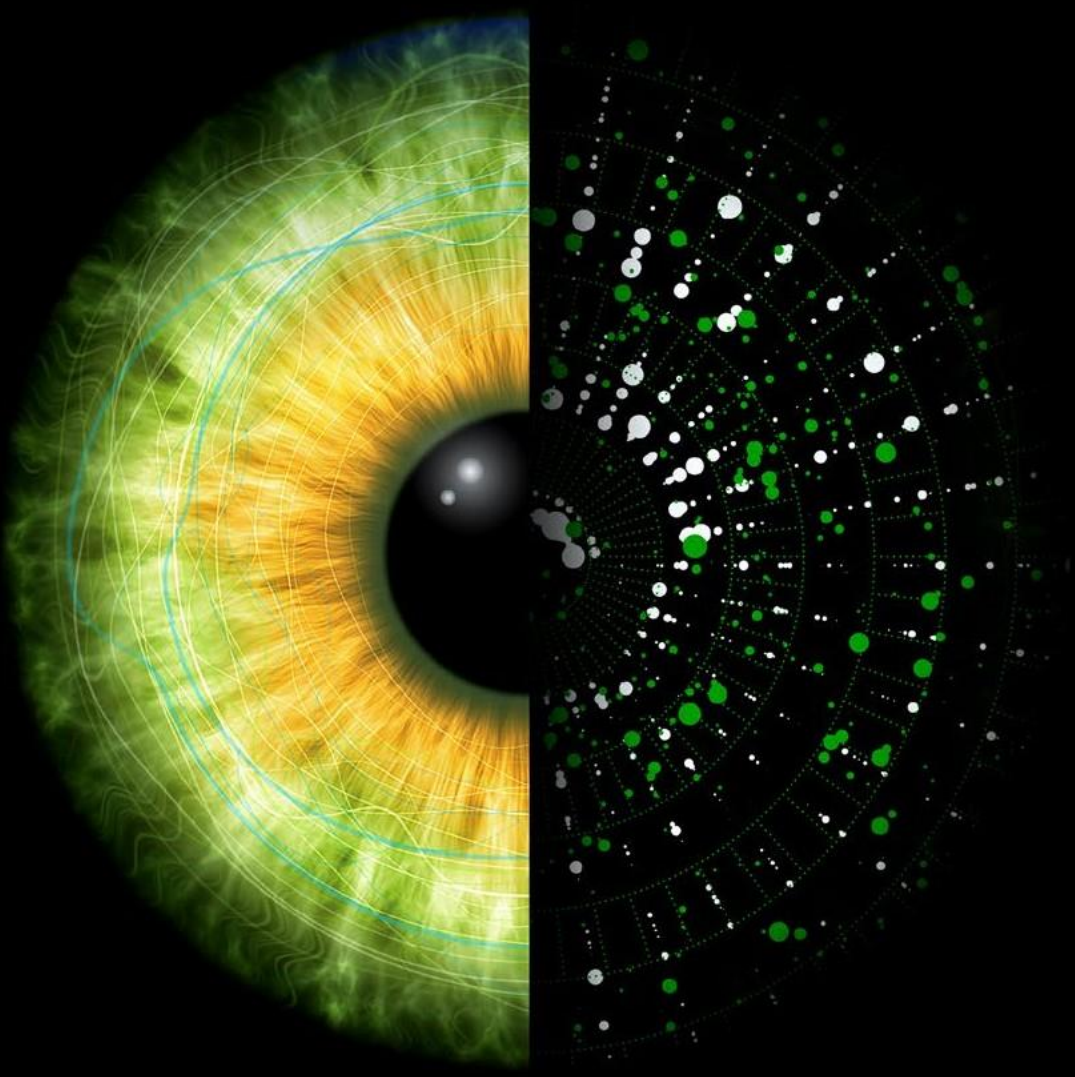
Infomerica

The Right Mind Set



Making ML models Trust worthy

About us

Infomerica is a Software Development and Systems Integration company, one of the fastest growing companies that is committed to help our customers grow and move forward in every aspect of their business.

Our Mission

Infomerica’ s mission is to use our extensive IT & Industry experience to deliver tangible business results enabling clients in the industry to profit from the proven advanced use of technology.

We are leading global provider of Business process related services, Enterprise Portals, IT related services.

Our vision is unchanging:

We aspire to be the Standard of Excellence, the first choice of the most sought-after clients and talent. Our shared values are timeless. They succinctly describe the core principles that distinguish the culture.

