The Echo Chamber of Algorithm Bias

Joseph O. Chan, Ph.D.
Roosevelt University
425 S. Wabash Ave., Chicago, IL 60605, USA
Tel: 312-281-3279, Email: jchan@roosevelt.edu

Abstract

Cognitive bias has affected many aspects of society, many have grave consequences. They include recent incidents about police shooting of minority, the rise of Asian crimes, quota against admission of Asian students by elite institutions, employment discrimination, exclusive immigration policies, discriminatory voting rights and more. These actions can be attributed to the manifestation of cognitive biases. The emergence of machine learning, a branch of artificial intelligence, attempts to automate the human process of learning and decision making. The human experience is programmed from datasets into algorithms that assists in decision making. The unfortunate side effect is that the human experiences resulting from cognitive bias are also programmed into the algorithms. Social media through the echo chamber, exacerbates the propagation of misinformation. It contributes to the perfect storm by adding the oxygen and fuel to the expansion and spread of bias. This paper examines cognitive bias in the age of machine learning and social media. It describes the life cycle for the proliferation of algorithmic bias through the echo chamber, and devises mitigation strategies during the different stages from initiation to consumption of bias.

Keywords: Machine learning, algorithmic bias, social media, misinformation, echo chamber

1. *Introduction*

Tversky and Kahneman (1974) described heuristics employed in making judgement under uncertainty that can lead to systematic and predictable errors. It described the types of cognitive bias that affect human decisions and predictions. It illustrated the effects of biases including representativeness (using existing stereotype), availability (using information most readily available), anchoring (bias towards initial values), simulation (mentally undoing of past events), etc. The emergence of artificial intelligence and machine learning utilizes algorithms that can be programmed in the decision-making process from datasets that are compiled from human experiences. Inevitably, these algorithms that reflect human experiences would carry over the inherent cognitive biases. Algorithmic bias occurs when erroneous assumptions due to human cognitive bias are used in the machine leaning process. The effect of social media as an echo chamber where like-minded people repeat and exaggerate information in a closed loop is adding the ingredient that stirs this perfect storm of spreading bias. This paper examines the effect of bias as exacerbated by machine learning and social media. It proposes a framework of mitigation strategies for algorithmic bias. The paper is organized as followed. Section 2 discusses bias and social impacts. Section 3 describes machine learning models and algorithmic bias. Section 4examines the effect of social media in spreading biases through echo chambers. Section 5 discusses the algorithmic bias cycle and describes mitigation strategies.

2. *Bias and Social Impacts*

2.1. *Types of Biases*

According to Psychology Today (2022), bias is a tendency, inclination, or prejudice toward or against something or someone. It originates from a person's own experience, perception, and viewpoints. It influences how a person thinks, behaves, perceives, and judges others. Biases can be conscious or subconscious. Conscious bias is intentional, whereas unconscious or implicit bias is not consciously held or deliberately planned or carried out (Merriam-Webster 2022). According to Moule (2009), unconscious bias refers to unconscious forms of discrimination and stereotyping based on race, gender, sexuality, ethnicity, ability, age, and so on. Unconscious bias arises from assumptions and perceptions based on a person's background and life experience.

The concept of cognitive bias was first proposed by Tversky and Kahneman (1974). The article described heuristics employed in making judgement under uncertainty, while usually effective, can lead to systematic and predictable errors. Humanhow (2021) described a comprehensive list of cognitive biases.

Here are a few examples: confirmation bias – tendency to favor information that confirm existing beliefs; hindsight bias, also known as the "I knew it all along" phenomenon - the tendency to see events as more predictable than they are after they happen; and the anchoring bias - the tendency to rely on the first piece of information that one receives. Cognitive biases arise from flawed judgement based on systematic thinking errors, creating mental shortcuts resulting in unconscious bias.

2.2. *Impact of Bias*

The impact of bias spans across different spectrum of society and underlying values. In the following, some examples related to racism, employment, law, and medicine are examined.

Racism: Recent events of racial injustice have awakened the souls of Americans to reflect on our values and question the intrinsic biases of our society. The last two years were bombarded with such events that drew attention nationally and internationally. They include the killings of black Americans - Ronald Greene (May 10, 2019, Louisiana), Breonna Taylor (March 13, 2020, Kentucky), George Floyd (May 25, 2020, Minnesota), Daunte Wright (April 11, 2021, Minnesota); the killings of Asian Americans – Asian spa killings (March 16, 2021, Georgia), Thai immigrant Vichar Ratanapakdee (January 28, 2021, San Francisco), and many recent hate crimes against Asian Americans. Unconscious bias plays a role in these incidents against minorities. Studies have shown that in the shooter task experiment, participants consistently shoot armed black targets faster than they shoot armed white targets. Correll et al. (2014) described participants in the original studies showed bias in both response latencies and error rates; shooting more quickly for black targets and erroneously shooting unarmed black targets more frequently than unarmed whites targets.

