Gap Between Ethics and Practice: Guidelines for Reliable, Safe, and Trustworthy Human-centered AI Systems. *ACM Transactions on Interactive Intelligent Systems*, *10*(4), 26:1-26:31. https://doi.org/10.1145/3419764

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. *ArXiv:1906.02243 [Cs]*. http://arxiv.org/abs/1906.02243

Stupp, C. (2022, January 19). EU Companies Face Fallout From Decision Against Google. *Wall Street Journal*. https://www.wsj.com/articles/eu-companies-face-fallout-from-decision-against-google-11642614679

Sunstein, C. R. (2018). *The cost-benefit revolution*. Cambridge, Massachusetts : The MIT Press.

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

The National New Generation Artificial Intelligence Governance Specialist Committee. (2021, September

25). Ethical Norms for New Generation Artificial Intelligence Released. *Center for Security and Emerging Technology*. https://cset.georgetown.edu/publication/ethical-norms-for-new-generation-artificial-intelligence-released/

The OECD Artificial Intelligence (AI) Principles. (n.d.). Retrieved February 8, 2022, from https://oecd.ai/en/ai-principles

The precautionary principle: Decision-making under uncertainty. (2017). *issue 18, September 2017*.

Ting, D. S. W., Liu, Y., Burlina, P., Xu, X., Bressler, N. M., & Wong, T. Y. (2018). AI for medical imaging goes deep. *Nature Medicine*, *24*(5), 539–540. https://doi.org/10.1038/s41591-018-0029-3

Toner, H., Creemers, R., & Triolo, P. (2021, August 27). Experts Examine China's Pioneering Draft Algorithm Regulations. *DigiChina*. https://digichina.stanford.edu/work/experts-examine-chinas-pioneering-draft-algorithm-regulations/

UNESCO. (2021). *Report of the Social and Human Sciences Commission (SHS)* (p. 39). UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000379920.page=14

United Nations General Assembly. (2015, September 25). *Transforming our world: The 2030 Agenda for Sustainable Development*. United Nations. https://sdgs.un.org/goals

Ünver, H. A. (2018). *Artificial Intelligence, Authoritarianism and the Future of Political Systems*. Centre for Economics and Foreign Policy Studies. http://www.jstor.org/stable/resrep26084

US EPA, O. (2015, December 29). *Sources of Greenhouse Gas Emissions* [Overviews and Factsheets]. https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions

van Berkel, N., Papachristos, E., Giachanou, A., Hosio, S., & Skov, M. B. (2020). A Systematic Assessment of National Artificial Intelligence Policies: Perspectives from the Nordics and Beyond. *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, 1–12. https://doi.org/10.1145/3419249.3420106

van Wynsberghe, A. (2021). Sustainable AI: AI for sustainability and the sustainability of AI. *AI and Ethics*, 1(3), 213–218. https://doi.org/10.1007/s43681-021-00043-6

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications*, *11*(1), 233. https://doi.org/10.1038/s41467-019-14108-y

Vought, R. T. (2020). *Guidance for Regulation of Artificial Intelligence Applications*. Office of Management and Budget. https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf

Wakefield, J. (2022, January 27). Musk: Robots to be bigger business than Tesla cars. *BBC News*. https://www.bbc.com/news/technology-60154782

Yan, Y. (2009). The Good Samaritan's new trouble: A study of the changing moral landscape in contemporary China1: THE GOOD SAMARITAN'S NEW TROUBLE. *Social Anthropology*, *17*(1), 9–24. https://doi.org/10.1111/j.1469-8676.2008.00055.x

Yayboke, E., & Carter, W. A. (2020, April 10). *The Need for a Leapfrog Strategy*. Center for Strategic and International Studies. https://www.csis.org/analysis/need-leapfrog-strategy

Yoo, C. S. (2013). Protocol Layering and Internet Policy. *University of Pennsylvania Law Review*, *161*, 66.

Yu, R., & Alì, G. S. (2019). What's Inside the Black Box? AI Challenges for Lawyers and Researchers. *Legal Information Management*, *19*(1), 2–13. https://doi.org/10.1017/S1472669619000021

Zeng, Y., Lu, E., & Huangfu, C. (2018). Linking Artificial Intelligence Principles. *ArXiv:1812.04814 [Cs]*. http://arxiv.org/abs/1812.04814

Zeng, Y., Lu, E., & Ruan, Z. (2022). *Linking Artificial Intelligence Principles (LAIP)*. Linking AI Principles LAIP. https://www.linking-ai-principles.org/keywords

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

Zhang, D., Mishra, S., Brynjolfsson, E., Etchemendy, J., Ganguli, D., Grosz, B., Lyons, T., Manyika, J., Niebles, J. C., & Sellitto, M. (2021). *The AI Index 2021 Annual Report* (Artificial Intelligence Index Report 2021). Stanford University Human-Centered Artificial Intelligence. https://aiindex.stanford.edu/wp-content/uploads/2021/03/2021-AI-Index-Report-_Chapter-7.pdf

Appendix

Table 3.

Overview of AI Principles developed worldwide.

	For	Sustainabil-	Collaborat-			Transpare-				Accountab	Long-Term
Source	Human	ity	ion	Share	Fairness	ncy	Privacy	Security	Safety	-ility	AI
Beijing 2019	17	1	2	5	3	5	3	2	3	2	4
NGCNGAI 2019	10	1	4	5	6	3	6	1	7	4	1
AIIA 2019	11	1	1	3	7	7	4	6	10	4	
WHO 2021	4	3	1	8	12	23	12	2	11	16	
UNICEF 2020	5	1	2	2	4	6	2	1	4	2	
UNESCO 2021	48	8	3	9	16	29	17	5	9	12	
Telia 2019	2	1	3	2	6	4	1	1	4	5	
Smart Dubai 2019	13		3	1	4	4	6	6	7	2	6
OECD 2019	7	2	2	6	8	3	5	3	7	7	
NGCNGAI 2021	4	3	1	4	14	6	9	6	13	5	
Montreal 2018	6	3	2	7	3	4	8	2	5	7	

Source	For Human	Sustainability	Collaboration	Share	Fairness	Transparency	Privacy	Security	Safety	Accountability	Long-Term AI
ITI 2017	10	1	5	3	3	1	3	8	6	12	
G20 2019	8	2	2	6	8	3	5	3	7	8	
FLI 2017	9		2	4	1	4	4	2	6	2	4
EGE 2018	15	2	3	5	9	2	10	5	8	6	
Beijing Children 2020	7	1	3	3	3	3	5	1	4	5	
Cabinet Office 2018	22	6	5	5	12	2	13	6	4	3	
EC 2019	2	2		4	11	9	6	3	9	5	
Sony 2018	3	2	1		2	2	2	2	1	1	
Australia 2019	9	1		3	5	4	6	5	7	14	
Shanghai YoungAI 2019	4	1		1	5	7	3	2	5	6	
SHAIISEAC 2019	2		1	1	2	2	4	11	7	3	
Russia 2021	12		3	2	7	3		2	5	9	
Nadella 2016	8		2	1	2	2	1	1	1	3	
MIC 2018	7		8	3	5	7	17	13	4	5	
MIC 2017	6		4	2	4	4	15	12	23	3	

	For	Sustainabil-	Collaborat-			Transpare-				Accountab	Long-Term
Source	Human	ity	ion	Share	Fairness	ncy	Privacy	Security	Safety	-ility	AI
Internet Society 2017	1		3	1	2	4	3	7	10	7	
UNI Global Union 2017	7	1		2	2	12	5	2	2	11	
HLEG 2018	15	3		6	35	13	11	4	10	8	
Google 2018	4		1	1	8	1	5	2	7	1	
US OSTP 2020	1			2	11	10	4	8	7	4	
Telefonica 2018	1	1	1		8	4	10	5	2		
Aotearoa 2020	4	1		1	5	5	3	2		2	
Deutsche Telekom 2018	4		4	2	4	4	4	6		9	
Tencent 2018	5			2	6	8	4	4	7		2
PDPC Compilation 2020	6	1		1	5	4		2	2	8	
Tsinghua CISS 2019	4		1	3	1	3	1	2	1		
Montreal 2017	5		1	2	4	2	9		1	6	
JSAI 2017	8			2	5	1	3	2	7	3	
Intel 2017			1	1	4	2	8	5	6	3	
HAIP 2018	15		2	4	5		5		7	3	2

Source	For Human	Sustainability	Collaboration	Share	Fairness	Transparency	Privacy	Security	Safety	Accountability	Long-Term AI
US OSTP 2020	1			2	11	10	4	8	7	4	
US IC 2020	3		1	1	1	3	2	5		2	
ICDPPC 2018	12		1	2	9	10	16		2	3	
IEEE 2017	13		1	1		8	5	3	6	8	
IA Latam 2019	3		4		2		2	1	2	1	
DoDDIB 2019	1			1	1	3		1	2	2	
SAP 2018	2		6		6	2	7	2	6		
The Public Voice 2018	1				12	4	4	5	8	13	
Tieto 2018	2			2	3	2		1	1	1	
GER DEC 2019	4	3		1	3		3	3	1		
CIGI 2018				1	4	5	3	3	12	5	
Megvii 2019	1				3	1	2	4	2	3	
COMEST 2019	2	1			1	4		1	1	3	
Alan Turing Inst 2019		2		2	4	3		1	1	3	
House of Lords 2018	4			1	4	4	2	3		1	
DeepMind 2017	1		4	1	1	2				2	

