DUAL EDGES
AI & DEI

Virginia Inclusion Summit
September 5, 2024
Richmond, Virginia

Visit Our Website
piercecreativeconsulting.com

Ida Pierce – Strategist | Innovative Problem Solver | Consultant

Here's what I value: Inclusion, Creativity, Growth

This is what I do: I partner with mission driven organizations and businesses to identify AND deploy the right tools, resources and supports necessary to achieve results.

My background includes:

- Project Management
- Change Management
- Operations Leadership
- Experience Design
- Professional Development
- Robotic Process Automation
- Strategy & Innovation
- Facilitation
- Data & Analytics
- AI for Business Transformation

Pierce Creative
& CONSULTING LLC

POWER HOUR:

- Understanding AI and it’s affects on DEI
- Learn strategies to identify and mitigate AI biases
- Explore real-world examples and actionable practices.

Engage openly
Share insights
Ask questions

What is Artificial Intelligence (AI)?

AI is about creating systems that can perform tasks normally requiring human intelligence. This includes learning, reasoning, and self-correction.

Components of AI

Algorithms

Step-by-step instructions that guides the computer through the steps necessary to make a decision or calculation

Neural Networks

Neural networks are inspired by the human brain and designed to recognize patterns

Natural Language Processing

Applies algorithms to identify and extract the rules of natural language enabling computers to understand and process human languages

Machine Learning

Teaching computers to learn from and interpret data without being explicitly programmed for every task

AI learns from data by using **algorithms** that identify **patterns and relationships**, adjust predictions, and optimize performance based on **feedback and continuous input**.

The more data and feedback it receives, the better it can **learn and perform tasks** accurately.

Important Distinction

Generative AI

Text Generation (GPT's)

Summarization and/or translation

Images, style transfer, videos

Music, voice cloning, text-to-speech

Automated code writing, bug fixing

Art and design

Chatbots, virtual assistants, learning systems

Other Types of AI

Weather prediction models

Robotic control

Self-driving cars

Forecasting (stock market, housing)

Recommendation systems (Netflix, Amazon)

Detection (network security, unusual transactions)

“AI systems are only as effective, accurate and inclusive as the **data** that they are trained on - if the data provided to a model for training purposes is incomplete, biased or unrepresentative, then the system will be as well”.

-Kasia Chmielinski

2023 DCSL/CCSRE Technology & Racial Equity Practitioner Fellow

AI & DEI Intersection

More accessible digital content and services

Reduce bias with anonymized data, focusing on skills and objective criteria

Enable more informed DEI strategies by analyzing large datasets to uncover disparities and trends in demographics, pay equity and promotion rates

Potential for unfair outcomes due to perpetuating or amplify biases present in training data

Exclusion of diverse populations when data used to train AI models is not representative of all groups

Limited awareness and education within teams might perpetuate biases, overlook the nuances of equitable AI, or fail to meet regulatory and ethical standards.

Reflect & Share

What opportunities can you identify for using AI to enhance DEI efforts in your organization or work?

What risks might you need to consider or address?

Partner and discuss.

CASE STUDIES

Review case; name potential AI Biases

Evaluate fairness and ethical implications

Discuss potential mitigation strategies

Importance of Equitable AI Development

BENEFITS

- Reduce Bias and Discrimination
- Enhance Trust and Adoption
- Promote Innovation and Inclusivity

RISKS

- Amplify Existing Inequities
- Legal and Ethical Consequences
- Undermines AI's Potential

Share one key takeaway from today’s session.

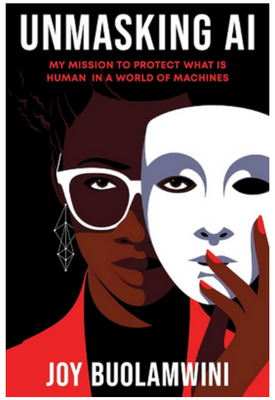
COMMUNITY

AI Ready RVA

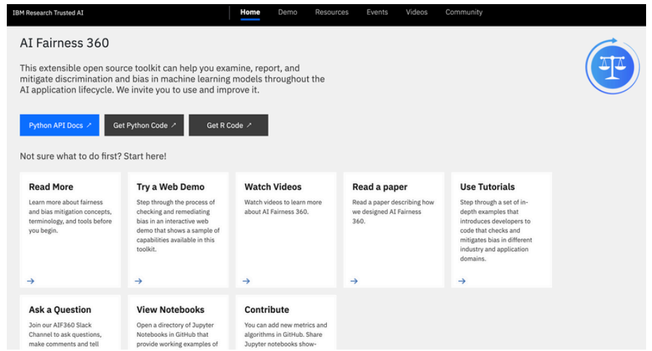
Timnit Gebru, DAIR
Distributed AI Research

