Vol.3,Issue 01, March 2024
Journal of Artificial Intelligence General Science JAIGS

Home page http://jaigs.org

Examining Ethical Aspects of AI: Addressing Bias and Equity in the Discipline

Jeff Shuford

Nationally Syndicated Business & Technology Columnist,USA

*Corresponding Author: Jeff Shuford

ARTICLEINFO
Article History:
Received:
05.03.2024
Accepted:
10.03.2024
Online: 07.04.2024
Keyword: artificial intelligence, bias, fairness, discrimination, mitigation strategies

ABSTRACT

The rapid progress in implementing Artificial Intelligence (AI) across various domains such as healthcare decision-making, medical diagnosis, and others has raised significant concerns regarding the fairness and bias embedded within AI systems. This is particularly crucial in sectors like healthcare, employment, criminal justice, credit scoring, and the emerging field of generative AI models (GenAI) producing synthetic media. Such systems can lead to unfair outcomes and perpetuate existing inequalities, including biases ingrained in the synthetic data representation of individuals.This survey paper provides a concise yet comprehensive examination of fairness and bias in AI, encompassing their origins, ramifications, and potential mitigation strategies. We scrutinize sources of bias, including data, algorithmic, and human decision biases, shedding light on the emergent issue of generative AI bias where models may replicate and amplify societal stereotypes. Assessing the societal impact of biased AI systems, we spotlight the perpetuation of inequalities and the reinforcement of harmful stereotypes, especially as generative AI gains traction in shaping public perception through generated content.Various proposed mitigation strategies are explored, with an emphasis on the ethical considerations surrounding their implementation. We stress the necessity of interdisciplinary collaboration to ensure the effectiveness of these strategies. Through a systematic literature review spanning multiple academic disciplines, we define AI bias and its various types, delving into the nuances of generative AI bias. We discuss the adverse effects of AI bias on individuals and society, providing an overview of current approaches to mitigate bias, including data preprocessing, model selection, and post-processing. Unique challenges posed by generative AI models are highlighted, underscoring the importance of tailored strategies to address them effectively.Addressing bias in AI necessitates a holistic approach, involving diverse and representative datasets, enhanced transparency, and accountability in AI systems, and exploration of alternative AI paradigms prioritizing fairness and ethical considerations. This survey contributes to the ongoing discourse on developing fair and unbiased AI systems by outlining the sources, impacts, and mitigation strategies related to AI bias, with a particular focus on the burgeoning field of generative AI.

Introduction

The increasing utilization of AI systems has intensified discussions regarding fairness and bias in artificial intelligence, as potential biases and discrimination become more evident. This survey investigates the origins, consequences, and methods to mitigate fairness and bias issues in AI. Several studies have uncovered biases against certain groups in AI systems, such as the facial recognition systems scrutinized by Buolamwini and Gebru (2018), and hiring algorithms examined by Dastin (2018) and Kohli (2020). These biases can perpetuate systemic discrimination and inequality, adversely affecting individuals and communities in hiring, lending, and criminal justice domains (O'Neil, 2016; Eubanks, 2018; Barocas and Selbst, 2016; Kleinberg et al., 2018).

Researchers and practitioners have proposed various mitigation strategies, including enhancing data quality (Gebru et al., 2021) and developing explicitly fair algorithms (Berk et al., 2018; Friedler et al., 2019; Yan et al., 2020). This paper offers a comprehensive examination of bias sources and impacts in AI, scrutinizing data, algorithmic, and user biases, along with their ethical implications. It surveys ongoing research on mitigation strategies, discussing their challenges, limitations, and the importance of interdisciplinary collaboration.

The significance of fairness and bias in AI is widely acknowledged by researchers, policymakers, and the academic community (Kleinberg et al., 2017; Caliskan et al., 2017; Buolamwini and Gebru, 2018; European Commission, 2019; Schwartz et al., 2022; Ferrara, 2023). This survey paper delves into the complex and multifaceted issues surrounding fairness and bias in AI, encompassing bias sources, their impacts, and proposed mitigation strategies. Overall, the paper aims to contribute to ongoing efforts to develop more responsible and ethical AI systems by shedding light on the sources, impacts, and mitigation strategies of fairness and bias in AI.

Sources Of Bias In Ai

Artificial intelligence (AI) holds immense potential to revolutionize numerous industries and enhance people's lives in various ways. However, a significant challenge in the development and deployment of AI systems is the presence of bias. Bias refers to systematic errors in decision-making processes that result in unfair outcomes. In the context of AI, bias can emerge from multiple sources, including data collection, algorithm design, and human interpretation. Machine learning models, a type of AI system, can learn and replicate biases present in the data used to train them, leading to unfair or discriminatory outcomes. In this section, we will delve into the different sources of bias in AI, including data bias, algorithmic bias, and user bias, and explore real-world examples of their impact.

Definition Of Bias In Ai And Its Different Types

Bias is defined as systematic errors in decision-making processes that lead to unfair outcomes. In the context of AI, bias can arise from various sources, including data collection, algorithm design, and human interpretation. Machine learning models, being a type of AI system, can learn and replicate biases present in the data used to train them, resulting in unfair or discriminatory outcomes. It is crucial to identify and address bias in AI to ensure fairness and equity for all users. In the next sections, we will explore the sources, impacts, and mitigation strategies of bias in AI in more detail.

Sources Of Bias In Ai, Including Data Bias, Algorithmic Bias, And User Bias

Bias in AI can originate from different stages of the machine learning pipeline, including data collection, algorithm design, and user interactions. This survey discusses the various sources of bias in AI and provides examples of each type, including data bias, algorithmic bias, and user bias (Selbst et al., 2016; Crawford & Calo, 2016).

Data bias occurs when the data used to train machine learning models is unrepresentative or incomplete, leading to biased outputs. This can happen when the data is collected from biased sources or when it is incomplete, missing crucial information, or contains errors. Algorithmic bias, on the other hand, occurs when the algorithms used in machine learning models have inherent biases that are reflected in their outputs. This can happen when algorithms are based on biased assumptions or when they use biased criteria to make decisions. User bias occurs when the people using AI systems introduce their biases or prejudices into the system, consciously or unconsciously. This can happen when users provide biased training data or when they interact with the system in ways that reflect their biases.