	For	Sustainabil-	Collaborat-			Transpare-				Accountab	Long-Term
Source	Human	ity	ion	Share	Fairness	ncy	Privacy	Security	Safety	-ility	AI
ITechLaw 2019	1				5	4	3		2	6	
Microsoft 2018					1	1	2	2	2	2	
Samsung 2019				1	3	3	1	1		2	
US DoD 2020				1	1	3		1	2	2	
Vodafone 2019	2				1	1	2	2		2	
Rome Call 2020	1				2	3	2	2		2	
NATO 2021	1				2	2		1	4	5	
PAI 2016	1		2			1	1	2		2	
OpenAI 2018	4		2	1				1	8		12
The Future Society 2017	4		2	2		3				3	
ADP 2018						4	8	1	1	1	
USACM 2017					4	4	2		5	3	
Unity 2018	2				2	1	1			2	
IEEE 2019	4					1	1		2	1	
FATML 2016				1	3	3	3			7	
Canada 2019				2		1	1	1		2	
IBE 2018	1				5	5	5			7	

	For	Sustainabil-	Collaborat-			Transpare-				Accountab	Long-Term
Source	Human	ity	ion	Share	Fairness	ncy	Privacy	Security	Safety	-ility	AI
NYTimes 2019	1				2	2	1			2	
South Korea 2020	2					1	1		1	1	
TBS Canada 2018	1					2	1	1		1	
PDPC 2019	1				2	4			1		
Russia 2019	2		1			3		2			
Adobe 2021					4	2			1	3	
IBM 2018b	1				2	1				2	
IBM 2018a			1	1	1	5					
IBM 2017			1			1		1	2		
GE Healthcare 2018					1	2	1		2		
OP Financial 2018						2	4			1	
Baidu 2018	2			1					2		
US AI Initiative 2019							1	1	2		
Sage 2017	1				1					3	
Etzioni 2017							1	1			
Stanford 2018	1										

	For	Sustainabil-	Collaborat-			Transpare-				Accountab	Long-Term
Source	Human	ity	ion	Share	Fairness	ncy	Privacy	Security	Safety	-ility	AI
	414	55	103	150	374	348	334	223	334	329	31

Note. *Descriptions in the source column link to their original documents. From Linking Artificial Intelligence Principles (Zeng et al., 2022).

Table 4.

Overview of key governance frameworks from China.

Title	Year	Governing Body	Description
New-Generation AI Development Plan (AIDP)	2017	The State Council of People's Republic of China	First national level legislative effort that explicitly focused on AI development as a unified strategy .
Cybersecurity Law of the People's Republic of China	2017	The State Council of People's Republic of China	To ensure cybersecurity; safeguard cyberspace sovereignty and national security, and social and public interests; protect the lawful rights and interests of citizens, legal persons, and other organizations; and promote the healthy development of the informatization of the economy and society.
Three-Year Action Plan to Promote the Development of a New Generation of Artificial Intelligence Industry	2018 - 2020	Minister of Industry and Information Technology (MIIT)	China's 'Three-year Guidance for Internet Plus Artificial Intelligence Plan (2016-2018)' focuses on: enhancing AI hardware capacity, ii) strong platform ecosystems, iii) AI applications in important socioeconomic areas, and iv) AI's impact on society.
AI Standardization	2018	China Electronics Standardization Institute	Outlines the national AI standardization framework and plan for AI capability development
Governance principles for the new generation artificial intelligence—Developing responsible artificial intelligence	2019	Ministry of Science and Technology (MOST)	This initiative highlights the theme of developing responsible artificial intelligence, emphasizing the eight principles of harmony, friendliness, fairness, inclusiveness, respect for privacy, security and controllability, shared responsibility, open collaboration,

			and agile governance.
Ethical Norms for the New Generation Artificial Intelligence	2021	Ministry of Science and Technology (MOST)	Lays out ethical norms for the use of AI in China. The norms cover areas such as the use and protection of personal information, human control over and responsibility for AI, and the avoidance of AI-related monopolies. The document does not specify how these norms are to be enforced; nor does it mention any punishments for those who violate the norms.
Internet Information Service Algorithmic Recommendation Management Provisions (Opinion-seeking Draft)	2021	Cyberspace Administration of China	To standardize Internet information service algorithmic recommendation activities
China Personal Information Protection Law	2021	Standing Committee of the National People's Congress	To protect personal information rights and interests, standardize personal information handling activities, and promote the rational use of personal information.

Guiding Opinions on Strengthening Overall Governance of Internet Information Service Algorithms	2021	Cyberspace Administration of China Central Propaganda Department Ministry of Education Ministry of Science and Technology Ministry of Industry and Information Technology Ministry of Public Security Ministry of Culture and Tourism State Administration of Market Regulation National Radio and Television Administration	Over the next three years, to gradually establish a comprehensive algorithm security governance structure with a robust governance mechanism, a refined supervisory system, and a standardized algorithm ecosystem.

Note. The highlighted document was chosen for the comparative analysis. Sourced from Bal & Gill, 2020; OECD, n.d.; Roberts et al., 2021; Sheehan, 2022.

Table 5.

Overview of key governance frameworks from the European Union.

Title	Year	Governing Body	Description
General Data Protection Regulation (GDPR)	2018	European Union (European Union)	The GDPR is a regulation in European Union law on data protection and privacy in the European Union.
European Union Strategy for Artificial Intelligence	2018	European Commission (EC)	The AI strategy proposed measures to streamline research, as well as policy options for AI regulation, which fed into work on the AI package.
Policy and Investment Recommendations of Trustworthy AI	2019	European Commission (EC)	Provides recommendations that can guide Trustworthy AI towards sustainability, growth and competitiveness, as well as inclusion – while empowering, benefiting and protecting human beings.
Data Governance Act	2020	European Commission (EC)	The instrument aims to foster the availability of data for use by making certain public sector data re-usable, increasing trust in data intermediaries, by promoting data altruism and by setting in place a governance mechanism for certain aspects of standardization.

Digital Services Act Package	2020	European Commission (EC)	The new Digital Services Act package aims to modernize the current legal framework for digital services by proposing (i) clear rules framing the responsibilities of digital services to address the risks faced by their users and to protect their rights, and (ii) ex ante rules covering large online platforms acting as gatekeepers, which now set the rules of the game for their users and their competitors.
AI Legislative Package (AI Act)	2021	European Commission (EC)	The“AI legislative package” comprises: i) a Proposal for a Regulation on a European approach for Artificial Intelligence; ii) an updated Coordinated Plan with Member States, and iii) a Proposal for a Regulation on Machinery Products. The draft legislation follows a horizontal and risk-based regulatory approach that differentiates between uses of AI that generate i) minimal risk; ii) low risk; iii) high risk; and iv) unacceptable risk, for which the EC proposes a strict ban.

Note. The highlighted document was chosen for the comparative analysis. Sourced from the European Commission and the OECD AI Policy Observatory.

Table 6.

Overview of key governance frameworks from the United States.

Title	Year	Governing Body	Description
National AI R&D Strategic Plan	2018	National Science and Technology Council (NSTC)	Identifies the critical areas of AI R&D that require Federal investments.
Federal Data Strategy	2019	Federal Geospatial Data Committee President's Management Council General Services Administration National Center for Education Statistics Department of Education and Training Federal Statistical Research Data Center Program Management Office U.S. Census Bureau Department of Commerce Federal Committee on Statistical Methodology Interagency Council on Statistical Policy Department of Education Office of Management and Budget Interagency Council on Statistical Policy	The Federal Data Strategy (FDS) encompasses a 10-year vision for how the Federal government will accelerate the use of data to deliver on its mission, serve the public, and steward resources while protecting security, privacy, and confidentiality.
A Plan for Federal Engagement in Developing Technical Standards and Related Tools	2019	National Institute of Standards and Technology	

Title	Year	Governing Body	Description
American Artificial Intelligence Initiative: Year One Annual Report	2020	Office of Science and Technology	This document provides both a summary of progress and a continued long-term vision for the American AI Initiative.
Guidance for Regulation of Artificial Intelligence Applications	2020	Office of Management and Budget (OMB) Office of Science and Technology Policy Domestic Policy Council National Economic Council	A memorandum that provides guidance to all Federal agencies to inform the development of regulatory and nonregulatory approaches regarding technologies and industrial sectors that are empowered or enabled by artificial intelligence (AI) and consider ways to reduce barriers to the development and adoption of AI technologies
National Security Commission on AI	2021	National Security Commission on AI	The NSCAI Final Report presents an integrated national strategy to reorganize the government, reorient the nation, and rally our closest allies and partners to defend and compete in the coming era of AI-accelerated competition and conflict.

Note. The highlighted document was chosen for the comparative analysis. Sourced from National Science and Technology Council, 2019; National Security Commission on Artificial Intelligence, 2021; OECD AI Policy Observatory, n.d.; Office of Science and Technology Policy, 2020; and Vought, 2020.

