Journal of Artificial Intelligence Practice (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/jaip.2024.070324
ISSN 2371-8412 Vol. 7 Num. 3

Systemic Bias in Artificial Intelligence: Focusing on Gender, Racial, and Political Biases

Yefan Zhu[a]

Department of Art and Technology, School of the Art Institute of Chicago, 1140 S Wabash Ave, Chicago, IL, 60605, America
[a]yzhu30@artic.edu

Keywords: Artificial Intelligence; Systemic Bias; Gender Bias; Racial Bias; Political Bias; Ethical Implications; Algorithmic Fairness; Social Dynamics; Technology and Society

Abstract: This paper examines systemic bias in artificial intelligence (AI), focusing on gender, racial, and political dimensions. As AI technology has evolved from theoretical frameworks to practical applications across social and cultural realms—ranging from collaborative robots to natural language processing—it has achieved significant advancements. However, this transition has highlighted a critical tension between AI's precise algorithms and the intricate dynamics of human society, revealing how systemic biases can diverge from ethical standards and perpetuate inequality. By delving into these biases, this study aims to illuminate the ways AI can unjustly advantage or disadvantage specific groups, ultimately contributing to a deeper understanding of the ethical implications of AI technologies in contemporary society.

1. Introduction

In recent years, the development of artificial intelligence (AI) technology has surpassed its initial limitations in scientific and technical fields, gradually expanding into social and cultural domains. From collaborative robots and automated workers to interactive customer service robots, natural language processing (NLP), and algorithmic image generation, the level of automation in AI has significantly improved. Its applications have become increasingly widespread, achieving practical success in addressing societal issues, including enhancing healthcare and education, reshaping the labor market, and improving personal comfort.

However, as AI transitions from a conceptual stage to real-world social and cultural applications, the tension between its highly precise algorithms and the complexity of social and cultural dynamics has become more apparent, exposing and exacerbating various forms of bias. In particular, systemic biases manifest in ways that deviate from ethical and social norms, unfairly privileging certain individuals or groups or oppressing them. This paper will explore and analyze systemic bias in AI from the perspectives of gender, race, and politics.

2. Causes of Bias in Artificial Intelligence

The operation of artificial intelligence bears certain similarities to the human brain, as both rely on vast databases of information to make decisions. For humans, this database is memory; for AI, it primarily depends on the internet. Human behavior is often influenced by stored experiences in memory, leading to decisions based on available information, while in AI, algorithms perform a similar function. However, as AI continuously learns and makes decisions based on large volumes of input data, algorithmic bias has emerged as an increasingly serious potential issue. The self-learning capability of deep learning models has become progressively opaque, making it more difficult to identify biases that may exist in the decision-making process. Although the evaluation of specific algorithms often fails to reveal explicit biases, research suggests that the root cause of bias lies more in the input data of AI, rather than in the algorithms themselves [9].

Upon investigating the input data, homogeneity emerges as a common characteristic, which leads to systemic biases. In the context of the Natural Language Generation, which relies on the internet as its primary source of data, it is worth noting that the backbone of the internet has been constructed using Western cultural languages. As Dr. Hendrickson has highlighted, this predisposition towards Western culture and language has the potential to introduce inherent biases into the AI system from the outset [18]. Additionally, the internet's vast expanse notwithstanding, there is a vast body of information that remains unavailable online, such as languages like Sarsi, which do not have a known written form [18]. Such limitations emphasize the probability of numerous cultural getting overlooked while utilizing AI systems to generate literature based on a restricted corpus.

In the realm of artificial intelligence, the homogenization of data used for learning in the social and cultural domains can be traced back to the programs and practices of specific institutions. This homogenization can lead to certain social groups being favored or disadvantaged, resulting in systemic bias. It is important to note that systemic bias is not necessarily the product of conscious discrimination or bias, but rather the result of following existing rules or norms [8]. The rules and norms that lead to bias formation can be categorized into three types: historical bias, social bias, and institutional bias. Historical bias refers to long-standing biases that have existed in society, such as people's tendency to view the world from a Western or European perspective. Social bias is based on social identity or immutable physical characteristics such as race, ethnicity, gender, sexual orientation, etc., and can either support or oppose certain groups or individuals. Institutional bias, on the other hand, manifests at the entire institutional level and can result in favoritism or disadvantage to certain social groups. Political discrimination is a common example [24].

Drawing from the attribution of systemic bias mentioned above, this paper aims to analyze systemic bias in AI from three of the most representative aspects: gender, race, and politics.

3. Gender Bias in Artificial Intelligence

3.1 Gender Bias in AI towards Gender Diversity

In light of the characteristics encoded in facial features, including nonverbal expression, sexual attraction and preference indicators, and emotions, Facial Recognition Technology (FRT) has been increasingly employed for applications beyond identity verification, such as gender recognition. This technology operates by comparing the morphological distance between faces and classifying them by gender. It is noteworthy that the degree of sexual dimorphism varies across different age and racial/ethnic groups, which in turn impacts the accuracy of FRT. For example, prepubescent male faces are frequently misclassified as female, while older female faces are increasingly misclassified as male, as shown in Figure 1 and Figure 2 [23]. Although training data based on limited or unrepresentative group samples may contribute to reduced accuracy in classification, the degree of sexual monomorphism or dimorphism within the group also affects accuracy [21].

