Navigating the AI Ethics Frontier: A Cross-national Comparison of AI Policy Documents for Developing Responsible AI Workforce

Emad Ali, Virginia Polytechnic Institute and State University

Emad Ali is a Ph.D. student in the Department of Engineering Education at Virginia Tech. His research focuses on the integration of artificial intelligence in engineering education with particular emphasis on its ethical implications. He holds a master's degree in Electrical Engineering with specialization in Embedded Systems, from the Information Technology University (ITU Lahore), as well as a Bachelor's degree in Avionics Engineering from the National University of Sciences and Technology (NUST), Islamabad. He has worked as TA of multiple undergraduate and graduate courses during his Masters in ITU and as a Lecturer in the Superior University in Lahore. He aims to contribute to the advancement of educational practices in engineering by addressing both the opportunities and challenges presented by the emerging technologies.

Bailey Kathryn McOwen, Virginia Polytechnic Institute and State University

Bailey McOwen is a Ph.D. student in Engineering Education at Virginia Tech with an academic foundation in physics and industrial engineering. Her research focuses on workforce development, professional training for engineering practitioners, and engineering ethics, with an emphasis on how emerging technologies can enhance continued education. Through her research, service, and academic work, she aims to bridge engineering education and industry needs to support ethical, lifelong learning across diverse engineering pathways.

Arsalan Ashraf, Virginia Polytechnic Institute and State University

Arsalan Ashraf is a Ph.D. student in the Department of Engineering Education at Virginia Tech. His research interests include AI ethics, ethics and social responsibility, and lifelong learning. He has broad experience in academia and industry, which motivates him to research these vibrant areas. He is a first-generation student from a small village in Punjab, Pakistan. He completed his B.S. in Aviation Management from Lahore in 2017 and then worked in the aviation industry. After gaining experience, he changed his career path and entered academia as a lecturer at The University of Faisalabad in 2022. He joined Virginia Tech in the fall of 2024.

Dr. Dayoung Kim, Virginia Polytechnic Institute and State University

Dr. Dayoung Kim is an Assistant Professor of the Department of Engineering Education (College of Engineering) at Virginia Tech and a Director of the LABoratory for Innovative and REsponsible ENgineering workforce (LAB-IREEN). She conducts research in engineering practice and workforce development (e.g., practices and experiences of, and competencies required for, engineers in various employment settings, such as business organizations and government agencies; workforce challenges and opportunities from emerging technologies, such as Artificial Intelligence), engineering ethics (e.g., social responsibility of engineering professionals and its education), and related policy concerns. Dr. Kim's research aims to identify effective strategies to cultivate an innovative and responsible engineering workforce through educational initiatives and science & technology policy. She received her Ph.D. in Engineering Education at Purdue University (2022) and received her B.S. and M.S. in Chemical Engineering at Yonsei University (2017) and Purdue University (2021) respectively. She was the 2022 Christine Mirzayan Science & Technology Policy Graduate Fellow of the National Academies of Sciences, Engineering, and Medicine and received the 2022 College of Engineering Outstanding Graduate Student Research Award from Purdue University. She was also a Bilsland Fellow of the College of Engineering at Purdue University when she was a Ph.D. student.

Navigating the AI Ethics Frontier: A Cross-national Comparison of AI Policy Documents for Developing Responsible AI Workforce

1 Abstract

With the rapid growth of Artificial Intelligence (AI), and the world becoming a global village, the ethical implications of AI are gathering concerns from the technologists, policymakers, and ethicists around the globe alike. Given the powerful impact of AI and its ability to shape the lives of human beings in the coming years, it is of utmost importance to nurture responsible AI engineers who can design ethical AI systems. To be able to nurture such engineers, it would be helpful for educators to understand existing efforts to govern AI ethically and responsibly. In this paper, we focus on the five stakeholders of AI, namely: the USA, Great Britain, European Union, China, and India. The reasons for selection of these countries are the significant role they play on the global level, their population diversity (India and China present one-third of the world's population, a population that migrates in huge numbers to the other three regions mentioned above), and varying cultural perspectives.

This paper analyzes the documents that discusses the AI policy of each mentioned country, to take a deep dive into the frameworks that were developed for regulation of AI technology. The paper will introduce the policy documents presented by the governments and then conduct comparative analysis. We will answer the following question: *What are the similarities and differences in ethical concerns and mitigation strategies discussed in the policy documents across these five countries?* The findings of this study will allow the engineering faculties to see a comprehensive picture of the existing efforts to govern AI ethically, and incorporate these in education of engineers of tomorrow. Naturally, the educational implications of the findings will also be discussed.

Keywords— AI Policies, AI Ethics Frameworks, Global AI Policies, AI Ethics Education

2 Introduction

The concept of Artificial Intelligence (AI) has existed for years, with first being published by Alan Turing, in his paper "Computer Machinery and Intelligence" [1]. From the time of simple "Turing Test", the domain of AI has seen a massive boost, specifically after the advent of Large Language Models (LLMs) [2], that powers the Generative AI tools, such as Chat GPT. With such a rapid rise, AI has become a game changer in many industries [3], including healthcare [4], education [5], finance [6], and national security [7], due to its ability to automate tasks [8], improve decision making [9], and enable innovation [10]. The rise of AI and its integration into such important factors have raised concerns of

algorithmic biases [11], data privacy [12], accountability [13] and regulatory oversight [14]. Today's discussions on the global scale of AI focus on how best to foster innovation responsibly.

Given these challenges, governments around the world have created AI policies and regulatory frameworks to mitigate risks and promote the development of responsible AI [15, 16]. The United States, United Kingdom, European Union, China, and India are among the world's most influential AI economies [17], and contribute most to research of AI [18], and each has adopted different governance models according to their socio-political, economic and cultural priorities. Understanding current policy initiatives of those five countries will help engineering educators to be better prepared to nurture AI engineers who can innovate responsibly.

Therefore, this research paper seeks to identify the similarities and differences in the policy initiatives, especially for ethical AI systems, across the five countries - The United States, United Kingdom, European Union, China, and India. Based on the analysis, we will identify implications for engineering education, specifically, how AI ethics can be integrated into engineering curricula to enable future AI practitioners to create fair, transparent, and accountable AI systems. We answer the following research question: *What are the similarities and differences in ethical concerns and mitigation strategies discussed in the policy documents across these five countries?*

The justification for choosing the five countries is as follows: 1) These five countries/regions (namely: USA, UK, China, EU and India) provide a large portion of the world's AI research, investment, and market influence. For example, the US, UK, EU, and China combined contributed 101 notable Machine Learning (ML) algorithms, whereas India was the second largest contributor to Github's AI project [19]. 2) The U.S. and China are leading in the development and application of commercial AI, and control the biggest AI driven economies [20]. 3) The EU and UK are at the forefront in the discourse on AI regulation and ethics and are developing the norms of the international community with their policies [21]. 4) India is unique in its AI policy focus on inclusive AI development, making it a useful country to examine when it comes to AI in developing economies [22].