Table 7. *A Comparative Analysis of the Ethical Norms, the AI Act, and the Guidance.*

AI Principles	Ethical Norms (China)	AI Act (European Union)	Guidance (United States)
For Human	"- (INTRO p. 1) Ethical Norms puts forward six basic ethical requirements, namely: the advancement of human welfare, the promotion of fairness and justice, the protection of privacy and security, the assurance of controllability and trustworthiness, the strengthening of accountability, and improvements to the cultivation of ethics. - (SEC 1 ART 3. p. 2)All types of AI activities shall comply with the basic ethical norms below. (I) Advancement of Human Welfare. Persist in being people-centered (以人为 本), abide by shared human values, respect human rights and appeals to fundamental human interests, and comply with national or regional ethics. Persist in giving priority to the public interest, promote human-computer harmony and friendliness, improve the people's livelihoods, enhance the sense of gain and the sense of well-being, advance economic, social, and ecological sustainable development, and jointly build a community of common destiny for humanity (人类命运共同	"- (1.1. p. 1) It is in the Union interest to preserve the EU's technological leadership and to ensure that Europeans can benefit from new technologies developed and functioning according to Union values, fundamental rights (mentioned 33 times) and principles. - (1.1. p. 1) the Commission would put forward legislation for a coordinated European approach on the human and ethical implications of AI. - (1.1. p. 1) AI should be a tool for people and be a force for good in society with the ultimate aim of increasing human well-being. Rules for AI available in the Union market or otherwise affecting people in the Union should therefore be human centric, so that people can trust that the technology is used in a way that is safe and compliant with the law, including the respect of fundamental rights. - (1.1. p. 2) Council further highlighted the importance of ensuring that European citizens' rights	"- (Intro. p. 1) When considering regulations or policies related to AI applications, agencies should continue to promote advancements in technology and innovation, while protecting American technology, economic and national security, privacy, civil liberties, and other American values, including the principles of freedom, human rights, the rule of law, and respect for intellectual property. - (7.1 p. 13) The analysis of these alternatives should also evaluate, where relevant and appropriate and consistent with Executive Order 13859, impacts to equity, human dignity, fairness, potential distributive impacts, privacy and civil liberties, personal freedom, and other American values. - (4.1 p. 11) Accordingly, agencies should engage in dialogues to promote compatible regulatory approaches to AI and to promote American AI innovation, while protecting privacy, civil rights, civil liberties, and American values." "- (2. p. 2) While narrowly tailored and evidence based regulations that address specific and identifiable risks could provide an enabling environment for U.S.

	体)." - (ART 7 p.3) Fully respect and assure the privacy, freedom, dignity, security, and rights and other lawful interests of relevant entities. Prohibit infringement of the lawful rights and interests of natural persons, legal persons, and other organizations by the improper exercise of authority.	are fully respected and called for a review of the existing relevant legislation to make it fit for purpose for the new opportunities and challenges raised by AI. - (1.1. p. 3) the Commission puts forward the proposed regulatory framework on Artificial Intelligence with the following specific objectives: ensure that AI systems placed on the Union market and used are safe and respect existing law on fundamental rights and Union values; enhance governance and effective enforcement of existing law on fundamental rights and safety requirements applicable to AI systems; - (1.3. p. 5)It lays down a coherent, effective and proportionate framework to ensure AI is developed in ways that respect people's rights and earn their trust, making Europe fit for the digital age and turning the next ten years into the Digital Decade - (2.3. p. 7) The proposal builds on existing legal frameworks and is proportionate and necessary to achieve its objectives, since it follows a risk-based approach and imposes regulatory burdens only when an AI system is likely to pose high risks to fundamental rights and safety.	companies to maintain global competitiveness, agencies must avoid a precautionary approach that holds AI systems to an impossibly high standard such that society cannot enjoy their benefits and that could undermine America's position as the global leader in AI innovation. - (2. p. 2) Where AI entails risk, agencies should consider the potential benefits and costs of employing AI, as compared to the systems AI has been designed to complement or replace. - (4. p. 3) Agencies should calibrate approaches concerning these principles and consider case-specific factors to optimize net benefits. - (4.5. p. 5) For example, while the broader legal environment already applies to AI applications, the application of existing law to questions of responsibility and liability for decisions made by AI could be unclear in some instances, leading to the need for agencies, consistent with their authorities, to evaluate the benefits, costs, and distributional effects associated with any identified or expected method for accountability. - (4.5. p. 5) Executive Order 12866 calls on agencies to ""select those approaches that maximize net benefits (including potential economic, environmental, public health

		- (3.3. p. 10) (selecting policy option 3+) By requiring a restricted yet effective set of actions from AI developers and users, the preferred option limits the risks of violation of fundamental rights and safety of people and foster effective supervision and enforcement, by targeting the requirements only to systems where there is a high risk that such violations could occur. - (3.5. p.11) With a set of requirements for trustworthy AI and proportionate obligations on all value chain participants, the proposal will enhance and promote the protection of the rights protected by the Charter: the right to human dignity (Article 1), respect for private life and protection of personal data (Articles 7 and 8), nondiscrimination (Article 21) and equality between women and men (Article 23). It aims to prevent a chilling effect on the rights to freedom of expression (Article 11) and freedom of assembly (Article 12), to ensure protection of the right to an effective remedy and to a fair trial, the rights of defence and the presumption of innocence (Articles 47 and 48), as well as the general principle of good administration. Furthermore, as	and safety, and other advantages; distributive impacts; and equity)."" - (4.5. p. 5) Agencies should, when consistent with law, carefully consider the full societal costs, benefits, and distributional effects when considering regulations related to the development and deployment of AI applications. - (4.5. p. 5) Such consideration will include the potential benefits and costs of employing AI, when compared to the systems AI has been designed to complement or replace; whether implementing AI will change the type of errors created by the system; and comparison to the degree of risk tolerated in other existing systems. - (7.1. p. 12) In conducting such retrospective reviews, agencies can determine whether regulatory changes are necessary to remove barriers to the adoption of net beneficial AI systems by identifying and promulgating deregulatory actions, consistent with Executive Orders 13771, ""Reducing Regulation and Controlling Regulatory Costs,""23 and 13777, ""Enforcing the Regulatory Reform Agenda. "" - (7.1 p. 13) After identifying a set of potential regulatory approaches, the agency should conduct a benefit-cost analysis that estimates the benefits and costs associated

		applicable in certain domains, the proposal will positively affect the rights of a number of special groups, such as the workers’ rights to fair and just working conditions (Article 31), a high level of consumer protection (Article 28), the rights of the child (Article 24) and the integration of persons with disabilities (Article 26). The right to a high level of environmental protection and the improvement of the quality of the environment (Article 37) is also relevant, including in relation to the health and safety of people. - (3.5. p. 11)In case infringements of fundamental rights still happen, effective redress for affected persons will be made possible by ensuring transparency and traceability of the AI systems coupled with strong ex post controls. - (5.2.2. p. 12) The list of prohibited practices in Title II comprises all those AI systems whose use is considered unacceptable as contravening Union values, for instance by violating fundamental rights." - (1.1. p. 2) The EP Resolution on a Framework of Ethical Aspects of Artificial Intelligence, Robotics and Related Technologies specifically	with each alternative approach. The benefits and costs should be quantified and monetized to the extent possible and appropriate, and presented in both physical units (e.g., number of accidents avoided) and monetary terms." -(7.1. p. 13) The analysis of these alternatives should also evaluate, where relevant and appropriate and consistent with Executive Order 13859, impacts to equity, human dignity, fairness, potential distributive impacts, privacy and civil liberties, personal freedom, and other American values.

		recommends to the Commission to propose legislative action to harness the opportunities and benefits of AI, but also to ensure protection of ethical principles. - (5.2.7. TITLE IX p. 16) Those codes may also include voluntary commitments related, for example, to environmental sustainability, accessibility for persons with disability, stakeholders' participation in the design and development of AI systems, and diversity of development teams. "- (3.5. p.11) This proposal imposes some restrictions on the freedom to conduct business (Article 16) and the freedom of art and science (Article 13) to ensure compliance with overriding reasons of public interest such as health, safety, consumer protection and the protection of other fundamental rights ('responsible innovation') when high-risk AI technology is developed and used. Those restrictions are proportionate and limited to the minimum necessary to prevent and mitigate serious safety risks and likely infringements of fundamental rights. - (5.2.2. TITLE II p. 13) Other manipulative or exploitative practices affecting adults that might be	

		facilitated by AI systems could be covered by the existing data protection, consumer protection and digital service legislation that guarantee that natural persons are properly informed and have free choice not to be subject to profiling or other practices that might affect their behaviour."	