Integrity

We believe that nothing is more important than our reputation, and behaving with the highest levels of integrity is fundamental to who we are. We demonstrate a strong commitment to sustainable, responsible business practices.

Outstanding value to markets & clients

We play a critical role in helping both the capital markets and our member firm clients operate more effectively. We consider this role a privilege, and we know it requires constant vigilance and unrelenting commitment.

Commitment to each other

We believe that our culture of borderless collegiality is a competitive advantage for us, and we go to great lengths to nurture it and preserve it. We go to extraordinary lengths to support our people.

Strength from cultural diversity

Our member firm clients’ business challenges are complex and benefit from multidimensional thinking. We believe that working with people of different backgrounds, cultures, and thinking styles helps our people grow into better professionals and leaders.

Executive Summary

A LARGE REGIONAL BANK uses a newly developed fraud detection artificial intelligence (AI) algorithm to identify potential cases of bank fraud including anomalous patterns of financial transactions, loan applications, and new account applications. The algorithm is trained on an initial set of data to give an idea of what normal versus fraudulent transactions look like. However, the training data becomes biased by oversampling applicants over 45 years of age for examples of fraudulent behavior. This oversampling continues over a period of months, with the bias growing and remaining undetected. The model becomes more likely to think an older person is committing fraud than reality suggests. Customers are increasingly turned down for loans. Some begin to feel alienated while regulators start to ask questions. Trust is lost, the brand's reputation suffers, and the bank faces significant consequences to its bottom line.

We know model bias is potentially a problem, but do we really know how pervasive it is? Certainly, media outlets write stories that capture the public imagination, such as the AI hiring model that is unfairly biased against women[1] or the AI health insurance risk algorithm that unfairly assigns higher risk scores based on racial identity.[2] But as bad as such examples may be, the AI model bias story hardly ends with what we read in the popular press.

Our research indicates that model bias could be more prevalent than many organizations are aware and that it can do much more damage than we may assume, eroding the trust of employees, customers, and the public. The costs can be high: expensive tech fixes, lower revenue and productivity, lost reputation, and staff shortages, to say nothing of lost investments. In fact, 68% of executives surveyed in recent *State of AI in the Enterprise, 4th Edition* report reported that their functional group invested US$10 million or more in AI projects in the past fiscal year alone.[3] Even internal-facing models can do significant harm and potentially put those millions of dollars of investment at risk.

To solve this problem, we need to go beyond empathy and good intentions. Understanding, anticipating, and, as much as possible, avoiding the occurrence of model bias can be critical to advance the use of AI models across the organization in a way that preserves stakeholder trust. The good news is that there are approaches that organizations can adopt—including technology-based solutions—that can help.

ORGANIZING THE “WILD WEST” OF MODEL BIAS

Several classes or archetypes of model bias emerged during our research. We identify two main groups of biases based on the type of action that impacts the model: “Passive” bias—where bias is not the result of a planned act—and “active” bias—where the bias occurs because of human action, either with or without intent and, even when intentional, often without *negative* intent. Both types of bias can manifest in different ways, and both should be considered when developing strategies to mitigate model bias risk. In characterizing bias in the classification that follows, we use our own terms as well as terms that are commonly observed in social science and technology literature.[12]

Passive bias

Examples of passive bias may include:

- *Selection bias*: Over inclusiveness or under inclusiveness of a group; insufficient data; poor labeling.An example of selection bias may be found in an AI model trained on data in which a particular group is identified with a certain characteristic at a higher rate than objective reality justifies.
- *Circumstantial bias*: Training data staleness; changing circumstances. An example of circumstantial bias may include a predictive AI model trained on data that was accurate originally but is no longer accurate because of changing realities or “facts on the ground.”
- *Legacy or associational bias*: AI models trained on terms or factors associated with legacies of bias based on race, gender, and other grounds, even though unintentionally. One example is found in a hiring algorithm trained on data that, while not overtly gender-biased, refers to terms that carry a legacy of male association.

Active bias

Examples of active bias may include:

- *Adversarial bias*: Data poisoning; post-deployment adversarial bias. A hostile actor, for example, gains access to a model’s training data and introduces a bias for nefarious objectives.
- *Judgment bias*: Model is trained properly, but bias is introduced by a model user during implementation by way of misapplication of AI decision output. For example, a model may produce objectively correct results, but the end user misapplies those results in a systemic fashion. In that sense, judgment bias differs from other model biases in that it is not the direct result of flawed training data.[13]

The above grouping is far from exhaustive or definitive; other bias characterizations exist. Such speaks to the evolving and still nascent understanding of what model bias is and how it occurs.