Employment: Tversky and Kahneman (1974) described the representativeness heuristic that causes judgement errors. It plays a role in identifying occupation where a person is assessed by the degree to which he is representative of oris like the stereotype of a certain occupation. This cognitive bias often occurs in the hiring process to determine where a candidate is right for a position. Segrest et al. (2006) examined implicit sources of bias in employment interview judgments and decisions. It concluded that "interviewers are still allowing illegal and often irrelevant factors, such as the combined effects of ethnicity and accent, to affect judgments and decisions about job applicants, instead of focusing only on job-related qualifications." Stereotypes based on race, gender, sexuality, ethnicity, ability, age, etc. may come into play.

Law: According to the United States Sentencing Commission (2017), sentences of black male offenders were longer than those of white male offenders for all periods studied and were 19.1 percent longer in the post-report period. Hyman (2014) pointed out that African Americans continue to be arrested, convicted, and sentenced to prison at much higher rates than Caucasians. It attributed implicit bias as a lead contributing factor to inequality across various legal contexts. It further discussed the dispositions affecting judges in making rulings, jurors in deliberating, and prosecutors in deciding how aggressively to pursue a defendant.

Medicine: O'Sullivan and Schofield (2018) explored the pervasiveness of cognitive bias in clinical practice. It described that up to 75% of errors in internal medicine can be cognitive errors that can be identified throughout the diagnostics process. It illustrates that availability bias may result in unnecessary CT scans leading to more radiation exposure for the patient. Confirmation bias may lead to misinterpretation of information to fit preconceived diagnosis. Representativeness bias may cause the misrepresentation of the likelihood of an event based on the population characteristics instead of individual characteristics. Sullivan and Schofield (2018) provided examples of clinical errors based on various cognitive biases. It suggested methods of debiasing to enhance better clinical decision making.

3. *Machine Learning and Algorithmic Bias*

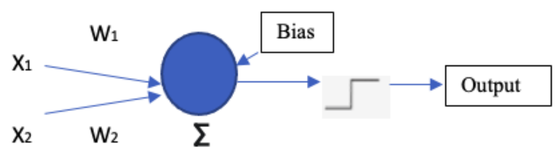
3.1. *Artificial Intelligence and Machine Learning*

Artificial intelligence is an area of computer science that concerns the study, the representation and duplication of human thought process (Sharda et al. 2015). It refers to the ability of a computer to exhibit human intelligence and carry out tasks commonly associated with intelligent beings. Machine learning is a branch of artificial intelligence that automates the learning process of the machine from data and experience, capable of making decisions with minimal human intervention. The learning algorithms are trained to make classifications or predictions. They can be classified as either supervised or unsupervised based on the way in which the patterns are extracted from the historical data. Sharda et al. (2015) shows a taxonomy for data mining tasks, along with the learning methods, and the respective popular algorithms.

Supervised learning is an approach to classify data based on a labeled set. Unsupervised learning is an approach to classify data without labeled sets, rather, data is clustered based on affinity, similarity, or proximity. Accuracy of machine learning relies on the datasets the algorithms deployed. As such, human bias could be transmitted in the design of datasets and learning algorithms. Taniguchi et al. (2018) described some common machine learning models to include the perceptron, the neural networks and support vector machine, logistic regression, the nearest-neighbor rule, naïve Bayes, and random forests.

The following illustrates the concept of bias in neural networks by examining the perceptron. The brain has anywhere from 50 billion to 150 billion neurons partitioned into networks (Sharda et al. 2015). An artificial neural network (ANN) is a type of machine learning model which emulates a biological neural network. It receives inputs from and sends output signals to other neurons, which can be adjusted by weights. The study of a perceptron, an early neural network structure that has no hidden layer, reveals how the ANN works. The perceptron is an algorithm for supervised learning that uses a binary classifier that can decide whether an input belongs to some specific class.

Figure 1 – The Perceptron



The perceptron is trained to recognize the input patterns to give the corresponding output (Figure 1). The summation function $\sum X_i W_i$ computes the activation level of the neuron. The output is produced based on the activation or transfer function, which could be linear or non-linear, and maps the resulting values in the range of values in an interval (a,b). For example, a commonly used activation function, the Sigmoid function defined by $1/(1+e^{-x})$ normalizes the output values between 0 and 1. In the context of the bias function in neural networks, it allows the activation function to shift by adding a constant to the input and is often tuned to train models to better fit the data. It is important to note that the inherent bias inside the data is used to train the models. Therefore, one cannot expect fair treatment from algorithms built from biased data. In order to create models that are ethically unbiased and fair, the model needs to be fine-tuned to mitigate bias in the training data sets. Korteling et al. (2018) described the four basic neural network principles: association, compatibility, retainment, and focus, which are inherent to neural networks which were originally optimized to perform concrete biological, perceptual, and motor functions. Human decisions are ingrained with cognitive biases. Machine learning algorithms which make use of training data based on human judgements inherently extend these biases.