Fairness	"- (INTRO p. 1) As the Ethical Norms passed through stages such as special investigation, focused drafting, and solicitation of opinions, full consideration was given to the ethical concerns of all sectors of today's society about privacy, bias, discrimination, fairness, and so on…Ethical Norms puts forward six basic ethical requirements, namely: the advancement of human welfare, the promotion of fairness and justice, the protection of privacy and security, the assurance of controllability and trustworthiness, the strengthening of accountability, and improvements to the cultivation of ethics. - (SEC 1 ART 1 p. 2) These norms aim to incorporate ethics into the entire AI life cycle and to promote fairness, justice, harmony, and security while avoiding such problems as bias, discrimination, and privacy and information leaks. - (SEC 1 ART 3. p. 2) (II) Promotion of Fairness and Justice. Uphold inclusivity and tolerance, truly protect the lawful rights and interests of each relevant entity, promote fair sharing of AI benefits by all of society, and advance social fairness and justice and equality of opportunity. When	- (1.1. p. 2) The most recent Conclusions from 21 October 2020 further called for addressing the opacity, complexity, bias, a certain degree of unpredictability and partially autonomous behaviour of certain AI systems, to ensure their compatibility with fundamental rights and to facilitate the enforcement of legal rules. - (3.5. p.11) The obligations for ex ante testing, risk management and human oversight will also facilitate the respect of other fundamental rights by minimising the risk of erroneous or biased AI-assisted decisions in critical areas such as education and training, employment, important services, law enforcement and the judiciary." - (1.2. p.4) Consistency is also ensured with the EU Charter of Fundamental Rights and the existing secondary Union legislation on data protection, consumer protection, non-discrimination and gender equality. Furthermore, the proposal complements existing Union law on non-discrimination with specific requirements that aim to minimise the risk of algorithmic discrimination, in particular in relation to the design and the quality of data sets used for the	"""- (3. p. 2) The deployment of AI holds the promise to improve efficiency, effectiveness, safety, fairness, welfare, transparency, and other economic and social goals, and America's continued status as a global leader in AI development is important to preserving our economic and national security. - (4.7 p. 6) Agencies should consider in a transparent manner the impacts that AI applications may have on discrimination. AI applications have the potential of reducing present-day discrimination caused by human subjectivity. At the same time, applications can, in some instances, introduce real-world bias that produces discriminatory outcomes or decisions that undermine public trust and confidence in AI or be used in other ways that violate antidiscrimination statutes. When considering regulations or non-regulatory approaches related to AI applications, agencies should consider, in accordance with law, issues of fairness and nondiscrimination with respect to outcomes and decisions produced by the AI application at issue, as well as whether the AI application at issue may reduce levels of unlawful, unfair, or otherwise unintended discrimination as compared to existing processes. - (7.1 p. 13) The analysis of these

	providing AI products and services, fully respect and help vulnerable groups and special groups, and provide appropriate alternatives as necessary." - (SEC 3 ART 13 p. 4) Avoid Bias and Discrimination. In data collection and algorithm development, strengthen ethics investigations, and fully consider differentiated claims (差异化诉求). Prevent any data bias and algorithmic bias issues that may emerge, and strive to achieve AI system inclusivity, fairness, and non-discrimination.	development of AI systems complemented with obligations for testing, risk management, documentation and human oversight throughout the AI systems' lifecycle - (5.2.2. TITLE II p. 13)The proposal also prohibits AI-based social scoring for general purposes done by public authorities. Finally, the use of 'real time' remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement is also prohibited unless certain limited exceptions apply.	alternatives should also evaluate, where relevant and appropriate and consistent with Executive Order 13859, impacts to equity, human dignity, fairness, potential distributive impacts, privacy and civil liberties, personal freedom, and other American values."" - (4.3 p. 4) When an agency regulates AI applications, it should, as relevant, transparently articulate the strengths and weaknesses of the applications; intended optimizations or outcomes; bias and risk mitigations; potential impacts on competition, privacy and personal decisionmaking; any national security implications; and appropriate uses of the AI application's results. - (7.1. p. 12) Agencies should explain whether the action (need for regulations) is intended to address a market failure (e.g., asymmetric information), clarify uncertainty related to existing regulations, or address another factor, such as protecting privacy or civil liberties, preventing unlawful discrimination, or advancing the United States' economic and national security."

Transparency	"- (SEC 3 ART 12 p. 4) Enhance Security and Transparency. In the algorithm design, implementation, and application stages, improve transparency, explainability, comprehensibility, reliability, and controllability; enhance AI systems' toughness, adaptiveness, and ability to resist interference. Gradually achieve verifiability, auditability, supervisability, traceability, predictability, and reliability. - (SEC 4 ART 16 p. 4) Safeguard User Rights and Interests. Users should be clearly informed of the use of AI technology in products and services. The features and limitations of AI products and services should be indicated. Guarantee users' rights to be informed and 4 to consent. Provide simple and easy-to-understand solutions so that users can choose to use or exit AI modes. Do not place barriers to the equal use of AI by users."	"- (1.1. p. 3) For some specific AI systems, only minimum transparency obligations are proposed, in particular when chatbots or 'deep fakes' are used. - (2.3. p. 7) For other, non-high-risk AI systems, only very limited transparency obligations are imposed, for example in terms of the provision of information to flag the use of an AI system when interacting with humans. - (2.3. p. 7) For high-risk AI systems, the requirements of high quality data, documentation and traceability, transparency, human oversight, accuracy and robustness, are strictly necessary to mitigate the risks to fundamental rights and safety posed by AI and that are not covered by other existing legal frameworks. - (3.3. p.9) (selecting policy option 3+) The requirements will concern data, documentation and traceability, provision of information and transparency, human oversight and robustness and accuracy and would be mandatory for high-risk AI systems. - (3.4. p. 10) For national public administrations, it will promote public trust in the use of AI and strengthen enforcement mechanisms (by introducing a European coordination mechanism, providing for appropriate	"- (3. p. 2) The deployment of AI holds the promise to improve efficiency, effectiveness, safety, fairness, welfare, transparency, and other economic and social goals, and America's continued status as a global leader in AI development is important to preserving our economic and national security. - (4.3. p. 4) Agencies should hold information, whether produced by the government or acquired by the government from third parties, that is likely to have a clear and substantial influence on important public policy or private sector decisions (including those made by consumers) to a high standard of quality and transparency. When an agency regulates AI applications, it should, as relevant, transparently articulate the strengths and weaknesses of the applications; intended optimizations or outcomes; bias and risk mitigations; potential impacts on competition, privacy and personal decisionmaking; any national security implications; and appropriate uses of the AI application's results. - (4.4. p. 4) Agencies should be transparent about their evaluations of risk and re-evaluate their assumptions and conclusions at appropriate intervals so as to foster accountability. - (4.7. p. 6) Agencies should consider in a

		capacities, and facilitating audits of the AI systems with new requirements for documentation, traceability and transparency). Create legal certainty and trust in businesses in the EU. - (3.5. p.11) In case infringements of fundamental rights still happen, effective redress for affected persons will be made possible by ensuring transparency and traceability of the AI systems coupled with strong ex post controls. - (3.5. p. 11) The increased transparency obligations will also not disproportionately affect the right to protection of intellectual property (Article 17(2)), since they will be limited only to the minimum necessary information for individuals to exercise their right to an effective remedy and to the necessary transparency towards supervision and enforcement authorities, in line with their mandates. - (5.2.3. TITLE III p.13) Chapter 2 sets out the legal requirements for high-risk AI systems in relation to data and data governance, documentation and recording keeping, transparency and provision of information to users, human oversight, robustness, accuracy and security.	transparent manner the impacts that AI applications may have on discrimination. - (4.8. p. 6) In addition to improving the rulemaking process, transparency and disclosure can increase public trust and confidence in AI applications by allowing (a) non-experts to understand how an AI application works and (b) technical experts to understand the process by which AI made a given decision. Such disclosures, when required, should be written in a format that is easy for the public to understand and may include identifying when AI is in use, for instance, if appropriate for addressing questions about how the application impacts human end users. Disclosures may be required to preserve the ability of human end users and other members of the public to make informed decisions, although agencies should be aware that some applications of AI could improve or assist human decision-making. Agencies should carefully consider the sufficiency of existing or evolving legal, policy, and regulatory environments before contemplating additional measures for disclosure and transparency. What constitutes appropriate disclosure and transparency is context-specific, depending on assessments of potential harms (including those resulting from the exploitation of disclosed information), the

		- (5.2.3. TITLE III p. 14) After the provider has performed the relevant conformity assessment, it should register those stand-alone high-risk AI systems in an EU database that will be managed by the Commission to increase public transparency and oversight and strengthen ex post supervision by competent authorities. - (5.2.4. TITLE IV p. 14) Title IV concerns certain AI systems to take account of the specific risks of manipulation they pose. Transparency obligations will apply for systems that (i) interact with humans, (ii) are used to detect emotions or determine association with (social) categories based on biometric data, or (iii) generate or manipulate content ('deep fakes'). When persons interact with an AI system or their emotions or characteristics are recognised through automated means, people must be informed of that circumstance. If an AI system is used to generate or manipulate image, audio or video content that appreciably resembles authentic content, there should be an obligation to disclose that the content is generated through automated means, subject to exceptions for legitimate purposes (law enforcement,	magnitude of those harms, the technical state of the art, and the potential benefits of the AI application. - (5. p. 7) If this information is of significant public interest, agencies should consider periodically informing the general public about emerging trends to help coordinate research efforts, new or emerging changes that will affect particular stakeholders (e.g., consumers), and transparency about how specific AI applications generate net benefits and, if relevant, distributional effects. - (5. p. 8) Increasing such access to government data must be done in a manner consistent with the Open, Public, Electronic, and Necessary Government Data Act; 11 0MB Circular No. A-130 ""ManagingInformation as a Strategic Resource""; 0MB Memorandum M-13-13, ""Open Data PolicyManaging Information as an Asset""; and other relevant authorities that require agencies to collect and create information in a way that supports public transparency as well as downstream, secondary information dissemination and processing by third parties, thereby making government information accessible, discoverable, and usable. - (5. p. 9) Agencies should communicate this information (RFIs) transparently by describing the underlying assumptions and

		freedom of expression). This allows persons to make informed choices or step back from a given situation." "- (1.1. p. 1) By improving prediction, optimising operations and resource allocation, and personalising service delivery, the use of artificial intelligence can support socially and environmentally beneficial outcomes and provide key competitive advantages to companies and the European economy. - (1.1. p. 2) The most recent Conclusions from 21 October 2020 further called for addressing the opacity, complexity, bias, a certain degree of unpredictability and partially autonomous behaviour of certain AI systems, to ensure their compatibility with fundamental rights and to facilitate the enforcement of legal rules"	uncertainties regarding expected outcomes, both positive and negative. - (7.2. p. 13) In soliciting public input on Notices of Proposed Rulemaking (NPRMs) that relate to AI applications, agencies will benefit from the perspectives and expertise of stakeholders engaged in the design, development, deployment, operation, and impact of AI applications, and facilitate a decisionmaking process that is more transparent and accountable."