“By progressing new ethical frameworks for AI and thinking critically about the quality of our datasets and how humans perceive and work with AI, we can accelerate the [AI] field in a way that will benefit everyone. IBM believes that [AI] actually holds the keys to mitigating bias out of AI systems – and offers an unprecedented opportunity to shed light on the existing biases we hold as humans.

AI must be designed to minimize bias and promote inclusive representation.

– Bias in AI: How we Build Fair AI Systems and Less-Biased Humans[18]

AI provides deeper insight into our personal lives when interacting with our sensitive data. As humans are inherently vulnerable to biases, and are responsible for building AI, there are chances for human bias to be embedded in the systems we create. It is the role of a responsible team to minimize algorithmic bias through ongoing research and data collection which is representative of a diverse population.

12 use cases.

Companies are misaligned with policymakers at this critical moment in contrast, a much weaker consensus exists among companies. Their prioritization of ethical principles is relatively undifferentiated, with evenly distributed responses across use cases and principles. Their top choices are also preferred by narrower margins than those of policymakers. More importantly, companies are focused on the principles prioritized by existing regulations such as GDPR (e.g., Privacy and cybersecurity) rather than on emerging issues that will become critical in the age of AI (e.g., explain ability, fairness and non-discrimination). For instance, companies place “privacy and data rights” among the top three ethical principles in 10 out of 12 use cases and rank it the top concern in seven of those use cases. Policymakers also rate this ethical principle highly, but to a lesser degree — ranking it the most important use case just twice, and instead placing it second or third in nine use cases

Companies rate “**safety and security”** more highly than do policymakers

	Rank assigned by **policymakers:**			Rank assigned by **companies:**		
	Most important	Second-most important	Third-most important	Most important	Second-most important	Third-most important
Algorithmic financial planners			✓✓			
Algorithmic health care providers	✓✓			✓		
Algorithmic recruiting						✓
Behavioral modification					✓	
Facial recognition check-ins					✓	
Fully autonomous vehicles	✓✓			✓✓		
Home virtual voice assistants		✓✓			✓	
Human emotion analysis				✓		
Information curation and distribution						
Law enforcement surveillance					✓	
Personalized algorithmic pricing					✓	
Social credit and underwriting				✓		

Companies rate "**privacy and data rights**" the most important issue

	Rank assigned by **policymakers:**			Rank assigned by **companies:**		
	Most important	Second-most important	Third-most important	Most important	Second-most important	Third-most important
Algorithmic financial planners		✓		✓		
Algorithmic health care providers		✓				✓
Algorithmic recruiting			✓	✓		
Behavioral modification		✓		✓		
Facial recognition check-ins		✓		✓		
Fully autonomous vehicles						
Home virtual voice assistants	✓			✓		
Human emotion analysis	✓				✓	
Information curation and distribution		✓		✓		
Law enforcement surveillance		✓		✓		
Personalized algorithmic pricing		✓				
Social credit and underwriting		✓				✓