3.2. *The Negative Effects of AI*

Francis (2020) cited the warning by the late legendary physicist Stephen Hawking in 2017 that "Success in creating effective AI could be the biggest event in the history of our civilization. Or the worst." Professor Hawking posits that creating thinking machines poses a threat to our very existence. In an interview with BBC (Cellan-Jones 2014) Prof Stephen Hawking, said that efforts to create thinking machines pose a threat to our very existence. He fears the consequences of creating something that can match or surpass humans would take off on its own and re-design itself at an ever-increasing rate. While artificial intelligence promises positive advancements in many areas in science and humanity, it can have negative impacts causing social injustice and discrimination.

Another negative side effect is bias programmed in machine learning into algorithms that assist decision making. They are designed from datasets representing human experiences and inevitably inherit cognitive biases from the human mind resulting in systematic errors and flawed judgements.Algorithmic bias is at the root of many social ills as illustrated in the following examples. Buolamwini & Bebru (2018) studied the discrimination based on race and gender in machine learning algorithms. In the study, three commercially available facial-recognition technologies made by Microsoft's Cognitive Services Face API, IBM's Watson Visual Recognition API, and Megvii (Face++) misclassified darker-skinned females with error rates of up to 34.7% while the maximum error rate for lighter-skinned males is 0.8%. Horwitz (2021) reported that Facebook provides a platform whose algorithm learns and perpetuates the existing difference in employee demographics.

For example, ads to recruit delivery drivers for Domino's Pizza Inc. were disproportionately shown to men, and that they were more likely to show to women a technical job at Netflix Inc. Larson et al. (2016) pointed out that the COMPAS (Correctional Offender Management Profiling for alternative Sanctions), an algorithm used in state court systems throughout the United States, found that black defendants were far more likely than white defendants to be incorrectly judged to be at a higher risk of recidivism. White defendants on the other hand, were found more likely than black defendants to be incorrectly flagged as low risk.Dastin (2018) reported that Amazon scrappedits AI recruiting tool that showed bias against women and not rating candidates in a gender-neutral way. It attributed to Amazon's models were trained to vet applicants by observing patterns in resumes submitted over a 10-year period, where most came from men.

4. *How Social Media Exacerbates Bias*

Social media refers to the platform that enables social interactions among people, machines, and things, in which they can create, share, and exchange information. The Internet emerged in the late 1960s evolved from the development of ARPNET by the U.S. Defense Department, was designed to link computers at Pentagon-funded research institutes over telephone lines. The growth of the Internet enabled online communication services such as America Online, CompuServe and Yahoo. Some significant social media that emerged included LinkedIn (2002), Myspace (2003), Facebook (2004), Reddit (2005), Twitter (2006), Instagram (2010), Snapchat (2011), and more recently, TikTok (2016).

Social media creates an echo chamber propagating beliefs across like-minded people (Acemoğlu et al. 2021) and exacerbates the spread of misinformation and disinformation. It asserted that echo chambers and the viral spread of misinformation are more likely when articles contain extreme content. Social media reinforces the confirmation bias of individuals and groups of existing beliefs. Acemoğlu et al. (2021) suggests that the optimal platform algorithm is to recommend extreme content that aligns with the most extremist users, leading to the viral spread of misinformation. The effect of social media reaches significant breadth, depth, and distance; and carries all characteristics of Big Data in volume, variety, and velocity.

The major forces fueling the echo-chamber consist of homophily, the "six degrees of separation", and the ubiquity of the Internet. Homophily, the love of sameness, creates the tendency of clustering among like-minded people. Karinthy (1929) described the concept of "six degrees of separation" which suggests that any two people are connected by six or fewer steps of connection. The ubiquity of the Internet transcends temporal, geographic, national, and cultural boundaries. Combined, they create a perfect storm with the intensity and speed in the spread of misinformation and bias. The Internet provides the universal standard allowing disparate technologies to communicate with each other. Social media contributes to the high volume, velocity and variety of data that characterize Big Data. It is estimated that 6,000 tweets are sent every second (Sayce 2020). Worldwide, there are over 2.90 billion monthly active users (MAUs) of Facebook as of June 30, 2021, representing a 7 percent increase in Facebook MAUs year-over-year (Noyes 2021). Social media generates rich messages with text, audio, and video and contributes to the high variety of data. The Internet provides a scalable platform to simultaneously deliver Big Data to large number of people. Adding to this arsenal are mobility and intelligence. The proliferation of mobile devices allows user to send or receive information anytime and anyplace. According to O'Dea (2021), the number of smartphone subscriptions worldwide today surpasses six billion (about 75% of the world population) and is forecast to further grow by several hundred million in the next few years.Artificial intelligence has become a disruptive technology in the digital transformation of business enterprises. In a negative way, its algorithms can enable and spread bias on social media across society.

In order to address the bias on social media, it is necessary to identify the root cause. According to an Indiana University study (Indiana University 2022), political bias on social media emerges from users, not platform. It cited Professor Menczer describing that online influence is affected by the echo-chamber characteristics and that more partisan news sources received more politically aligned followers, embedded in denser echo chambers. In the study by Chen et al. (2021), social bots ("drifters") with neutral (unbiased) and random behavior were deployed as instruments to probe exposure biases in social media. It found that online influence is affected by the echo-chamber characteristics of the social network, which are correlated with partisanship. Drifters following more partisan news sources receive more politically aligned followers, becoming embedded in denser echo chambers and gaining influence within those partisan communities. Therefore, while social media platforms exacerbate the spread of bias, it is the influence of associations in echo-chambers that have the major impact.