For a rigorous and representative analysis, this study reviews official AI strategy and regulatory documents. Below is a list of documents chosen and reasoning for the choice:

- **U.S.'s Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People** [23]: A non-binding document from the Office of Science and Technology Policy (OSTP) which represents US government's ethical AI principles , that demonstrates the industry-driven approach and focuses on the transparency, accountability and human rights, making it an essential read to understand AI self-regulation.
- **UK's National AI Strategy** [24]: An official policy document by UK government, that showcases innovation through sector-led governance approach, and AI Assurance framework provides guidance for the development of ethical AI.
- **India's National Strategy for Artificial Intelligence** [25]: An official policy document that presents a development-oriented approach, and highlight AI's goal in emerging economies making it as valuable case for AI accessibility and digital inclusion.
- **EU's Artificial Intelligence Act** [26]: Perhaps one of the most comprehensive and legally binding document that regulates the development of AI in 27 nations of the world, hence setting benchmarks for ethical AI's global standardization.

- **China's Next Generation Artificial Intelligence Development Plan** [27]: Official document from the Chinese government that outlines China's long-term development goals related to AI, and helps understand the AI's role in government oversight and digital sovereignty.

This paper will present an overview of all these policy documents and then focus on the ethical concerns and mitigation strategies presented in these policy documents. Toward the end, the discussion section will present similarities and differences across the ethical concerns and their mitigation strategies discussed in the documents and how the findings can be used to integrate AI ethics into engineering education.

3 Overview of AI Policy Documents

3.1 US's Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People

The "Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People" [23], hereafter mentioned as the AI Bill of Rights, is a white paper published by the White House Office of Science and Technology Policy (OSTP) in October 2022, outlines five principles intended to guide the ethical and responsible development and use of automated systems in the United States. While it is not an official policy document, it seeks to protect the American public from potential harms arising from the use of AI while fostering innovation and leveraging their benefits. While acknowledging the transformative potential of AI to improve lives, the document underscores concerns about its misuse, citing cases where automated systems have proven unsafe, ineffective, biased, and intrusive. The AI Bill of Rights states that these principles should be considered whenever automated systems have the potential to meaningfully impact the public's rights, opportunities, or access to essential resources and services. The following paragraph presents the five principles of the AI Bill of Rights together with their descriptions and expectations and practical steps for stakeholders to implement these protections.

The first principle, *Safe and Effective Systems*, emphasizes the importance of pre-deployment testing, risk identification and mitigation, ongoing monitoring and clear organizational oversight for the safe and effective operation of automated systems. It stresses the need for robust evaluations to verify that systems are operating as intended and do not present an undue risk to public safety or well-being. The second principle, *Algorithmic Discrimination Protections*, deals with preventing algorithmic bias and ensuring fairness through proactive measures. This includes the use of representative data, protections against proxies for demographic features, integrated accessibility for people with disabilities, and thorough disparity testing and mitigation efforts. These measures are designed to protect against discrimination and ensure equitable outcomes. *Data Privacy*, the third principle, highlight the importance of privacy protection for responsible AI usage, and advocates for measures such as privacy-by-design, data minimization, and transparency in data usage. This principle supports limiting the collection and usage of data, and enhanced oversight on the surveillance technologies, specifically in sensitive domains, such as education, employment, housing. Fourth principle, titled *Notice and Explanation*, underscores the need to provide clear and easy-to-read and understand information on the use of automated systems. It includes informing the users about the system changes, and providing explanations of the working of said systems. The purpose of this principle is to promote transparency and provide sufficient information to users to make informed choices. The fifth principle, *Human Alternatives, Consideration, and Fallback*, emphasizes the right of individuals to opt out of automated systems in favor of human alternatives where appropriate, while highlighting the importance of accessible, and effective fallback mechanisms. It places the accountability of decision to humans overseeing the working of said systems.

The AI Bill of Rights is not a legally binding document and the values reflected in it are not enforced law. Instead, it is a guideline that can assist in the development of policies and practices that protect civil rights and uphold democratic values in the age of AI systems.

3.2 UK's National AI Strategy

The "National AI Strategy", hereafter mentioned as NAS-UK outlines the United Kingdom's 10-year plan to maintain its position as a global leader in AI, emphasizing ethical considerations, risk mitigation, and strategies to ensure AI benefits all sectors of society [24]. The document acknowledges the transformative potential of AI while recognizing the unique challenges it presents, which require a comprehensive approach to governance and regulation. Three core pillars of investing in the AI ecosystem mentioned in NAS-UK are discussed briefly.

The first pillar of the plan is *Investing in the Long Terms Needs of the AI Ecosystem*. This is a call to action to build on the UK's potential for sustained AI leadership by addressing critical areas of fostering skills, research and development, data collection, computation power, financial resources, and AI commercialization. The document also emphasizes the need to attract and train AI talent, support innovative research, and ensure high quality data availability. The document outlines how the British government will increase compute capacity, support chip design, provide private and public funding for AI innovators, and expand market access. This pillar also supports the promotion of collaboration between AI developers and businesses to create useful and ethically sound AI solutions. The second pillar is *Ensuring AI Benefits All Sectors and Regions* which is an equitable deployment of AI across all sectors and regions. It includes growing the adoption of AI in industries, supporting public sector leadership, and protecting intellectual property and innovation in AI. The pillar emphasizes the use of AI to tackle social challenges, including climate change and health. The last pillar is *Governing AI Effectively* which provides a framework for the proper regulation and governance of the general principles of AI governance, safety, enhancing fairness, and trustworthiness. The initiative includes the release of a White Paper on AI governance together with improved regulatory capabilities and AI assurance tool development. The UK government seeks to establish leadership in AI standard development while promoting worldwide governance collaboration to ensure democratic value alignment. NAS-UK establishes measures to tackle AI risks through efforts against misinformation, algorithm transparency and advanced AI safety measures.

The strategy envisions a future where the UK leverages its strengths in research, innovation, and regulation to maximize the benefits of AI while upholding ethical and democratic principles, and safeguarding the public interest.

3.3 India's National Strategy for Artificial Intelligence

The "National Strategy for Artificial Intelligence #AIFORALL" [25], hereafter mentioned as NSAI, presents India's ambitious vision for leveraging AI to drive inclusive growth and position itself as a global leader in the AI revolution. This strategy has identified a multi-faceted approach to the development of AI, research, commercialization, workforce readiness, ethical AI, and governance frameworks. Following are the core pillars of this strategy detailed out in respective sections.

To boost India's AI research ambitions, the NSAI has divided R&D into two tiered centers. Sensory AI, Physical AI, Cognitive AI, General AI, Explainable AI (XAI), and high-precision learning from small datasets will all be subareas of focus for basic research at the Centers of Research Excellence (COREs). They will also engage in the mentoring of research institutions and fundamental AI innovation. The

International Centers of Transformational AI (ICTAIs) will link the country's academic and industry sectors and concentrate on application-based research and commercialization of AI technologies. ICTAIs will develop prototypes to deployable AI solutions and have a structured relationship with COREs for technology transfer. All these activities will be supervised by an umbrella organization that will be involved in international cooperation, socioeconomic impact assessment, and financial stability. The strategy also encourages 'moonshot' projects, under which ICTAIs have the freedom to initiate high-risk, high-reward AI innovations.