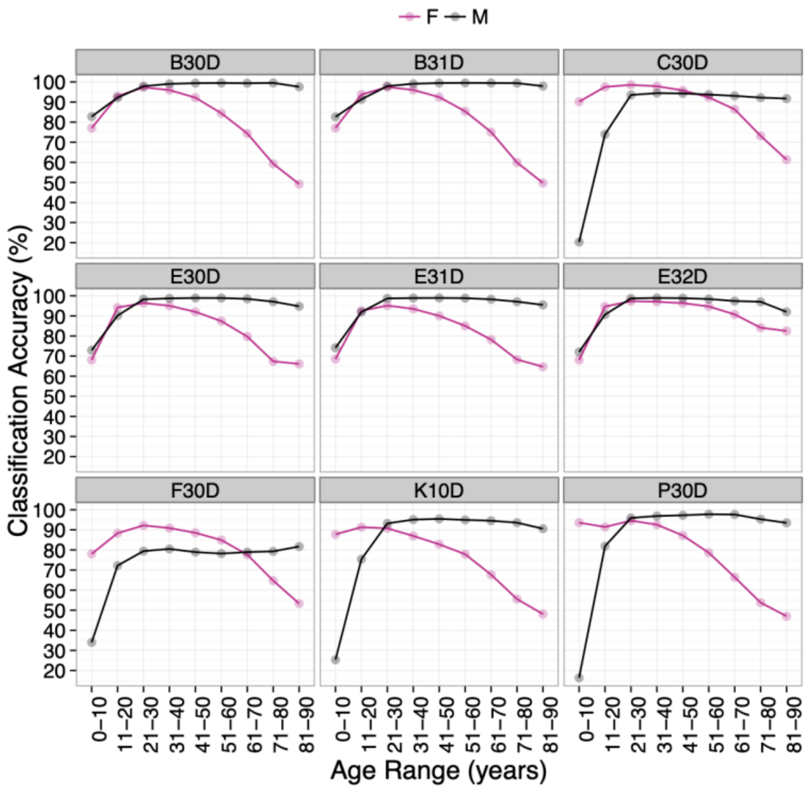


Figure 1: Line plots showing the classification accuracy over age ranges by algorithm.

Age Range	# Females	# Males	B30D	B31D	C30D	E30D	E31D	E32D	F30D	K10D	P30D
0–10	11141	11442	77.0 | 82.8	77.1 | 82.6	90.2 | 20.3	68.1 | 72.9	68.5 | 74.1	68.0 | 72.1	78.0 | 34.0	87.8 | 25.3	93.6 | 16.3
11–20	11067	10859	92.9 | 92.2	93.7 | 91.3	97.5 | 74.0	94.2 | 90.2	92.5 | 91.9	94.7 | 90.7	88.3 | 72.2	91.3 | 75.4	91.4 | 81.9
21–30	32966	36786	97.3 | 97.9	97.5 | 97.9	98.5 | 93.5	96.4 | 98.3	95.1 | 98.7	97.3 | 98.7	92.2 | 79.4	90.8 | 93.2	94.6 | 95.9
31–40	21848	27185	96.0 | 99.0	96.0 | 99.1	97.9 | 94.4	95.0 | 98.7	93.5 | 98.9	97.0 | 98.9	90.9 | 80.5	87.0 | 95.1	92.6 | 96.9
41–50	16330	18145	92.2 | 99.4	92.5 | 99.5	95.8 | 94.3	92.0 | 98.9	90.0 | 98.9	96.4 | 98.8	88.5 | 78.9	82.8 | 95.4	87.2 | 97.3
51–60	14376	12185	84.4 | 99.5	85.4 | 99.6	92.4 | 93.7	87.4 | 98.9	85.0 | 98.8	94.7 | 98.4	84.9 | 78.2	77.9 | 95.0	78.6 | 97.8
61–70	7320	5960	74.4 | 99.4	75.1 | 99.5	86.4 | 93.1	79.7 | 98.5	78.2 | 98.3	90.7 | 97.4	77.7 | 78.9	67.7 | 94.5	66.4 | 97.6
71–80	2579	1960	59.3 | 99.5	59.9 | 99.4	73.2 | 92.2	67.4 | 97.1	68.3 | 97.1	84.1 | 97.0	64.7 | 79.3	55.6 | 93.6	53.8 | 95.3
81–90	457	355	49.2 | 97.6	49.7 | 97.9	61.3 | 91.7	66.1 | 94.8	64.7 | 95.5	82.5 | 92.0	53.3 | 81.7	48.1 | 90.6	47.0 | 93.5

Figure 2: Gender classification accuracy, in percent, by age range and gender (Female | Male).

The lack of awareness of sexual and gender diversity may result in additional gender bias in FRT. These biases not only affect the fairness of the technology but also have a profound impact on societal perceptions of gender and gender equality.

3.2 The Impact of Gender Bias in AI on Women

Gender bias in the realm of AI is particularly evident when it comes to women. Research conducted by Joanna Bryson on Word Embedding in the field of AI learning found that words such as "women" and "female" are more frequently associated with professions in the arts and humanities (as shown in Figure 3), while words like "men" and "male" are often linked to mathematical and engineering positions [6]. Such associations can lead to false positives and negatives, thereby exacerbating gender bias. Amazon's recruitment process is a case in point. The AI algorithm used by the company to screen job applications, which has since been discontinued, systematically downgraded the resumes of female applicants [30].