To mitigate these sources of bias, various approaches have been proposed, including dataset augmentation, bias-aware algorithms, and user feedback mechanisms. Dataset augmentation involves adding more diverse data to training datasets to increase representativeness and reduce bias. Bias-aware algorithms involve designing algorithms that consider different types of bias and aim to minimize their impact on the system's outputs. User feedback mechanisms involve soliciting feedback from users to help identify and correct biases in the system.

Research in this area is ongoing, with new approaches and techniques being developed to address bias in AI systems. It is crucial to continue investigating and developing these approaches to create AI systems that are more equitable and fairer for all users.

Real-World Examples of Bias In Ai

Numerous instances of bias in AI systems have been observed across various industries, ranging from healthcare to criminal justice. One well-known example is the COMPAS system utilized in the United States criminal justice system, which predicts the likelihood of a defendant reoffending. A study by ProPublica revealed bias against African-American defendants in this system, as they were more likely to be labeled as high-risk even without prior convictions. Similar biases were found in a comparable system used in the state of Wisconsin (Angwin et al., 2016).

In healthcare, an AI system used to predict patient mortality rates was found to be biased against African-American patients. Research conducted by Obermeyer et al. (2019) indicated that the system tended to assign higher-risk scores to African-American patients, even when other factors, such as age and health status, were identical. Such biases can lead to African-American patients being denied access to healthcare or receiving inferior treatment.

Another example of bias in AI systems is the facial recognition technology employed by law enforcement agencies. A study by the National Institute of Standards and Technology (NIST) revealed that facial recognition technology exhibited significantly lower accuracy rates for individuals with darker skin tones, resulting in higher rates of false positives (Schwartz et al., 2022). This bias can have severe consequences, including wrongful arrests or convictions.

With the emergence of generative AI systems (GenAI), the risk of harmful biases amplifies. An alarming instance of GenAI bias was reported, wherein text-to-image models like StableDiffusion, OpenAI's DALL-E, and Midjourney exhibited racial and stereotypical biases in their outputs (Nicoletti & Bass, 2023). When tasked with generating images of CEOs, these models predominantly produced images of men, reflecting gender bias. Moreover, when prompted to generate images of criminals or terrorists, the models' output overwhelmingly depicted more people of color.

This incident underscores the risk of generative AI perpetuating societal biases. GenAI models trained on internet-sourced images are likely to inherit such biases, as the data reflects existing disparities. This example highlights the critical need for diverse and balanced training datasets in AI development to ensure fair and representative outputs from generative models.

These examples underscore the serious consequences of bias in AI systems and emphasize the need for careful evaluation and mitigation strategies to address such biases.

Typeof Bias	**Description**	**Examples**
SamplingBias	Occurs when the training data is not representativeofthepopulationitserves, leadingtopoorperformanceandbiased predictionsforcertaingroups.	Afacialrecognitionalgorithmtrained mostly on white individuals that performs poorly on people of other races.
Algorithmic Bias	Resultsfromthedesignandimplementation ofthealgorithm,whichmayprioritizecertain attributesandleadtounfairoutcomes.	An algorithm that prioritizes age or gender,leadingtounfairoutcomesin hiring decisions.
Representation Bias	Happenswhenadatasetdoesnotaccurately representthepopulationitismeanttomodel, leading to inaccurate predictions.	Amedical dataset that under-represents women, leading to less accuratediagnosisforfemalepatients.
Confirmation Bias	MaterializeswhenanAIsystemisusedto confirmpre-existingbiasesorbeliefsheldby its creators or users.	AnAIsystemthatpredictsjob candidates'successbasedonbiases held by the hiring manager.
Measurement Bias	Emerges when data collection or measurementsystematicallyover-orunder-representscertaingroups.	Asurveycollectingmoreresponses fromurbanresidents,leadingtoan under-representationofruralopinions.
Interaction Bias	OccurswhenanAIsysteminteractswith humansinabiasedmanner,resultingin unfairtreatment.	Achatbotthatrespondsdifferentlyto men and women, resulting in biased communication.
Generative Bias	OccursingenerativeAImodels,likethose usedforcreatingsyntheticdata,images,or text. Generative bias emerges when the model's outputs disproportionately reflect specificattributes, perspectives, orpatterns presentinthetrainingdata,leadingtoskewed or unbalanced representations in generated content.	Atextgenerationmodeltrained predominantly on literature from Western authors may over-represent Western cultural norms and idioms, under-representing or misrepresenting othercultures.Similarly,animage generation model trained on datasets with limited diversity in human portraitsmaystruggletoaccurately representabroadrangeofethnicities.

Impacts of Bias In Ai

The rapid advancement of artificial intelligence (AI) has brought about numerous benefits, yet it also poses potential risks and challenges. One of the paramount concerns is the negative impacts of bias in AI on individuals and society. Bias in AI can perpetuate and even exacerbate existing inequalities, resulting in discrimination against marginalized groups and restricting their access to essential services. In addition to reinforcing gender stereotypes and discrimination, it can also give rise to new forms of discrimination based on factors such as skin color, ethnicity, or physical appearance. To ensure fairness, equity, and inclusivity in AI systems, it is crucial to identify and mitigate bias. Moreover, the use of biased AI raises numerous ethical implications, including the potential for discrimination, the responsibility of developers and policymakers, erosion of public trust in technology, and limitations on human agency and autonomy. Addressing these ethical concerns will necessitate a concerted effort from all stakeholders involved and the development of ethical guidelines and regulatory frameworks promoting fairness, transparency, and accountability in the development and deployment of AI systems.

Negative Impacts Of Bias In Ai On Individuals And Society, Including Discrimination And Perpetuation Of Existing Inequalities

The negative impacts of bias in AI can be profound, affecting both individuals and society. Discrimination is a key concern associated with biased AI systems, as they can perpetuate and exacerbate existing inequalities (Sweeney, 2013). For instance, biased algorithms employed in the criminal justice system can result in unfair treatment of certain groups, particularly people of color, who are more likely to face wrongful convictions or harsher sentences (Angwin et al., 2016).