To develop the skills necessary for the AI age, the strategy emphasizes the need to realign the education sector, especially STEM. It proposes a shift from knowledge-based to skill-based education. It also laments the need to incentivize the IT sector to re-skill its employees. In this regard, the use of online tools like Coursera and edX will be promoted. The NSAI also proposes the creation of a formal marketplace, the National AI Marketplace (NAIM), for raw data, annotated data, and deployable AI models. This idea aims at reducing information asymmetry, opening up data sources, and resolving the ethical issues of data sharing.

NSAI recommends an "Accelerating Adoption" strategy to build an ecosystem with public-private partnership, start-up support, and regulatory integration. It will focus on sectors with high socio-economic impact for AI adoption, such as agriculture, healthcare, smart cities, and transportation. The Indian government will offer incentives to enterprises to adopt and implement AI-based solutions and reduce the bureaucratic burden through regulatory framework, for seamless integration.

NSAI laments the need for responsible AI systems, with transparent and no algorithmic biases, which it considers as key to address the ethical concerns. It also proposes development of Consortium of Ethics Council at each CORE to ensure that the AI technology is developed ethically. It also proposes a framework to address the damages caused by AI. The NSAI declares government's role as that of a facilitator and promoter, and provides support through finances, awareness, startups and data privacy laws.

3.4 EU's Artificial Intelligence Act

The European Union (EU) AI Act [26] outlines a regulatory framework for artificial intelligence (AI) systems within the European Union. This framework aims to ensure the responsible and ethical development, deployment and use of AI systems while protecting fundamental rights, safety, and public interests. The Act focuses primarily on high-risk AI systems, defined as those with significant potential for harm to health, safety, or fundamental rights. These systems are subject to specific requirements, including risk management systems, data quality standards, technical documentation, human oversight, transparency measures, and conformity assessment procedures. The EU AI Act provides a detailed breakdown of these requirements and the obligations of various stakeholders in the AI value chain, including providers, importers, and distributors.

The EU AI Act establishes high-risk system regulations, but acknowledges situations where AI systems operating in potentially high-risk areas do not present dangers. The document establishes specific criteria to determine when exceptions apply to profiling AI systems operating in critical sectors such as hiring, credit scoring, and law enforcement.

The EU AI Act addresses the problems of new generative general-purpose AI models which can be integrated into multiple systems and generate systemic risks. The Act provides a framework to identify systemic risks through an assessment of general purpose AI model capabilities and their impact on the internal market. The deployers must document all characteristics together with performance details and

training data for these models. The Act creates the European AI Office (AI Office) to enforce compliance and provide guidance while promoting collaboration.

The Act has a major focus on transparency and accountability. It mandates transparency in the development of high risk AI systems such that their functionality and limitations are understandable by their deployers. This includes: detailed instructions for use that describe performance characteristics, potential risks, and human oversight measures. The Act also has requirements for AI developers to explain AI decisions clearly. The Act details a robust enforcement mechanism, including tiers of fines, post-market monitoring systems and powers granted to authorities protecting fundamental rights. The Act calls for regular reviews and evaluations to ensure its continued relevance in light of technological advancements.

3.5 China's Next Generation Artificial Intelligence Development Plan

The "Next Generation Artificial Intelligence Development Plan (NGAIDP)" issued by China's State Council in July 2017 [27], lays out the country's ambitious road map for becoming a global leader in artificial intelligence by 2030. The plan outlines a three-step approach to achieve the objectives, that are: catch up to global AI leadership by 2020, make major breakthroughs by 2025, and establish China as the world's primary AI innovation center by 2030.

To achieve these objectives, it outlines four key principles that are briefly explained below. The *technology-led* approach focuses on driving advancements through scientific breakthroughs and innovation, ensuring that technological progress aligns with broader societal and economic goals. The *systems-layout* principle emphasizes a strategic and coordinated approach to advancing AI, recognizing its complexity and multifaceted nature, and advocates for a targeted development strategy that tailors efforts to the distinct requirements of foundational research, technological research and development (R&D), industrial growth, and commercial applications. The *market-dominant* strategy emphasizes that the market should be the primary driver of AI development and commercialization, with government providing policy guidance, regulatory support, and ethical oversight. Lastly, the *open-source and open* principle encourages a collaborative ecosystem through shared resources and global partnerships, fostering innovation while integrating civilian and military technological advancements.

The plan also identifies several areas of focus, such as building an open and coordinated AI innovation system, fostering highly efficient economy systems, development of a safe intelligent society, strengthening civilian-military integration in AI domain, building safe and efficient intelligent infrastructure, and implementation of major science and technology projects for a new generation of AI. These key areas comprehensively cover wide range of R&D, industry, societal and infrastructural aspects to achieve the objectives set forward by the plan. The plan further calls for significant investment in basic and applied research, development of key AI technologies, establishment of AI innovation platforms, and development of an AI skilled workforce to achieve these objectives. The plan also places importance on the need for collection, storage, processing and sharing of datasets required for comprehensive AI systems development. The plan also outlines measures for support of AI systems and society, such as developing laws, regulations, and ethical norms for addressing the responsibility, data privacy, and accountability.

Overall NGAIDP is a reflection of China's comprehensive vision of leveraging AI to drive economic growth, enhance social well-being and boost national security. By 2030, the NGAIDP envisions a China where AI has become deeply integrated into all aspects of life, from industry and commerce to education, healthcare, and social governance, transforming the country into a global leader in the AI revolution.

4 Ethical concerns and their mitigation strategies

This section discusses the ethical concerns that were discussed in the policy documents, which are summarized in the previous section, and the mitigation strategies they discuss to overcome those challenges. A brief summary of these concerns and their mitigation strategy is provided in table 1.