RESEARCH AND EXPLORE

Bipartisan Roadmap for AI Innovation



Joy Buolamwini, AI Bias Expert
The Algorithmic Justice League



IBM's AI Fairness 360

THANK YOU

APPENDIX

Customer Service

A large retail chain implements AI-powered chatbots and virtual assistants to handle customer inquiries and complaints 24/7. However, customers who speak dialects or use non-standard grammar find the chatbot responses inaccurate and unhelpful, leading to frustration and negative feedback. In contrast, customers with standard English receive efficient service. This creates a divide in customer experience based on language use.

Identify the potential AI system biases:

How does this scenario impact diversity, equity, and inclusion (DEI) in customer service?

What strategies could the company use to make AI-powered tools more inclusive for all customers?

Marketing

A major e-commerce company uses AI predictive analytics to create targeted marketing campaigns. The algorithm favors promotions to high-income neighborhoods based on past purchasing patterns. As a result, marketing efforts disproportionately reach affluent customers, neglecting lower-income communities who may also benefit from the promotions.

Identify the potential AI system biases:

What are the DEI implications of this AI strategy?

How can the company adjust its use of AI to promote more equitable outreach?

Human Resources

A government agency uses AI for talent acquisition, automatically screening resumes for certain keywords. However, the AI system disproportionately rejects candidates from underrepresented groups who may use different terminology or lack certain credentials but have equivalent experience.

Identify the potential AI system biases:

How could this affect diversity in hiring?

What measures can be taken to ensure the AI system promotes fair and equitable recruitment?

Operations and Supply Chain

A nonprofit organization relies on AI for inventory management of donated goods, using historical data to predict demand. The AI algorithm suggests reducing stock of items that are in less frequent demand in wealthier neighborhoods but needed in lower-income areas, potentially creating disparities in resource allocation.

Identify the potential AI system biases:

How does this impact the organization's commitment to equity?

What could be done to ensure the AI-driven decisions reflect the nonprofit’s mission of serving all communities equally?

Finance

A financial institution uses AI for fraud detection, flagging transactions as suspicious based on patterns. However, the AI disproportionately flags transactions from certain zip codes with predominantly minority populations, leading to higher rates of account freezes in those areas.

Identify the potential AI system biases:

What are the DEI concerns associated with this AI practice?

How can the financial institution adjust its AI model to avoid biased outcomes?

Product Development

A tech company uses AI to analyze customer feedback for product development. The feedback predominantly comes from a demographic that is not representative of the company's entire customer base, skewing product enhancements toward a narrow user group.

Identify the potential AI system biases:

What are the DEI implications of this feedback analysis approach?

How can the company ensure that its product development process is inclusive of all customer voices?

Business Intelligence

A multinational corporation uses AI for predictive business analytics, but the data used is largely historical data from countries with established markets, excluding emerging markets. This limits the company's strategic planning to well-established regions and neglects potential growth areas in underserved markets.

Identify the potential AI system biases:

How might this bias affect the company’s global strategy and inclusion efforts?

What changes could be made to include diverse data sources?

Cyber Security

A university deploys AI for threat detection and response, but the system is biased towards detecting threats from specific regions or countries, based on historical data. This leads to increased monitoring of international students from those regions, creating a sense of distrust and exclusion.

Identify the potential AI system biases:

What are the potential DEI impacts of this AI-based security approach?

How can the university modify its AI system to ensure fair treatment of all students and staff?