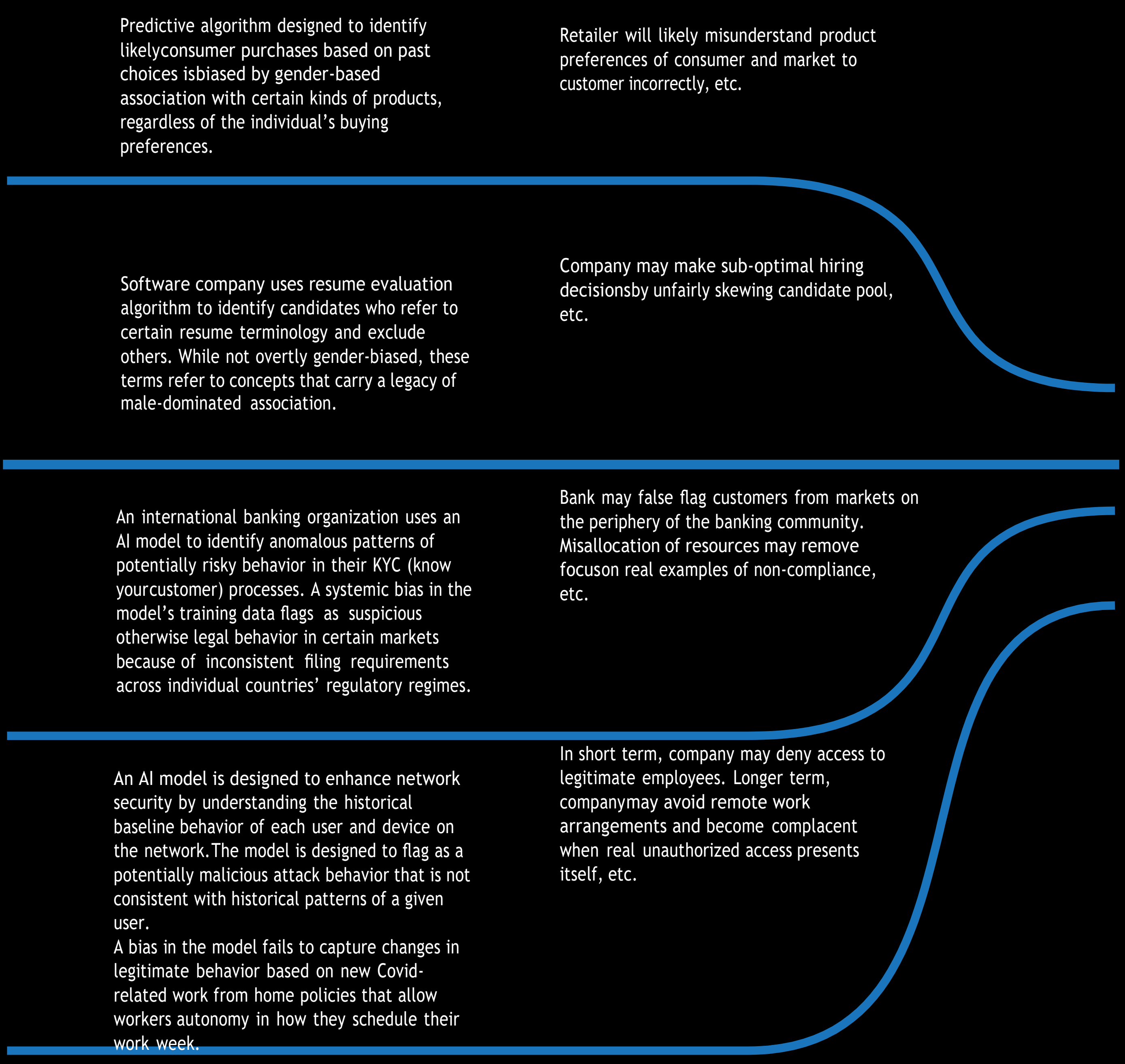
12 use cases with descriptions

	Use case	Description
1	Algorithmic health care providers	AI makes diagnoses and prescribes medical interventions
2	Behavioral modification	AI “nudges” people in highly customized ways to change behaviors (e.g., health, saving/spending)
3	Home virtual voice assistants	Voice assistants that help residents (elders, children, adults etc.) in the home on a variety of tasks
4	Social credit and underwriting	AI uses big data to get 360-degree view of individuals’ behaviors and make decisions about loan applications, premium pricing, etc.
5	Algorithmic financial planners (Robo-advisory)	AI agents monitor personal finances and conduct financial/retirement planning on behalf of clients
6	Fully autonomous vehicles	Vehicles that drive themselves without human input (Level 4 or 5)
7	Facial recognition check-ins	Facial recognition speeds check in at airports, hotels, banks, etc.
8	Law enforcement surveillance	Widespread use of facial recognition, license plate scanners, drones etc. by law enforcement and security services
9	Personalized algorithmic pricing	Algorithms change prices for individual consumers based on data about their habits, preferences and circumstances
10	Algorithmic recruiting	Using algorithms to screen applicants and make hiring decisions
11	Information curation and distribution	Algorithm’s curate, analyze, prioritize and amplify digital content (algorithms that pick and choose what news/information to show you)
12	Human emotion analysis	Sensors and computer vision combined with algorithms that predict and track human emotions

FIGURE

Model bias scenarios and their potential impact on decision-making and trust

Customers Employees Suppliers Regulators Community Potential customers Investors



Example of biased model | **Potential impact of bias on decision-making**

Predictive algorithm designed to identify likelyconsumer purchases based on past choices isbiased by gender-based association with certain kinds of products, regardless of the individual's buying preferences.

Retailer will likely misunderstand product preferences of consumer and market to customer incorrectly, etc.