Table 1: Ethical concerns and their mitigation strategies

Year	Policy	Country	Ethical Concerns	Mitigation Strategies
2022	Blueprint for an AI Bill of Rights [23]	USA	Unsafe and ineffective automated system	Consultation with diverse stakeholders, Rigorous pre-deployment testing, risk identification, ongoing monitoring, Use of relevant & high-quality data, establishing clear organizational oversight & governance structures, & independent evaluation by a third party.
			Algorithmic discrimination	Proactive assessment of equity, representative & robust datasets usage. Predeployment and on going disparity testing. Independent evaluation, and plain language reporting. Guarding against the use of proxies
			Abusive data practices and privacy violations	Privacy implementation by design implementation, providing individuals with clear and understandable notice. Heightened oversight of surveillance technologies, and enhanced protections and restrictions for sensitive data. No continuous and unchecked surveillance.
			Lack of transparency and explainability	Clear, timely, and understandable notice, Generally accessible documentation, and up-to-date notices and explanations.
2023	National AI Strategy [24]	UK	Lack of diversity and inclusion	Increasing diversity in the AI ecosystem. Ensuring diverse teams are involved in the development and deployment of AI systems. Providing up to 1,000 scholarships for postgraduate students
			Algorithmic Bias	Diverse and high quality data. Algorithmic Transparency. Regulatory Frameworks and Guidance, and International Collaboration.
			Lack of transparency and explainability	Algorithmic Transparency Standard, AI Assurance Ecosystem, and technical standards.
2018	National Strategy for AI [25]	India	Security Concerns	The strategy includes a framework for addressing damages caused by AI, including negligence tests, safe harbors, and apportionment of damages.
			Lack of transparency and explainability	Addresses the need to tackle biases in AI systems and open the ”black box” of AI to ensure transparency
			Algorithmic Bias	Diverse, well-represented datasets. Identify the bias. Assess it’s impact. Find ways to reduce it, including rigorous testing, audits, external oversight.
			Data privacy violations	The strategy calls for the establishment of a data protection framework with legal backing, sectoral regulatory frameworks, and adherence to international standards.
2024	EU AI Act [26]	EU	Harm to fundamental rights	Continuous risk management system to be established to identify, assess, & mitigate potential harms. Fundamental rights impact assessment must be done to identify and mitigate risks. Specific requirements and considerations to protect vulnerable groups.
			Bias and discrimination	Data providers must take steps to identify, document, and mitigate potential biases in AI datasets and systems. Training, validation, and testing datasets for high-risk AI systems must be relevant, representative, error-free, & comprehensive to minimize bias. Providers can process special categories of personal data to ensure bias detection and correction in high-risk AI systems.
			Lack of transparency and explainability	Humans have the right to know they are interacting with AI, and right to get explanation of the decision taken by AI. The AI must have documentation that outlines all its characteristics and limitations.
			Surveillance and privacy violations	Processing of personal data must comply with the EU’s General Data Protection Regulation (GDPR) and Law Enforcement Directive. Scraping of facial images without consent is also banned.
			Emotion recognition and manipulation	Use of emotion recognition is prohibited in educational settings & workplace, along with the use of manipulation techniques.
2017	Next-Generation AI Development Plan & Ethical Norms [27, 28]	China	Bias and discrimination	Strengthen ethics investigation, consider the needs of different groups, strive to achieve AI system inclusivity, fairness, and non-discrimination.
			Information leaks and violation of personal privacy	Handle personal information with principles of legality, propriety, & necessity. Guarantee personal privacy rights, & data security. Improve security protection system, and implement design accountability, and supervision.
			Safety risks	Ensure the deployment of AI in a safe, reliable, and controllable manner. Strengthen research on legal, ethical, and social issues related to AI. Clarify the roles of AI, rights, obligations, and responsibilities.
			Transformation of employment structures	Skills training. Re-employment training and guidance. New job creation.

4.1 US’s AI Bill of Rights

The AI Bill of Rights discusses various ethical concerns that are raised by the use of AI systems such as unsafe or ineffective systems, stating that people should be protected from them. It discussed some cases in

which the use of technology can cause unintended damage. To mitigate this concern the AI Bill of Rights suggests consulting with diverse stakeholders, including impacted communities and domain experts, and it implores the need for rigorous pre-deployment testing and risk identification and mitigation, along with ongoing monitoring of the systems to ensure systems operate as intended. It further suggests using third-party experts for independent evaluation of AI systems for safety, effectiveness, and fairness of AI systems.

The AI Bill of Rights also discusses algorithmic discrimination and defines that algorithmic discrimination occurs when an automated system contributes to unjustified treatment or disfavoring of people based on race, color, ethnicity, sex, religion, age, national origin, disability, genetic information, or any other classification that is protected by the law. To mitigate these biases and reduce algorithmic discrimination, the AI Bill of Rights stressed on the importance of proactive assessment of equity in the design phase, and using representative dataset that reflect the diverse population while incorporating the historical, and societal context accurately. Furthermore, it also guards against the use of proxies (attributes that are highly correlated with demographic features), as it can lead to algorithmic discrimination which is prohibited by law.

The document also addresses the ethical concern of abusive data practices and privacy violations. It highlights the misuse of data, such as unauthorized collection, use, sharing and storage. The AI Bill of Rights suggests implementing privacy-by-design practice, which include providing clear and understandable notices about data collection and its use. Another suggestion the document provides is that sensitive data such as health, financial, and criminal records need to be kept under enhanced protection, and the person whose data have been used, must have access to the information of who has access to their data, how it is being used, and the right to withdraw consent for the usage of data. It further discusses protecting the American people from unchecked surveillance, by suggesting to limit the surveillance unless strictly necessary, and conducting pre-deployment potential risk assessment, along with heightened oversight during use.

Another ethical issue the document tackles is the lack of transparency and explainability, which are cornerstone for ensuring public's trust in the AI systems. The document suggests that providing clear, timely, and understandable notice when an automated system is being used and how it contributes to the outcome, along with accessible documentation that can describe the overall system functionality, can make automated systems explainable, and thus foster public's trust.

AI Bill of Rights also discusses the over-reliance on automated systems, specifically in high-risk areas, such as healthcare, finance, and justice, where AI-driven errors can lead to unfair, opaque, and unchallengeable decisions, and may limit the avenues to readdress them. To mitigate these concerns, the white paper recommends having accessible human intervention channels, and presenting people with choice to request alternatives to automated systems, when the automated decision infringe upon their rights. It also recommends organization to develop clear accountability frameworks to take responsibility of the decisions taken by their AI-systems.

4.2 UK's National AI Strategy

NAS-UK identifies following ethical concerns in the development of AI, mainly in the second and third pillar: lack of diversity and inclusion, algorithmic bias, and lack of transparency and explainability. With the growth of AI, NAS-UK determines that it is imperative to increase the diversity in AI ecosystem, the government is doing that by providing scholarships to under-represented postgraduate student in the domain of AI and Data Science, as part of its long term goal. According to the data, 1200 students enrolled

in the first year, out of which 40% were women out of which 76% received scholarship, 25% were black out of which 45% received scholarship, and 15% of the total cohort were disabled students, out of which 24% received scholarships. It highlights the government's interest in attracting talent from around the globe in order to have developers with diverse opinions, skills, background and experiences.

NAS-UK acknowledges that AI systems' autonomy and data-driven decision making approach can introduce new pathways for bias, impacting fairness and might perpetuate existing inequalities, therefore it highlights the need to mitigate the algorithmic bias, along with the need of fairness in all the decisions made by AI systems. To ensure this, the document necessitates the collection of diverse and high quality data, which will be collected, stored and shared under strict regulatory frameworks. It also highlights the government interest in collaborating with international partners to mitigate bias.

One of the biggest factor of public trust in the AI systems, is its ability to be transparent and explainable of how it got to the decision it presented. In this regard, NAS-UK outlines the development of a mature AI Assurance Ecosystem, along with the technical standards that will be upheld. The policy document also highlights that Central Digital and Data Office (CDDO) is in the process of conducting research, with aim of developing a cross-government standard for algorithmic transparency. This transparency is vital for building trust and ensuring accountability in AI applications.

4.3 India's National Strategy for AI

NSAI regards AI as the tipping point in technological evolution, and discusses the approach towards responsible AI, and discusses the concerns related to biases, transparency, privacy, and security. NSAI places the issue of fairness at forefront of discussion in academic, research and policy fora. It recommends identifying the biases in the system, assessing their impact, and find ways to reduce them as the best way to overcome biases in the dataset.