ida@piercecreativeconsulting.com

Types of Bias in AI Systems

TYPE OF BIAS	DESCRIPTION	EXAMPLE
Historical Bias	AI models are trained on past data that reflect previous discriminatory practices or patterns, causing the model to perpetuate those biases.	An AI model predicting job performance might be biased if trained on historical data from a company that has a history of underrepresenting certain groups in high-performing roles.
Sampling Bias	Data used to train the AI model does not adequately represent the entire population, leading to skewed predictions.	An AI tool designed for diagnosing medical conditions may perform poorly if it is trained primarily on data from a specific ethnic group and fails to account for genetic differences in others.
Measurement Bias	The variables or measurements used in the model are biased or inaccurately represent the phenomenon being studied.	An AI system that uses social media activity as a measure of customer engagement may misinterpret the engagement levels of individuals who do not use social media frequently.
Algorithmic Bias	Involves biases that are unintentionally embedded within the AI algorithm itself due to the design, data processing, or optimization methods. Amplifies or introduces bias.	An AI system that recommends promotions based on historical sales data may favor certain product lines over others, not because they are inherently better, but because of the algorithm's weighting of certain features.
Confirmation Bias	Arises when the AI model is trained in a way that reinforces pre-existing beliefs or assumptions, often by prioritizing data that supports these beliefs.	If a predictive policing tool is fed data from neighborhoods with a high police presence, it may reinforce the assumption that these areas are more crime-prone, regardless of actual crime rates.
Feature Selection Bias	Occurs when the features (variables) selected for use in a model are themselves biased or exclude important variables that affect outcomes.	A credit scoring AI that excludes income stability as a feature may disproportionately disadvantage applicants with irregular income, such as freelancers or gig workers.
Label Bias	Happens when the outcomes (labels) used to train a model are biased, reflecting biased human judgment or systemic issues.	A hiring AI trained on data where past successful hires were predominantly from a certain university may learn to prefer candidates from that university.
Group Attribution Bias	Arises when an AI model makes generalizations based on group data, attributing characteristics or behaviors of the group to all individual members.	A risk assessment tool for loans that classifies an entire demographic group as high-risk based on aggregated data, ignoring individual creditworthiness.

Types of Bias in AI Systems - page 2

TYPE OF BIAS	DESCRIPTION	EXAMPLE
Overfitting Bias	Happens when an AI model is excessively trained on a specific dataset, causing it to perform well on that data but poorly on new, unseen data.	An AI model predicting employee attrition based on one company's historical data may not generalize well to another company with a different culture and practices.
Automation Bias	Occurs when decision-makers overly rely on AI-generated recommendations, assuming the AI is always correct, leading to errors or biased outcomes.	In healthcare, doctors may over-rely on AI diagnostic tools and overlook symptoms that the AI does not flag as significant.
Survivorship Bias	Arises when only data from successful cases is used, ignoring data from failures, which can lead to misleading conclusions.	A study of business success that only includes data from companies that survived past five years, ignoring those that failed.
Observer Bias	Happens when the training data reflects the subjective opinions or prejudices of the human labelers.	An AI model for content moderation on social media might be biased if the data reflects moderators' subjective interpretations of offensive content.
Exclusion Bias	Takes place when certain groups or types of data are excluded from the dataset, leading to biased outcomes.	An AI model predicting customer satisfaction might be biased if it only includes feedback from customers who completed a survey online, excluding those who could not access the survey.
Proxy Bias	Arises when a seemingly neutral variable is used as a proxy for sensitive or protected characteristics, leading to biased outcomes.	Using ZIP codes as a feature for loan approvals may indirectly result in racial discrimination if certain ZIP codes correlate strongly with race or ethnicity.
Interaction Bias	Occurs when biases are introduced through the way users interact with an AI system, often influenced by their expectations or the design of the interface.	In AI-driven customer service chatbots, biased interactions can occur if the bot has learned from previous biased conversations, reinforcing stereotypes or misunderstanding diverse dialects.

AI Mitigation Strategies

STRATEGY	DESCRIPTION
Inclusive AI Governance Framework	Establish governance frameworks, including diverse perspectives and ethical guidelines for AI dev and deployment.
Diverse and Representative Data Collection	Ensure that the training data is diverse and representative of the entire population the AI is intended to serve.
Bias Auditing and Regular Testing	Conduct regular bias audits and tests on AI models to identify and correct biases.
Human-in-the-Loop (HITL) Review	Involve human experts in the decision-making process to verify and override AI-generated outputs when necessary.
Algorithmic Transparency, Documentation and Explainability	Make AI models more transparent and explainable so that their decision-making processes can be understood, challenged, and improved. Document the AI development process, including data sources, model assumptions, and potential biases, to increase accountability and trust.
Regularly Update and Retrain Models (including sensitivity analysis and feature engineering)	Regularly update and retrain AI models with new data to ensure they reflect current realities and reduce the impact of outdated biases.
Inclusive Design Practices	Design AI systems with input from diverse stakeholders to ensure that the AI serves all users fairly and inclusively.
Education and Bias Awareness Training	Educate all on principles, practices and foundations of AI with your organization. Train AI developers, data scientists, and users on bias awareness and mitigation strategies.
Scenario and Stress Testing	Use scenario analysis and stress testing to evaluate how AI models perform under various conditions and with different types of data.
Accountability Measures	Implement mechanisms to allow affected individuals or groups to report biases and seek redress. Set internal expectations.