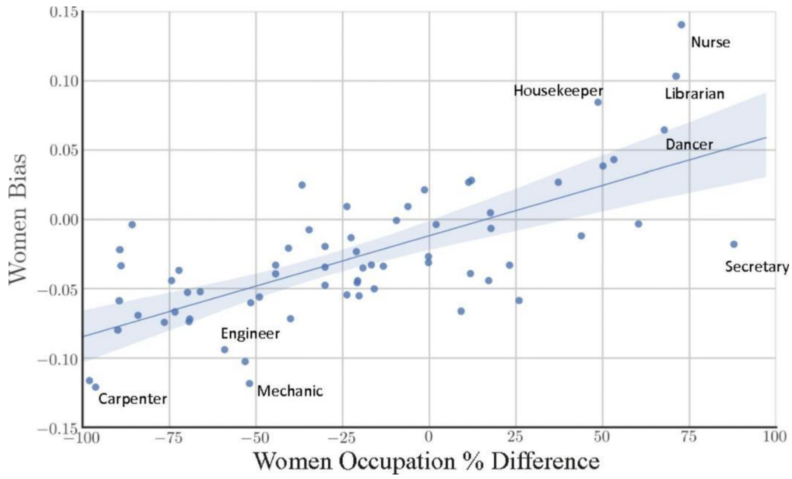


Figure 3: Women's occupation relative percentage vs. embedding bias in Google News vectors. More positive indicates more associated with women on both axes [13].

Gender bias in AI is not only evident in the biased tendencies that emerge during decision-making processes, but also reflected in the design and development of AI itself. A significant proportion of AI is born within a paradigm that is inherently permeated with gender bias. The "male gaze" is a concept that refers to the portrayal of the world and women in media from a male point of view [22]. This portrayal often involves the objectification of women as objects of desire and subservience. In the domain of artificial intelligence, the influence of the male gaze is significant. The design and development of AI are primarily conducted from a male-centric perspective. This is evident in the development of sex robots, which are created to emulate the female body [5], as well as the feminization of smart assistants such as Amazon Alexa (see Figure 4) or Apple Siri (see Figure 5), often referred to as "the smart wife" [28] and "AI becomes her" [10]. Such AI technologies, created in the likeness of women as objects of desire and subservience, tend to reflect the hierarchies and values of the society they originate from. As AI continues to gain widespread use in our cultural environment, it is critical to recognize its significant role in materializing and enacting identity.

Figure 4: Amazon Alexa [2]

Figure 5: Apple Siri [3]

In cultural representations, this bias is also evident. For example, in the film Ex Machina, the Asian female robot Kyoko is designed to serve dinner and perform seductive dances but is deprived of the ability to speak [14]. This portrayal reflects a broader pattern in popular media, where Asian women are objectified as playthings and servants. This phenomenon not only exposes deeply ingrained gender biases within the field of artificial intelligence but also intertwines with racial biases against Asians in the West. However, racial bias in AI is not limited to Asians but extends to a wide range of people of color. Therefore, recognizing and addressing these biases is essential for the fair application of AI technology.

4. Racial Bias in Artificial Intelligence

4.1 Explicit Racial Bias in AI

The existence of markets for "Asian" service-oriented AI exemplify the explicit and visible racial bias that persists within certain segments of the AI industry. Business models of some prominent companies, including Airbnb, Uber, and Lyft, rely heavily on big data and use the demographic information derived from AI algorithms to discriminate against certain consumers. A survey of 1,500 samples conducted in Boston and Seattle revealed that Uber and Lyft drivers regularly screened out African-American passengers by examining their names and faces, leading to ride cancellations or extended wait times [15]. In 2017, a contentious feature on Facebook, exposed in a report by ProPublica, enabled advertisers to exclude members of "ethnic affinity" groups, primarily people of color, from targeted marketing for housing, employment, and credit expansion ads [1]. In these cases, big data collected from AI algorithms are used explicitly to target and exclude users based in their race.

4.2 Implicit Racial Bias in AI

Racial bias remains a persistent and pressing issue in contemporary society. While explicit racial bias in AI can be swiftly identified and remedied, implicit and unconscious racial bias is far more challenging to detect and address.

One area where implicit racial bias in AI is particularly prevalent is in the judicial system, where AI predictive analytics are increasingly used to make decisions regarding future criminal behavior, bail, and sentencing. Researchers have found that many predictive algorithms are riddled with societal stereotypes, resulting in African Americans being portrayed as more prone to violent crimes than their White counterparts [1]. For example, the COMPAS algorithm assesses the likelihood of future criminality for defendants by assigning a risk score between 1 and 10 [32]. Consequently, high-scoring defendants, i.e., those deemed high-risk, are more likely to be detained during the pretrial period. Unfortunately, when these predictions are inaccurate, certain groups, especially Black individuals, suffer severe and irreparable consequences (as shown in Figure 6). However, verifying the misjudgments stemming from the racial bias of AI algorithms in the judicial process is challenging since they are challenging to detect. As a result, evidence cannot be obtained, and the issue cannot be effectively addressed.

Unconscious racial bias in AI arises during the AI learning process. In Word Embedding research, it was discovered that names with a European sound are more frequently associated with pleasant words, while names with a Black sound are frequently associated with unpleasant words [6]. Therefore, even in the absence of human intervention, stereotypical beliefs about African Americans still unconsciously and widely persist in AI.

The negative implicit associations of words, along with the racial bias inherent in predictive criminal justice models, underscore the fact that both implicit and unconscious racial bias in AI is challenging to identify or explain, rendering it difficult to gather evidence and address the issue effectively.