5. *Mitigating Algorithmic Bias*

As humankind is tackling critical global problems pertaining to the pandemic, climate change, political change, poverty, social injustice, and so on, bias on social media is taking on equal significance that needs to be dealt with. In the following, the steps to mitigate its impact are examined. The algorithmic bias cycle is illustrated in Figure 2, and the bias mitigation strategies are illustrated in Figure 3.

Figure 2: The Algorithmic Bias Cycle

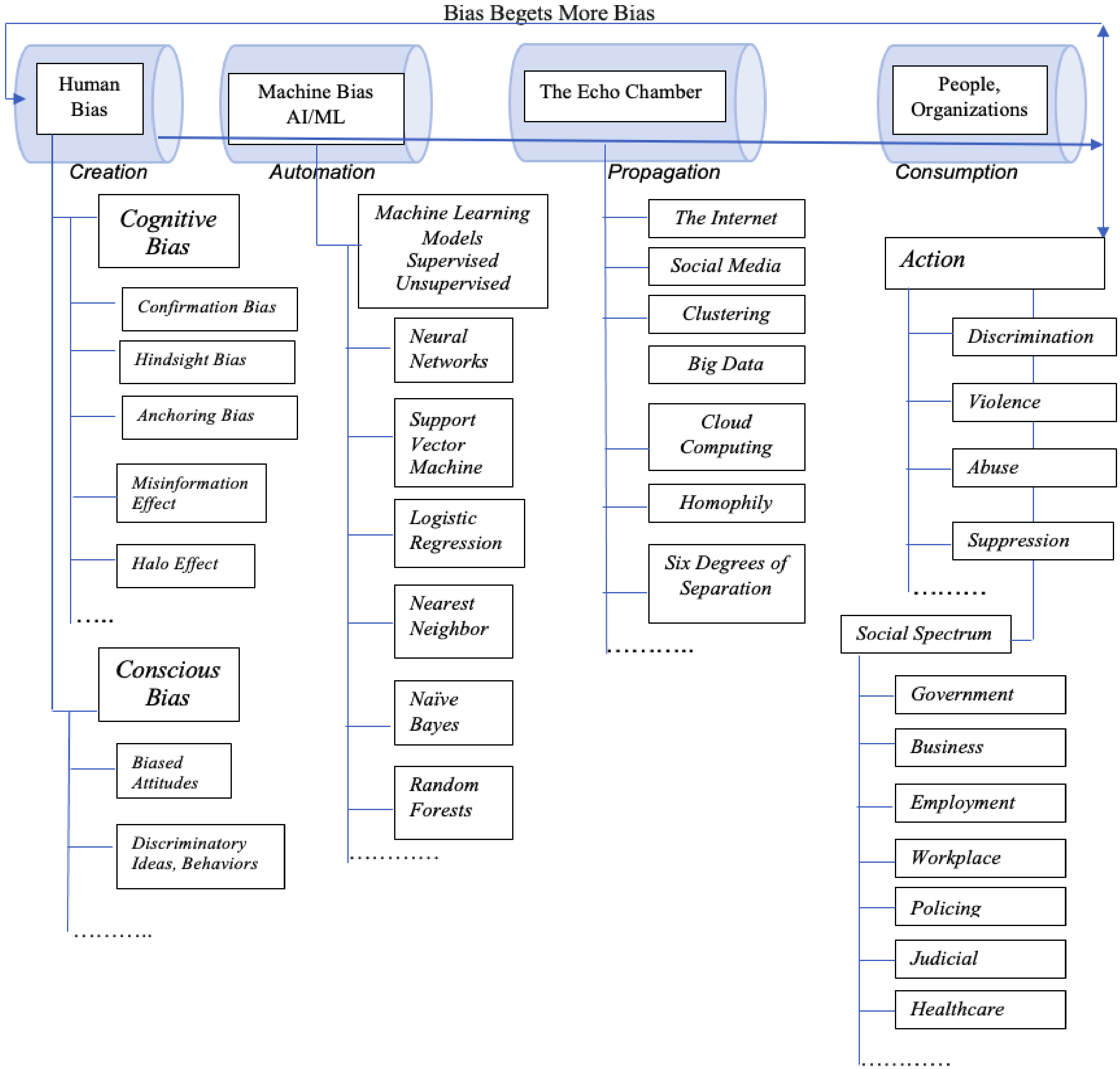
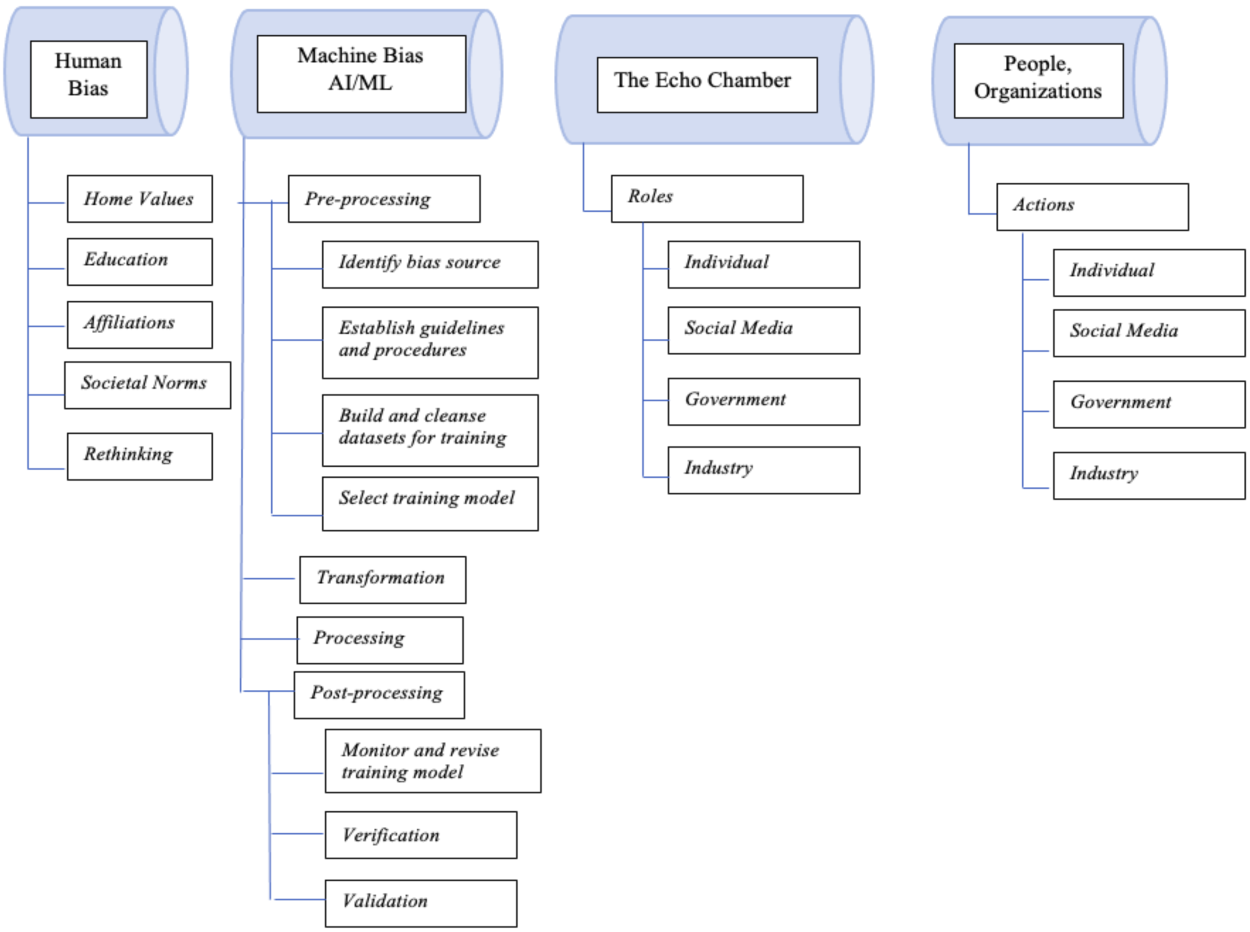


