

What GCs Need to Know About Algorithmic Bias

Legal Dive

October 4, 2023 | Publications

Bradley Merrill Thompson, Strategic Advisor with EBG Advisors and Member of the Firm at Epstein Becker Green, co-authored an article in *Legal Dive*, titled “What GCs Need to Know About Algorithmic Bias.”

Following is an excerpt:

General counsel are aware of the long and growing list of stories: An employment screening tool that doesn’t account for accents. Facial recognition software that struggles with darker skin tones. A selection app that shows a preference for certain backgrounds, education, or experience.

As local, state and federal agencies race to implement regulations to address these issues, general counsel face potentially costly legal liabilities and significant reputational harm if their organization attempts to use new AI-powered technologies in their hiring.

Can’t data scientists just “fix it”? Isn’t there a simple change to the math involved that could address the error?

Unfortunately, the answer is no, in many cases. There are two intertwined layers to the complexity that make remediation a matter of judgment requiring expert knowledge in data science and law. Those layers are (1) the technical challenges around finding and correcting bias, and (2) the legal complexities around defining what bias is and when it is acceptable.

Technical obstacles

Bias can creep into algorithmic decision-making in many ways, including by training an algorithm on data that contains encoded prior human bias, failing to ensure that

People



Bradley Merrill Thompson
Strategic Advisor
Artificial Intelligence
202-861-1817
bthompson@ebglaw.com

Focus Areas

Services

Artificial Intelligence

the training data includes adequate representation of smaller groups, assumptions or mistakes made in the design and coding of the algorithm and measurement bias where the data collected for training differ from that collected in the real world, to name just a few.

Can we overcome those problems simply by making sure that the data analyzed by the algorithm excludes data on sensitive attributes like age, sex and race? Unfortunately, sensitive information can be inferred from other information—information that has value that we do not want to lose.

You would think at least finding discrimination in an algorithm would be easy, but no. A key part in testing algorithms is having a metric to determine whether unlawful discrimination exists. But we don't even have a uniformly agreed upon measure of what constitutes bias in, for example, qualitative outputs from large language models.

When bias is found, often it cannot simply be eradicated. Unless it is based on finding new, suitable data to supplement the existing training set, improving equality in the overall output often means reducing overall accuracy.