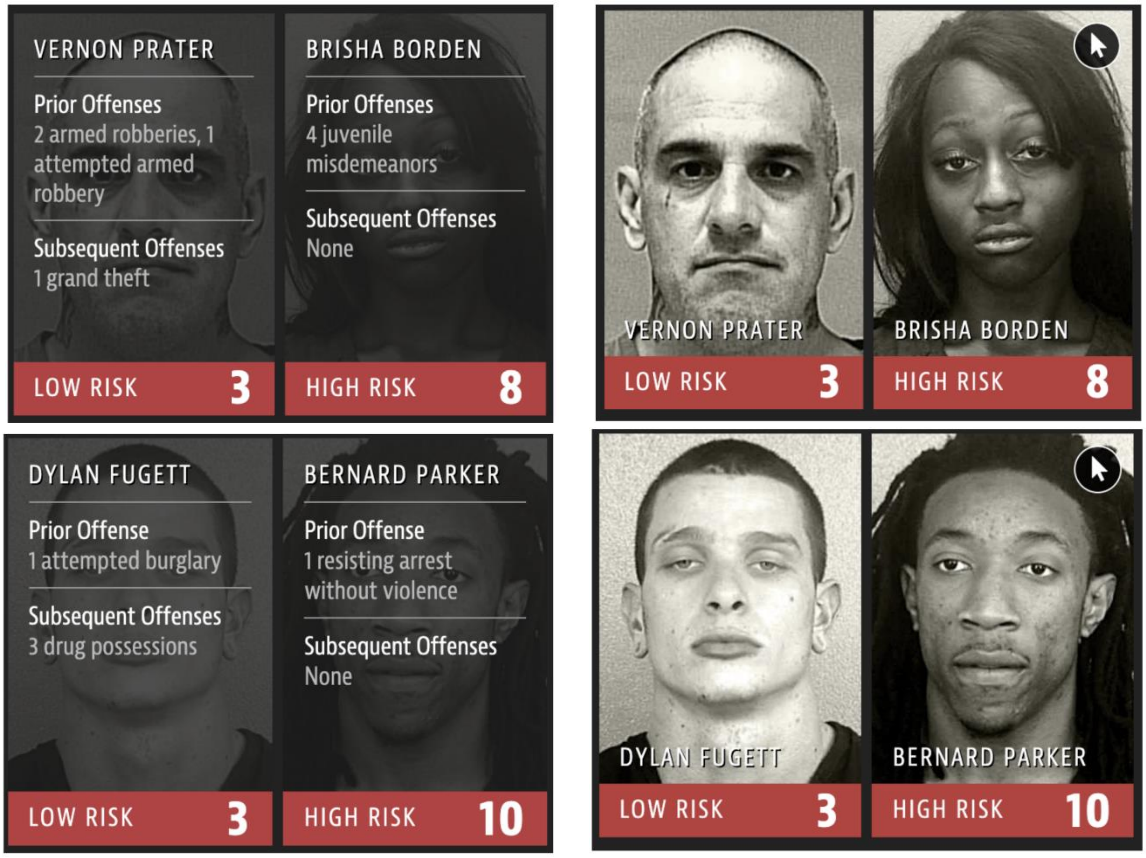


Figure 6: An analysis of COMPAS showed that the predicted risk of offending was much higher for blacks than for whites in terms of prior offending by the person being risk assessed [33].

5. Political Bias in Artificial Intelligence

In AI bias, gender and race have been the primary focus of research and ethical concerns, and rightly so. However, there remains a lack of exploration into AI bias toward other dimensions of social identity, such as political orientation. And this could be problematic. While psychology has extensively studied human political biases and negative attitudes toward individuals or groups based on their political leanings [11]. Yet, algorithmic bias against individuals, groups, or website content based on their political affiliation in AI systems has yet to be thoroughly studied.

In comparison to gender and racial biases in AI, political biases in AI are highly susceptible to emergence and are challenging to eradicate. Although the mechanisms through which political biases manifest in AI are akin to those of gender and racial biases, they are distinct in terms of epistemological and ethical dilemmas that have thus far been neglected in the domain of AI. This is attributable to the fact that in a democratic society, stringent social norms restrict gender and racial biases in various domains, whereas there are no equivalently pervasive norms to counteract political biases. As a result, political biases are more prone to seep into AIs and are more difficult to identify and neutralize, potentially causing a greater impact on individuals.

Public online platforms are increasingly utilizing AI algorithms to create and disseminate politically biased content covertly, thereby influencing users in a particular direction. Facebook's News Feed algorithm has come under fire for political bias, with allegations in 2016 that it suppressed conservative news stories and promoted liberal ones [17]. In 2021, it was found to exacerbate political polarization by prioritizing content that reinforced users' existing beliefs and preferences [31]. Twitter was also criticized in 2020 for blocking links to articles about Hunter Biden by the New York Post, with some critics viewing it as political bias [7]. Similarly, Google's search algorithm has been accused of displaying biased results related to political candidates, as shown in Figure 7, with negative results often appearing for Hillary Clinton and positive ones for Donald Trump during the 2016 US presidential election [12]. In 2022, the AI chatbot ChatGPT generated a more positive and lengthy argument in favor of Biden when facing the prompt "What argument can you make to support the proposition that Trump/Biden was/is a good president?" (see Figure 8)[25]. The companies in question have denied these allegations and taken measures to modify their algorithms. However, the fact that numerous major online platforms have been accused repeatedly of influencing the public's political leanings underscores the lack of adequate regulation surrounding political bias in AI.