Figure3: Bias Mitigation



5.1. *Mitigating Human Bias*

Human bias is the root of all biases. Wharton (2018) cited management professor Sigal Barsade, "bias is a function of how our brains work, coupled with evidence about the world formed in childhood and influenced by outside information gathered from sources like the media, and the resulting biases are often unconscious." Eradicating machine bias by itself cannot eliminate biases. While the accuracy of machine learning models reflects the design goals, it does not reflect the validity of the goals themselves. Lee et al. (2019) asserts that there is no simple metric to measure fairness that a software engineer can apply, and that fairness is human and requires human involvement. Thus, reducing human bias must start with the value systems developed in one's upbringing as affected by the family, education, and affiliations. Further developmental efforts can contribute to mitigating human bias behaviors through continuous education, societal norms, and rethinking of values that can challenge previous beliefs. Investigation into these efforts is beyond the scope of this paper.

5.2. *Mitigating Machine Bias*

Computer models can reproduce and amplify human biases, explicit or implicit, which are shaped by prejudices and erroneous beliefs. Silberg & Manyika (2019) suggests that the underlying data are most often the main source of machine bias. General mitigation approaches consist of pre-processing, transformation, processing, and post-processing of training datasets. In pre-processing of datasets, the data scientist is looking for the accuracy and completeness of data. Is the data a correct representation of the respective group and do they represent the views of multiple groups? The goal of pre-processing is to minimize the noise in data, choose a strategy for handling missing attribute values, use any suitable method for selecting and ordering attributes (features) according to their informativity, discretize/fuzzify numerical (continuous) attributes and eventually, process continuous classes (Bruha & Famili 2000).

