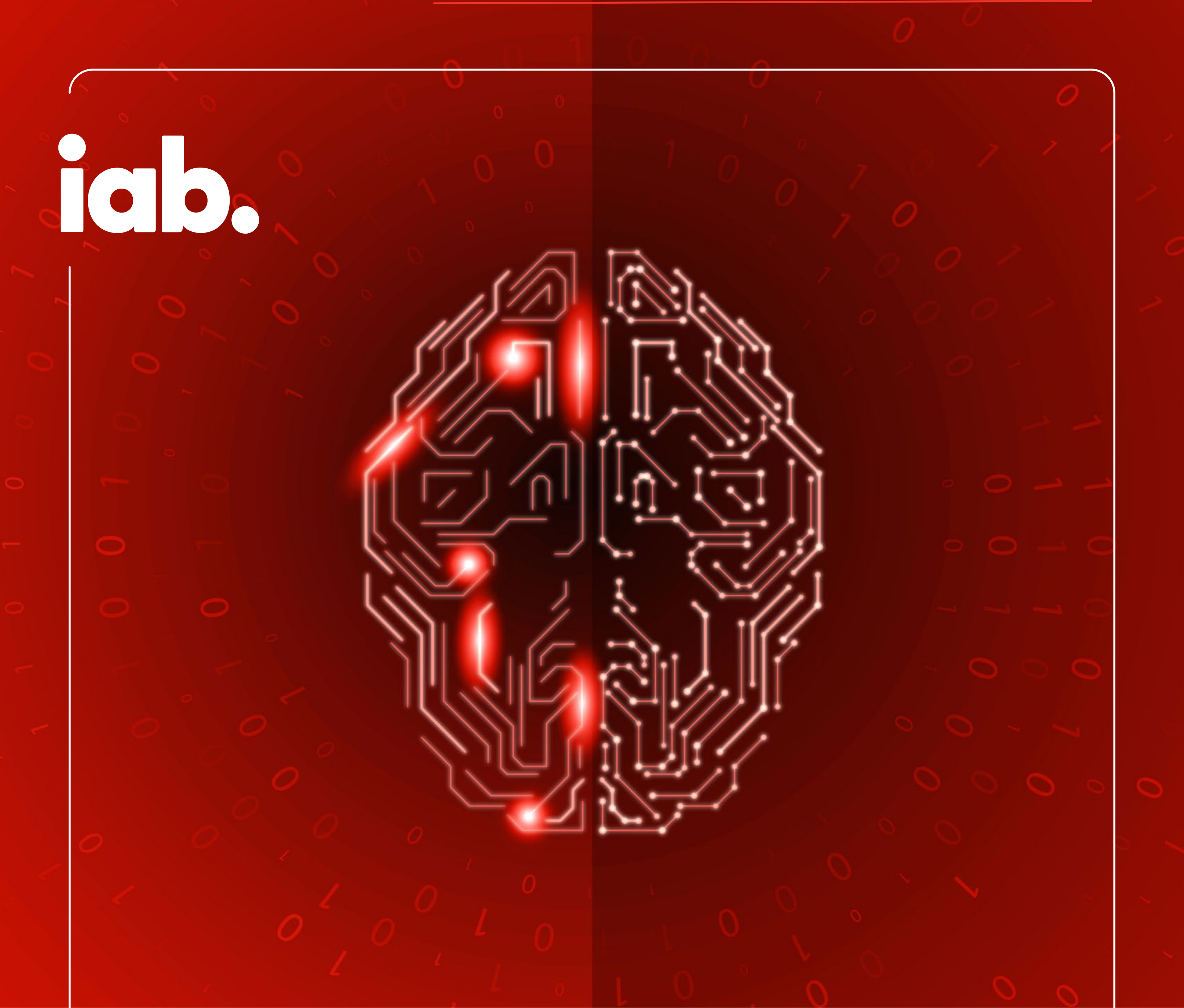


Understanding Bias in AI for Marketing

A Comprehensive Guide to Avoiding Negative Consequences with Artificial Intelligence

NOVEMBER 2021

Acknowledgments

This document was developed by the IAB AI Standards Working Group. IAB would like to thank the members of the IAB AI Standards Working Group for their time and effort.

AI Standards Working Group Leads

David Olesnevich, Head of Ad Product, IBM Watson Advertising (Chair)

Angelina Eng, Vice President, Measurement and Attribution, IAB

Main Contributors

Alliant: Malcolm Houtz, Vice President, Data Science

Dentsu: Brian Klochkoff, Head of Automation

Ericcson Emodo: Jake Moskowitz, Vice President of Data Strategy and Head of the Emodo Institute

IPONWEB: Troy Cunningham, Head of Information Security

Loeb & Loeb: Cathy Mulrow-Peattie, of Counsel

IBM Watson Advertising: Robert Redmond, Design Principal and Head of AI Ad Product Design

ViacomCBS: Imène Drir, Vice President, Measurement Strategy and Data Governance

Table of Contents

Introduction: The Growing Role (and Risks) of AI in Advertising and Marketing

Companies that seek to be successful in the future should develop genuine relationships with their customers, built on trust. Customers reward these companies with deep engagement around products and services and accompanying brand loyalty which often cannot be acquired through monetary means. Marketing and advertising are key practices for developing customer relationships.

> ***Great AI systems require fastidious design, development, deployment, and maintenance by teams forged through broad and diverse representation.***

The underlying mechanisms for marketing and advertising are increasingly enhanced by artificial intelligence (AI) to help achieve all the benefits in efficiency and effectiveness. Great AI systems require fastidious design, development, deployment, and maintenance by teams forged through broad and diverse representation. Throughout the lifecycle of a system, teams should strive to deliver AI that can be explained, trusted, and understood. Mission-critical is the understanding of unwanted or unintentional bias, how it originates, infiltrates systems, impacts models, and is deployed at scale in algorithms, affecting performance and potentially exacerbating societal inequities and eroding trust.

With proposed new regulations in the European Union[1] for trust and excellence in AI, U.S. state legislatures are passing bills[2] to study the impact of artificial intelligence on citizens. With increased regulatory scrutiny and prioritization by the Federal Trade Commission[3] on AI and increasing internal compliance governance on AI use forthcoming, this guide is a must-read and a starting point for companies to develop frameworks for better AI solutions. It is intended for the entire value chain, not just the solution developers. Focusing on bias, we pull from real-world experience by AI professionals to define key terminology and explore the roles and responsibilities of stakeholders: requestors, builders, end-users, compliance and legal teams, and consumers. Throughout four phases—awareness, exploration, development, and activation—we explore the role of key stakeholders and their associated responsibilities as AI champions and arbiters of bias.

Bias is generally introduced into AI systems unintentionally by humans, but the duality of humans and machines makes bias detectable—and the risk mitigated helps companies do the right thing for their businesses and society.

[1] https://ec.europa.eu/commission/presscorner/detail/en/IP_21_1682

[2] https://www.ncsl.org/research/telecommunications-and-information-technology/2020-legislation-related-to-artificial-intelligence.aspx

[3] https://www.ftc.gov/news-events/blogs/business-blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai

Roles and Responsibilities

From the initial business need to design and develop to maintaining a scaled system, individual participants are responsible for enacting practices to help mitigate bias. To help guide awareness of obligations, we have identified the common roles below to classify participants and their impact on the process. While your exact role might not be listed here, the intent is to showcase the relevance across varying roles of engagement from decision-making and strategic planning to execution. Within the advertising and marketing industry, these roles might touch various parts of the ecosystem:

- Ad Verification & Brand Safety
- Audience Activation
- Audience Segmentation
- Audience Prospecting
- Audience Targeting
- Behavioral Targeting
- Bid Optimization
- Campaign Optimization
- Closed-Loop Reporting
- Contextual Targeting
- Content Generation
- Conversations
- Creative Development
- Creative Optimization
- Data Migration
- Geotargeting
- Look-Alike Modeling
- Measurement & Attribution
- Personalization
- Predictive Audiences
- Process Discovery
- Remarketing / Retargeting
- Traffic Shaping
- Video Management
- And Many More...