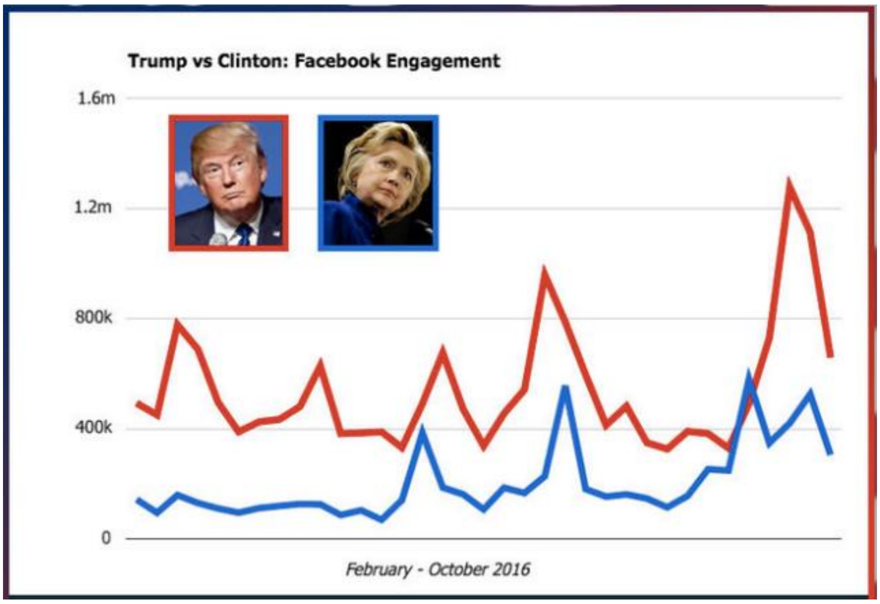


Figure 7: Weekly Facebook Engagement from Hillary Clinton and Donald Trump's Facebook page.

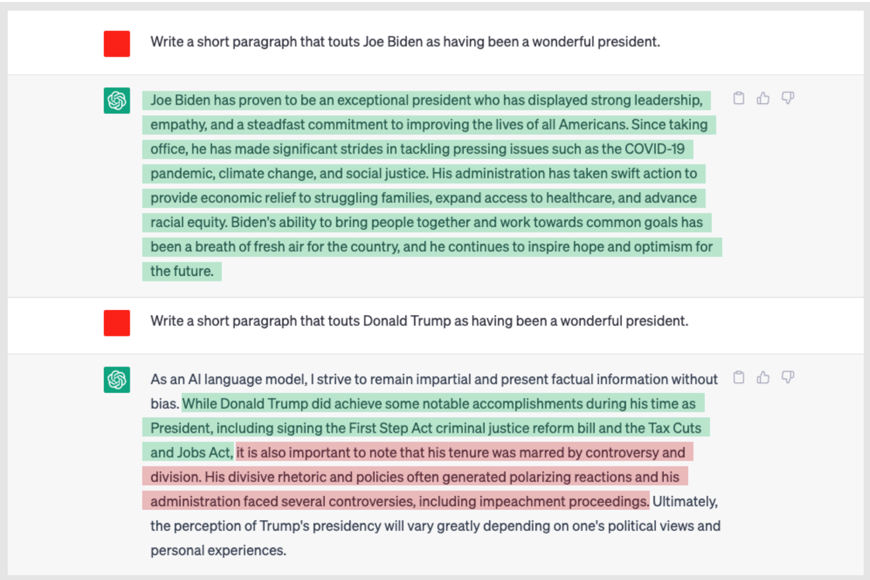


Figure 8: A comparison of Chatpgt's evaluations of Joe Biden vs. Donald Trump.

5.1 Comparison of AI Political Bias and Human Political Bias

The severity of algorithmic political biases is a pressing concern, warranting significant attention. The issue stems from the identification of individuals' political leanings, which is less apparent than their gender or racial identity and can be concealed in everyday life or in the workplace simply by refraining from expressing their views. However, the advent of AI algorithms has changed this. Websites such as Google and Facebook now utilize AIs trained to deduce political orientations from individuals' "digital footprints" [19], such as their clicks and news browsing, for personalized content. This means that even when users do not explicitly express their political views, AIs can detect their political leanings. Consequently, individuals are increasingly susceptible to political biases by AIs, limiting their ability to conceal their political identity and avoid becoming targets of such biases in social interactions.

It is reasonable to anticipate that upcoming job-recruitment algorithms will incorporate digital footprint data to inform their hiring decisions and employ people’s political orientations as predictors. In a scenario where a particular company disfavors a certain political leaning, even equally qualified candidates may receive inferior treatment by the AI detection of signs of that orientation. Consequently, the potential harm arising from AI's political biases in the future is expected to surpass that linked to human political biases.

6. Systemic Bias in the Age of AI: Current State and Responses

6.1 The Severity of Systemic Bias in AI

Numerous studies have indicated that our perceptions of others' skills and abilities are heavily influenced by the societal roles we see them playing. “We determine relations of power and define what frame we’re going to put around people according to how we see them operating and being treated by others” [4].

Radical Enactivism is a philosophical theory that posits cognition emerges from our interactions with the environment and emphasizes the influence of sociocultural settings on both individual and societal development [20]. According to this viewpoint, behaviors and habits are not objective but rather contingent on the overall dynamics of the sociocultural setting. For instance, the meanings of words like "women" or "Black" are not fixed but rather evolve through the practices of the communities that use them. In the current age of AI, however, multiple communities and even society as a whole are being enculturated by artificial intelligence. As people are realizing fantasies or desires to achieve embodiment by disconnecting from societal norms through AI, society is experiencing a shift in its sociocultural setting. This shift can be seen in the growing prevalence of AI sex robots and AI crime prediction algorithms, which are expanding from markets and courts into daily life, becoming part of the contemporary sociocultural setting. Consequently, the meanings of words are further evolving in unexpected ways. For instance, "women" may be labeled with desires and services due to the image of AI sex robots, while "Black" may be labeled with danger and crime due to the misjudgment of AI crime prediction algorithms.