To improve transparency of the AI systems, NSAI suggests opening the 'black box', of the decision making algorithm. It suggests making COREs and ICTAIs adopt ethical practices, and places obligation on Indian government to setup a consortium of ethics councils at each CORE and ICTAI to define the standard practices and monitor their adoption.

To safeguard the privacy of data, the NSAI recommends instituting a data privacy legal framework, creating sectoral regulatory guidelines, and collaborating on privacy preserving technology research in AI. NSAI puts onus on the Indian government, to address and implement data protection framework that protect human rights and privacy, without stifling the innovation, collaborate with industry to come out with sector specific guidelines on privacy, security and ethics, and support COREs to do research in new mathematical models and technology for preserving privacy.

NSAI promotes doing sustainable research to overcome the security issues in AI systems, and recommends setting up Centre for Studies on Technological Sustainability (CSTS), and recommends that Indian government sets up CSTS to address issues relating to ethics, privacy, legal aspects, and social sustainability.

4.4 EU's AI Act

The EU Act being a legal document, places special importance on the effects AI carries on the humans. It discusses in detail ethical concerns such as protection of fundamental rights (recital 48), biases (recital 67), transparency (recital 101 & 102), surveillance (recital 69), and emotion recognition (recital 18). Practices that violate fundamental rights are strictly prohibited by the EU Act. It safeguards the fundamental human

rights by stating the a continuous risk management system should be established, implemented, documented and maintained throughout the entire life-cycle of high-risk AI systems to identify, assess, and mitigate potential harms. Deployers of high-risk AI systems must conduct fundamental rights impact assessments to identify and mitigate risks to individuals or groups.

To mitigate the bias and reduce the discrimination, training, validating, and testing datasets for AI systems must be relevant, representative, and comprehensive. The data providers must take steps to identify, document, and mitigate potential biases in AI systems, and datasets. Furthermore the AI systems must be accompanied by documents that outline their characteristics, capabilities, limitations, and risks associated with them, to ensure transparency. Moreover, the humans have the right to obtain explanation of the decisions made by the AI systems.

The processing of personal data, including biometric details, must comply with EU's General Data Protection Regulation (GDPR), and Law Enforcement Directive. The EU Act bans the expanding of facial recognition database through untargeted scraping of facial images to protect the privacy of humans. Use of emotion recognition software in places like education sector, and workplace is also prohibited, along with the use of data manipulation techniques.

4.5 China's NGAIDP

NGAIDP and its additional ethics document, Ethical Norms for New Generation Artificial Intelligence [28], hereafter mentioned as Ethical Norms, lament on the challenges of biases (Ethical Norms, Section III, Article 13), information protection (Ethical Norms, Section I, Article 3(III)), safety risks (Ethical Norms, Section I, Article 3(V), Section IV, Article 17, and NGAIPD: Gurantee Measures (IV)), and the transformation of employment structures due to AI systems (NGAIPD: Strategic Objectives for 2030). It laments on the need to strengthen the ethics investigation during data collection and algorithm development, comprehensive considerations of various groups, and achieve inclusivity, fairness and non-discrimination. It stresses on the need for research on legal issues such as protection privacy and property, and information security utilization, and to protect the vital information from leaking, it outlines the need for personal data to be handled according to the principles of legality, propriety, necessity, and good faith, which guarantee personal privacy and data protection. It also warns against stealing, tampering with, leaking, illegal collection and utilization of personal information. To further increase public trust, it outlines a plan to establish an open and transparent AI supervision system, along with the establishment of a traceability and accountability system.

To govern the evolving field of AI systems, it laments on the need for solid research on ethical, legal, and social issues related to AI, along with establishing laws, regulations and ethical frameworks with the end goal of minimizing risks, and ensuring the development of AI systems that are safe, and reliable. With such a rapid change approaching, NGAIDP shows concern about the population that will loose their jobs to AI. In this regard, NGAIDP suggests developing lifelong learning and employment training systems, and supporting higher learning institutes, vocational schools, and training institutes to impart AI skills training.

5 Discussion

The analysis of various AI policy documents in the USA, UK, EU, India, and China shows that there are both similarities and differences in the approaches used to tackle the ethical issues pertaining to the AI development on the global level. After discussing the similarities and differences, we will discuss how engineering educators can utilize them to teach AI ethics to the engineers of tomorrow.

5.1 Similarities and Differences

Across all jurisdictions, AI policies underscore the need for responsible AI governance to mitigate the risks of bias, discrimination, lack of transparency, and privacy violations. Table 2 presents a comparative analysis of similarities and differences between the five documents, which are discussed below in detail.

Table 2: Similarities and Differences between Policy Documents

Countries	Bias & Fairness	Transparency	Privacy & Data	Accountability	Surveillance
USA	Voluntary fairness assessments, industry-driven	Industry self-regulation, explainability encouraged	Best practices recommended but no law	Encourages self-regulation and consumer rights	Discourages mass surveillance, but allows law enforcement use
EU	Mandated bias audits and mitigation	Mandatory documentation and traceability	GDPR-aligned, strict personal data collection	Legally binding risk management and liability	Prohibits untargeted facial recognition
UK	Encourages AI assurance framework	Sector-based transparency standard	GDPR-aligned, flexible for innovation	Flexible accountability across sectors	Encourages ethical AI but allows certain use
China	Acknowledges bias, focuses on state-led AI	Supports AI traceability in governance	AI data governance for security and state use	State-controlled AI governance	AI used in social credit system & public governance
India	Fairness through explainable AI models	Encourages explainability & trust building	Ethical AI data sharing through formal marketplaces	AI governance structures recommended	Ethical deployment of AI in governance

5.1.1 Shared Ethical Concerns

The five policies identify bias in AI-based decision-making as their main critical ethical concern. The U.S. AI Bill of Rights and the EU AI Act both stress proactive bias assessments and equitable data representation to prevent discriminatory AI models. NAS-UK adopts an assurance framework instead of strict legal mandates to address fairness. Both China and India acknowledge bias, but their focus is on delivering AI benefits equally to diverse population groups.

Another similar concern is around transparency, and AI explainability is a key theme in all documents; however, the implementation differs across all documents. EU Act mandates transparency through documentation and traceability requirements, while the U.S. AI Bill of Rights calls for industry-driven transparency measures. The UK's sector specific approach promotes explainability within its AI governance model. China and India stress AI transparency in public facing applications but don't have as strict of regulations as the EU.

Privacy, is another one of the concerns that finds its way in all these documents, however, only EU AI Act sets legal protection and aligns closely with GDPR. While the NAS-UK also aligns with GDPR, it allows flexibility for innovation. The AI Bill of Rights makes recommendation for best practices but lacks enforcement mechanisms. China's AI Ethical Norms discuss personal data security but also talks about integrating AI in surveillance applications. India's policy discusses lawful and ethical AI data sharing.

Along with these similar ethical concerns all the documents address the accountability of AI systems as well. The EU AI Act lays down clear accountability for AI developers through risk management and legal liability structures. The U.S. AI Bill of Rights encourages risk assessments by the industry, rather than legal obligations. UK's approach is sector-led, relying on individual industries to define AI accountability, while state oversight is emphasized by China and India for AI governance models.