Transformation is the process to organize data from disparate silos into an integrated form that can be used in machine learning models. Processing is the application of the machine algorithm to extract new knowledge. Post-processing includes verification and validation of the model's predictions.

It answers the questions whether the model was developed according to a correct methodology and that whether the model is doing what it is designed to do. Bruha & Famili (2000) described post-processing to consist of knowledge filtering, interpretation and explanation, evaluation, and knowledge integration. It is important to point out that these mitigation approaches do not answer the question whether the model produces an outcome that is fair and equitable. Here it depends on the fairness constraints that are designed into the models. Fairness is a human issue and must be addressed before the constraints are embedded in the algorithms.

5.3. *Mitigating the Echo Chamber Effect*

5.3.1. The Role of Individuals

A key factor in the echo chamber is confirmation bias. Mitigation techniques include allowing oneself to be wrong in the face of new data, testing one's hypothesis by searching out disconfirming evidence, and beware of repetition that doubles down misinformation (MasterClass 2020). Lee et al. (2019) suggests that individual operators of algorithms must abide by U.S. laws and statutes that forbid discrimination. It further describes the adoption of self-regulatory practices such as developing a bias impact statement and brainstorming initial assumptions. Individuals need better algorithmic bias literacy, which is crucial for mitigating bias.

5.3.2. The Role of Social Media

Social media can play a role in mitigating the echo chamber effect by controlling the news sources to suspend or restrict low-credibility or inauthentic sources (Chen et al. 2021), or in diversifying the sources to provide different viewpoints. Wagner (2017) reported that Facebook modified the trending page transitioning from single news source to multiple news sources. It's also going to stop personalizing trending topics for each user. Smith (2017) described the launch of "Outside Your Bubble" feature to give their audience a glimpse outside their own social media spaces to add a transparency that has been lost in the rise of social-media-driven filter bubbles.

5.3.3 The Role of Government and Industry

In reducing algorithmic bias, the government can play a role as a user or as a regulator (Pimentel 2021). As a user, governments have large market power and control over many important algorithmic use cases where it can set standards, provide guidance, and highlight practices to reduce algorithmic bias. As a regulator, government can adapt existing frameworks to incentivize ethical algorithms. It asserts that anti-discrimination as a legal requirement has a strong basis in the U.K and that the U.S. likewise has non-discrimination, civil rights, and sectoral laws that must be updated and connected to the digital world. For example, Title VII of the Civil Rights Act of 1964 prohibits employment discrimination based on race, color, religion, sex, and national origin. The First Amendment provides broad protection of speeches from government censorship with exceptions to specific situations for example when the hate speech incites imminent lawless action, or when pornography constitutes obscenity (Freedom Forum 2022). Such prohibition of government censorship of public speeches renders its ineffectiveness of any meaningful restriction of expressions by bad actors. However, private companies like Twitter, Facebook or Google are not bound and can set their own standards for regulation. For the sake of society and their own long-term preservation, these large companies should self-regulate and come up with industry standards, and enforcement. For example, while major social media have policy statements regarding hate speeches, its policy interpretation and enforcement standards vary. For example, Facebook (now Meta) may allow room for certain types of hate speech while requiring people to clearly indicate their intent before removal of content (Meta 2022). Wagner (2021) reported that Twitter Inc. "took action" on a record number of user accounts for violating the company's hate speech policies during the second half of 2020. Some notable permanent suspensions include former President Donald Trump's account due to the risk of further incitement of violence, and most recently Representative Marjorie Taylor Greene for violation of Covid-19 misinformation policies (Alba 2022). In the meantime, Facebook banned Greene for 24 hours over COVID misinformation (Pitofsky 2022).

5.3.4. Mitigating Biased Actions

On the receiving end are the people and organizations that will utilize and execute the information or misinformation into actions. What are the guardrails that can be put in place to mitigate bias in decision making? Carter et al. (2020) described anti-bias training increases the awareness of bias and provides strategies to change people's behavior. It recommends the coupling of anti-bias training with other diversity and inclusion initiatives. Wharton (2018) argued that there's lack of evidence that diversity training works. It stressed that in addition to educating people what bias is, it is important to prescribe strategies for individuals and organizations, so they know what to do in their decision making.

It suggests that employee resource groups creating micro-programs to continue conversations about bias and formal programs creating accountability are necessary follow-ups. It emphasizes that it's in the day-to-day reinforcement of desired behavior that behavior changes. Reducing bias in an organization is a cultural change. It must start with the top of the organization. Vinkenburg (2017) described the strategy of engaging gatekeepers, who are decision makers in organizational positions of power. It described engaging power holders as essential for making diversity interventions successful. Other strategies include setting standards and enforcement policies.