Moreover, bias in AI can hinder individuals' access to essential services such as healthcare and finance. Biased algorithms may lead to the underrepresentation of certain groups, such as people of color or those from lower socioeconomic backgrounds, in credit scoring systems, making it challenging for them to secure loans or mortgages (Dwork et al., 2012).

Furthermore, bias in AI can perpetuate gender stereotypes and discrimination. For example, facial recognition algorithms trained on primarily male data may struggle to accurately recognize female faces, perpetuating gender bias in security systems (Buolamwini & Gebru, 2018). When prompted to generate images of CEOs, some AI models tend to reinforce stereotypes by predominantly depicting CEOs as men (Nicoletti & Bass, 2023).

In addition to perpetuating existing inequalities, bias in AI can also lead to new forms of discrimination, such as those based on skin color, ethnicity, or physical appearance. The same AI models that exhibit gender bias may also depict criminals or terrorists as people of color.

The public deployment of these biased systems can have serious consequences, including denial of services, job opportunities, or even wrongful arrests or convictions. The risk is twofold: on an individual level, it affects people's perception of themselves and others, potentially influencing their opportunities and interactions. On a societal level, the widespread use of such biased AI systems can entrench discriminatory narratives and hinder efforts toward equality and inclusivity. As AI becomes more integrated into our daily lives, the potential for such technology to shape cultural norms and social structures becomes more significant, underscoring the importance of addressing these biases in the developmental stages of AI systems to mitigate their harmful impacts (Ferrara, 2023; Ferrara, 2023b).

Discussion of The Ethical Implications of Biased Ai

The use of biased AI raises numerous ethical implications that must be carefully considered. One of the primary concerns is the potential for discrimination against individuals or groups based on factors such as race, gender, age, or disability (Noble, 2018). Biased AI systems can perpetuate existing inequalities and reinforce discrimination against marginalized groups. This is especially concerning in sensitive areas such as healthcare, where biased AI systems can lead to unequal access to treatment or harm patients (Obermeyer et al., 2019).

Another ethical concern is the responsibility of developers, companies, and governments in ensuring that AI systems are designed and used in a fair and transparent manner. If an AI system is biased and produces discriminatory outcomes, the responsibility lies not only with the system itself but also with those who created and deployed it (Mittelstadt et al., 2016). As such, it is crucial to establish ethical guidelines and regulatory frameworks that hold those responsible for the development and use of AI systems accountable for any discriminatory outcomes.

Moreover, the use of biased AI systems may undermine public trust in technology, leading to decreased adoption and even rejection of new technologies. This can have serious economic and social implications, as the potential benefits of AI may not be realized if people do not trust the technology or if it is seen as a tool for discrimination.

Finally, it is important to consider the impact of biased AI on human agency and autonomy. When AI systems are biased, they can limit individual freedoms and reinforce societal power dynamics. For example, an AI system used in a hiring process may disproportionately exclude candidates from marginalized groups, limiting their ability to access employment opportunities and contribute to society.

Addressing the ethical implications of biased AI will require a concerted effort from all stakeholders involved, including developers, policymakers, and society at large. It will be necessary to develop ethical guidelines and regulatory frameworks that promote fairness, transparency, and accountability in the development and use of AI systems (Ananny & Crawford, 2018). Additionally, it will be important to engage in critical discussions about the impact of AI on society and to empower individuals to participate in shaping the future of AI in a responsible and ethical manner.

Mitigation Strategies For Bias In Ai

Researchers and practitioners have proposed various approaches to mitigate bias in AI. These approaches encompass preprocessing data, model selection, and post-processing decisions. However, each approach encounters limitations and challenges, such as the lack of diverse and representative training data, the difficulty of identifying and measuring different types of bias, and the potential trade-offs between fairness and accuracy. Additionally, there are ethical considerations regarding how to prioritize different types of bias and which groups to prioritize in bias mitigation efforts.

Despite these challenges, mitigating bias in AI is crucial for creating fair and equitable systems that benefit all individuals and society. Ongoing research and development of mitigation approaches are necessary to overcome these challenges and ensure that AI systems are used for the benefit of all.

Overview of Current Approaches To Mitigate Bias In Ai, Including Pre-Processing Data, Model Selection, And Post-Processing Decisions

Mitigating bias in AI poses a complex and multifaceted challenge. However, several approaches have been proposed to address this issue. One common approach is to pre-process the data used to train AI models to ensure that it is representative of the entire population, including historically marginalized groups. This can involve techniques such as oversampling, undersampling, or synthetic data generation (Koh & Liang, 2017). For example, a study by

Buolamwini and Gebru (2018) demonstrated that oversampling darker-skinned individuals improved the accuracy of facial recognition algorithms for this group. Pre-processing data involves identifying and addressing biases in the data before the model is trained. This can be done through techniques such as data augmentation, which involves creating synthetic data points to increase the representation of underrepresented groups, or through adversarial debiasing, which involves training the model to be resilient to specific types of bias (Zhang et al., 2018). Documenting such dataset biases and augmentation procedures is of paramount importance (Gebru et al., 2021).

Another approach to mitigate bias in AI is to carefully select the models used to analyze the data. Researchers have proposed using model selection methods that prioritize fairness, such as those based on group fairness (Yan et al., 2020) or individual fairness (Zafar et al., 2017). For example, a study by Kamiran and Calders (2012) proposed a method for selecting classifiers that achieve demographic parity, ensuring that the positive and negative outcomes are distributed equally across different demographic groups. Another approach is to use model selection techniques that prioritize fairness and mitigate bias. This can be done through techniques such as regularization, which penalizes models for making discriminatory predictions, or through ensemble methods, which combine multiple models to reduce bias (Dwork et al., 2018).

Post-processing decisions are another approach to mitigate bias in AI. This involves adjusting the output of AI models to remove bias and ensure fairness. For example, researchers have proposed post-processing methods that adjust the decisions made by a model to achieve equalized odds, which ensures that false positives and false negatives are equally distributed across different demographic groups (Hardt et al., 2016).

While these approaches hold promise for mitigating bias in AI, they also have limitations and challenges. For example, pre-processing data can be time-consuming and may not always be effective, especially if the data used to train models is already biased. Additionally, model selection methods may be limited by the lack of consensus on what constitutes fairness, and post-processing methods can be complex and require large amounts of additional data (Barocas & Selbst, 2016). Therefore, it is crucial to continue exploring and developing new approaches to mitigate bias in AI.