Privacy	"- (INTRO p. 1) As the Ethical Norms passed through stages such as special investigation, focused drafting, and solicitation of opinions, full consideration was given to the ethical concerns of all sectors of today's society about privacy, bias, discrimination, fairness, and so on...Ethical Norms puts forward six basic ethical requirements, namely: the advancement of human welfare, the promotion of fairness and justice, the protection of privacy and security, the assurance of controllability and trustworthiness, the strengthening of accountability, and improvements to the cultivation of ethics. - (SEC 1 ART 1 p.2) These norms aim to incorporate ethics into the entire AI life cycle and to promote fairness, justice, harmony, and security while avoiding such problems as bias, discrimination, and privacy and information leaks. - (SEC 1 ART 3 p. 2) (III) Protection of Privacy and Security. Fully respect everyone's right to know the extent of the use of, and to consent to the use of, their personal information. Handle personal information according to the principles of legality, propriety, necessity, and good faith, and	- (3.5. p.11) With a set of requirements for trustworthy AI and proportionate obligations on all value chain participants, the proposal will enhance and promote the protection of the rights protected by the Charter: the right to human dignity (Article 1), respect for private life and protection of personal data (Articles 7 and 8), nondiscrimination (Article 21) and equality between women and men (Article 23). "- (1.1. p. 2) In 2017, the European Council called for a 'sense of urgency to address emerging trends' including 'issues such as artificial intelligence ..., while at the same time ensuring a high level of data protection, digital rights and ethical standards - (1.2. p. 4) Consistency is also ensured with the EU Charter of Fundamental Rights and the existing secondary Union legislation on data protection, consumer protection, non-discrimination and gender equality. The proposal is without prejudice and complements the General Data Protection Regulation (Regulation (EU) 2016/679) and the Law Enforcement Directive (Directive (EU) 2016/680) with a set of harmonised rules applicable to the	"- (1. p. 1) When considering regulations or policies related to AI applications, agencies should continue to promote advancements in technology and innovation, while protecting American technology, economic and national security, privacy, civil liberties, and other American values, including the principles of freedom, human rights, the rule of law, and respect for intellectual property. - (4.1. p. 3) Regulatory approaches may also be needed to protect reasonable expectations of privacy on the part of individuals who interact with AI and to ensure that AI does not compromise the ability of individuals to make their own informed decisions. The appropriate regulatory or nonregulatory response to privacy and other risks must necessarily depend on the nature of the risk presented and the tools available to mitigate those risks. - (4. 3. p. 4) When an agency regulates AI applications, it should, as relevant, transparently articulate the strengths and weaknesses of the applications; intended optimizations or outcomes; bias and risk mitigations; potential impacts on competition, privacy and personal decisionmaking; any national security implications; and appropriate uses of the AI application's results.

	guarantee personal privacy and data security. Do not harm individuals' legal data rights and interests; do not steal, tamper, leak, or otherwise illegally collect or use personal information; and do not infringe upon personal privacy rights. - (SEC 4 ART 15 p. 4) Strengthen Quality Control. Strengthen quality monitoring and use assessment of AI products and services, and avoid harm to health, property, and user privacy caused by problems such as design and product defects. Do not operate, sell, or provide products and services that do not comply with quality standards." - (SEC 3 ART 11 p. 4) Improve Data Quality. In data collection, storage, use, processing, transmission, provision, disclosure, and other such stages, strictly comply with data-related laws, standards, and norms. Improve data integrity, timeliness, consistency, normative compliance, and accuracy.	design, development and use of certain high-risk AI systems and restrictions on certain uses of remote biometric identification systems. - (1.2. p. 4) Furthermore, the proposal complements existing Union law on non-discrimination with specific requirements that aim to minimise the risk of algorithmic discrimination, in particular in relation to the design and the quality of data sets used for the development of AI systems complemented with obligations for testing, risk management, documentation and human oversight throughout the AI systems' lifecycle. - (1.3. p.5) Furthermore, the promotion of AI-driven innovation is closely linked to the Data Governance Act19, the Open Data Directive20 and other initiatives under the EU strategy for data21, which will establish trusted mechanisms and services for the re-use, sharing and pooling of data that are essential for the development of data-driven AI models of high quality. - (2.1. p. 6) In addition, considering that this proposal contains certain specific rules on the protection of individuals with regard to the processing of personal data, notably restrictions of the use of AI systems for	- (4.6. p. 5) Targeted agency conformity assessment schemes, to protect health and safety, privacy, and other values, will be essential to a successful, and flexible, performance-based approach. - (4.10. p. 7) Consistent with Executive Order 12866, agencies should coordinate with each other to share experiences to ensure consistency and predictability of AI-related policies that advance American innovation and adoption of AI, while appropriately protecting privacy, civil liberties, national security, and American values and allowing sector-and application-specific approaches. - (5. p. 8) many existing (voluntary) frameworks-including those specific to safety, cybersecurity and privacy-have been developed with AI considerations in mind or are otherwise applicable to AI. - (6. p. 10) To promote innovation, use, and adoption of AI applications, standards could address many technical aspects, such as AI performance, measurement, safety, security, privacy, interoperability, robustness, trustworthiness, and governance. - (6. p. 11) Accordingly, agencies should engage in dialogues to promote compatible regulatory approaches to AI and to promote American AI innovation, while protecting privacy, civil rights, civil

		‘real-time’ remote biometric identification in publicly accessible spaces for the purpose of law enforcement, it is appropriate to base this regulation, in as far as those specific rules are concerned, on Article 16 of the TFEU. - (2.3. p. 7) For high-risk AI systems, the requirements of high quality data, documentation and traceability, transparency, human oversight, accuracy and robustness, are strictly necessary to mitigate the risks to fundamental rights and safety posed by AI and that are not covered by other existing legal frameworks - (3.3. p. 9) (selecting policy option 3+) The requirements will concern data, documentation and traceability, provision of information and transparency, human oversight and robustness and accuracy and would be mandatory for high-risk AI systems. - (5.2.2. TITLE II p. 13) Other manipulative or exploitative practices affecting adults that might be facilitated by AI systems could be covered by the existing data protection, consumer protection and digital service legislation that guarantee that natural persons are properly informed and have free choice not to	liberties, and American values. (at least two other such mentions) - (7.1 p. 13) The analysis of these alternatives should also evaluate, where relevant and appropriate and consistent with Executive Order 13859, impacts to equity, human dignity, fairness, potential distributive impacts, privacy and civil liberties, personal freedom, and other American values." "- (6.1. p.8) Access to data. Increasing such access to government data must be done in a manner consistent with the Open, Public, Electronic, and Necessary Government Data Act; 11 0MB Circular No. A-130 ""Managing Information as a Strategic Resource""; 12 0MB Memorandum M-13-13, ""Open Data PolicyManaging Information as an Asset""; 13 and other relevant authorities that require agencies to collect and create information in a way that supports public transparency as well as downstream, secondary information dissemination and processing by third parties, thereby making government information accessible, discoverable, and usable. - (6.1. p.9) Agencies may also review their existing disclosure protocols to determine if it is appropriate to make more data public, as well as provide more granular data, rather than aggregate data. In

		be subject to profiling or other practices that might affect their behaviour. - (5.2.3. TITLE III p. 13) Chapter 2 sets out the legal requirements for high-risk AI systems in relation to data and data governance, documentation and recording keeping, transparency and provision of information to users, human oversight, robustness, accuracy and security. The proposed minimum requirements are already state-of-the-art for many diligent operators and the result of two years of preparatory work, derived from the Ethics Guidelines of the HLEG29, piloted by more than 350 organisations"	increasing data access, agencies should not lose sight of the legal and policy requirements regarding the protection of sensitive information and vital public interests, such as privacy, security, and national economic competitiveness. - (7.3. p.13) In addition, because components of AI applications, such as algorithms or the data they are trained on and use, may be sensitive or subject to legal protections (e.g., privacy or intellectual property), agencies should consider the risks of inadequate protections to algorithms and data throughout the design, development, deployment, and operation of an AI system, given the level of sensitivity of the algorithms and data."