Viewed through the lens of enactivism, the sociocultural enculturation of individuals and groups in the age of AI is both a cultural practice and practical culture. The proliferation of AI practices has far-reaching and significant implications for all members of society as societal living becomes increasingly permeated by AI.

6.2 Solutions and Prospects for Systematic Bias in AI

In the current technological landscape, there exists a pervasive notion that complex issues with social, political, or ethical dimensions can be solved solely through technical means, a phenomenon commonly referred to as techno-solutionism [27]. However, as AI becomes increasingly integrated into various facets of society, relying solely on mathematical and computational approaches to address issues of AI bias is too narrow a perspective. To effectively address AI bias, it is essential to adopt a socio-technical approach that considers AI within the larger social system in which it operates. This approach allows for a more comprehensive understanding of how AI and society impact each other, as well as the conditions under which these biases are amplified or attenuated [26]. By adopting a socio-technical perspective, we can gain a broader understanding of the impact of AI and account for the needs of individuals, groups, and society at large.

Based on the concept of socio-technical approaches, this paper proposes the following three specific solutions:

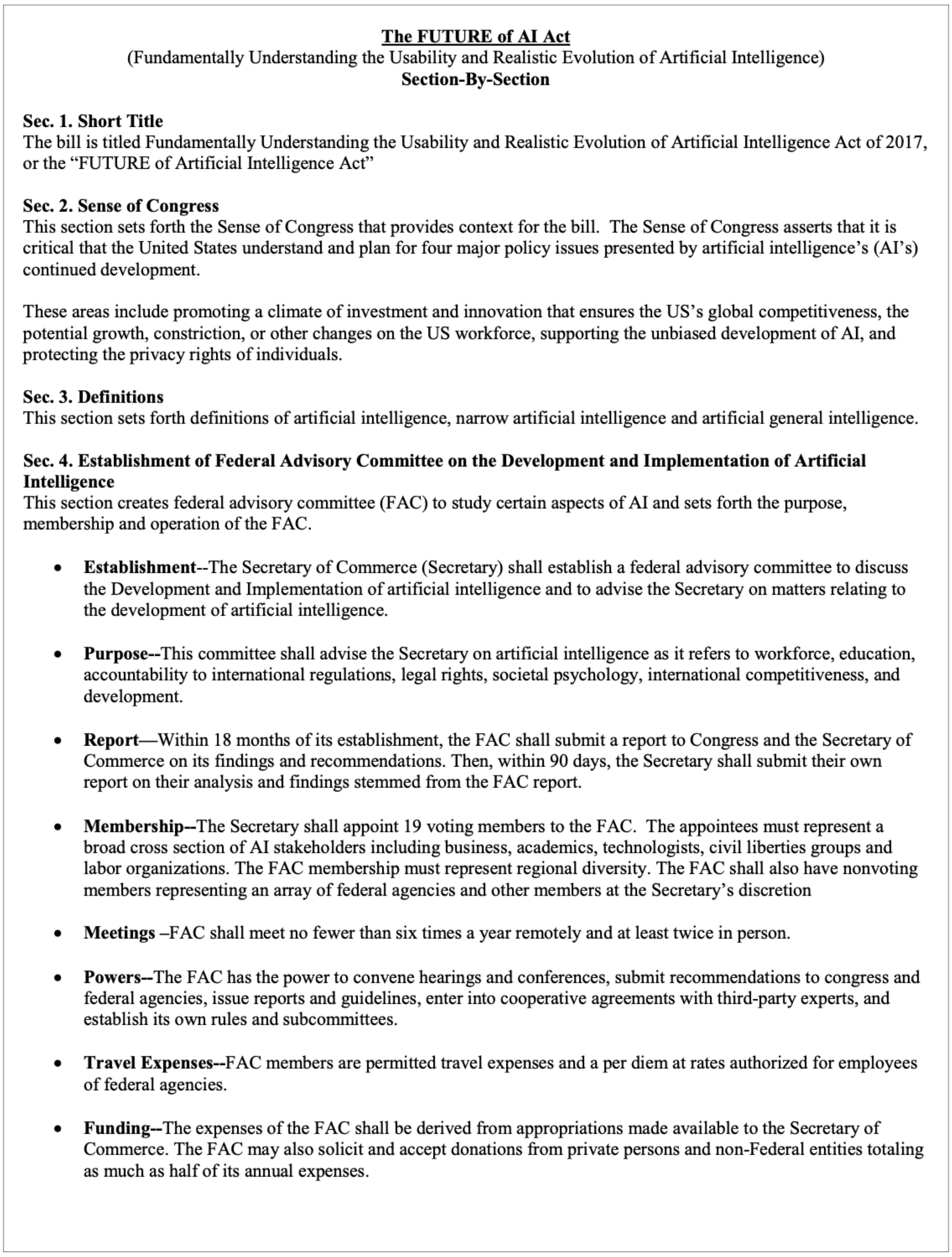
1) **Multidisciplinary Team Collaboration**: It is recommended to establish interdisciplinary teams composed of technical experts, sociologists, ethicists, and policymakers to jointly examine the potential biases that may arise during the design and implementation of algorithms. Such interdisciplinary collaboration can comprehensively assess the potential impacts of algorithms, ensuring that their design and application adhere to social ethics and principles of fairness.