5.1.2 Differences in Ethical Concerns

While there exists a substantial overlap between the ethical concerns, there is also some difference among them as well. For instance, the AI Bill of Rights emphasizes on the safe and effective autonomous systems, so that the AI systems work as intended. While safety is an implicit concern in other policies, the AI Bill of Rights is the only one to frame it explicitly, as a primary ethical challenge.

The EU AI Act holds distinction in introducing ethical concern related to emotion recognition, subliminal manipulation, and social scoring, labeling them as practices in violation of fundamental rights. None of the other policy documents discuss these concerns in detail, making them unique to EU AI Act. NAS-UK holds its distinctiveness as it discusses sustainability, the environmental effect of AI systems.

China's AI policy addresses the social stability and governance as ethical concerns, emphasizing the use of AI for social credit systems, predictive policing, and public surveillance. The NGAIPD lays this out as ethical challenge to maintain societal harmony and state security. India differs from others by identifying digital inclusion as an ethical concern, underscoring the need to provide access of AI systems to under-served communities. China and India also highlight the ethical concern regarding the workforce displacement by introduction of AI systems, a concept less touched upon by other policy documents.

5.1.3 Similarities and Differences in Mitigation Strategies

The AI policy documents of the USA, UK, EU, China, and India contain both common approaches to ethical issues and distinct approaches that are aligned with regional interests. Even though all documents address similar concerns such as bias, transparency, privacy, accountability, and surveillance, the ways in which these challenges can be addressed differ significantly across the frameworks.

To guarantee the fairness of AI systems, all five frameworks emphasize bias mitigation. A common approach shared in the policies is the use of high-quality and representative datasets to help minimize discriminatory outcomes. For example, the EU AI Act makes legally binding the audits and dataset reviews for high-risk AI systems, although it does not mandate fairness assessments; the USA AI Bill of Rights recommends that developers conduct voluntary fairness assessments. Similarly, India's NSAI promotes fairness through Explainable AI (XAI) and ethical oversight councils, and to ensure that AI systems serve the common good, China's policies promote general principles for unbiased data processing. In contrast, the UK has adopted a sector-focused approach with bias mitigation frameworks for industries such as healthcare and finance, which provide flexibility but lack comprehensive enforcement.

Transparency is universally recognized as an essential component for establishing public trust in AI systems. All frameworks stress the need for traceability, documentation, and explainability. This makes the EU AI Act stand out, as it mandates comprehensive documentation for high-risk AI system throughout the entire life cycle. By contrast, the USA and UK have adopted voluntary documentation practices, encouraging developers to adopt explainability standards without mandating them. India links transparency with trust building and has provided guidelines on how to make AI systems understandable for public-facing applications. China also supports transparency and integrates it with the compliance mechanisms, the latter of which are primarily for regulators rather than people.

The protection of personal data is a shared responsibility across all five documents and there is a collective emphasis on responsible data governance. The EU has the most stringent rules for the lawful processing of data with GDPR aligned rules. However, the USA has no federal AI-specific privacy law and instead counts on privacy-by-design, and state-level regulations. The UK has mirrored GDPR but with a more flexible approach, which allows for both innovation and data protection. India encourages ethical data sharing through measures like the NAIM, while China focuses on data protection within its state-controlled framework and sees privacy as secondary to national security.

From the documents, it is clear that the accountability of developers and deployers their AI systems is essential. The EU AI Act introduces legally binding risk management requirements, thus transforming accountability into a regulatory obligation for high-risk systems. In contrast, the USA and UK rely on voluntary accountability mechanisms; encouraging developers to do their own impact assessments and to maintain ethical oversight. India advocates for multi-stakeholder collaboration to build accountability, and China's centralized model places accountability under government oversight.

5.2 Integration of AI Ethics into Engineering Education

Engineering ethics education has taught professional responsibility of engineers (e.g., by introducing engineering codes of ethics), such as considering health, safety, and welfare of the public [29, 30]. However, with the advent of AI in various industries, there is a need for the engineering ethics curriculum to specifically include discussion around the ethical concerns related to AI [31]. As AI is increasingly incorporated into engineering practice, it is crucial to incorporate ethics education into AI education for engineers to prepare future engineers for the challenges of AI application [32, 33]. The commonalities across ethical concerns in the AI policy documents (bias mitigation, transparency, privacy, accountability, and risk management), which we identified through this study, can be prioritized to be emphasized in AI ethics curricula in engineering programs.

Specifically, engineering educators can: 1) teach how to mitigate bias in AI systems from various global regulatory perspectives, 2) teach privacy, transparency and accountability frameworks, and 3) prepare engineers to be familiar with AI governance and its impacts on engineering practice.

5.2.1 Teaching how to mitigate bias in AI systems

All five AI governance frameworks identify bias in AI systems as an important ethical concern. Therefore bias mitigation can be an important component of AI ethics education. Bias is an issue that engineers need to address particularly in the context of machine learning, predictive analytics, and automated decision-making. AI curricula should include bias audit methodologies used in the EU AI Act, UK AI Assurance Framework, and India's XAI initiative to expose students to how bias can be detected practically. Engineering education researchers have suggested that interdisciplinary collaboration and hands-on projects can be used to teach bias mitigation [34–37]. For instance, students should audit real-world AI datasets for fairness issues by analyzing hiring, healthcare, or lending application datasets to detect demographic imbalances or proxies that may result in discrimination. Students would need to evaluate datasets through fairness metrics such as disparate impact, equalized odds and demographic parity. Students would investigate bias origins by examining historical data inequalities and create solutions to address these issues through dataset re-balancing, and sensitive attribute removal.

After identifying biases in their audits using the methods prescribed by the EU AI Act and UK AI Assurance Framework, students could then apply statistical debiasing techniques such as re-weighting algorithms, adversarial debiasing, or pre-processing adjustments. They can thus correct the biases they

have identified in their audits when implementing these techniques, so as to develop AI models that meet the fairness criteria. For example, students might use open source libraries like Fairlearn [38] or IBM AI Fairness 360 [39] to adjust the training data or to apply constraints which ensure that AI decisions cannot effect specific demographic groups.

Students could take part in AI policymaking exercises which demonstrate how various governance models handle bias. Students would take on EU regulatory roles to implement legally binding fairness audits or function as USA, which advocates for support industry self-regulation. The exercises would require students to create policy suggestions and perform simulated bias impact evaluations and discuss the trade-offs between AI innovation and fairness during development. Such activities can encourage critical thinking, help students to recognize how bias mitigation strategies differ by regulatory framework, and how governance models can impact the practical application of ethics in AI.

5.2.2 Teaching privacy, transparency, and accountability frameworks

The transparency and explainability requirements in the EU AI Act, U.S. AI Bill of Rights, and India's AI Governance Strategy present the potential for introducing ethical AI documentation practices into engineering education. Researchers also note that engineering students should be trained to develop AI systems that are not only technically efficient, but also legally and ethically compliant [40–42].