Safety	- (SEC5 ART 20 p. 5) Prohibit Violations and Malicious Use. Prohibit the use of AI products and services that fail to comply with laws and regulations, ethics, standards, or norms, and prohibit the use of AI products and services to engage in illegal activities. Strictly forbid endangering national security, public safety (公共安全), and production safety, and strictly prohibit harm to the public interest. - (SEC 4 ART 15 p. 4) Strengthen Quality Control. Strengthen quality monitoring and use assessment of AI products and services, and avoid harm to health, property, and user privacy caused by problems such as design and product defects. Do not operate, sell, or provide products and services that do not comply with quality standards. - (ART 8 p. 3) Strengthen Risk Prevention. Strengthen the bottom-line mindset (底线 思维) and risk awareness, strengthen research on and assessment of potential risks in AI development, and promptly launch system risk monitoring and assessment. Establish an effective early warning mechanism for risks, and improve AI ethical risk control and handling capabilities.	"- (1.1. p. 1) Rules for AI available in the Union market or otherwise affecting people in the Union should therefore be human centric, so that people can trust that the technology is used in a way that is safe and compliant with the law, including the respect of fundamental rights. - (1.1. p. 3) the Commission puts forward the proposed regulatory framework on Artificial Intelligence with the following specific objectives: ensure that AI systems placed on the Union market and used are safe and respect existing law on fundamental rights and Union values; enhance governance and effective enforcement of existing law on fundamental rights and safety requirements applicable to AI systems; facilitate the development of a single market for lawful, safe and trustworthy AI applications and prevent market fragmentation. - (1.1. p. 3) The proposal lays down a solid risk methodology to define “high-risk” AI systems that pose significant risks to the health and safety or fundamental rights of persons. - facilitate the development of a single market for lawful, safe and trustworthy AI applications and	"- (3. p. 2) The deployment of AI holds the promise to improve efficiency, effectiveness, safety, fairness, welfare, transparency, and other economic and social goals, and America's continued status as a global leader in AI development is important to preserving our economic and national security. - (4.5. p 5) Executive Order 12866 calls on agencies to ""select those approaches that maximize net benefits (including potential economic, environmental, public health and safety, and other advantages; distributive impacts; and equity). - (4.6. p. 5) Targeted agency conformity assessment schemes, to protect health and safety, privacy, and other values, will be essential to a successful, and flexible, performance-based approach. - (4.9. p. 6) Agencies should promote the development of AI systems that are safe, secure, and operate as intended, and encourage the consideration of safety and security issues throughout the AI design, development, deployment, and operation process. Agencies should pay particular attention to the controls in place to ensure the confidentiality, integrity, and availability of the information processed, stored, and transmitted by AI systems. Agencies should also consider methods for providing systemic resilience, and for

	"- (SEC 4 ART 16 p. 4) Safeguard User Rights and Interests. Users should be clearly informed of the use of AI technology in products and services. The features and limitations of AI products and services should be indicated. Guarantee users' rights to be informed and 4 to consent. Provide simple and easy-to-understand solutions so that users can choose to use or exit AI modes. Do not place barriers to the equal use of AI by users. - (SEC 1 ART 3 p. 3) (IV) Assurance of Controllability and Trustworthiness. Ensure that humans have fully autonomous decision-making rights and that they have the right to accept or reject AI-provided services, the right to withdraw from AI interactions at any time, and the right to terminate AI system operations at any time. Ensure that AI is always under human control."	prevent market fragmentation. - (1.1. p. 3) Predictable, proportionate and clear obligations are also placed on providers and users of those systems to ensure safety and respect of existing legislation protecting fundamental rights throughout the whole AI systems' lifecycle - (1.2. p. 4) As regards high-risk AI systems which are safety components of products, this proposal will be integrated into the existing sectoral safety legislation to ensure consistency, avoid duplications and minimise additional burdens. - (1.2. p. 4) With regard to the interplay of requirements, while the safety risks specific to AI systems are meant to be covered by the requirements of this proposal, New Legislative Framework (NLF) legislation aims at ensuring the overall safety of the final product and therefore may contain specific requirements regarding the safe integration of an AI system into the final product. - (2.1. p. 6) Some Member States are already considering national rules to ensure that AI is safe and is developed and used in compliance with fundamental rights obligations.	preventing bad actors from exploiting AI systems, including cybersecurity risks posed by AI operation, and adversarial use of AI against a regulated entity. When evaluating or developing regulatory and non-regulatory approaches to AI applications, agencies should be mindful of any potential safety and security risks and vulnerabilities, as well as the risk of possible malicious deployment and use of AI applications. Moreover, agencies should consider, where relevant, any national security implications raised by the unique characteristics of AI and AI applications and take actions to protect national security as appropriate for their authorities. - (5.4. p. 8) many existing (voluntary) frameworks-including those specific to safety, cybersecurity and privacy-have been developed with AI considerations in mind or are otherwise applicable to AI. - (6.3. p. 10) To promote innovation, use, and adoption of AI applications, standards could address many technical aspects, such as AI performance, measurement, safety, security, privacy, interoperability, robustness, trustworthiness, and governance." - (4.1. p. 3) Since the continued adoption and acceptance of AI will depend significantly on public trust and validation, the government's regulatory and non-

		(Discouraged because pathhwork of potentially divergent national rules will hamper the seamless circulation of products and services related to AI systesm across the EU). - (2.3. p. 7) The proposal builds on existing legal frameworks and is proportionate and necessary to achieve its objectives, since it follows a risk-based approach and imposes regulatory burdens only when an AI system is likely to pose high risks to fundamental rights and safety - (2.4. p. 7) The direct applicability of a Regulation, in accordance with Article 288 TFEU, will reduce legal fragmentation and facilitate the development of a single market for lawful, safe and trustworthy AI systems - (3.3. p. 10) By requiring a restricted yet effective set of actions from AI developers and users, the preferred option limits the risks of violation of fundamental rights and safety of people and foster effective supervision and enforcement, by targeting the requirements only to systems where there is a high risk that such violations could occur. - (3.3 p. 10) Businesses or public authorities that develop or use AI	regulatory approaches to AI should contribute to public trust in AI by promoting reliable, robust, and trustworthy AI applications. - (5.1. p. 7) Agencies should consider using any existing statutory authority to issue non-regulatory policy statements, guidance, or testing and deployment frameworks, as a means of encouraging AI innovation in that sector. "- (3. p.3) While narrowly tailored and evidencebased regulations that address specific and identifiable risks could provide an enabling environment for U.S. companies to maintain global competitiveness, agencies must avoid a precautionary approach that holds AI systems to an impossibly high standard such that society cannot enjoy their benefits and that could undermine America's position as the global leader in AI innovation. Where AI entails risk, agencies should consider the potential benefits and costs of employing AI, as compared to the systems AI has been designed to complement or replace. - (4.1. p. 3) The appropriate regulatory or nonregulatory response to privacy and other risks must necessarily depend on the nature of the risk presented and the tools available to mitigate those risks. - (4.4. p. 4.) Regulatory and non-regulatory

		applications that constitute a high risk for the safety or fundamental rights of citizens would have to comply with specific requirements and obligations - (3.5. p. 11) The right to a high level of environmental protection and the improvement of the quality of the environment (Article 37) is also relevant, including in relation to the health and safety of people. - (3.5. p. 11) This proposal imposes some restrictions on the freedom to conduct business (Article 16) and the freedom of art and science (Article 13) to ensure compliance with overriding reasons of public interest such as health, safety, consumer protection and the protection of other fundamental rights (‘responsible innovation’) when high-risk AI technology is developed and used. Those restrictions are proportionate and limited to the minimum necessary to prevent and mitigate serious safety risks and likely infringements of fundamental rights. - (5.2.3. TITLE III p. 13) Title III contains specific rules for AI systems that create a high risk to the health and safety or fundamental rights of natural persons. - (5.2.3. TITLE III p. 13) The	approaches to AI should be based on a consistent application of risk assessment and risk management across various agencies and various technologies. It is not necessary to mitigate every foreseeable risk; in fact, a foundational principle of regulatory policy is that all activities involve tradeoffs. Instead, a risk-based approach should be used to determine which risks are acceptable and which risks present the possibility of unacceptable harm, or harm that has expected costs greater than expected benefits. Agencies should be transparent about their evaluations of risk and re-evaluate their assumptions and conclusions at appropriate intervals so as to foster accountability. Correspondingly, the magnitude and nature of the consequences should an AI tool fail, or for that matter succeed, can help inform the level and type of regulatory effort that is appropriate to identify and mitigate risks. Specifically, agencies should follow the direction in Executive Order 12866, ""Regulatory Planning and Review,""6 to consider the degree and nature of the risks posed by various activities within their jurisdiction. Such an approach will, where appropriate, avoid hazard-based and unnecessarily precautionary approaches to regulation that could unjustifiably create anticompetitive effects or inhibit

		classification of an AI system as high-risk is based on the intended purpose of the AI system, in line with existing product safety legislation. - (5.2.3. TITLE III p. 13) Chapter 1 of Title III sets the classification rules and identifies two main categories of highrisk AI systems: AI systems intended to be used as safety component of products that are subject to third party ex-ante conformity assessment; - (5.2.3. TITLE III p. 14) for reasons of consistency with the existing product safety legislation, the conformity assessments of AI systems that are safety components of products will follow a system with third party conformity assessment procedures already established under the relevant sectoral product safety legislation." - (5.1. p. 12) This registration will also enable competent authorities, users and other interested people to verify if the high-risk AI system complies with the requirements laid down in the proposal and to exercise enhanced oversight over those AI systems posing high risks to fundamental rights. "- (1.2. p. 4) Union law on non-discrimination with specific	innovation.7 Whenever practical and consistent with applicable law, agencies should seek to apply consistent risk assessment and risk management frameworks and approaches to similar AI functionalities across sectors. Any assessment of risk should compare that risk to risk presented by the situation that would obtain absent the AI application at issue; if an AI application lessens risk that would otherwise obtain, any relevant regulations presumably should permit that application. - (7.4. p. 14) The management of risks created by AI applications should be appropriate to, and commensurate with, the degree of risk that an agency determines in its assessment. For AI applications, agencies should adopt a tiered approach in which the degree of risk and consequences of both success and failure of the technology determines the regulatory approach, including the option of not regulating. For AI applications that pose lower risks, agencies can rely on less stringent and burdensome regulatory approaches--or non-regulatory approachessuch as requiring information disclosures or consumer education. For higher risk AI applications, agencies should consider, for example, the effect on individuals, the environments in which the