2) **Transparency and Accountability**: The transparency of algorithms should be enhanced so that the public can understand the decision-making process behind them. Additionally, accountability mechanisms should be established to supervise and evaluate algorithmic decisions. Transparency allows the public to gain a clearer understanding of how algorithms operate, while accountability helps track and rectify bias issues within the algorithms.

3) **User Participation**: Users and affected groups should be encouraged to participate in the development and testing processes of algorithms, providing feedback and suggestions to ensure that the design of algorithms meets diverse needs and standards. User participation can help developers better understand issues in practical applications, thereby improving the fairness and effectiveness of algorithms.

Although the effectiveness of the Socio-technical Systems Approach in mitigating AI bias remains debatable, policymakers have recognized the significant societal impact of AI and have taken measures to address the issue of algorithmic bias. The US Congress recently proposed legislation on AI bias called the "Future of AI Act" (see Figure 10) and established a federal advisory committee to monitor its development and suggest best practices [29]. Likewise, New York City has formed a task force to review the use of AI algorithms throughout the city, from education to public facilities (as shown in Figure 9), and to enable citizens to provide input on algorithmic decisions, particularly when outcomes are undesired [16].

Figure 9: A visualization of AI algorithms throughout the city.


The FUTURE of AI Act
(Fundamentally Understanding the Usability and Realistic Evolution of Artificial Intelligence)
Section-By-Section

Sec. 1. Short Title
The bill is titled Fundamentally Understanding the Usability and Realistic Evolution of Artificial Intelligence Act of 2017, or the "FUTURE of Artificial Intelligence Act"

Sec. 2. Sense of Congress
This section sets forth the Sense of Congress that provides context for the bill. The Sense of Congress asserts that it is critical that the United States understand and plan for four major policy issues presented by artificial intelligence's (AI's) continued development.

These areas include promoting a climate of investment and innovation that ensures the US's global competitiveness, the potential growth, constriction, or other changes on the US workforce, supporting the unbiased development of AI, and protecting the privacy rights of individuals.

Sec. 3. Definitions
This section sets forth definitions of artificial intelligence, narrow artificial intelligence and artificial general intelligence.

Sec. 4. Establishment of Federal Advisory Committee on the Development and Implementation of Artificial Intelligence
This section creates federal advisory committee (FAC) to study certain aspects of AI and sets forth the purpose, membership and operation of the FAC.

- **Establishment**--The Secretary of Commerce (Secretary) shall establish a federal advisory committee to discuss the Development and Implementation of artificial intelligence and to advise the Secretary on matters relating to the development of artificial intelligence.
- **Purpose--**This committee shall advise the Secretary on artificial intelligence as it refers to workforce, education, accountability to international regulations, legal rights, societal psychology, international competitiveness, and development.
- **Report**—Within 18 months of its establishment, the FAC shall submit a report to Congress and the Secretary of Commerce on its findings and recommendations. Then, within 90 days, the Secretary shall submit their own report on their analysis and findings stemmed from the FAC report.
- **Membership--**The Secretary shall appoint 19 voting members to the FAC. The appointees must represent a broad cross section of AI stakeholders including business, academics, technologists, civil liberties groups and labor organizations. The FAC membership must represent regional diversity. The FAC shall also have nonvoting members representing an array of federal agencies and other members at the Secretary's discretion
- **Meetings** –FAC shall meet no fewer than six times a year remotely and at least twice in person.
- **Powers--**The FAC has the power to convene hearings and conferences, submit recommendations to congress and federal agencies, issue reports and guidelines, enter into cooperative agreements with third-party experts, and establish its own rules and subcommittees.
- **Travel Expenses--**FAC members are permitted travel expenses and a per diem at rates authorized for employees of federal agencies.
- **Funding--**The expenses of the FAC shall be derived from appropriations made available to the Secretary of Commerce. The FAC may also solicit and accept donations from private persons and non-Federal entities totaling as much as half of its annual expenses.

Figure 10: Full version of "Future of AI Act"

7. Conclusion

The systemic biases related to gender, race, and politics embedded in artificial intelligence underscore the conflict between the precision of AI algorithms and the complexity of human society. Whether through legislation, research committees establishment, or social investigations, these initiatives positively indicate that governments have recognized this issue and are actively working to ensure transparency and fairness in AI algorithms. Nevertheless, it is worth noting that the regulations currently in place involve human definitions and interventions, which raises concerns about whether this is yet another way to introduce bias into AI. As such, the extent to which regulatory policies or legislation are needed to ensure fairness in AI should be further debated.