In the realm of generative AI, addressing bias is even more challenging as it requires a holistic strategy (Ferrara, 2023). This begins with the pre-processing of data to ensure diversity and representativeness. This involves the deliberate collection and inclusion of varied data sources that reflect the breadth of human experience, thus preventing the overrepresentation of any single demographic in training datasets. Model selection must then prioritize algorithms that are transparent and capable of detecting when they are generating biased outputs. Techniques such as adversarial training, where models are continually tested against scenarios designed to reveal bias, can be beneficial. Post-processing involves critically assessing the AI-generated content and, if necessary, adjusting the outputs to correct for biases. This might include using additional filters or transfer learning techniques to refine the models further. Regular audits, continuous monitoring, and incorporating feedback loops are essential to ensure that generative AI systems remain fair and equitable over time. These efforts must be underpinned by a commitment to ethical AI principles, actively engaging diverse teams in AI development, and fostering interdisciplinary collaboration to address and mitigate AI bias effectively.

Furthermore, implementing these approaches requires careful consideration of ethical and societal implications. For example, adjusting the model's predictions to ensure fairness may result in trade-offs between different forms of bias and may have unintended consequences on the distribution of outcomes for different groups (Kleinberg et al., 2018; Ferrara, 2023c).

Approach	Description	Examples	Limitations and Challenges	Ethical Considerations
Pre-processing Data	Involvesidentifyingand addressingbiasesinthe databeforetrainingthe model.Techniques such asoversampling, undersampling,or syntheticdata generation are used to ensurethedatais representativeofthe entire population, includinghistorically marginalizedgroups.	1.Oversamplingdarker-skinnedindividualsina facialrecognitiondataset (Buolamwiniand Gebru, 2018).2.Data augmentationtoincrease representationof underrepresentedgroups. 3.Adversarialdebiasing to trainthemodeltobe resilienttospecifictypes ofbias(Zhangetal., 2018).	1.Time-consuming process.2. Maynot alwaysbe effective, especially if thedataused totrain models is already biased.	1.Potentialfor over-or underrepresentation ofcertaingroupsin thedata,whichcan perpetuateexisting biasesorcreate new ones.2.Privacy concernsrelatedto datacollectionand usage,particularly forhistorically marginalized groups.
Model Selection	Focusesonusingmodel selectionmethodsthat prioritizefairness. Researchershave proposedmethods basedongroupfairness orindividualfairness. Techniquesinclude regularization,which penalizesmodelsfor makingdiscriminatory predictions,and	1.Selectingclassifiersthat achievedemographic parity(Kamiranand Calders,2012).2.Using modelselectionmethods basedongroupfairness (Yanetal., 2020) or individualfairness(Zafar etal.,2017). 3. Regularizationtopenalize discriminatorypredictions. 4.Ensemblemethodsto	Limitedby thepossible lackof consensuson what constitutes fairness.	1.Balancing fairnesswithother performance metrics,suchas accuracyor efficiency.2. Potentialfor modelstoreinforce existingstereotypes orbiasesiffairness criteriaarenot carefully

	ensemblemethods, which combine multiple modelstoreducebias.	combinemultiplemodels andreducebias(Dworket al.,2018).		considered.
Post-processing Decisions	Involvesadjustingthe outputofAImodelsto removebiasandensure fairness.Researchers haveproposedmethods thatadjustthedecisions madebyamodelto achieveequalizedodds, ensuringthatfalse positives andfalse negativesareequally distributedacross differentdemographic groups.	Post-processingmethods that achieve equalized odds(Hardtetal.,2016).	Canbe complexand requirelarge amountsof additional data(Barocas &Selbst, 2016).	1.Trade-offs betweendifferent formsofbiaswhen adjusting predictionsfor fairness.2. Unintended consequenceson thedistribution of outcomesfor differentgroups.

Discussion of The Limitations And Challenges of These Approaches

Various approaches have been proposed to address bias in AI, but they also face limitations and challenges.

One of the main challenges is the lack of diverse and representative training data. As mentioned earlier, data bias can lead to biased outputs from AI systems. However, collecting diverse and representative data can be challenging, especially when dealing with sensitive or rare events. Additionally, there may be privacy concerns when collecting certain types of data, such as medical records or financial information. These challenges can limit the effectiveness of dataset augmentation as a mitigation approach.

Another challenge is the difficulty of identifying and measuring different types of bias in AI systems. Algorithmic bias can be difficult to detect and quantify, especially when the algorithms are complex or opaque. Additionally, the sources of bias may be difficult to isolate, as bias can arise from multiple sources, such as the data, the algorithm, and the user. This can limit the effectiveness of bias-aware algorithms and user feedback mechanisms as mitigation approaches.

Moreover, mitigation approaches may introduce trade-offs between fairness and accuracy. For example, one approach to reducing algorithmic bias is to modify the algorithm to ensure that it treats all groups equally. However, this may result in reduced accuracy for certain groups or in certain contexts. Achieving both fairness and accuracy can be challenging and requires careful consideration of the trade-offs involved.

Finally, there may be ethical considerations around how to prioritize different types of bias and which groups to prioritize in the mitigation of bias. For example, should more attention be paid to bias that affects historically marginalized groups, or should all types of bias be given equal weight? These ethical considerations can add complexity to the development and implementation of bias mitigation approaches.

Despite these challenges, addressing bias in AI is crucial for creating fair and equitable systems. Ongoing research and development of mitigation approaches are necessary to overcome these challenges and to ensure that AI systems are used for the benefit of all individuals and society.