6. Conclusion

The Orwellian society described in George Orwell's 1984 (Orwell 1949) is well and alive in 2022withrumors and misinformation abound. It described the family had in effect an extension of the Thought Police, a device by means everyone could be surrounded night and day by informers who know them intimately. The device now takes a different form, but still is ubiquitous around people's lives. These effects are exacerbated by modern technology with machine learning algorithms and the social media. People are embedded inside the tunnels of the echo chamber where misinformation is growing and spreading. There are consequences in society that include injustice, discrimination, inequity, and social unrest. In order to solve problem, one must acknowledge and identify the root of the problem. This paper exposes significant sources contributed by cognitive biases exacerbated by advance in technologies in machine learning and social media. It examines the nature of cognitive biases, machine learning algorithms, and the echo chamber effect. It provides a framework for mitigation strategies at the human, machine, echo chamber and organization levels. The research areas involve multiple disciplines, that include psychology, economics, mathematics, computer, and social sciences. Future research can focus on specific areas such as the interplay between mathematical psychology and computer algorithms in addressing cognitive biases in society.

References

Acemoğlu, D., Ozdaglar, A., and Siderius, J. (2021). "Misinformation: Strategic sharing, homophily, and endogenous echo chambers". VoxEU & CEPR, 30 June 2021. Available: https://voxeu.org/article/misinformation-social-media

Alba, D. (2022). "Twitter Permanently Suspends Marjorie Taylor Greene's Account". The New York Times. Jan. 2, 2022. Available: https://www.nytimes.com/2022/01/02/technology/marjorie-taylor-greene-twitter.html

Bruha, I., & Famili, A. (2000). Postprocessing in machine learning and data mining. ACM SIGKDD Explorations Newsletter, 2(2), 110-114. Available: https://dl.acm.org/doi/abs/10.1145/380995.381059

Buolamwini, J. and Gebru, T. (2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification". Proceedings of Machine Learning Research 81:1–15, 2018 Conference on Fairness, Accountability, and Transparency. Available: http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Carter, E., Onyeador, I.N. and Lewis, N.A. Jr. (2020). "Developing & delivering anti-bias training: Challenges and recommendations". Behavioral science & policy, volume 6, issue 1, 2020.Available: https://behavioralpolicy.org/wp-content/uploads/2020/08/Developing-delivering-effective-anti-bias-training-Challenges-.pdf

Cellan-Jones, R. (2014). "Stephen Hawking warns artificial intelligence could end mankind". BBC News, Technology. 2 December 2014. Available: https://www.bbc.com/news/technology-30290540

Chen, W., Pacheco, D., Yang, KC., and Menczer (2021). Neutral bots probe political bias on social media. Nat Commun 12, 5580 (2021). https://doi.org/10.1038/s41467-021-25738-6

Correll, J., Hudson, S.M., Guillermo, S., & Ma, D.S. (2014), "The Police Officer's Dilemma: A Decade of Research on Racial Bias in the Decision to Shoot: The Police Officer's Dilemma." *Social and personality psychology compass* 8.5 (2014): 201–213. Available: https://doi.org/10.1111/spc3.12099

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

Francis, D. (2020). “All bets are off if AI becomes smarter than people”. Financial Post, Innovation. Feb 03, 2020. Available: https://financialpost.com/diane-francis/diane-francis-all-bets-are-off-if-ai-becomes-smarter-than-people

Freedom Forum (2022). “FREE EXPRESSION ON SOCIAL MEDIA”. FREEDOM FORUM INSTITUTE. Available: *https://www.freedomforuminstitute.org/first-amendment-center/primers/free-expression-on-social-media/*

Horwitz, J. (2021). Facebook Algorithm Shows Gender Bias in Job Ads. The Wall Street Journal, Tech, April 9, 2021. Available:https://www.wsj.com/articles/facebook-shows-men-and-women-different-job-ads-study-finds-11617969600

Humanhow (2021). “The Ultimate List of Cognitive Biases: Why Humans Make Irrational Decisions”. Available: http://humanhow.com/list-of-cognitive-biases-with-examples/

Hyman, M.B. (2014). “Implicit Bias in the Courts”. Illinois Bar Journal. January 2014, Volume 102 • Number 1.

Indiana University (2022). “Political bias on social media emerges from users, not platform, IU study says”. Research Impact, Informatics and Data Science. Available: https://research.impact.iu.edu/our-strengths/informatics-data-science/social-media-platform-bias.html#:~:text=Political%20bias%20on%20social%20media%20emerges%20from,not%20platform%2C%20IU%20study%20says&text=Political%20biases%2C%20the%20researchers%20found,our%20very%20first%20online%20friends.

Karinthy, F. (1929), “CHAIN-LINKS”. 1929, Everything is Different. Available: https://djjr-courses.wdfiles.com/local--files/soc180%3Akarinthy-chain-links/Karinthy-Chain-Links_1929.pdf

Korteling, J., Brouwer, A.M., andToet, A. (2018). “HYPOTHESIS AND THEORY”. Front. Psychol., 03 September 2018. Available: https://doi.org/10.3389/fpsyg.2018.01561

Larson, J., Mattu, S., Kirchner, L., and Angwin, J. (2016). “How We Analyzed the COMPAS Recidivism Algorithm”. ProPublica, May 23, 2016. Available: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

Lee, N. T., Resnick, P., & Barton, G. (2019). “Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms”. Brookings Institute: Washington, DC, USA. Available: https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/