For some of these use cases, please refer to the Artificial Intelligence Use Cases and Best Practices for Marketing guide published by the IAB AI Standards Working Group in March 2021.

What follows are the key participants in the bias detection and mitigation process, and an understanding of their roles and responsibilities.

Requestor

The requestor is typically the individual or team that has defined a business need, helping to inform the strategy and vision of the solution. Often, this role will help craft the direction but will likely not be the day-to-day consumer of the system and will probably offer guidance but not own the final decisions.

Typical job functions include: *C-level and senior executives, strategists, research leaders, client services leaders, and product managers.*

Main responsibilities include

- Identify a business need
- Articulate specific needs and/or requirements
- Consider factors that may limit or prohibit progress such as data limitations, financial costs, regulatory considerations, bias risk factors, etc.

Builders

Builders are the team that will guide the entirety of the problem-solving, solution-creation process and guide the execution path. They consist of the groups that work on discovery to identify the feasibility of the solution, then follow through with planning, requirements, research, architecture, design, development, and quality assurance. This group is critical in defining and implementing bias mitigation strategies.

Typical job functions include: *Product managers, designers, engineers, architects, developers, researchers, data scientists, and quality assurance.*

End User/Beneficiary

Whether internal or external, every solution will have an end-user or beneficiary whose needs and expectations are critical to success. Often these participants will provide feedback on performance and flag inconsistencies and issues. They usually will also request improvements or new features to meet their needs. This role is often aligned with the requestor and can be considered as a secondary stakeholder.

Typical job functions: *Can vary widely and consist of anyone within an organization, a partner, or a consumer.*

Legal and Compliance

Corporate legal and compliance responsibilities vary from company to company and industry to industry. It's important to have a team of experts who understand the legal and regulatory rules which apply to your AI application and help weigh the risks and potential liabilities. The function of this role is to identify product outcomes or results that may be directly and indirectly out of compliance with government and industry regulations, as well as with their company's anti-bias practices. They advise on whether the use of data is consistent and in line with privacy and related product regulations, while always keeping in mind possible trends in regulatory compliance rules that may affect their business.

Typical job functions: *Can vary widely from paralegal, legal counsel, executive, officer, advisor, or auditor that is focused on regulations, law, policies, privacy, governance, risk, and/or compliance.*

Consumer

The human connection with a brand and any outcomes, engagements, or KPIs are the typical focus of most advertising technology tools. In this context, the consumer's experience should be considered in designing the data or algorithms. Their experience will be affected, whether observable or not, by any unwanted bias in a system.

Bias could affect their perception, consideration, and awareness. It might even affect them socially or economically without their understanding.

A consumer is anyone who might be exposed to the outcome of marketing or advertising technology.

Understanding Bias in AI: Key Considerations

As AI adoption grows across the advertising and marketing ecosystem, we must establish practices that help minimize bias in the technologies and campaigns we create and execute. The trust of the consumer, the partner, and the industry is at stake. It is possible to define practices that support trusted AI-driven tools, but we must consider the entire process.

We have gathered insights and best practices from leading agencies, brands, advertising technology providers, and legal and compliance leaders to help you transform your practices.

At a high level, here are a few considerations:

To err is human. To err in a system is a choice to not audit.

As we seek to understand bias, we should know that it is a human cognitive condition earned over a lifetime of experiences. We blindly educate systems to employ our biases when we don't question our approaches and assumptions. We have a choice in the matter, and we just need practices to help us identify and mitigate issues.

AI is not inherently biased.

There are perceptions that AI itself is biased. But the reality is that algorithms inherit biases from humans throughout the project lifecycle and can be trained to harbor biases and scale them. This guide helps you consider ways to create AI systems where biases are minimized.

Unwanted outcomes can cascade.

Many systems engage with and interact with others in our data-rich industry, weaving value across the marketer's landscape and providing insights. As we consider how bias might affect our products and processes, we should also reflect on how those interactions could impact downstream systems.

Knowledge is accepting.

Bias can creep into our processes and strategies from many different directions. While this document helps you consider approaches, the knowledge that bias will be present is a huge step in establishing practices to mitigate it. Ensure that your teams understand types of biases, where they occur in project life cycles, and develop organizational strategies to minimize their effects.

Use this guide and the definitions at the end of this document as opportunities to kickstart your organizational knowledge. Ask questions and enact changes that make your products and consumer relationships better while helping to increase trust and understanding for marketing and advertising.

Important Factors and Considerations

A variety of forms of bias may appear in algorithms. Even when sensitive variables such as gender, ethnicity, and sexual orientation are excluded, AI systems still make decisions in the context of training data that may reflect human prejudices or historical inequalities.

Ideally, companies should be aware of all types of bias that can be introduced at any point in the process or in the outcome of the algorithm, evaluate the relevance of each bias to their business, determine their risk tolerances, and establish a plan to mitigate each. Generally, this task is better accomplished by a group of people rather than a few individuals. And instead of managing the process in silos, it should be managed throughout the organization.

While there is no guarantee that bias risks will not occur, it is more important to be able to manage and reduce bias in a way that delivers fair results and increases trust in the process.

While there is no guarantee that bias risks will not occur, it is more important to be able to manage and reduce bias in a way that delivers fair results and increases trust in the process. Considering any potential consequences that aren't evident immediately is key, and to prepare effectively, organizations must anticipate and weigh all possible outcomes, then implement the policies and strategies appropriate for each one.

Beyond the necessity to question both our conscious and unconscious assumptions, intentions, and our proposed applications of AI, there are several other factors and challenges you should consider:

1. **Data volume.** The internet has made it possible for marketers to obtain lots of information about consumer habits, likes, dislikes, activities, and preferences that were previously impossible to know. As the digital advertising ecosystem develops, and as more data is ingested from multiple sources like mobile phones, social media, connected TV, programmatic, search, and display, it becomes increasingly difficult for businesses to acquire, sort, link, normalize, and use the data.

   To analyze, train, and minimize bias in AI systems, enough data is as important as the data itself. When you don't have enough data for your models it results in a variety of uncertain outcomes. Companies that realize that they do not have enough data to make informed decisions can become frustrated as a result. This is often referred to as underfitting, where a statistical model or machine learning algorithm fails to capture the underlying trend in the data.

On the other hand, the opposite is true when you have too much data. Although it might seem that the more data you have, the better your results will be, that is not the case. Statistical models that have too many parameters can result in overfitting. Overfitting occurs when a model is extremely complex, such as when it has too many parameters relative to how many observations are available. When a model has been overfit, it means the algorithm is specifically set up to be accurate against the training data set, but it isn't necessarily transferable to other data sets in the wild. Thus, it renders poor predictive results, since it overreacts to small changes in the training data. Consider reducing the dimensions of the data so that it removes any noise and still allows you to generate the same results or outcomes.

Since data sets for AI vary in size depending on the type of information you need, your organization will have to decide whether you have the appropriate data sets and data points needed. What matters most is not how many data points there are, but what the data contains. It is important to have standard terminology, standard definitions, and standard approaches to data definitions throughout your organization. You should also determine your data volume requirements and to provide specific data requirements to avoid any issues.

2. **Data quality.** Data sets are generally regarded as being high quality if they are suitable for the intended use in operations, decision making, and planning as well as if they accurately represent the real-world construct they refer to. Diversity, scale, and quality of input data are fundamental characteristics that determine the predictive effectiveness of AI models.