Typeof Fairness	**Description**	**Examples**
Group Fairness	Ensures that different groups are treated equallyorproportionallyinAIsystems. Can be further subdivided into demographicparity,disparatemistreatment, or equal opportunity.	1. Demographic parity: Positive and negativeoutcomesdistributedequally acrossdemographicgroups(Kamiran& Calders, 2012). 2. Disparate mistreatment: Defined in terms of misclassification rates (Zafar et al., 2017). 3.Equalopportunity:Truepositiverate (sensitivity)andfalsepositiverate(1-specificity)areequalacrossdifferent demographicgroups(Hardtetal.,2016).
Individual Fairness	Ensuresthatsimilarindividualsaretreated similarlybyAIsystems,regardlessoftheir group membership. Can be achieved throughmethodssuchassimilarity-based or distance-basedmeasures.	Using similarity-based or distance-based measurestoensurethatindividualswith similarcharacteristicsorattributesare treatedsimilarlybytheAIsystem (Dworketal., 2012).
Counterfactual Fairness	Aims to ensure that AI systems are fair even in hypothetical scenarios. Specifically, counterfactualfairnessaimstoensurethat an AI system would have made the same decisionforanindividual,regardlessof theirgroupmembership,eveniftheir attributes had been different.	EnsuringthatanAIsystemwouldmake thesamedecisionforanindividual,even if their attributes had been different (Kusner et al., 2017).
Procedural Fairness	Involvesensuringthattheprocessusedto makedecisionsisfairandtransparent.	Implementingatransparentdecision-makingprocessinAI systems.
Causal Fairness	Involvesensuringthatthesystemdoesnot perpetuatehistoricalbiasesandinequalities.	DevelopingAIsystemsthatavoid perpetuating historical biases and inequalities(Kleinbergetal.,2018).

Real-World Examples Of Fairness In Ai

Several real-world instances illustrate the potential benefits of integrating fairness into AI systems. One such example is the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) system, used to forecast recidivism likelihood among criminal defendants. Studies revealed bias against African American defendants, with a higher likelihood of falsely predicting their reoffending compared to white defendants (Angwin et al., 2016). To address this bias, the Northpointe COMPAS was adjusted to include a "race-neutral" algorithm version, maintaining similar accuracy while reducing racial bias (Larson et al., 2016).

Another instance pertains to AI deployment in recruitment processes. Studies found AI recruitment systems biased against women, who were less likely to be chosen for male-dominated roles (Dastin, 2018). To mitigate this bias, some companies implemented "gender decoder" tools analyzing job postings and suggesting changes to reduce gender bias (Crawford, 2019).

In healthcare, AI systems used to forecast healthcare outcomes were found biased against certain groups like African Americans (Obermeyer et al., 2019). To tackle this, researchers proposed employing techniques such as subgroup analysis to identify and address biases in the data used for training AI models (Lamy et al., 2020).

These real-world cases underscore the advantages of embedding fairness into AI systems. By addressing bias and ensuring fairness, AI systems can become more accurate, ethical, and equitable, thus fostering social justice and equality.

Mitigation Strategies For Fairness In Ai

As artificial intelligence (AI) utilization expands, ensuring fairness in decision-making becomes increasingly crucial. AI's application in pivotal domains like healthcare, finance, and law holds significant potential to impact people's lives, necessitating fair and unbiased decisions. To address this challenge, various approaches have emerged, including group fairness and individual fairness. However, these approaches face limitations and challenges, such as trade-offs between different fairness types and defining fairness itself.

Group fairness aims to ensure fair treatment of different demographic groups, such as genders, races, or ethnicities, to prevent systematic discrimination. Techniques like re-sampling, pre-processing, or post-processing data can rectify biased datasets used for AI model training. Individual fairness, conversely, seeks to prevent biased decisions against individuals irrespective of their group membership, achieved through methods like counterfactual fairness or causal fairness.

Despite their promise, these approaches encounter hurdles like trade-offs between fairness types and the difficulty of consensus on fairness definitions. Additionally, current methods may not consider intersectionality, leading to incomplete fairness assessments. Concerns about unintended consequences also loom large, with some mitigation attempts inadvertently worsening disparities.

Addressing these challenges necessitates a multi-disciplinary approach involving experts from various fields. By continually refining mitigation strategies, AI systems can evolve to be unbiased, transparent, and accountable, ensuring equitable outcomes for all.

Approach	Description	Examples	Limitationsand Challenges
Group Fairness	EnsuresthatAIsystemsarefair to different groups of people, such as people of different genders, races, or ethnicities. Aims to prevent the AI system fromsystematicallydiscriminating against any group. Can be achievedthroughtechniquessuch asre-sampling,pre-processing,or post-processing the data.	1. Re-sampling techniquestocreatea balanced dataset. 2. Pre-processing or post-processing to adjust AI model output.	1. May result in unequal treatment of individuals within a group. 2. May not addresssystemicbiases that affect individual characteristics. 3. Groupfairnessmetrics maynotconsider intersectionality.
Individual Fairness	Ensures that AI systems are fair toindividuals,regardlessoftheir group membership. Aims to prevent the AI system from making decisions that are systematicallybiasedagainst certain individuals. Can be achieved through techniques such as counterfactual fairness or causalfairness.	1. Counterfactual fairnessensuringthe same decision regardlessofraceor gender.	1. May not address systemic biases that affectentiregroups.2. Difficultydetermining whichtypesoffairness are appropriate for a givencontextandhow to balance them.
Transparency	InvolvesmakingtheAIsystem's decision-makingprocessvisibleto users.	MakingAI system's decisions and processes understandableto users.	Differentdefinitionsof fairnessamongpeople and groups, and changing definitions overtime.

Accountability	Involvesholdingthesystem's developersresponsibleforany harm caused by the system.	Developersheld responsibleforunfair decisionsmadebyAI systems.	Determining responsibility and addressingpotential harm.
Explainability	InvolvesmakingtheAIsystem's decisions understandable to users.	Providing clear explanationsofAI system'sdecisions.	Addressing the complexity of human behavioranddecision-making.
Intersectionality *(not explicitly mentioned as an approach, but it is an aspecttoconsider)*	Considers the ways in which differentdimensionsofidentity (such as race, gender, and socioeconomicstatus)interact and affect outcomes.	Developing AI systemsthatconsider the interaction of differentdimensions of identity.	Addressing the complexity of intersectionality and ensuringfairnessacross multipledimensionsof identity.

Conclusions

In summary, this paper has shed light on the diverse sources of biases in AI and ML systems and their profound societal repercussions, with a detailed exploration of the emerging concerns surrounding generative AI bias. It is evident that these powerful computational tools, if not meticulously designed and audited, possess the potential to perpetuate and even exacerbate existing biases, particularly those related to race, gender, and other societal constructs. We have examined numerous instances of biased AI systems, with a specific emphasis on the complexities of generative AI, highlighting the critical necessity for comprehensive strategies to detect and mitigate biases across the entire AI development pipeline.