MasterClass (2020). “How to Identify Confirmation Bias: 3 Ways to Reduce Bias”. MasterClass, Science & Tech, Nov 8, 2020. Available:https://www.masterclass.com/articles/how-to-identify-confirmation-bias#what-is-confirmation-bias

Merriam-Webster (2022). Merriam-Webster Dictionary. Available: https://www.merriam-webster.com/dictionary/unconscious

Meta (2022). “Hate Speech”. Available: https://transparency.fb.com/policies/community-standards/hate-speech/

Moule, J. (2009). “Understanding Unconscious Bias and Unintentional Racism”. Sage Journals, Phi Delta Kappan, Volume: 90, issue: 5, page(s): 320-326, January 1, 2009.Available: https://doi.org/10.1177/003172170909000504

Noyes, D. (2021). “The Top 10 Valuable Facebook Statistics – Q2 2021”. Zephoria, Digital Marketing Available: https://zephoria.com/top-15-valuable-facebook-statistics/

O’Dea, S. (2021) “Smartphone users worldwide 2016-2021”. Statista. Available:https://www.statista.com/statistics/330695/number-of-smartph one-users-worldwide/

Orwell, George. 1984. London: Secker and Warburg, 1949. Print.

O’Sullivan, ED., and Schofield, SJ. (2018). “Cognitive bias in clinical medicine”. J R Coll Physicians Edinb 2018; 48: 225–232 | doi: 10.4997/JRCPE.2018.306. Available: https://www.rcpe.ac.uk/sites/default/files/jrcpe_48_3_osullivan.pdf

Pimentel, H. (2021). “Should the government play a role in reducing algorithmic bias?” Brookings Institution, March 19, 2021. Available: https://www.brookings.edu/events/should-the-government-play-a-role-in-reducing-algorithmic-bias/

Pitofsky, M. (2022). “Rep. Marjorie Taylor Greene barred from Facebook for 24 hours over COVID misinformation”. USA TODAY Jan 3, 2022. Available: https://www.usatoday.com/story/news/politics/2022/01/03/marjorie-taylor-greene-facebook-ban-covid/9081567002/

Psychology Today (2022). “Bias”. Available: https://www.psychologytoday.com/us/basics/bias

Sayce, D. (2020). "The Number of tweets per day in 2020". Digital Marketing Consultant. Available:https://www.dsayce.com/social-media/tweets-day/

Segrest, S. L., Perrew, P. L., Gillespie, T. L., Mayes, B. T., & Ferris, G. R. (2006). "Implicit sources of bias in employment interview judgments and decisions". Organizational Behavior and Human Decision Processes, 101(2), 152-167. doi:10.1016/ j.obhdp.2006.06.005. Available: https://digital.stpetersburg.usf.edu/cgi/viewcontent.cgi?article=1334&context=fac_publications

Sharda, R., Delen, D. & Turban, Efram (2015). "Business Intelligence and Analytics: Systems for Decision Support".10th Edition. Pearson Education. EISBN13: 9780133868906.

Silberg, J., and Manyika, J. "Tackling bias in artificial intelligence (and in humans)". McKinsey Global Institute, June 6, 2019. Available: https://www.mckinsey.com/~/media/mckinsey/featured%20insights/artificial%20intelligence/tackling%20bias%20in%20artificial%20intelligence%20and%20in%20humans/mgi-tackling-bias-in-ai-june-2019.pdf

Smith, B. (2017). "Helping You See Outside Your Bubble". BuzzFeed, Feb 17, 2017. Available: https://www.buzzfeed.com/bensmith/helping-you-see-outside-your-bubble

Tversky, A. and Kahneman, D. (1974). "Judgment under Uncertainty: Heuristics and Biases". Science, New Series, Vol. 185, No. 4157. (Sep. 27, 1974), pp. 1124-1131. Available: https://www2.psych.ubc.ca/~schaller/Psyc590Readings/TverskyKahneman1974.pdf

United States Sentencing Commission (2017). "Demographic Differences in Sentencing". Available: https://www.ussc.gov/sites/default/files/pdf/research-and-publications/research-publications/2017/20171114_Demographics.pdf

Vinkenburg, C.J. (2017). "Engaging Gatekeepers, Optimizing Decision Making, and Mitigating Bias: Design Specifications for Systemic Diversity Interventions". J Appl Behav Sci. 2017 Jun; 53(2):212-234. Available:https://doi.org/10.1177/0021886317703292

Wagner, K. (2017). "Facebook is changing its Trending section to fight the spread of fake news". Vox, Jan 25, 2017. Available: https://www.vox.com/2017/1/25/14376734/facebook-trending-topics-update-fake-news

Wagner, K. (2021). "Twitter Penalizes Record Number of Accounts for Posting Hate Speech". BLOOMBERG JULY 14, 2021. Available: https://time.com/6080324/twitter-hate-speech-penalties/

Wharton (2018). "How Companies Can Develop Anti-bias Strategies that Work". Knowledge@Wharton, Management, May 11, 2018. Available: https://knowledge.wharton.upenn.edu/article/how-companies-can-develop-anti-bias-strategies-that-work/