   Unfortunately, some algorithms are found to contain prejudicial data definitions, or what is called "prejudicial algorithm data" used for training a system and generating an algorithm that lacks adequate data to reflect reality. The truth is that most data sets contain inaccurate, duplicate, mislabeled, and missing data. It can be especially problematic if the data is unstructured, unorganized, or outdated.

   In this context, monitoring the quality and accuracy of data is essential, as is obtaining and implementing appropriate data sets regularly to minimize bias, so that they can be accessed and analyzed to properly train models.

   Furthermore, you should assess and evaluate any outliers, subcategories, or data points you do not want to bring into your systems as they may introduce bias. Think about what types of data you need to avoid or what anomalies need to be eliminated. It is also vital to consider gaps in the data that could potentially exclude a certain audience segment or negatively affect the user experience.

3. **Computing power.** An AI system's performance can be adversely affected by technology and process problems. Modern data-intensive systems often have a combination of data streams, complicated ETL processes, post-processing logic, and a wealth of analytical and cognitive components. As a result, advanced analytical models, real-time decision-making, and knowledge extraction demands an ever-increasing number of cores and graphics processing units (GPUs) to work efficiently.

In most cases, this requires the computational power of a supercomputer. Hardware for AI-based systems differs from standard computer hardware. Unfortunately, these machines are quite expensive. Despite the ease and effectiveness of cloud computing and parallel processing systems, they also come with a cost.

Furthermore, microprocessors, microchips, general-purpose chips such as GPUs, or more specialized chips such as tensor processing units (TPUs) or vision processing units (VPUs) used in AI applications to accelerate data processing face a shortage of supplies and higher prices. In an era when unprecedented quantities of data are flowing in and algorithms are becoming increasingly complex, not everyone can afford these investments. It is a perennial challenge to make computing power more cost effective as well as more efficient.

A good place to start when choosing AI hardware is to decide how improving customer or supplier interaction will affect your business. The next step is to choose a software and hardware solution that supports these changes.

The decision of whether to go with general-purpose chips, more specialized chips, or to look for startups that offer more innovative solutions will depend on what AI tasks your company needs to accomplish.

As we’ve seen in the rapidly evolving AI hardware market, out of the three most important components of hardware infrastructure—computing, storage, and networking—it is computing that has seen the most progress and makes up most of the market. As a team, you should determine the right systems that will best meet the organization’s needs. It is important to consider both the short-term and long-term financial implications as well as the longevity of the hardware. Think about any potential issues that may arise as you collaborate with your team on developing specs and requirements.

4. **Data privacy and security risks.** Machine learning and artificial intelligence systems possess the same potential for misuse and misconfiguration as any other technology, but they also pose unique risks. These risks have increased as enterprises implement AI-powered digital transformations. It is imperative to ensure the algorithmic model satisfies risk-management and regulatory requirements throughout the development and outcome process.

   AI systems can contain sensitive information, such as user geolocation, health data, financial details, and personal information that can be exploited by hackers or disingenuous users or which require transparency and appropriate notices. While companies can be unsuspecting accomplices, they can still be subject to consumer backlash and regulatory consequences as a result. The sensitivity of this data, to both an organization and to individuals means that the unknowing collection and release of such data can have vast regulatory consequences. Security needs to be a critical layer of your AI systems.

Another security risk associated with AI systems is the possibility of compromising the decision-making processing so that the systems behave in a way that is different from what their design principles would expect or desire. One of the most frequent attacks on machine learning systems is designed to make high-volume algorithms make incorrect or false predictions by supplying a malicious input to the system. An attacker uses this method to present data that does not exist in the real world, forcing them to make decisions based on false data. This is one of the most damaging types of attacks since its effects can last for quite some time and extend widely, which makes it more dangerous than many other security threats.

These risks illustrate the importance of integrity checking the training data sets, which are used to develop AI models, and securing the inputs and outputs of the systems that underlie the AI solution. Regulation is now widely recognized as necessary for both the development of AI and the management of the hazards associated with it.

Because the stakes can be high with the use of AI, it is prudent to make a concerted effort to secure your AI project in a way that's consistent with the best practice for security overall. Some of the best ways to do that is to threat model your system with frameworks like STRIDE or DREAD.[4] These threat modelling frameworks are approaches one can use to systematically evaluate whether threats exist in and for the system, as well as how severe they are.

For example, STRIDE evaluates a system for places where you can:

  - **Spoof** an identity
  - **Tamper** with data
  - challenge the **Repudiation** of an event
  - create instances of **Information disclosure**
  - perform a **Denial** of service
  - or **Elevate** your privileges

- The DREAD framework is similar, with a focus on ascertaining:
  - how much **Damage** an attack could do
  - how **Reproducible** the attack is
  - how **Exploitable** it is
  - how many users would be **Affected**
  - and how easy it is to **Discover** the treat

Many threat and risk modelling frameworks exists, and STRIDE and DREAD are just two examples of threat modelling frameworks which can be used.

[4] OWASP Advanced Threat Modeling

Alongside threat modelling, you should plan to include security controls around your AI systems, or your organization. At the minimum, you should ensure that you have strong identity and access management controls, and you should consider adopting privileged access management (PAM) controls and policies to restrict access from unwanted parties. Additionally, consider making use of system-to-system communication policies (like software-defined perimeters) to make sure that data isn't injected mid-stream by untrusted systems. You may also want to put in place intrusion detection and prevention systems (IDS/IPS) to detect anomalous input (like a sudden spike).

The details of what you should do to secure a system can be overwhelming, but fortunately there are lots of common frameworks and standards that you can use. Consider a few of the following:

- OWASP Top 10 for your application security
- CIS Top 18 controls for organizational security
- NIST Privacy Framework for risk management
- NCSC guidance on cyber security

5. **Legal and regulatory risks.** The ever-growing regulatory landscape and scrutiny of consumer, business, and employee data also affect how data is and can be used in AI. Companies need to be mindful of the type of data that is needed, be sensitive to the data you have, minimize the data that you use, and try not to exclude any groups of people. Intentional or unintentional bias could potentially lead to the launch of a regulatory investigation and litigation. Regulations on bias are not new to the advertising industry as certain sectoral regulations exist to prevent bias from occurring, such as the Equal Credit Opportunity Act (ECOA) and Fair Credit Reporting Act (FCRA).

The core approach should consider the dangers and biases of AI's underlying technology, such as machine-learning algorithms, source of the data, algorithm testing, the decision model, and the outcomes—as well as whether users can understand and developers can explain the technology.

Systems have developed processes to examine human behavior, but they struggle to develop an analytical framework to examine the decisions made by artificial intelligence, which is often marked by a lack of transparency to the data subject into the process of how algorithms are developed and the biases of the programmers working on them.

The U.S. government's National Security Commission on Artificial Intelligence (NSCAI) has emphasized the need to build systems of artificial intelligence that can be audited through rigorous, standardized systems of documentation. As part of its recommendations, the commission supports the development of a robust design documentation process and standards for artificial intelligence modeling, including an explanation of what data the model uses, what its parameters and weights are, how it is trained and tested, and the results it produces. The commission offers several recommendations concerning transparency related to AI technology but has not yet extended those recommendations to explain how the documentation will be used for purposes of accountability and auditing.

The Federal Trade Commission (FTC) has exercised its authority to regulate the private sector's use of consumer information and the algorithms that affect the consumer market under Section 5 of the FTC Act, which prohibits unfair and deceptive practices. In 2021, while the FTC acknowledges the potential benefits of AI, it stresses the importance of harnessing these benefits without unintentionally introducing bias. Several examples are provided in the FTC's most recent guidance[5] regarding ways in which AI can be considered deceptive or unfair. And the FTC also has provided updated guidance, regarding its expectations for organizations using AI. This guidance is below and the FTC has indicated that AI fairness is one of its regulatory enforcement priorities for this year and beyond.