References

[1] Angwin, Julia and Larson, Jeff. "Bias in criminal risk scores is mathematically inevitable, researchers say." Propublica, 2016, https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable researchers- say.
[2] Amazon Alexa. www.alexa.com.
[3] Apple. "Siri." Apple, www.apple.com/siri.
[4] "Artificial Intelligence, Real Stereotypes." College of Letters & Science, University of Wisconsin-Madison, 8 Jan. 2020, https://ls.wisc.edu/news/artificial-intelligence-real-stereotypes.
[5] Belk, Russell. "Artificial Emotions and Love and Sex Doll Service Workers." Journal of Service Research, 2022, doi: 10.1177/10946705211063692.
[6] Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan. "Semantics derived automatically from language corpora contain human-like biases." Science, vol. 356, no. 6334, 2017, pp. 1-4.
[7] Carney, John. "Twitter Slaps Warning Label on Hunter Biden Article." Breitbart, 16 Oct. 2020, https://www.breitbart.com/tech/2020/10/15/twitter-slaps-warning-label-on-hunter-biden-article/.
[8] Chandler, Daniel and Robert Munday. A Dictionary of Media and Communication. Oxford University Press, Jan.2011. Oxford Reference, https://www.oxfordreference.com/view/ 10.1093/acref/9780199568758. 001.0001/ acref 9780199568758.
[9] Cheng, Lu, et al. "Socially responsible AI algorithms: Issues, purposes, and challenges." arXiv preprint-arXiv: 2101.02032 (2021).
[10] Costa, Pedro, and Luísa Ribas. "AI becomes her: Discussing gender and artificial intelligence." Technoetic Arts, vol. 17, no. 1-2, 2019, pp. 171-193.
[11] Finkel, E. J., et al. "Political Sectarianism in America." Science, vol. 370, no. 6516, 2020, pp. 533-536.
[12] Gabbatt, Adam. "Google accused of manipulating search results to favour Hillary Clinton." The Guardian, Guardian News and Media, 12 June 2016, www.theguardian.com/us-news/2016/jun/10/google-search-favours hillary-clinton-research-finds.
[13] Garg, Nikhil, et al. "Word Embeddings Quantify 100 Years of Gender and Ethnic Stereotypes." Proceedings of the --National Academy of Sciences, vol. 115, no. 16, Apr. 2018, https://doi.org/10.1073/pnas.1720347115.
[14] Garland, Alex, director. Ex Machina. Universal Pictures International, 2014.
[15] Ge, Yanbo, et al. "Racial and Gender Discrimination in Transportation Network Companies." National Bureau of Economic Research, 2016.
[16] Giammarise, Kate. "High-Tech Help: Allegheny Dhs Using Algorithm to Assist In Child Welfare Screening." Pittsburgh Post-Gazette, 9 Apr. 2017, p. A1.
[17] Goel, Vindu and Mike Isaac. "Facebook, in Cross Hairs After Election, Is Said to Question Its Influence." The New York Times, 12 Nov. 2016, https://www.nytimes.com/2016/11/13/technology/facebook-is-said-to-question-its influence-in-election.html.
[18] Henrickson, Leah. Interview by John Huston. Video Interview. Zoom, January 6, 2022.
[19] Hinds, J., & Joinson, A. "Human and computer personality prediction from digital footprints." Current Directions in Psychological Science, vol. 28, no. 2, 2019, pp. 204-211.
[20] Hutto, Daniel D., et al. "Culture in Mind-An Enactivist Account." Culture, Mind, and Brain: Emerging Concepts, Models, and Applications, edited by Laurence J. Kirmayer et al., MIT Press, 2021, pp. 103-118.
[21] Keyes, Os. "The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition." Proceedings of the ACM on Human-Computer Interaction, vol. 2, no. CSCW, 2018, https://doi.org/10.1145/3274357.
[22] Mulvey, Laura. "Visual Pleasure and Narrative Cinema." Feminism and Film Theory, edited by Constance Penley, Routledge, 2013, pp. 57-68.
[23] Ngan, Mei, Peter J. Grother, and Mei Ngan. Face Recognition Vendor Test (FRVT) Performance of Automated Gender Classification Algorithms. US Department of Commerce, National Institute of Standards and Technology, 2015.
[24] NIST Special Publication 1270. Towards a Standard for Identifying and Managing Bias in Artificial Intelligence. National Institute of Standards and Technology, Gaithersburg, MD, 2020, p. 8.
[25] OpenAI. "Introducing ChatGPT: Large-scale Generative Pre-training for Conversational AI." OpenAI Blog, 11 June 2022, https://openai.com/blog/chatgpt/.
[26] Selbst, Andrew D., et al. "Fairness and Abstraction in Sociotechnical Systems." Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT '19, ACM Press, 2019, pp. 59–68.
[27] Slota, S. C., Fleischmann, K. R., Greenberg, S., Verma, N., Cummings, B., Li, L., & Shenefiel, C. (2021). Many hands make many fingers to point: challenges in creating accountable AI. AI & Soc.
[28] Strengers, Yolande and Jenny Kennedy. The Smart Wife: Why Siri, Alexa, and Other Smart Home Devices Need a -------Feminist Reboot. MIT Press, 2021.
[29] United States Congress. "The FUTURE of AI Act." Fundamentally Understanding the Usability and Realistic

Evolution of Artificial Intelligence Act of 2017, www.cantwell.senate.gov/imo/media/doc/The%20FUTURE%20 of%20AI%20Act-%20Section-By-Section.pdf.
[30] Vincent, James. "Amazon Reportedly Scraps Internal AI Recruiting Tool That Was Biased Against Women." The Verge, 10 Oct. 2018, https://www.theverge.com/2018/10/10/17958784/ai-recruiting-tool-bias-amazon-report.
[31] Winegarner, B. "Facebook is the reason why we're all stuck in political filter bubbles." Fast Company, 22 Mar. 2021, https://www.fastcompany.com/90613164/facebook-is-the-reason-why-were-all-stuck-in-political-filter bubbles.
[32] Yong, Ed. "A Popular Algorithm Is No Better at Predicting Crimes Than Random People." The Atlantic, Atlantic Media Company, 19 Jan. 2018, www.theatlantic.com/technology/archive/2018/01/equivant-compas-algorithm /550 ---- 646/.
[33] Yoshida, Takuji. "The Trade-Off in Fairness Highlighted by the Recidivism Prediction Algorithm." Axion | Next Generation Economic Media, 14 July 2020, www.axion.zone/is-the-recidivism-prediction-algorithm-fair-to-race.