To achieve this, AI curricula can integrate AI documentation and explainability best practices (e.g., model cards [43] and algorithmic impact assessments [44]). Model cards are the standardized documentation templates that contain information about the AI model, such as what is the model supposed to do, how well does it perform, what are the limitations of the model, and what are the potential risks, so that users and stakeholders will know what the system is capable of and what are the ethical implications of the system. Similarly, algorithmic impact assessments are tools used to analyze the potential societal, ethical, and regulatory effects that may result from deploying an AI system, which helps developers identify risks and how best to mitigate them before the system is deployed. These best practices should be incorporated into the curriculum to teach students how to practically document and evaluate AI systems with high transparency and accountability in mind.

AI curricular can also include comparative exercises analyzing how different regulatory systems approach AI transparency (e.g., GDPR in Europe vs. voluntary disclosure in the U.S.). Educators can include mock regulatory compliance exercises that let students evaluate AI systems through the EU's high-risk AI classification framework which determines system risk levels for individual or societal harm. Students can analyze facial recognition system for law enforcement to determine if it meets EU standards for risk management and transparency and bias mitigation before proposing modifications to meet regulatory requirements. Through these exercises students can learn how compliance frameworks operate in real-world scenarios while developing their ability to analyze ethical and legal challenges of AI deployment.

The analysis of privacy regulations across different global AI governance models offers an excellent educational case study about how engineers need to modify AI systems to meet local legal requirements. For instance, the contrast between GDPR-aligned AI privacy frameworks and China's AI-driven state data governance model can demonstrate how innovation flexibility relates to strict regulation.

5.2.3 Preparing engineers to be familiar with AI governance and its impacts

As more engineers become to contribute to AI policy-making efforts [45, 46], they must also develop policy and regulatory literacy. Thus, engineering students need to develop policy literacy, be familiar with

technology policy-making processes, and be able to interact with various stakeholders of policymaking, such as policymakers, ethics boards, and other interdisciplinary teams, in order to tackle real world AI-governance issues. Therefore, AI ethics education can include policy simulation exercises in which students evaluate AI ethics dilemmas in the context of global ethical AI laws. A potential example of a policy simulation exercise might have students examine an AI system used for hiring decisions, and ask them to determine whether or not it complies with the EU AI Act by identifying risks related to bias and transparency, or the U.S. AI Bill of Rights by evaluating fairness and explainability. The students could act as policymakers, regulators or developers and in groups discuss the trade offs between ethical obligations and business interests as they develop recommendations that would cause the system to comply with global ethical AI laws.

AI ethics education can also include cross-disciplinary projects, integrating computer science, law, philosophy, and social sciences to prepare engineers for ethical AI leadership. Such projects can also serve as venues for interdisciplinary learning in which engineering students would work either with their own peers or interact with guest lecturers to understand the societal, legal, and ethical issues of AI systems. To help students situate their learning in real-world challenges, AI ethics educators can invite professionals such as compliance officers, AI ethics committee members or policymakers to deliver guest lectures or lead workshops on real-world governance challenges. They can also work with companies to develop capstone projects or internships in which students take on the task of AI governance, such as auditing algorithms for fairness or developing compliance strategies for regulatory frameworks. Engineering ethics educators can also include real world case studies to illustrate how AI governance is effecting engineering practice.

6 Conclusion

This paper presents a comparative analysis of the AI policy documents of the United States, the United Kingdom, the European Union, China, and India to identify both ethical concerns and their mitigation strategies. The study identified key themes of algorithmic bias and discrimination, lack of transparency, and privacy violations, and how these can be addressed through different regulatory and policy approaches. By comparing these AI policies, this study shows where there is convergence and where there is divergence in AI governance. The analysis further explains how these findings can be incorporated into engineering ethics education to enable future AI professionals to develop as ethically competent, which is required for responsible AI development.

This research concludes that the usage of harmonized AI-governance strategies is becoming inevitable. Despite the differences in regional priorities in AI policymaking, the universal ethical principles of fairness, accountability, and transparency are not different. To this end, academic institutions can play a pivotal role in training AI practitioners for the future by aligning engineering education with global AI governance frameworks, so that those developing and deploying AI prioritize ethical responsibility in its use. As a result, it will be critical for the AI workforce to be well informed and ethics conscious in its development of AI that benefits all of humanity in an equitable and responsible way as AI continues to shape societies.

References

[1] A. M. Turing, *Computing machinery and intelligence*. Springer, 2009.

[2] A. Gokul, "Llms and ai: Understanding its reach and impact," 2023.

[3] S. Anwer, M. S. Hosen, D. S. Khan, E. Oluwabusayo, M. Folorunso, and H. Khan, "Revolutionizing the global market: An inclusion of ai the game changer in international dynamics," *Migration Letters*, vol. 21, no. S13, pp. 54–73, 2024.

[4] S. Venkatraman and M. Miah, "Artificial intelligence in healthcare: A potential game changer." *Global Journal of Business Disciplines*, vol. 6, no. 1, 2022.

[5] X. Zhai, "Chatgpt and ai: The game changer for education," *Zhai, X.(2023). ChatGPT: Reforming Education on Five Aspects. Shanghai Education*, pp. 16–17, 2023.

[6] M. S. Farahani and G. Ghasemi, "How ai changes the game in finance business models," *International Journal of Innovation in Management, Economics and Social Sciences*, vol. 4, no. 1, pp. 35–46, 2024.

[7] D. S. Hoadley and N. J. Lucas, "Artificial intelligence and national security," 2018.

[8] D. Acemoglu and P. Restrepo, "Artificial intelligence, automation, and work," in *The economics of artificial intelligence: An agenda*. University of Chicago Press, 2018, pp. 197–236.

[9] S. Choi, H. Kang, N. Kim, and J. Kim, "How does ai improve human decision-making? evidence from the ai-powered go program," *Evidence from the AI-Powered Go Program (October 1, 2023). USC Marshall School of Business Research Paper Sponsored by iORB, No. Forthcoming*, 2023.

[10] R. Verganti, L. Vendraminelli, and M. Iansiti, "Innovation and design in the age of artificial intelligence," *Journal of product innovation management*, vol. 37, no. 3, pp. 212–227, 2020.

[11] N. Kordzadeh and M. Ghasemaghaei, "Algorithmic bias: review, synthesis, and future research directions," *European Journal of Information Systems*, vol. 31, no. 3, pp. 388–409, 2022.

[12] N. Khalid, A. Qayyum, M. Bilal, A. Al-Fuqaha, and J. Qadir, "Privacy-preserving artificial intelligence in healthcare: Techniques and applications," *Computers in Biology and Medicine*, vol. 158, p. 106848, 2023.

[13] C. Novelli, M. Taddeo, and L. Floridi, "Accountability in artificial intelligence: what it is and how it works," *Ai & Society*, vol. 39, no. 4, pp. 1871–1882, 2024.

[14] H. B. Harvey and V. Gowda, "Regulatory issues and challenges to artificial intelligence adoption," *Radiologic Clinics*, vol. 59, no. 6, pp. 1075–1083, 2021.

[15] D. Kim, Q. Zhu, and H. Eldardiry, "Toward a policy approach to normative artificial intelligence governance: Implications for ai ethics education," *IEEE Transactions on Technology and Society*, vol. 5, no. 3, pp. 325–333, 2024.