To address bias, this paper has underscored strategies such as robust data augmentation, the application of counterfactual fairness, and the urgent need for diverse, representative datasets alongside unbiased data collection methods. Furthermore, we have considered the ethical implications of AI in safeguarding privacy and stressed the importance of transparency, oversight, and continuous evaluation of AI systems.

Looking ahead, research in fairness and bias in AI and ML should prioritize diversifying training data and tackling the nuanced challenges of bias in generative models, particularly those employed for synthetic data creation and content generation. It is imperative to develop comprehensive frameworks and guidelines for responsible AI and ML, encompassing transparent documentation of training data, model choices, and generative processes. Equally crucial is diversifying the teams engaged in AI development and evaluation, as it brings a multitude of perspectives capable of better identifying and rectifying biases.

Lastly, the establishment of robust ethical and legal frameworks governing AI and ML systems is paramount, ensuring that privacy, transparency, and accountability are foundational elements rather than afterthoughts in the AI development lifecycle. Research must also delve into the implications of generative AI, ensuring that as we progress in creating ever more sophisticated synthetic realities, we remain vigilant and proactive in safeguarding against the subtle encroachment of biases that could shape society in unintended and potentially harmful ways.

References

[1]. Rehan, H. (2024). Revolutionizing America's Cloud Computing the Pivotal Role of AI in Driving Innovation and Security. *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, *2*(1), 189-208. DOI: https://doi.org/10.60087/jaigs.v2i1.p208

[2]. Rehan, H. (2024). AI-Driven Cloud Security: The Future of Safeguarding Sensitive Data in the Digital Age. *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, *1*(1), 47-66. DOI: https://doi.org/10.60087/jaigs.v1i1.p66

[3]. Li, Z., Huang, Y., Zhu, M., Zhang, J., Chang, J., & Liu, H. (2024). Feature Manipulation for DDPM based Change Detection. *arXiv preprint arXiv:2403.15943*.

https://doi.org/10.48550/arXiv.2403.15943

[4]. Ramírez, J. G. C. (2023). Incorporating Information Architecture (ia), Enterprise Engineering (ee) and Artificial Intelligence (ai) to Improve Business Plans for Small Businesses in the United States. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, *2*(1), 115-127. **DOI:** https://doi.org/10.60087/jklst.vol2.n1.p127

[5]. Ramírez, J. G. C. (2024). AI in Healthcare: Revolutionizing Patient Care with Predictive Analytics and Decision Support Systems. *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, *1*(1), 31-37. **DOI:** https://doi.org/10.60087/jaigs.v1i1.p37

[6]. Ramírez, J. G. C. (2024). Natural Language Processing Advancements: Breaking Barriers in Human-Computer Interaction. *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, *3*(1), 31-39.**DOI:** https://doi.org/10.60087/jaigs.v3i1.63

[7]. Ramírez, J. G. C., &mafiqul Islam, M. (2024). Application of Artificial Intelligence in Practical Scenarios. *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, *2*(1), 14-19. **DOI:** https://doi.org/10.60087/jaigs.v2i1.41

[8]. Ramírez, J. G. C., & Islam, M. M. (2024). Utilizing Artificial Intelligence in Real-World Applications. *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, *2*(1), 14-19.

DOI: https://doi.org/10.60087/jaigs.v2i1.p19

[9]. Ramírez, J. G. C., Islam, M. M., & Even, A. I. H. (2024). Machine Learning Applications in Healthcare: Current Trends and Future Prospects. *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, *1*(1). **DOI:** https://doi.org/10.60087/jaigs.v1i1.33

[10]. RAMIREZ, J. G. C. (2023). How Mobile Applications can improve Small Business Development. *Eigenpub Review of Science and Technology*, *7*(1), 291-305. **https://studies.eigenpub.com/index.php/erst/article/view/55**

[11]. RAMIREZ, J. G. C. (2023). From Autonomy to Accountability: Envisioning AI's Legal Personhood. *Applied Research in Artificial Intelligence and Cloud Computing*, *6*(9), 1-16. https://researchberg.com/index.php/araic/article/view/183

[12]. Ramírez, J. G. C., Hassan, M., & Kamal, M. (2022). Applications of Artificial Intelligence Models for Computational Flow Dynamics and Droplet Microfluidics. *Journal of Sustainable Technologies and Infrastructure Planning*, *6*(12).https://publications.dlpress.org/index.php/JSTIP/article/view/70

[13]. Ramírez, J. G. C. (2022). Struggling Small Business in the US. The next challenge to economic recovery. *International Journal of Business Intelligence and Big Data Analytics*, *5*(1), 81-91. https://research.tensorgate.org/index.php/IJBIBDA/article/view/99

[14]. Ramírez, J. G. C. (2021). Vibration Analysis with AI: Physics-Informed Neural Network Approach for Vortex-Induced Vibration. *International Journal of Responsible Artificial Intelligence*, *11*(3).https://neuralslate.com/index.php/Journal-of-Responsible-AI/article/view/77

[15]. Shuford, J. (2024). Interdisciplinary Perspectives: Fusing Artificial Intelligence with Environmental Science for Sustainable Solutions. *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, *1*(1), 1-12.**DOI:** https://doi.org/10.60087/jaigs.v1i1.p12

[16]. Islam, M. M. (2024). Exploring Ethical Dimensions in AI: Navigating Bias and Fairness in the Field. *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, *1*(1), 13-17.**DOI:** https://doi.org/10.60087/jaigs.v1i1.p18

[17]. Khan, M. R. (2024). Advances in Architectures for Deep Learning: A Thorough Examination of Present Trends. *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, *1*(1), 24-30. **DOI:** https://doi.org/10.60087/jaigs.v1i1.p30

[18]. Shuford, J., & Islam, M. M. (2024). Exploring the Latest Trends in Artificial Intelligence Technology: A Comprehensive Review. *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, *2*(1). **DOI:** https://doi.org/10.60087/jaigs.v2i1.p13

[19]. Islam, M. M. (2024). Exploring the Applications of Artificial Intelligence across Various Industries. *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, *2*(1), 20-25.**DOI:** https://doi.org/10.60087/jaigs.v2i1.p25

[20]. Akter, S. (2024). Investigating State-of-the-Art Frontiers in Artificial Intelligence: A Synopsis of Trends and Innovations. *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, *2*(1), 25-30.**DOI:** https://doi.org/10.60087/jaigs.v2i1.p30

[21]. Rana, S. (2024). Exploring the Advancements and Ramifications of Artificial Intelligence. *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, *2*(1), 30-35.**DOI:** https://doi.org/10.60087/jaigs.v2i1.p35

[22]. Sarker, M. (2024). Revolutionizing Healthcare: The Role of Machine Learning in the Health Sector. *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, *2*(1), 35-48.

