PUBLIC SUBMISSION

Received: May 22, 2025
Tracking No. maz-9n72-8kbx
Comments Due: May 28, 2025 **Submission Type:** API

Docket: NSF-2025-OGC-0001
NITRD_FRDOC_0001

Comment On: NSF-2025-OGC-0001-0001
Request for Information: Development of a 2025 National Artificial Intelligence Research and Development Strategic Plan

Document: NSF-2025-OGC-0001-DRAFT-0094
Comment on FR Doc # 2025-07332

Submitter Information

Organization: Epistamai

General Comment

See attached file(s)

Attachments

NTIRD RFI

Response to RFI on the Development of a 2025 National Artificial Intelligence (AI) Research and Development (R&D) Strategic Plan, Docket ID No. NSF-2025-OGC-0001

Chris Lam, Epistamai

Earlier this year, President Trump signed Executive Order 14179. It states that maintaining American leadership in AI innovation requires developing AI systems that are free from ideological bias. A few months later, over 200 academic researchers signed a statement on aibiasconsensus.org that argued that there was already a scientific consensus on AI bias. The paper that they cited made no mention of ideological bias, yet claimed to be bipartisan. In the meanwhile, this academic community has failed to produce the guidance the AI industry needs to develop AI systems that are in compliance with US antidiscrimination law (e.g. disparate treatment). Indeed, many of the universities that these AI researchers represent are currently being investigated by the Trump administration for illegal discrimination.

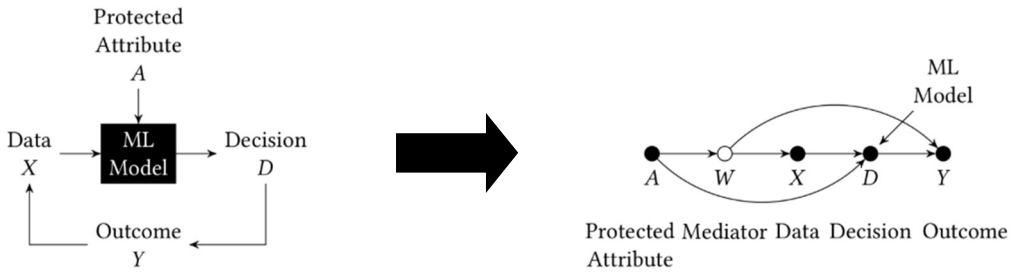
As an AI researcher who has submitted papers on ideological bias to this academic community in the past, I know from firsthand experience that this community has been rejecting papers that discuss the ideological differences between how Republicans and Democrats approach the AI bias problem. When it comes to racial discrimination, we all know that Republicans prefer racially blind systems that evaluate people solely on their merits and treats everyone equally. On the other hand, Democrats prefer affirmative action and DEI programs that give preference to certain minority groups in order to help level the playing field. And yet this academic community has canceled these types of ideological debates in order to advance a radical left-wing agenda. They have systematically excluded conservative viewpoints to prevent them from balancing out liberal viewpoints. As a result, the emerging and important field of AI ethics has been overrun by radical left-wing AI activists who have been trying to mislead the public into falsely believing that DEI is diverse, inclusive, or even bipartisan.

My startup has been doing research on novel approaches to prevent AI bias by modeling how AI systems interact with society. By going back to first principles, performing root cause analysis, and applying systems thinking, we can develop a truly nonpartisan approach to the AI bias problem that can actually help AI developers comply with the law. We discovered that the AI bias problem is actually ***symmetrical*** (there's a left and a right side), and that we needed to model a diversity of viewpoints in order to get a deep understanding of ideological bias. This is what makes our approach to AI bias more powerful, robust, and flexible than the existing scientific consensus. A symmetrical and balanced framework for AI bias can serve whichever policymakers are in power, whether they are Republicans or Democrats.

When we think about most AI systems today, we generally think of machine learning (ML) models like neural networks. These are black boxes that contain many hidden "inner" layers.

But what if these neural networks also have transparent "outer" layers? That is, we can use these outer layers to model how AI systems interact with society. This is achieved by encoding knowledge, assumptions, and beliefs about cause-and-effect relationships between different variables. In particular, these outer layers are very useful for modeling ideological biases and explaining how policies like DEI could actually backfire and perpetuate racial disparities over time. We can use these causal models to cut through the ideological dogma so that we can explain reality, such as why some minorities (e.g. Asians) were able to close the racial wealth gap with Whites while other minorities (e.g. Blacks) continue to struggle. Once we can understand the truth, then we can prevent AI systems from causing illegal discrimination.

The diagram below shows how to convert supervised ML into causal inference. On the left is a block diagram showing a black box ML model. Data X is being fed into the model, which renders a decision D. This in turn leads to some outcome Y, which gets fed back in as future data X. The key question is what to do with a protected attribute A (e.g. race), which if improperly used could trigger illegal discrimination like disparate treatment. On the right is a causal Bayesian network (CBN) that shows the causal relationship between each of these variables. We can place the ML model inside the node for decision D. Using this CBN, we can model illegal discrimination and understand what interventions are needed to ensure that an ML model is compliant with antidiscrimination law.