[16] ——, "Exploring approaches to artificial intelligence governance: From ethics to policy," in *Proceedings of the 2023 IEEE International Symposium on Ethics in Engineering, Science, and Technology (ETHICS)*, 2023.

[17] C. Feijóo, Y. Kwon, J. M. Bauer, E. Bohlin, B. Howell, R. Jain, P. Potgieter, K. Vu, J. Whalley, and J. Xia, "Harnessing artificial intelligence (ai) to increase wellbeing for all: The case for a new technology diplomacy," *Telecommunications Policy*, vol. 44, no. 6, p. 101988, 2020.

[18] S. Lab, "Sjr - international science ranking." [Online]. Available: https://www.scimagojr.com/countryrank.php?category=1702&area=1700&year=2023&order=it&ord=desc#google_vignette

[19] N. Maslej, L. Fattorini, R. Perrault, V. Parli, A. Reuel, E. Brynjolfsson, J. Etchemendy, K. Ligett, T. Lyons, J. Manyika, and et al., "The ai index 2024 annual report," Apr 2024. [Online]. Available: https://aiindex.stanford.edu/report/

[20] K.-F. Lee, *AI superpowers: China, Silicon Valley, and the new world order*. Houghton Mifflin, 2018.

[21] D. Almeida, K. Shmarko, and E. Lomas, "The ethics of facial recognition technologies, surveillance, and accountability in an age of artificial intelligence: a comparative analysis of us, eu, and uk regulatory frameworks," *AI and Ethics*, vol. 2, no. 3, pp. 377–387, 2022.

[22] M. Gupta and S. Kaul, "Ai in inclusive education: A systematic review of opportunities and challenges in the indian context," *MIER Journal of Educational Studies Trends and Practices*, pp. 429–461, 2024.

[23] W. House, "Blueprint for an ai bill of rights," *Making Automated Systems Work for the American People (Ocotober 2022)(abrufbar unter https://www. whitehouse. gov/wpcontent/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights. pdf)*, 2022.

[24] H. Government, "National ai strategy," https://assets.publishing.service.gov.uk/media/614db4d1e90e077a2cbdf3c4/National_AI_Strategy_-_PDF_version.pdf, 2021, accessed: Dec. 18, 2024.

[25] N. Aayog, "National strategy for artificial intelligence," *Paper. June*, pp. 2019–01, 2018.

[26] E. A. I. Act, "The eu artificial intelligence act," 2024.

[27] N. America, "A next generation artificial intelligence development plan," *China), July, New America*, 2017.

[28] N. N. G. A. I. G. S. Committee *et al.*, "Ethical norms for new generation artificial intelligence released," *C enter for Security and Emerging Technology*, 2021.

[29] S. Chance, T. Børsen, D. A. Martin, R. Tormey, T. T. Lennerfors, and G. Bombaerts, "The routledge international handbook of engineering ethics education," 2025.

[30] D. Kim, "Promoting professional socialization: A synthesis of durkheim, kohlberg, hoffman, and haidt for professional ethics education," *Business and Professional Ethics Journal*, vol. 41, no. 1, pp. 93–114, 2022.

[31] S. G. Moreno, M. Decker, and C. Leicht-Scholten, "Implementing ai ethics education: Engineering competencies for sustainable societal impact," in *Proceedings of the 20th International CDIO Conference*, 2024.

[32] C. A. Eden, O. N. Chisom, I. S. Adeniyi *et al.*, "Integrating ai in education: Opportunities, challenges, and ethical considerations," *Magna Scientia Advanced Research and Reviews*, vol. 10, no. 2, pp. 006–013, 2024.

[33] J. Weichert, D. Kim, Q. Zhu, and H. Eldardiry, "'do i have to take this class?': A review of ethics requirements in computer science curricula," in *Proceedings of the ACM Technical Symposium on Computer Science Education, Conference Proceedings*, 2025.

[34] S. Isaac, N. Kotluk, and R. Tormey, "Educating engineering students to address bias and discrimination within their project teams," *Science and engineering ethics*, vol. 29, no. 1, p. 6, 2023.

[35] Y. S. J. Aquino, S. M. Carter, N. Houssami, A. Braunack-Mayer, K. T. Win, C. Degeling, L. Wang, and W. A. Rogers, "Practical, epistemic and normative implications of algorithmic bias in healthcare artificial intelligence: a qualitative study of multidisciplinary expert perspectives," *Journal of Medical Ethics*, 2023.

[36] A. W. Fazil, M. Hakimi, and A. K. Shahidzay, "A comprehensive review of bias in ai algorithms," *Nusantara Hasana Journal*, vol. 3, no. 8, pp. 1–11, 2023.

[37] R. Ghani, K. T. Rodolfa, P. Saleiro, and S. Jesus, "Addressing bias and fairness in machine learning: A practical guide and hands-on tutorial," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023, pp. 5779–5780.

[38] [Online]. Available: https://fairlearn.org/

[39] [Online]. Available: https://aif360.res.ibm.com/

[40] S. O'Sullivan, N. Nevejans, C. Allen, A. Blyth, S. Leonard, U. Pagallo, K. Holzinger, A. Holzinger, M. I. Sajid, and H. Ashrafian, "Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (ai) and autonomous robotic surgery," *The international journal of medical robotics and computer assisted surgery*, vol. 15, no. 1, p. e1968, 2019.

[41] N. Díaz-Rodríguez, J. Del Ser, M. Coeckelbergh, M. L. de Prado, E. Herrera-Viedma, and F. Herrera, "Connecting the dots in trustworthy artificial intelligence: From ai principles, ethics, and key requirements to responsible ai systems and regulation," *Information Fusion*, vol. 99, p. 101896, 2023.

[42] A. F. Winfield and M. Jirotka, "Ethical governance is essential to building trust in robotics and artificial intelligence systems," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 376, no. 2133, p. 20180085, 2018.

[43] J. Metcalf, E. Moss, E. A. Watkins, R. Singh, and M. C. Elish, "Algorithmic impact assessments and accountability: The co-construction of impacts," in *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, 2021, pp. 735–746.

[44] E. Moss, E. A. Watkins, J. Metcalf, and M. C. Elish, "Governing with algorithmic impact assessments: six observations," in *Watkins, Elizabeth and Moss, Emanuel and Metcalf, Jacob and Singh, Ranjit and Elish, Madeleine Clare, Governing Algorithmic Systems with Impact Assessments: Six Observations (May 14, 2021). AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society (AIES)*, 2020.

[45] Q. Zhu, D. Kim, H. Eldardiry, and M. C. Ausman, "A preliminary investigation of the ethics policy concerns of artificial intelligence: Insights from ai professionals working in policy-related roles," in *Proceedings of the 2023 IEEE Frontiers in Education Conference*, 2023.

[46] J. Weichert, Q. Zhu, D. Kim, and H. Eldardiry, "Perceptions of ai ethics policies among scientists and engineers in policy-related roles: An exploratory investigation," *Digital Society*, vol. 4, no. 1, pp. 1–25, 2025.