DOI: https://doi.org/10.60087/jaigs.v2i1.p47

[23]. Akter, S. (2024). Harnessing Technology for Environmental Sustainability: Utilizing AI to Tackle Global Ecological Challenges. *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, *2*(1), 49-57.**DOI:** https://doi.org/10.60087/jaigs.v2i1.p57

[24]. Padmanaban, H. (2024). Revolutionizing Regulatory Reporting through AI/ML: Approaches for Enhanced Compliance and Efficiency. *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, *2*(1), 57-69.**DOI:** https://doi.org/10.60087/jaigs.v2i1.p69

[25]. Padmanaban, H. (2024). Navigating the Role of Reference Data in Financial Data Analysis: Addressing Challenges and Seizing Opportunities. *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, *2*(1), 69-78.**DOI:** https://doi.org/10.60087/jaigs.v2i1.p78

[26]. Camacho, N. G. (2024). Unlocking the Potential of AI/ML in DevSecOps: Effective Strategies and Optimal Practices. *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, *2*(1), 79-89.**DOI:** https://doi.org/10.60087/jaigs.v2i1.p89

[27]. PC, H. P., & Sharma, Y. K. (2024). Developing a Cognitive Learning and Intelligent Data Analysis-Based Framework for Early Disease Detection and Prevention in Younger Adults with Fatigue. *Optimized Predictive Models in Health Care Using Machine Learning*, 273.

[28]. Padmanaban, H. (2024). Quantum Computing and AI in the Cloud. *Journal of Computational Intelligence and Robotics*, *4*(1), 14–32. Retrieved from https://thesciencebrigade.com/jcir/article/view/116

[29]. Sharma, Y. K., & Harish, P. (2018). Critical study of software models used cloud application development. *International Journal of Engineering & Technology, E-ISSN*, 514-518.https://www.researchgate.net/profile/Harish-Padmanaban-2/publication/377572317_Critical_study_of_software_models_used_cloud_application_development/links/65ad55d7ee1e1951fbd79df6/Critical-study-of-software-models-used-cloud-application-development.pdf

[30]. Padmanaban, P. H., & Sharma, Y. K. (2019). Implication of Artificial Intelligence in Software Development Life Cycle: A state of the art review. *vol*, *6*, 93-98.https://www.researchgate.net/profile/Harish-Padmanaban-2/publication/377572222_Implication_of_Artificial_Intelligence_in_Software_Development_Life_Cycle_A_state_of_the_art_review/links/65ad54e5bf5b00662e333553/Implication-of-Artificial-Intelligence-in-Software-Development-Life-Cycle-A-state-of-the-art-review.pdf

[31]. Harish Padmanaban, P. C., & Sharma, Y. K. (2024). Optimizing the Identification and Utilization of Open Parking Spaces Through Advanced Machine Learning. *Advances in Aerial Sensing and Imaging*, 267-294. **https://doi.org/10.1002/9781394175512.ch12**

[32]. PC, H. P., Mohammed, A., & RAHIM, N. A. (2023). *U.S. Patent No. 11,762,755*. Washington, DC: U.S. Patent and Trademark Office.https://patents.google.com/patent/US20230385176A1/en

[33]. Padmanaban, H. (2023). Navigating the intricacies of regulations: Leveraging AI/ML for Accurate Reporting. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, *2*(3), 401-412.DOI: https://doi.org/10.60087/jklst.vol2.n3.p412

[34]. **PC, H. P. Compare and analysis of existing software development lifecycle models to develop a new model using computational intelligence.https://shodhganga.inflibnet.ac.in/handle/10603/487443**

[35]. Camacho, N. G. (2024). Unlocking the Potential of AI/ML in DevSecOps: Effective Strategies and Optimal Practices. *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, *2*(1), 79-89. DOI: https://doi.org/10.60087/jaigs.v2i1.p89

[36]. Camacho, N. G. (2024). The Role of AI in Cybersecurity: Addressing Threats in the Digital Age. *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, *3*(1), 143-154.

DOI: https://doi.org/10.60087/jaigs.v3i1.75

[37]. Islam, M. S., Ahsan, M. S., Rahman, M. K., &AminTanvir, F. (2023). Advancements in Battery Technology for Electric Vehicles: A Comprehensive Analysis of Recent Developments. *Global Mainstream Journal of Innovation, Engineering & Emerging Technology*, *2*(02), 01-28.

https://doi.org/10.62304/jieet.v2i02.63

[38]. Ahsan, M. S., Tanvir, F. A., Rahman, M. K., Ahmed, M., & Islam, M. S. (2023). Integration of Electric Vehicles (EVs) with Electrical Grid and Impact on Smart Charging. *International Journal of Multidisciplinary Sciences and Arts*, *2*(2), 225-234.

https://doi.org/10.47709/ijmdsa.v2i2.3322

[39]. Rahman, M. K., Tanvir, F. A., Islam, M. S., Ahsan, M. S., & Ahmed, M. (2024). Design and Implementation of Low-Cost Electric Vehicles (Evs) Supercharger: A Comprehensive Review. *arXiv preprint arXiv:2402.15728*.

https://doi.org/10.48550/arXiv.2402.15728

[40]. Latif, M. A., Afshan, N., Mushtaq, Z., Khan, N. A., Irfan, M., Nowakowski, G., ... &Telenyk, S. (2023). Enhanced classification of coffee leaf biotic stress by synergizing feature concatenation and dimensionality reduction. *IEEE Access*.
DOI: https://doi.org/10.1109/ACCESS.2023.3314590

[42]. Irfan, M., Mushtaq, Z., Khan, N. A., Mursal, S. N. F., Rahman, S., Magzoub, M. A., ... & Abbas, G. (2023). A Scalo gram-based CNN ensemble method with density-aware smote oversampling for improving bearing fault diagnosis. *IEEE Access*, *11*, 127783-127799.
DOI: https://doi.org/10.1109/ACCESS.2023.3332243

[43]. Irfan, M., Mushtaq, Z., Khan, N. A., Althobiani, F., Mursal, S. N. F., Rahman, S., ... & Khan, I. (2023). Improving Bearing Fault Identification by Using Novel Hybrid Involution-Convolution Feature Extraction with Adversarial Noise Injection in Conditional GANs. *IEEE Access*.