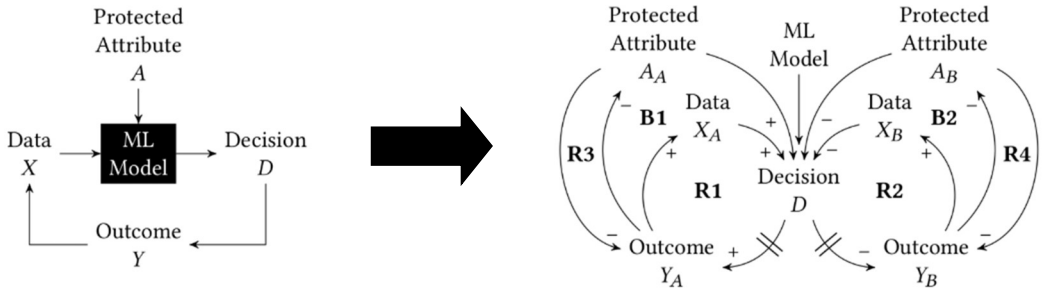
Supervised Machine Learning
Block Diagram

Causal Inference
Causal Bayesian Network

This provides a general framework for modeling fairness in multiple high-stakes applications:

Application	**Protected Attribute A**	**Mediator W**	**Data X**	**Decision D**	**Outcome Y**
Credit scoring	Race, gender	Creditworthiness	Income, credit history	Deny loan?	Loan default?
Resume screening	Race, gender	Qualifications	Experience, education	Screen out resume?	Employee turnover?
College admissions	Race, gender	Merit	Grades, test scores	Reject applicant?	Student failure?

A key limitation of CBNs is that they are directed acyclic graphs and thus cannot be used to model feedback loops. But we can instead convert supervised ML into system dynamics, as shown in the following diagram. On the left is the same block diagram as before. On the right is a causal loop diagram that shows how these same variables may perpetuate disparities between two protected groups (A and B) over time. It shows how the use of a protected attribute A in a decision D (e.g. affirmative action) could have unintended consequences on an outcome Y.



Supervised Machine Learning
Block Diagram

System Dynamics
Causal Loop Diagram

This is because over the long term, minority groups may become dependent on affirmative action and DEI programs. They could develop a sense of victimhood and learned helplessness that impairs their ability to compete on their own merit over the long term. Instead, they might develop an unhealthy addiction to big government interventions, which could reduce their ability to flourish. These are counterintuitive AI patterns that the DEI community does not want the Trump administration or the general public to know about.

Modeling AI bias as a complex systems problem is a novel area of fundamental AI research. This could lead to a paradigm shift in AI from a purely machine learning-based approach to a hybrid approach that combines symbolic AI and machine learning. By bridging together supervised machine learning, causal inference, and system dynamics, we can help the Trump administration remove ideological bias in AI systems.

Our approach reduces regulatory risk and uncertainty, provides strong mathematical guardrails to protect all consumers, and enables the development of next generation high-stakes AI systems while maintaining compliance with the law. Whereas the European Union's approach to AI bias involves bureaucracy and red tape, our approach involves causality and math. This provides a uniquely American approach to AI bias that is deeply rooted to American values.

We are working with IEEE to codify these causal models into the world's first AI standard to address ideological bias: IEEE P3591 (Standard for Fair Decision Making Through Causal Analysis). We are also working with Accredited Standards Committee X9 to develop ISO/ANSI

standards for using AI in highly regulated fields like fair lending. These voluntary consensus standards can play a key role in maintaining US leadership in AI. Our hope is that the federal government will continue to support the development of these new AI standards, as recommended by OMB Memorandum M-21-06.

According to an analysis led by U.S. Senate Commerce Committee Chairman Ted Cruz, over 3,400 grants awarded by the NSF totaling over $2 billion were made to support questionable DEI programs or neo-Marxist agendas. Reversing the damage that this research has caused will require significant federal funding in fundamental AI research and development. Taking on the DEI establishment to build a new scientific consensus on AI bias will not be easy. Even though our systems-based approach is technically superior, we know that these radical left-wing activists in academia will not be swayed through superior math, logic, or the truth. They will try to thwart AI innovation as long as it is not aligned to their ideological agenda, even if it means breaking the law. We need help from the federal government to take on these AI activists.

Americans deserve a new scientific consensus on AI bias that models the politics of the real world instead of the politics of the ivory tower. This new consensus must reflect the will of the American people, instead of the ideological viewpoints of a small number of academic elites who are out of touch with reality. We hope to work with the Trump administration to take on ideological bias in AI systems so that the US remains at the forefront of AI innovation.

This document is approved for public dissemination. The document contains no business-proprietary or confidential information. Document contents may be reused by the government in developing the 2025 National AI R&D Strategic Plan and associated documents without attribution.