- Start with the right data sets: Validate, revalidate and ensure whether there are gaps. Ask questions before you start.
- Beware of discriminatory outcomes: Test your results.
- Protect your algorithm from unauthorized use.
- Embrace transparency and independent review: Tell consumers how their data is being used in your algorithm and conduct independent audits.
- Tell the truth about the data you use and the algorithm results: If you are denying something of value, explain why and explain the results of the algorithm.
- Do more good than harm: Ask if your AI model meets this standard.

In addition to knowing the regulations and guidance for the U.S., companies should also consider how those regulations and guidance are applied on a local and international level. They may differ slightly, or completely contradict one another in some cases. Your company should evaluate the risks associated with each governance level.

6. **Public trust and reputational risk**. Generally, most people are unaware of how artificial intelligence is used in digital advertising and marketing. A vital question to consider is: How will the public (such as consumers, customers, journalists, etc.) respond if they were aware certain bias exists in your systems?

   An organization's reputation can suffer a lot if it is perceived that data has been misused or there are errors contained in the algorithms or models. AI distrust comes from a belief that biases may be hidden within AI technologies and that they can exacerbate harms in a more robust, extensive, and systematic way than human and societal biases can. With AI becoming more mainstream, awareness increasing, and public opinion converging, companies could find themselves not in favor if bias ever arises, and risk their reputation.

[5] Federal Trade Commission, Aiming for truth, fairness, and equity in your company's use of AI, April 2021

Reputational risks could be a result of:

- Discriminatory or unfair algorithms
- A lack of reliable or unintended outcomes
- Data mishandling or misuse of data
- Increased exposure to cyberattacks

A company's reputation and brand can be seriously damaged if any situation is handled wrong. This can even have serious ramifications on entire industries. Take a systematic approach to assessing a technology's benefits and downsides. How it relates to business strategy, objectives, culture, and people is crucial.

In addition to these factors, the remainder of the document outlines a few important concepts to consider, a checklist, and a list of questions for each phase of an AI system's lifecycle to help identify and mitigate bias.

Phase: Awareness and Discovery

Primary Stakeholders: Requestor and builder
Secondary Stakeholder: Consumer

What Happens in This Stage

Understanding and controlling for bias in your AI project starts at the beginning. This is where you set the initial scope of your project, and where you decide what you are doing. As a requestor, you should be aware that there are many biases, both positive and negative, that are present upon the inception of your project. Whatever good or bad input you have now, it will reflect through your entire project.

Like statistical work, most AI projects come with presuppositions that carry biases around the intended outcome, or about the group that's being studied.

Asking the Right Questions

Like statistical work, most AI projects come with presuppositions that carry biases around the intended outcome, or about the group that's being studied. What is it that you'd like to learn, and why are you trying to learn it? Are you trying to prove a specific point, or create a specific function or capability? Does the project unconsciously assume a potentially harmful bias?

Even the way you ask a question they can inform how you set up and study a problem with AI. For example, the question "How bad is driving for you?" already assumes that driving is bad and that it needs to be quantified, potentially excluding the conclusion that it's not bad at all. Question the goals and the desired output of the project before you engage, to avoid confirmation biases down the road.

Also consider the effects of time on your questions. A study done on driving based on data from the 1950s would be entirely different from one done now, or in the future. Consider how many safety laws have changed, how road surface and tire quality have changed, or even trends in commuting. The relevance of your project may be time limited, and it's worth understanding that to avoid using the learnings past their expiry date.

The Limitations of AI

Once you have a clear and specific understanding of your project's objectives, you can examine how you will use AI. Involve your AI builder, as you will require an understanding of the power and limitations of AI technology. This is key to understanding if it can be applied to your project.

AI is ultimately an input/output system, and your project will have a poor output if the input is poor. AI, like any other computer system, relies on an extremely narrow set of instructions, and capabilities to make decisions. AI cannot appreciate the complexity of a data set in so far as understanding what that data set means without being told. Without the multiple layers of context most people can natively imply and apply, AI is unable to make a value judgment and it's certainly not 100% accurate or correct, especially when stretched to its limit. Poor instructions can also lead to poor output.

The limits go beyond instruction sets, but also with technical capability. For example, automated facial recognition has been notorious for its inaccuracy in determining who a person is reliably. In one study for South Wales Police the automated facial recognition had as much as 50% false positives.[6] While poor input contributes results like this, it can also be due to a lack of sophistication with the available algorithms. But it can also be caused by basic limitations, like poor image quality, or physical limitations on the speed of data input and processing power.

Explore these limitations with your AI builder. The complexity, power, and availability of AI Technology, and data gathering technology has a large part in determining whether you can apply AI to your project.

Understanding the Impact

As you discover what you want to study, and preliminarily explore how you'd like to approach it with your AI Builder, you should more deeply explore the who and the why? Who is this project benefiting and why are you doing it? What are your core drivers?

Consider the impact on your consumers. Model how the outputs of AI could affect them and ask questions like: Could the output support a negative belief or bias about them? How does the data set (who's included and who's excluded) change the results? If data is excluded, could the results be harmful?

When dealing with any data around individuals, it's also important to consider privacy regulations. You need a legitimate basis for processing information that's clear and specific to the purpose and data required. It's best to consult with your organization's data protection officer, legal/product and compliance team, or privacy experts to get an early assessment. This will help you to get an outside view, but also ensure that you're complying with regulations and reducing your project risk in the long term. This will also help you to set limits and boundaries on the use and application of your AI project, which will pay dividends in later stages of your project.

Finding a Model That Works

Once you've been critical with the approach, impact, and possibilities of your AI project, it's time to work with your AI builder to assess and refine the project's goals, develop success metrics, and to evaluate any gaps between available data and desired data. Your conversations with the Builder should result in a well-scoped project which includes an achievable and carefully defined model objective, an understanding of how data will be procured and made available to the data science team, and how the model will be implemented and evaluated.

[6] 2018 Cardiff University, An Evaluation of South Wales Police's Use of Automated Facial Recognition, Statewatch

Let's switch roles and focus on the builder. As a builder, you understand that effective models must be trained and scored on relevant data. This data might be an existing, familiar data set, or it might be new and need to be vetted, procured, and joined to existing data. Although the discovery stage is too early to determine if biases exist in data sets, it is a good time to consider what biases may be inherent in the data, and what, if anything, can be done about it. As you work with the requestor, remind them of the limitations of AI, the need to be watchful of bias, and the need for relevant training data. You also know that AI/ML is not appropriate for every business problem. Although you probably have an idea of how the solution will be built, it is too early to know the model's exact architecture.

Builders and requestors, your conversations must include a discussion of bias. Bias can arise from improperly defined model objectives, third-party data sources, internal data collection methods, human interpretation, and prejudices. Thoughtful planning in the Discovery stage may make it easier to avoid bias during Exploration and Design. You both need to come to the table with an open mind tempered with your separate experience and expertise. AI is a tool, not a magic wand.

Biases in the Data

AI is an input/output system, and the quality of the data is at the heart of that. Inherent to our understanding of biased data is the assumption that we know what unbiased data looks like. To assess bias, analysts may compare distributions of various metrics from a development file to those of a file known to be unbiased, sometimes referred to as a "truth set". This unbiased baseline file is a standard used to measure bias in other files. If a comparison of the files reveals there are dimensions on the development sample which are biased, the analyst may be able to remove the bias by selectively sampling down the development file.

> ***Dimensions or attributes are rarely independent of each other, so sampling one dimension can change the distribution of other dimensions.***

Sampling a file to remove bias in multiple dimensions is typically an iterative process. Dimensions or attributes are rarely independent of each other, so sampling one dimension can change the distribution of other dimensions. If a data set contains twice as many records with a given attribute as the baseline file, the analyst can devise a sampling scheme that keeps only 50% of the records with that attribute. Once this is done, the distributions of all attributes need to be re-checked. If another dimension is now too high or too low, the sampling scheme is adjusted to also balance the second attribute, and another sample is drawn. Iteratively resampling is time consuming but should eventually lead to a data set with attributes that resemble the baseline file.