DOI: https://doi.org/10.1109/ACCESS.2023.3326367

[44]. Rahman, S., Mursal, S. N. F., Latif, M. A., Mushtaq, Z., Irfan, M., & Waqar, A. (2023, November). Enhancing Network Intrusion Detection Using Effective Stacking of Ensemble Classifiers With Multi-Pronged Feature Selection Technique. In *2023 2nd International Conference on Emerging Trends in Electrical, Control, and Telecommunication Engineering (ETECTE)* (pp. 1-6). IEEE.

DOI: https://doi.org/10.1109/ETECTE59617.2023.10396717

[45]. Latif, M. A., Mushtaq, Z., Arif, S., Rehman, S., Qureshi, M. F., Samee, N. A., ... & Al-masni, M. A. Improving Thyroid Disorder Diagnosis via Ensemble Stacking and Bidirectional Feature Selection.
https://doi.org/10.32604/cmc.2024.047621

[46]. Ara, A., &Mifa, A. F. (2024). INTEGRATING ARTIFICIAL INTELLIGENCE AND BIG DATA IN MOBILE HEALTH: A SYSTEMATIC REVIEW OF INNOVATIONS AND CHALLENGES IN HEALTHCARE SYSTEMS. *Global Mainstream Journal of Business, Economics, Development & Project Management*, *3*(01), 01-16.

DOI: https://doi.org/10.62304/jbedpm.v3i01.70

[47]. Bappy, M. A., & Ahmed, M. (2023). ASSESSMENT OF DATA COLLECTION TECHNIQUES IN MANUFACTURING AND MECHANICAL ENGINEERING THROUGH MACHINE LEARNING MODELS. *Global Mainstream Journal of Business, Economics, Development & Project Management*, *2*(04), 15-26.
DOI: https://doi.org/10.62304/jbedpm.v2i04.67

[48]. Bappy, M. A. (2024). Exploring the Integration of Informed Machine Learning in Engineering Applications: A Comprehensive Review. *American Journal of Science and Learning for Development*, *3*(2), 11-21.
DOI: https://doi.org/10.51699/ajsld.v3i2.3459

[49]. Uddin, M. N., Bappy, M. A., Rab, M. F., Znidi, F., &Morsy, M. (2024). Recent Progress on Synthesis of 3D Graphene, Properties, and Emerging Applications.
DOI: https://doi.org/10.5772/intechopen.114168

[50]. Hossain, M. I., Bappy, M. A., &Sathi, M. A. (2023). WATER QUALITY MODELLING AND ASSESSMENT OF THE BURIGANGA RIVER USING QUAL2K. *Global Mainstream Journal of Innovation, Engineering & Emerging Technology*, *2*(03), 01-11.
DOI: https://doi.org/10.62304/jieet.v2i03.64

[51]. Zhu, M., Zhang, Y., Gong, Y., Xing, K., Yan, X., & Song, J. (2024). Ensemble Methodology: Innovations in Credit Default Prediction Using LightGBM, XGBoost, and LocalEnsemble. *arXiv preprint arXiv:2402.17979*.

https://doi.org/10.48550/arXiv.2402.17979

[52]. Yafei, X., Wu, Y., Song, J., Gong, Y., &Lianga, P. (2024). Generative AI in Industrial Revolution: A Comprehensive Research on Transformations, Challenges, and Future Directions. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, *3*(2), 11-20.

DOI: https://doi.org/10.60087/jklst.vol.3n2.p20

[53]. Xu, J., Wang, H., Zhong, Y., Qin, L., & Cheng, Q. (2024). Predict and Optimize Financial Services Risk Using AI-driven Technology. *Academic Journal of Science and Technology*, *10*(1), 299-304.

https://drpress.org/ojs/index.php/ajst/article/view/19205

[54]. Ness, S., Sarker, M., Volkivskyi, M., & Singh, N. (2024). The Legal and Political Implications of AI Bias: An International Comparative Study. *American Journal of Computing and Engineering*, *7*(1), 37-45.
DOI: https://doi.org/10.47672/ajce.1879

[55]. Sarker, M. (2022). Towards Precision Medicine for Cancer Patient Stratification by Classifying Cancer By Using Machine Learning. *Journal of Science & Technology*, *3*(3), 1-30.
DOI: https://doi.org/10.55662/JST.2022.3301

[56]. Manoharan, A., &Sarker, M. REVOLUTIONIZING CYBERSECURITY: UNLEASHING THE POWER OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING FOR NEXT-GENERATION THREAT DETECTION.

DOI :https://www.doi.org/10.56726/IRJMETS32644

[57]. Lee, S., Weerakoon, M., Choi, J., Zhang, M., Wang, D., & Jeon, M. (2022, July). CarM: Hierarchical episodic memory for continual learning. In *Proceedings of the 59th ACM/IEEE Design Automation Conference* (pp. 1147-1152).

https://doi.org/10.1145/3489517.3530587

[58]. Lee, S., Weerakoon, M., Choi, J., Zhang, M., Wang, D., & Jeon, M. (2021). Carousel Memory: Rethinking the Design of Episodic Memory for Continual Learning. *arXiv preprint arXiv:2110.07276*.

https://doi.org/10.48550/arXiv.2110.07276

[59]. Weerakoon, M., Heaton, H., Lee, S., & Mitchell, E. (2024). TopoQual polishes circular consensus sequencing data and accurately predicts quality scores. *bioRxiv*, 2024-02.

doi: https://doi.org/10.1101/2024.02.08.579541