Software company uses resume evaluation algorithm to identify candidates who refer to certain resume terminology and exclude others. While not overtly gender-biased, these terms refer to concepts that carry a legacy of male-dominated association.

Company may make sub-optimal hiring decisionsby unfairly skewing candidate pool, etc.

An international banking organization uses an AI model to identify anomalous patterns of potentially risky behavior in their KYC (know yourcustomer) processes. A systemic bias in the model's training data flags as suspicious otherwise legal behavior in certain markets because of inconsistent filing requirements across individual countries' regulatory regimes.

Bank may false flag customers from markets on the periphery of the banking community. Misallocation of resources may remove focuson real examples of non-compliance, etc.

An AI model is designed to enhance network security by understanding the historical baseline behavior of each user and device on the network. The model is designed to flag as a potentially malicious attack behavior that is not consistent with historical patterns of a given user.

A bias in the model fails to capture changes in legitimate behavior based on new Covid-related work from home policies that allow workers autonomy in how they schedule their work week.

In short term, company may deny access to legitimate employees. Longer term, companymay avoid remote work arrangements and become complacent when real unauthorized access presents itself, etc.

FIG: illustrate the individual character of model bias,we depict a few different case scenarios in which the nature of model bias could manifest and how decision-making and trust might be affected as a result.

Potential impact of bias onstakeholder trust	Potential impact of bias on stakeholder behaviors and firm metrics
Shopper could question retailer's understanding of his or her needs or interest in serving his or her needs.	Customer may be less likely to purchase from the retailer or recommend the brand to a friend resulting in **loss of sales.**
Workers might question how retailer's dependence on (or ability to leverage) emerging technologies will affect their futures and whether retailer even cares.	Workers may be less motivated to work for the organization or less likely to recommend others work for the organization, driving **lower engagement** and **lower productivity.**
Vendors, once aware of the bias, may reevaluate whether retailer has ability to promote their products or even cares that this misunderstanding of shoppers' needs may reflect negatively on vendor.	Fewer suppliers may want to work with retailer resulting in fewer product offerings and potentially **lower revenue.**
Employees may question company's inability to hire diverse candidates or dedication to do so.	Current professionals may leave organization—and would-be professionals may avoid it altogether—leading to possible **staff shortages, lower productivity, and lower profits.**
State and federal agencies, such as the EEOC, may question company's ability to carry out equal opportunity mandates and its commitment to do so.	Regulators may begin investigations that lead to agency civil litigation that could result in **fines and other penalties that hurt the bottom line.**
The public may perceive company as part of the "old boys club" and as indifferent to that emerging reputation.	A public reputation of callous indifference to gender equality could lead to consequences that go **well beyond the bottom line** as a reputation once established—even if unfairly—can endure for years and is very difficult to reverse.
Regulators might question company's ability to identify truly suspicious activity and their willingness to their willingness to address the unique regulatory needs of customers.	Regulators may add new additional compliance requirements **generating unnecessary costs.**
Staff personnel may wonder whether perceived inability to account for variability in regulations extends to variability in skills and career goals and whether it even matters to the organization.	Damage to employee morale that stems from the organization's inability to achieve its core vision and purpose on something so fundamental as KYC — and what that means to their own career goals — could **negatively impact worker engagement, hiring and retention.**
Employees may question company's ability to keep their systems safe and secure and the company's interest in providing a safe and seamless work at home environment.	Repeated denial of access could drive frustration among workers and **negatively impact morale and productivity.**
Future clients of the organization may perceive that if the company cannot maintain something as basic as network access, it may not be able to handle their own client needs and may not even make it a priority.	Customer growth may become difficult once this model bias issue becomes known, **stifling revenue and profit growth as a result.**
Security matters a great deal to the investment community. So they may be quick to presume the worst about the company in this area and, unless corrected immediately, they'll likely be quick to assume that it is an issue of relative unimportance to the company.	If investor sentiment turns negative for a prolonged time, it may become **more difficult to raise capital.**

Acknowledgements

This work would not have been possible without the diligent contributions of Krupal A of Infomerica India Pvt.Ltd

Your Infomerica Contacts

Krupal A
CTO – Infomerica India Pvt.ltd
(919) 655-0800

Swamy Sriperumbudur
President & CRO: Innovator in Customer Journey
+91 9195798428

Get in touch:

Address:
252 Towne Village Dr.
Cary, NC 27513

Phone:
(919) 655-0800

Email:
info@infomericainc.com