Some AI applications may require data balanced more granularly than single dimensions. Consider an application requiring an unbiased sample of people who identify as female and Black Americans. Although the sample may match national averages, a closer look may still be necessary. The intersection of these two dimensions, female Black Americans, may be underrepresented in the sample, and quite possibly overlooked in a single dimension analysis. An awareness of this concept of intersectionality allows architects and engineers to make thoughtful decisions while assessing data for bias. Intersectionality is also important to revisit in the "Tuning and Testing" stage discussed later. Even if a training data set is deemed to be unbiased, that doesn't necessarily mean results from the model won't be biased, so it's important to test for bias, particularly intersectionality, at this later stage.

Bias can enter a project in various ways. Here is an example of bias entering an ML application through third-party data, and how it was resolved.

An insurance company wanted to run a U.S. Federal Housing Administration (FHA) compliant acquisition campaign, promoting its insurance products to 'pre-movers'—households about to move or change address. The strategy was to build a look-alike model to qualify a large, promotable audience that most closely resembles pre-movers.

The file of known pre-movers was compiled from real estate listings in a particular geographic area, including households that had recently put their homes on the market. Owning a home in certain geographical locations may highlight a certain level of affluence, and persons with protected characteristics may be over- or under-represented in those selected geographic locations. Thus, a model built on this data would indirectly express bias regarding some ethnic and age groups.

Awareness of this problem encouraged comparing the pre-mover file to a national demographic file (considered representative of the U.S), which allowed for significant biases to be found. To remedy the bias inherent in the data, the input file was balanced on affected dimensions to express no bias compared to a national baseline file for protected characteristics, and modeling proceeded.

After modeling was complete, a validation set was checked for bias to ensure that none had been inadvertently added during the modeling. After production scoring and audience selections were made, the final output file was also compared to the baseline to ensure bias had been eliminated.

The awareness and discovery stage are very conceptual: You have to ask a lot of questions about the questions that you're asking. It's all about understanding your project and your data before you start with AI. Bias is inherent to nearly everything we do as human beings. Despite the connotation, bias is neither inherently negative nor positive. It just is, and you must ensure that you are aware of what bias you're introducing into your project. That awareness starts at inception and sets the tone for the rest of your project. The more you're aware of now, the better your results will be in terms of accuracy and applicability.

Checklist

- □ Establish who's ultimately accountable for the AI project and its use.
- □ Establish the core details:
    - □ Who is this project for?
    - □ What are you studying? Not only the question or problem but what kind of data that you are proposing to analyze.
    - □ Why are you doing this analysis?
    - □ When are you doing this analysis? Is the data or the analysis timebound, or is it continuously refreshed?
- □ Arrange exploratory meetings with an AI Builder and legal and compliance team to
    - □ Determine whether AI is appropriate and fit for purpose, or if its limitations preclude its use.
    - □ Where personal data is concerned, establish the legal basis for the study/processing of it, and perform a privacy impact assessment.
- □ Establish your key success metrics.
- □ Establish the other project roles and responsibilities.
- □ Document your strategy around biases at this stage:
    - □ Consider if there's unconscious or confirmation bias in the problem you're trying to solve (i.e., "How bad is driving for you?")
    - □ Determine if there's bias in the human-driven process you're trying to replace/improve upon, that you need to adjust for.
    - □ Consider the intersectionality and diversity of your potential data set.
    - □ Consider the potentially harmful effects of the output.
- □ Re-visit the core details if the above actions reveal issues.

Input Needed

- A clear objective for your project and what you're trying to achieve.
- Your project's core concept details: who, what, and why.
    - Who is this project for?
    - What are you studying? Not only the question/problem, but what kind of data that you are proposing to analyze.
    - Why are you doing this analysis? What is the value/benefit?
- Data resources that may be useful.
- Regulatory guidelines or requirements.

Output Delivered

- A well-defined, specific set of questions or problems, compatible for accurate measurement with AI methods available to you.
- A shortlist of data resources and a refresh cadence.
- Potential model types and architectures.
- An understanding of resource availability.
- Success metrics for your AI/ML model.
- Metrics/processes to evaluate bias.
- A project brief, or statement of work documenting the above and:
    - Purpose of the project (who, what, why, when).
    - Legal basis for the project (if necessary).
    - Roles, responsibilities, and beneficiaries involved, including who's accountable overall for the project.
    - The project's approach and initial bias considerations.
    - Timing and budget expectations.
- Decide whether the project can proceed to the next stage or be reviewed or clarified.

Bias Questions/Risk Level Assessment

At the outset of a project, you need to put every effort into discovering hidden and entrenched biases. Most of the biases in AI start with the questions we're asking and the decisions we're making before implementation. It is therefore extremely critical to be exhaustive in discovering and discussing biases in your project and your project team as many of these biases will become much more difficult to detect in the data and output later.

Ask yourself:

- Biases in the purpose:
    - Are you trying to prove a specific point, or create a specific function or capability?
    - Is there a bias towards a specific belief or outcome in the question you're asking?
    - Who benefits from the output of this project?
    - Is the output in the best interest of the data subjects, and the requestor, or does it benefit a specific group?
    - Biases in the technology:
    - Can AI meaningfully and accurately solve the type of problem or question you're asking?

- Is the technology powerful enough, and complex enough to support the inquiry?
    - Can you program enough instructions into the algorithm to ensure a meaningful human output?
    - Biases in the data:
    - Is there a conscious or unconscious bias in the intersectionality of the data?
    - Does your desired data set only represent a specific demographic or group? (E.g., is it demographically, geographically, or ethically skewed?)
    - Is this model likely to favor or work more accurately for a certain ethnic group?
    - Are older people likely to have different outcomes than younger people, and is this OK?
    - What will be used as a benchmark for unbiased results? What do you use as a basis for "unbias"?
    - Have you created reasonable metrics for finding and assessing potential biases in the results?
    - Does your benchmark need a wider (e.g., In a regional, national, or international) sample, and does it make sense to sample by geography or other subgroups?
    - How will you test for bias in the raw data? In the predictions?
    - Will you exclude or include data that will alter results, and if so, why?
- Biases in the output:
    - Can the results of your AI project create unanticipated, and unintentional effects for consumers?
    - Will bias be introduced by the matching process or match-keys?
    - Are there features/variables which violate any compliance regulations?
    - Can the use of AI do more harm than good? Could some of the outputs result in more negative outcomes than positive ones?
    - Could the desired outputs be used to imply or determine other information that wasn't in the original use case?
    - Could the output be mistakenly or purposefully used for other cases outside the original scope?
    - Do time boundaries in the data or the analysis change the relevance of the output?

Phase: Exploration, Solutions, and Design

Primary Stakeholders: Builder
Secondary Stakeholders: Requestor, end user/beneficiary, and legal and compliance

What Happens in This Stage

Within any solution lifecycle, it's important to prove the feasibility of the solution concept at hand given the work done in the high-level scoping/requirements gathering and discovery phase. When you look at problems through the lens of edge technologies such as automation/AI, it's imperative to kick the tires of the proposed technology within a solution design framework to refine the realm of the possible and manage stakeholder expectations accordingly. This phase enables your project team to ensure that protocols are in place to mitigate known biases and then identify and remediate previously unexpected biases.

An excellent starting point for how to tackle this phase is the traditional scientific method:

1. Observation/Understanding
2. Hypothesis
3. Materials
4. Observation
5. Analysis
6. Conclusion
7. terate

Industry terminology will be used throughout the rest of this section; please refer to the definitions section for reference. While the terminology will be technical, the reader should be able to follow along regardless of background.