		requirements that aim to minimise the risk of algorithmic discrimination, in particular in relation to the design and the quality of data sets used for the development of AI systems complemented with obligations for testing, risk management, documentation and human oversight throughout the AI systems' lifecycle. - (3.3. p. 10) (selecting policy option 3+) By requiring a restricted yet effective set of actions from AI developers and users, the preferred option limits the risks of violation of fundamental rights and safety of people and foster effective supervision and enforcement, by targeting the requirements only to systems where there is a high risk that such violations could occur. - (3.5. p. 11) The obligations for ex ante testing, risk management and human oversight will also facilitate the respect of other fundamental rights by minimising the risk of erroneous or biased AI-assisted decisions in critical areas such as education and training, employment, important services, law enforcement and the judiciary - (5.2.2. TITLE II p. 12) Title II establishes a list of prohibited AI. The regulation follows a risk-based	applications will be deployed, the necessity or availability of redundant or back-up systems, the system architecture or capability control methods available when an AI application makes an error or fails, and how those errors and failures can be detected and remediated."

		approach, differentiating between uses of AI that create (i) an unacceptable risk, (ii) a high risk, and (iii) low or minimal risk. The list of prohibited practices in Title II comprises all those AI systems whose use is considered unacceptable as contravening Union values, for instance by violating fundamental rights. - (5.2.3. TITLE III p. 13) Title III contains specific rules for AI systems that create a high risk to the health and safety or fundamental rights of natural persons. In line with a risk-based approach, those high-risk AI systems are permitted on the European market subject to compliance with certain mandatory requirements and an ex-ante conformity assessment. - (5.2.5. TITLE V p. 15) AI regulatory sandboxes establish a controlled environment to test innovative technologies for a limited time on the basis of a testing plan agreed with the competent authorities. - (5.2.6. TITLE VIII p. 15) When necessary for their mandate, existing supervision and enforcement authorities will also have the power to request and access any documentation maintained following this regulation and, where needed, request market	

		surveillance authorities to organise testing of the high-risk AI system through technical means. "- (2.3. p. 7) For high-risk AI systems, the requirements of high quality data, documentation and traceability, transparency, human oversight, accuracy and robustness, are strictly necessary to mitigate the risks to fundamental rights and safety posed by AI and that are not covered by other existing legal frameworks. - (3.3. p. 9) The requirements will concern data, documentation and traceability, provision of information and transparency, human oversight and robustness and accuracy and would be mandatory for high-risk AI systems. - (5.2.3. TITLE III p. 13) Chapter 2 sets out the legal requirements for high-risk AI systems in relation to data and data governance, documentation and recording keeping, transparency and provision of information to users, human oversight, robustness, accuracy and security."	

Accountability	"- (INTRO p. 1) Ethical Norms puts forward six basic ethical requirements, namely: the advancement of human welfare, the promotion of fairness and justice, the protection of privacy and security, the assurance of controllability and trustworthiness, the strengthening of accountability, and improvements to the cultivation of ethics. - (SEC 1 ART 3 p. 3) (V) Strengthening of Accountability. Insist that humans are the ultimately responsible entities. Clearly define the responsibilities of interested parties, comprehensively heighten awareness of responsibility, and exercise self-reflection and self-discipline at every link throughout the AI life cycle. Establish AI accountability mechanisms, do not avoid investigations into responsibility, and do not evade one's own responsibilities." - (INTRO p. 2) These norms are established in order to heighten society's ethical 2 awareness and behavioral consciousness of AI, actively guide responsible AI R&D and application activities, and promote healthy AI development."	"- (1.2. p. 4) As regards AI systems provided or used by regulated credit institutions, the authorities responsible for the supervision of the Union's financial services legislation should be designated as competent authorities for supervising the requirements in this proposal to ensure a coherent enforcement of the obligations under this proposal and the Union's financial services legislation where AI systems are to some extent implicitly regulated in relation to the internal governance system of credit institutions. - (5.2.6. TITLE VII p. 15) Title VII aims to facilitate the monitoring work of the Commission and national authorities through the establishment of an EU-wide database for stand-alone high-risk AI systems with mainly fundamental rights implications. - (5.2.6. TITLE VIII p. 15) Title VIII sets out the monitoring and reporting obligations for providers of AI systems with regard to post-market monitoring and reporting and investigating on AI-related incidents and malfunctioning."	"- (4.2. p. 3) Public participation, especially in those instances where AI uses information about individuals, will improve agency accountability and regulatory outcomes, as well as increase public trust and confidence. - (4.4. p. 4) Agencies should be transparent about their evaluations of risk and re-evaluate their assumptions and conclusions at appropriate intervals so as to foster accountability. - (4.5. p. 5) the application of existing law to questions of responsibility and liability for decisions made by AI could be unclear in some instances, leading to the need for agencies, consistent with their authorities, to evaluate the benefits, costs, and distributional effects associated with any identified or expected method for accountability. - (7.2. p. 13) In soliciting public input on Notices of Proposed Rulemaking (NPRMs) that relate to AI applications, agencies will benefit from the perspectives and expertise of stakeholders engaged in the design, development, deployment, operation, and impact of AI applications, and facilitate a decisionmaking process that is more transparent and accountable."

Security	"- (INTRO p. 1) Ethical Norms puts forward six basic ethical requirements, namely: the advancement of human welfare, the promotion of fairness and justice, the protection of privacy and security, the assurance of controllability and trustworthiness, the strengthening of accountability, and improvements to the cultivation of ethics. - (SEC 1 ART 1 p. 2) These norms aim to incorporate ethics into the entire AI life cycle and to promote fairness, justice, harmony, and security while avoiding such problems as bias, discrimination, and privacy and information leaks. - (SEC 1 ART 3 p. 2) (III) Protection of Privacy and Security. Fully respect everyone's right to know the extent of the use of, and to consent to the use of, their personal information. Handle personal information according to the principles of legality, propriety, necessity, and good faith, and guarantee personal privacy and data security. Do not harm individuals' legal data rights and interests; do not steal, tamper, leak, or otherwise illegally collect or use personal information; and do not infringe upon personal privacy rights.	"- (1.1. p. 1) It supports the objective of the Union being a global leader in the development of secure, trustworthy and ethical artificial intelligence as stated by the European Council3 and ensures the protection of ethical principles as specifically requested by the European Parliament. - (5.2.3. TITLE III p. 13) Chapter 2 sets out the legal requirements for high-risk AI systems in relation to data and data governance, documentation and recording keeping, transparency and provision of information to users, human oversight, robustness, accuracy and security." RB: ART 15 cybersecurity	"- (1. p. 1) When considering regulations or policies related to AI applications, agencies should continue to promote advancements in technology and innovation, while protecting American technology, economic and national security, privacy, civil liberties, and other American values, including the principles of freedom, human rights, the rule of law, and respect for intellectual property. - (3. p. 2) The deployment of AI holds the promise to improve efficiency, effectiveness, safety, fairness, welfare, transparency, and other economic and social goals, and America's continued status as a global leader in AI development is important to preserving our economic and national security. - (4.9. p. 6) Agencies should promote the development of AI systems that are safe, secure, and operate as intended, and encourage the consideration of safety and security issues throughout the AI design, development, deployment, and operation process. Agencies should pay particular attention to the controls in place to ensure the confidentiality, integrity, and availability of the information processed, stored, and transmitted by AI systems. Agencies should also consider methods for providing systemic resilience, and for preventing bad actors from exploiting AI

	- (ART 7 p. 3) Fully respect and assure the privacy, freedom, dignity, security, and rights and other lawful interests of relevant entities. Prohibit infringement of the lawful rights and interests of natural persons, legal persons, and other organizations by the improper exercise of authority. - (SEC 3 ART 12 p. 4) Enhance Security and Transparency. In the algorithm design, implementation, and application stages, improve transparency, explainability, comprehensibility, reliability, and controllability; enhance AI systems' toughness, adaptiveness, and ability to resist interference. Gradually achieve verifiability, auditability, supervisability, traceability, predictability, and reliability. - (SEC 5 ART 20 p. 5) Prohibit Violations and Malicious Use. Prohibit the use of AI products and services that fail to comply with laws and regulations, ethics, standards, or norms, and prohibit the use of AI products and services to engage in illegal activities. Strictly forbid endangering national security, public safety (公共安全), and production safety, and strictly prohibit harm to the public interest.		systems, including cybersecurity risks posed by AI operation, and adversarial use of AI against a regulated entity. When evaluating or developing regulatory and non-regulatory approaches to AI applications, agencies should be mindful of any potential safety and security risks and vulnerabilities, as well as the risk of possible malicious deployment and use of AI applications. Moreover, agencies should consider, where relevant, any national security implications raised by the unique characteristics of AI and AI applications and take actions to protect national security as appropriate for their authorities. - (4.10. p. 7) Consistent with Executive Order 12866, agencies should coordinate with each other to share experiences to ensure consistency and predictability of AI-related policies that advance American innovation and adoption of AI, while appropriately protecting privacy, civil liberties, national security, and American values and allowing sector-and application-specific approaches. - (5.4. p. 8) many existing (voluntary) frameworks-including those specific to safety, cybersecurity and privacy-have been developed with AI considerations in mind or are otherwise applicable to AI. - (6. p. 8) Executive Order 13859 requires 0MB to issue a memorandum to agencies