Observation/Understanding

Formally aligning on your project team's understanding of the observed problem enables the team to speak the same language and maintain a transparent dialogue to ensure accountability throughout the development/solution lifecycle. This helps to limit implied biases as a result of poor project team alignment. To do this, you need to ask tough questions here to translate the business requirements in a way that can be meaningfully applied to computational modeling with aligned definitions. You also need to define the variables which will act as the inputs for the algorithmic procedure(s) and become comfortable that these knobs may need to be adjusted by humans or by the solution given the outputs from the experiment or proof of concept (PoC). There is a lot of academic jargon which needs to be properly sourced and can risk the project team falling into a philosophical debate. So, it's important to illustrate this in a common denominator fashion consumable by everyone in the project team, align on the understanding, and move on. For guidance on definitions, you can refer to the standard terminology section of this paper.

Algorithmic bias exists in several nuanced ways which requires deep attention to how problems are framed, how data is collected, and how the data is cleaned and prepped for training.

A given algorithm is said to be fair or to have fairness if its results are independent of given variables, especially those considered sensitive such as the traits of individuals which should not correlate with the outcome including gender, ethnicity, sexual orientation, disability, veteran status, etc. Algorithmic bias exists in several nuanced ways which requires deep attention to how problems are framed, how data is collected, and how the data is cleaned and prepped for training. As a relevant example, Amazon attempted to create an AI which identified the optimal candidates for software developer jobs; however, the models were trained to vet applicants by observing patterns in resumes submitted to the company over 10 years and that training was male dominated. This caused the model to become inherently biased to favor male applicants over female applicants. To understand the programming and behavior of automated algorithmic solutions, your project team must inspect the algorithm to be built into the machine learning model as well as its associated training set. This can be difficult to achieve when looking to implement or adopt third-party solutions as the intellectual property behind these models is proprietary and essential to the firm's competitive advantage. But at the same time, it creates a black box for your project team.

Your project team should be on the lookout for these categories of implicit bias and define them accordingly throughout the project—and align on policies that benchmark the collective alignment accordingly.

1. **Unknown Unknowns:** You may not be aware of a model's bias until it presents itself in outputs which is why your project team should PoC AI solutions.
2. **Imperfect Processes:** Test training sets are the same training sets used to develop.
3. **Lack of Social Context:** The concept and definition of the problem will vary across users requiring training and social alignment which reinforces the need for continuous collective alignment.
4. **Definition of Fairness:** You need to define what a fair world looks like and assign a quotient against that ideal state based on real-world production level variables.

Align on and define the scoped problem statement. For the sake of this article, let's define the problem statement as such:

"Brian always loses at tic-tac-toe."

Your project team can now help Brian win at tic-tac-toe going forward.

With a problem statement clearly defined, you will be able to build a strong objective foundation as a precedent to avoid implicit biases as the solution development lifecycle progresses and the temptation of scope creep evolves. You'll be able to more easily identify and mitigate the risks for biases such as anchoring bias, bandwagon bias, confirmation bias, self-interest bias, in-group/out-group bias, etc.

Defining the problem statement and then aligning the variables associated with that problem facilitates the business requirements being well understood and wrapped around a common vernacular. Using the tic-tac-toe example, you can define variables and vernacular as such:

1. Tic-tac-toe: Defined as the use case within the problem statement and may be referred to as 'the game.'
2. Player: Defined as participant or user in the game of which there will be player1 and player2.
3. Grid: Defined as the nine (9) square interface in which the players engage in the game
4. Square: Defined as a subset of the grid in which a player can attribute a mark against during the game.
5. Mark: Defined as the label or inscription a player can allocate to a square during a turn. One player will be recognized as X and the other player will be recognized as O.
6. Turn/Move: Defined as the dedicated alternating opportunity for a player to make a mark against a square in the grid.
7. Threat: Defined as the same mark in consecutive squares.
8. Opponent: Defined as the opposing player. Therefore player1 is player2's opponent and vice versa.
9. Winning: Defined as if one player achieves three consecutive linear marks in the grid
10. Draw: Defined as if neither player achieves three consecutive linear marks and all squares have marks allocated against them.

Now that your project team has aligned on the foundational problem statement and associated variables, you will need to define the algorithmic procedural requirements for the desired output. In the tic-tac-toe example, you would want to define the rules of the game that those variables can operate in to achieve the desired outcome of winning the game.

1. If a player has two consecutive markets (a threat), take the remaining square. Else:
2. If a move creates two threats simultaneously, play that move. Else:
3. If the opponent plays an outer center, play the opposite. Else:
4. Take the center square if it is free. Else:
5. If the opponent has played a corner, take the opposite corner. Else:
6. Take an empty corner if one exists. Else:
7. Take any empty square. Else:
8. Draw.

This approach mitigates the risk of technical biases becoming introduced to the solution itself by effectively defining the accepted rules and constraints that the solution needs to operate within.

With the project team aligned on the problem statement, variables, and rules; you have mitigated the risks of introducing groupthink biases and misinterpretations into the requirements gathering process which sets the tone for the rest of the project.

Hypothesis

Given your problem statement, entity definitions, and procedural logic outlined above, you can safely posit that artificial intelligence can solve for your use case

“The hypothesis is that artificial intelligence can be leveraged to win against humans in the game tic-tac-toe.”

This will be refined as you iterate through the rest of the process and given the lean development methodology for AI solutions, you will refine many such hypotheses and definitions as your project team approaches a product state and even after the solution has been implemented.

Materials

While an experiment/PoC is not intended to scale as a full production solution, you will still need materials and resources to allocate against the workstream.

It’s important to align on the technology stack that will be used in solution development and validate that this technology, platform, model framework, and/or language can account for the nuances in the problem statement, variables, and procedures/rules aligned on by the team. The technologies leveraged in the solution will play to the strengths and weaknesses of the builders. This introduces a series of risks that have the potential to execute the procedures in such a way that skews the input/output relationship.

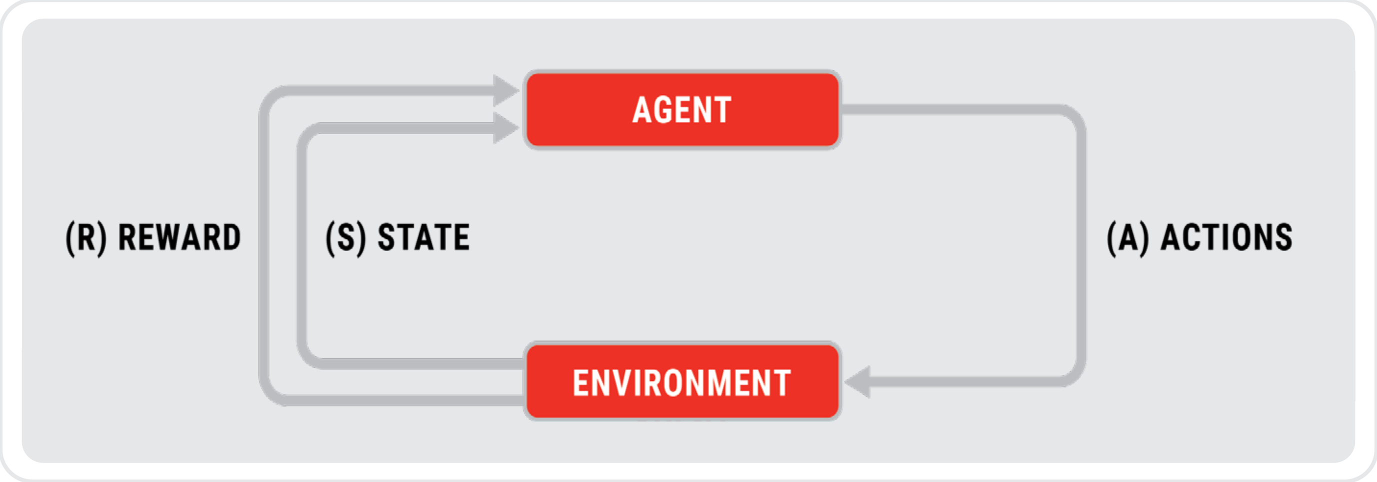
For example, a solution architect may be biased towards winning tic-tac-toe exclusively through non-diagonal consecutive marks. Transparency throughout this experiment and having accountability checks/balances in place through white-box modeling will ensure that these types of unknown biases are accounted for throughout the process.

This introduces a standard but typically overlooked responsibility for the builders of the solution: technical transparency. The builders of the solution hold the responsibility to ensure that the requestors, end users, and legal and compliance understand the functional implications of the technical build and how the various materials used in the build can affect output.

Using the tic-tac-toe example, the builders know that this is a zero-sum game with a manageable and finite number of moves. Therefore, the builders have a responsibility to explain their design approach and which technologies they would prefer to use and why so that the objective foundation set forth during project team alignment remains intact.

In the tic-tac-toe example, it's fair to expect the builders to bring a high-level understanding of reinforcement learning to the project team through visual aids and blueprints such as the one below:

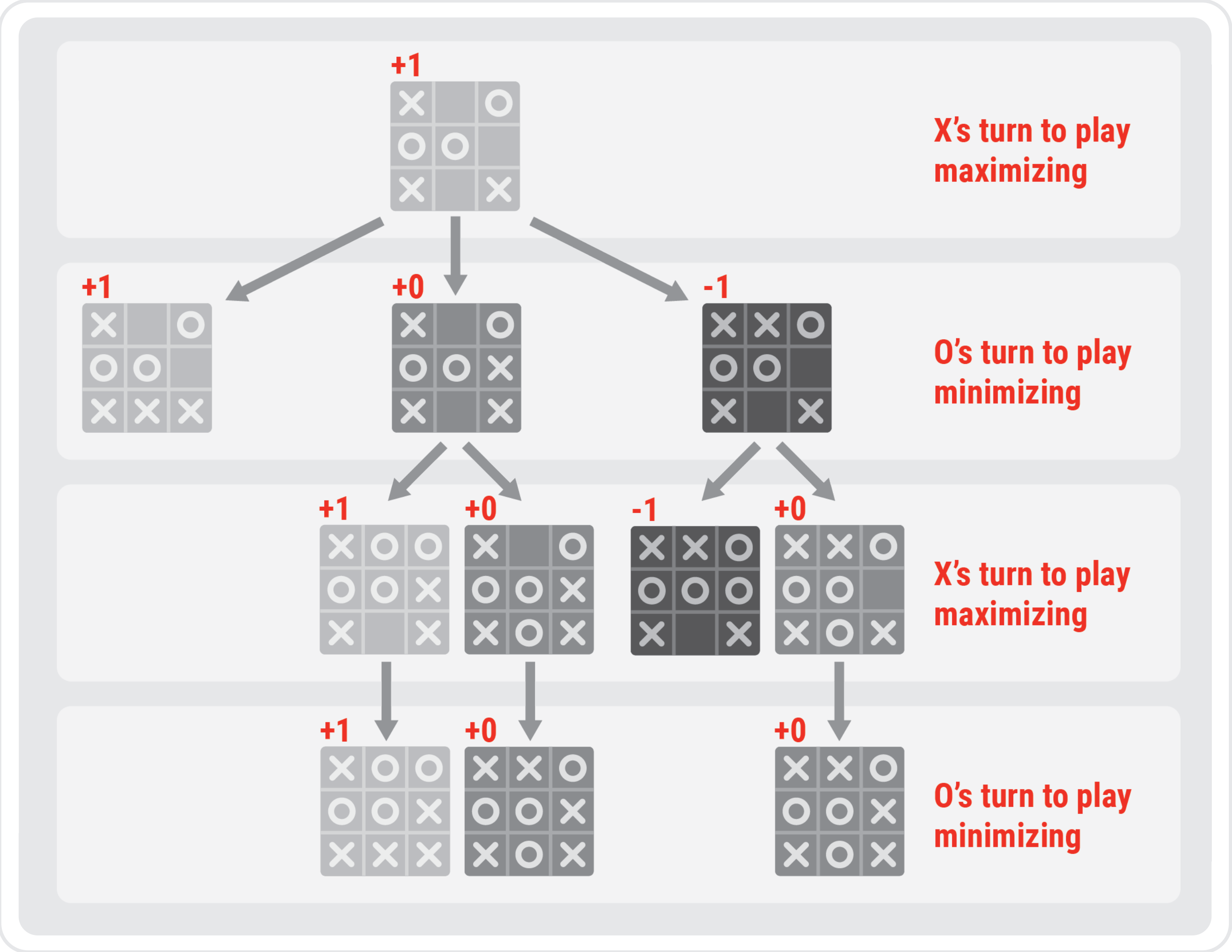
Tic-Tac-Toe Reinforcement Model Approach



Where there is a computer agent which takes actions (A) that act on an environment (E). That environment responds by providing a reward (R) for that action and bringing the system to the next state (S).

Upon aligning on the desired solution framework, the builders must explain the pros/cons of the materials and technologies they plan to use and how the limitations of those technologies could introduce technical biases such as compounding algorithmic biases. This approach also highlights the biases associated with third-party software and mitigates concepts such as the not invented here bias. This can be done by visualizing code. In the tic-tac-toe example, the builders would be expected to explain the minimax algorithm.

Tic-Tac-Toe Minimax Algorithm Tree Approach


+1
X's turn to play maximizing
+1
+0
-1
O's turn to play minimizing
+1
+0
-1
+0
X's turn to play maximizing
+1
+0
+0
O's turn to play minimizing

Looking at such a conceptual problem through the media/marketing lens, you can start to understand how these nested instructional procedures can amplify unintended biases and risk outputting skewed data which could then be used to make undesired decisions. For example, when leveraging AI/ML technologies to identify consumer insights based on a multitude of digital footprint data points, there is an implicit demographic exclusion risk if the solution requestors do not validate the visual workflows/diagram to understand how the algorithm operates when scoring certain groups higher than others. This is a critical point to consider when tuning the model during and after the proof of concept by maintaining transparency into the algorithmic procedures and foreseen programmatic consequences.

With the roles/responsibilities of the builder defined in how materials used can affect the design approach and introduce implicit technical biases into the algorithmic procedures, it's safe to assume that bases have been covered when it comes to putting up guardrails around introducing social biases into the technical solution. But it's critical to understand that there will always be risks of the defined solution retrospectively hosting undesired biases as social norms and definitions evolve. For this reason, it's important to host the solution within a flexible and widely accepted tooling which can be well versioned, annotated, and reworked over time.

It's a good rule of thumb to deploy the AI product or ML model through an ML infrastructure as a service in the cloud to avoid common challenges that come with these types of solutions such as migrating data from sources in the cloud to engineers' desktops for development which introduces both versioning risks as well as wait times. This means that waiting for a model to train based on a dynamic training set may result in missing bugs in code or even just the pain of waiting for models and sets to replicate.

Another best practice to adhere to when going through this type of solution architecture is to ensure that microservices are delimited for inference and business logic alike. This means that we're leveraging one microservice for taking the input of the solution and predicting the output, while the business logic for migrating and manipulating that data around the database exists as a separate microservice. This not only allows for scalability of the inference microservice from a compute resource perspective, but it also enables transparency for pinpointing bugs and undesired biases. For example, creating Docker containers for various microservices of the overarching service enables dedicated aspects of the model to be clearly annotated and understood during observation. This type of architecture is typically native within a cloud-based machine learning as a service or MLaaS system.