	- (SEC 5 ART 21 p. 5) Actively Provide Prompt Feedback. Actively participate in the practice of AI ethical governance. Promptly report to the relevant entities concerning such problems as technical security vulnerabilities, policy or regulatory vacuums, and regulatory lags discovered in the process of using AI products and services, and assist in solving the problems."		that shall ""consider ways to reduce barriers to the use of AI technologies in order to promote their innovative application while protecting civil liberties, privacy, American values, and United States economic and national security."" - (6.1. p. 9) In increasing data access, agencies should not lose sight of the legal and policy requirements regarding the protection of sensitive information and vital public interests, such as privacy, security, and national economic competitiveness. - (6.3. p. 10) To promote innovation, use, and adoption of AI applications, standards could address many technical aspects, such as AI performance, measurement, safety, security, privacy, interoperability, robustness, trustworthiness, and governance. - (7.1. p. 12) agencies should explain whether the action is intended to address a market failure (e.g., asymmetric information), clarify uncertainty related to existing regulations, or address another factor, such as protecting privacy or civil liberties, preventing unlawful discrimination, or advancing the United States' economic and national security."

Share	"- (SEC 1 ART 3. p. 2) (II) Promotion of Fairness and Justice. Uphold inclusivity and tolerance, truly protect the lawful rights and interests of each relevant entity, promote fair sharing of AI benefits by all of society, and advance social fairness and justice and equality of opportunity. When providing AI products and services, fully respect and help vulnerable groups and special groups, and provide appropriate alternatives as necessary. - (SEC 4 ART 16 p. 4) Safeguard User Rights and Interests. Users should be clearly informed of the use of AI technology in products and services. The features and limitations of AI products and services should be indicated. Guarantee users' rights to be informed and 4 to consent. Provide simple and easy-to-understand solutions so that users can choose to use or exit AI modes. Do not place barriers to the equal use of AI by users." - (SEC 4 ART 14 p. 4) Respect Market Rules. Strictly comply with the various rules and regulations governing market entry, competition, transactions, and other such activities. Actively uphold market order, and create a market environment that is	"- (1.2. p. 4) Consistency is also ensured with the EU Charter of Fundamental Rights and the existing secondary Union legislation on data protection, consumer protection, non-discrimination and gender equality. - (3.5. p. 11) With a set of requirements for trustworthy AI and proportionate obligations on all value chain participants, the proposal will enhance and promote the protection of the rights protected by the Charter: the right to human dignity (Article 1), respect for private life and protection of personal data (Articles 7 and 8), nondiscrimination (Article 21) and equality between women and men (Article 23)."	"- (4.5. p. 5) Executive Order 12866 calls on agencies to ""select those approaches that maximize net benefits (including potential economic, environmental, public health and safety, and other advantages; distributive impacts; and equity). - (7.1. p. 13) The analysis of these alternatives should also evaluate, where relevant and appropriate and consistent with Executive Order 13859, impacts to equity, human dignity, fairness, potential distributive impacts, privacy and civil liberties, personal freedom, and other American values." "- (5.3. p.8) Whenever relying on work done by private sector or other stakeholders or collaborating with them, agencies must ensure that their actions do not contribute to entrenchment by market incumbents or erect barriers to entry. - (7.4. p. 14) Agencies should also consider that an AI application could be deployed in a manner that yields anticompetitive effects that favors incumbents at the expense of new market entrants, competitors, or up-stream or down-stream business partners." "- (4.5. p. 5) the application of existing law to questions of responsibility and liability for decisions made by AI could be unclear in some instances, leading to the need for agencies, consistent with their authorities,

	conducive to AI development. Do not subvert orderly market competition through data or platform monopolies. Prohibit any method of intellectual property rights infringement by other entities.		to evaluate the benefits, costs, and distributional effects associated with any identified or expected method for accountability. Executive Order 12866 calls on agencies to ""select those approaches that maximize net benefits (including potential economic, environmental, public health and safety, and other advantages; distributive impacts; and equity). Agencies should, when consistent with law, carefully consider the full societal costs, benefits, and distributional effects when considering regulations related to the development and deployment of AI applications. - (5.2. p. 7) agencies should consider periodically informing the general public about emerging trends to help coordinate research efforts, new or emerging changes that will affect particular stakeholders (e.g., consumers), and transparency about how specific AI applications generate net benefits and, if relevant, distributional effects. - (7.1. p. 13) The analysis of these alternatives should also evaluate, where relevant and appropriate and consistent with Executive Order 13859, impacts to equity, human dignity, fairness, potential distributive impacts, privacy and civil liberties, personal freedom, and other American values."

Collaboration	- (ART 9. p. 4) Promote Tolerance and Openness. Fully value the rights, interests, and claims of every AI stakeholder. Encourage the application of diverse AI technologies to solve the actual problems of economic and social development. Encourage exchanges and cooperation across academic disciplines, areas of research, regions, and international boundaries. Promote the formation of AI governance frameworks and standards that have a far-reaching consensus.	"- (1.3. p. 5) The proposal also strengthens significantly the Union's role to help shape global norms and standards and promote trustworthy AI that is consistent with Union values and interests. It provides the Union with a powerful basis to engage further with its external partners, including third countries, and at international fora on issues relating to AI. - (3.1. p. 7) It targeted all interested stakeholders from the public and private sectors, including governments, local authorities, commercial and non-commercial organisations, social partners, experts, academics and citizens. After analysing all the responses received, the Commission published a summary outcome and the individual responses on its website. - (3.2. p. 8) The proposal builds on two years of analysis and close involvement of stakeholders, including academics, businesses, social partners, non-governmental organisations, Member States and citizens." - (1.1. p.3) The proposed rules wil be enforced through a governance system at Member States level, building on already existing structures, and a cooperation mechanism at Union level	"- (5.1. p. 7) This may also include work done in collaboration with industry, such as development of playbooks and voluntary incentive frameworks. - (5.3. p. 8) Whenever relying on work done by private sector or other stakeholders or collaborating with them, agencies must ensure that their actions do not contribute to entrenchment by market incumbents or erect barriers to entry." - (6.4. p. 11) They can also minimize the risk of unnecessary regulatory divergences from risk-based approaches implemented by key U.S. trading partners. In addition, agencies should consider existing international frameworks to which the United States has committed itself and the development·of strategic plans for coordination and cooperation with international partners. - (6.4. p. 10) Executive Order 13609, "Promoting International Regulatory Cooperation," calls on the Regulatory Working Group, which was established by Executive Order 12866, to consider "appropriate strategies for engaging in the development of regulatory approaches through international regulatory cooperation, particularly in emerging technology areas. - (6.4. p. 11) Accordingly, agencies should engage in dialogues to promote compatible

		with the establishment of a European Artificial Intelligence Board.	regulatory approaches to AI and to promote American AI innovation, while protecting privacy, civil rights, civil liberties, and American values.
Sustainability	"- (SEC 2 ART 5 p. 3) Promote Agile Governance. Respect the laws of AI development, fully understand AI's potential and limitations, and continually optimize governance mechanisms and approaches. In the processes of strategic decision-making, institution building, and resource allocation, do not deviate from reality, and do not be eager for quick success and short-term benefits. Promote the healthy and sustainable development of AI in an orderly manner. - (SEC 5 ART 18 p. 5) Promote Well-Intentioned Use. Strengthen pre-use demonstrations and assessments of AI products and services. Gain a full understanding of the benefits of AI	- (1.1. p. 1) By improving prediction, optimising operations and resource allocation, and personalising service delivery, the use of artificial intelligence can support socially and environmentally beneficial outcomes and provide key competitive advantages to companies and the European economy. - (1.1. p. 1) Such action is especially needed in high-impact sectors, including climate change, environment and health, the public sector, finance, mobility, home affairs and agriculture. - (3.5. p. 11) The right to a high level of environmental protection and the improvement of the quality of the	- (3. p. 2) As stated in Executive Order 13859, "the policy of the United States Government [is] to sustain and enhance the scientific, technological, and economic leadership position of the United States in AI. - (4.5. p. 5) Executive Order 12866 calls on agencies to "select those approaches that maximize net benefits (including potential economic, environmental, public health and safety, and other advantages; distributive impacts; and equity).

	products and services, and give full consideration to the lawful rights and interests of each stakeholder. More effectively promote economic prosperity, social progress, and sustainable development." - (SEC 1 ART 3 p. 3) (I) Persist in giving priority to the public interest, promote human-computer harmony and friendliness, improve the people's livelihoods, enhance the sense of gain and the sense of well-being, advance economic, social, and ecological sustainable development, and jointly build a community of common destiny for humanity (人类命运共同体).	environment (Article 37) is also relevant, including in relation to the health and safety of people. - (5.2.7. TITLE IX p. 16) Those codes may also include voluntary commitments related, for example, to environmental sustainability, accessibility for persons with disability, stakeholders' participation in the design and development of AI systems, and diversity of development teams.	
Long-term AI	NA.	NA.	NA.