The added value of observing the solution in the cloud is that additional high-performance computing power can be leveraged by busy stakeholders so that they remain engaged.

Observation

Now it's time to deploy this experiment into a development/testing environment with agile testing guidelines against the product. The project team and stakeholders must understand the design paradigm and architecture which has been implemented as a product through code so that they can perform user acceptance testing and penetration testing to understand where undesired biases may exist.

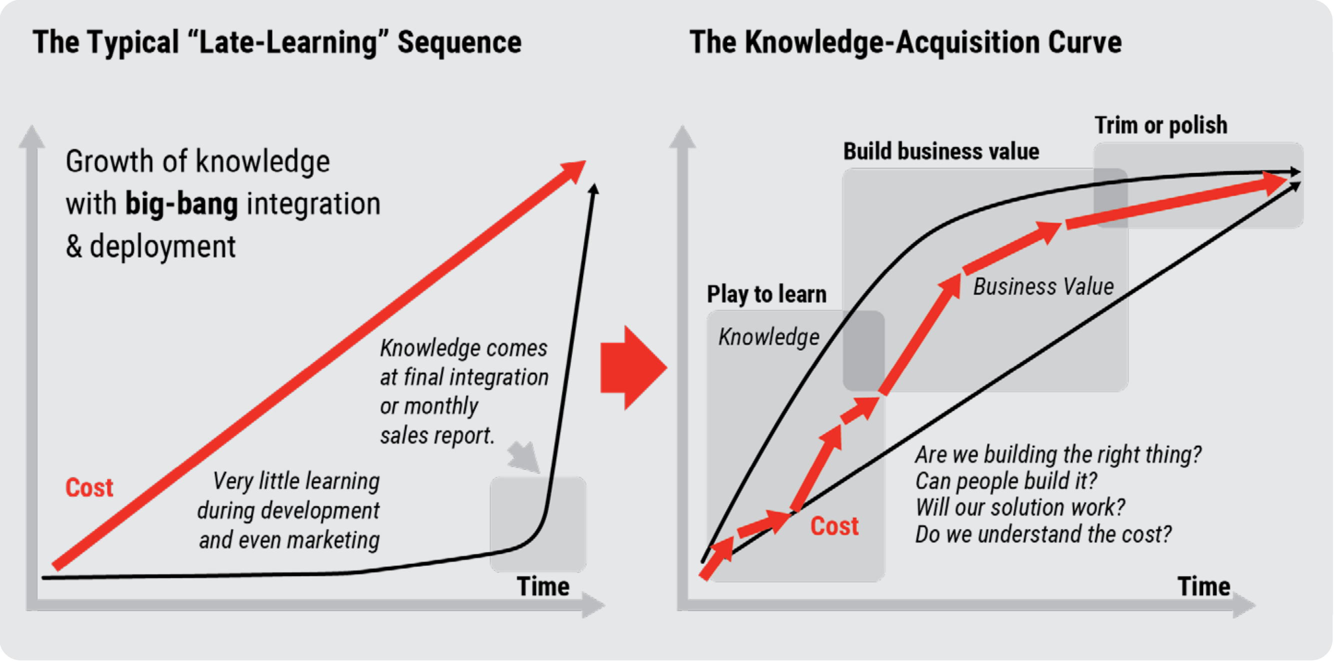
When you're working within this type of problem/solution, the stakeholders and development team alike must be interacting with the experiment/PoC and providing iterative feedback on a defined sprint schedule. This opens the dialogue within the project team to ensure that each team member is bringing their skills to this exploratory phase of the project and is accelerating up the learning curve as a team.

Typically, it's best practice to ensure that during any observation cycle that you see testing guidance provided through the lens of the problem statement and hypothesized solution with proper access given to the solution's front end across the project team. Commitments need to be made and accountability to those commitments needs to be adhered to. Approaching the observation cycle through daily scrums is an excellent way to ensure that this experimental phase is being observed properly while also pacing towards a conclusion.

The goal is to pay to learn upfront so that business value can be identified and realized through future iterations at an accelerated timetable.

Collective judgment is required from the group based on their testing against the developed solution and the challenged hypothesis. The goal is to pay to learn upfront so that business value can be identified and realized through future iterations at an accelerated timetable. Not only does this observation cycle approach test the development, maintenance, and release protocols of the solution, it also tests the will of the project team.

Risk Management


The Typical "Late-Learning" Sequence
Growth of knowledge with **big-bang** integration & deployment
Knowledge comes at final integration or monthly sales report.
Cost
Very little learning during development and even marketing
Time
The Knowledge-Acquisition Curve
Trim or polish
Build business value
Play to learn
Knowledge
Business Value
Cost
Are we building the right thing?
Can people build it?
Will our solution work?
Do we understand the cost?
Time

Using the tic-tac-toe problem, you would expect all project team members to become players against the AI and log their feedback of performance at a functional level as well as a user experience level in the daily scrum.

Throughout this observation exercise, we would expect the project team to be actively engaging with the solution to ensure that the categories of biases defined upfront are mitigated by the solution while also actively searching to turn unknown unknowns into known unknown biases and eventually known biases.

Analysis

Upon closing of the observation cycle, it's important to align on observations and analyze them collectively. What worked? What did not work? What could work better?

This will vary from project to project given the problem statement but consider that this swift phase is intended to be a proof of concept and not something that is intended to scale. Through the analysis of the collected feedback, the group must agree on if the solution concept was proven or does it require additional iterations before becoming comfortable with taking this into full-fledged and comprehensive development/tuning/testing phase.

At the end of the observation cycle, it should be clear given the frequent and ongoing conversations with the project team if the hypothesis was validated or not. And if not, it is worth running the experiment again with a refined hypothesis or different materials.

For example, in our tic-tac-toe use case and PoC solution, the user base would experience a slow response time from the model. In this case, the project team must ask itself: Is the user wait time an issue given our long-term goals? If the longer-term strategy for this type of a solution is to scale from tic-tac-toe to checkers and then to chess, the answer would likely be that this brute force minimax algorithmic approach is not scalable and you should debate whether to iterate this experiment using another approach such as depth-limited minimax or alpha-beta pruning or another more scalable approach. (See the appendix below for more details on these terms.)

Conclusion

Conclude based on the analysis and present those findings to each other as well as the sponsors/stakeholders at large. The possible conclusions are:

1. Proceed to full development
2. Refine the proof of concept
3. Hold
4. Kill

Consider the pros and cons for each conclusion scenario and posit a recommendation given the project team's experience in this phase and associated learnings considered against the available budget.

Iterate

If you refine the PoC and need to iterate through this phase again, it's critical to ensure that agile retrospectives are performed against the sprints identified throughout this phase. This ensures that learnings are accounted for as the experiment evolves accordingly so that a best-effort definition can be agreed to and the project team can reasonably expect a different outcome without bias.

Checklist

- □ Align on a problem statement, entities/variables, and solution rules.
- □ Align on a hypothesis.
- □ Analyze third-party technology and output to avoid liability for third party’s own biases.
- □ Propose materials and associated risks for technical biases.
- □ Build a proof-of-concept solution.
- □ Observe a PoC solution and deconstruct unforeseen biases.
- □ Conclude the future of the solution post PoC.
- □ Review data logs and outcome results.
- □ Test the data set of intended and unintended consequences.
- □ Test the outcome and results of intended and unintended consequences.
- □ Conduct third-party or independent audit testing of the above.
- □ Confirm the security practices governing sensitive data such as encryption, aggregation, and de-identification.
- □ Confirm the security practices governing all data.
- □ Establish a testing documentation process.
- □ Establish a change process and document it.
- □ Establish a feedback system and/or process.
- □ Confirm the accuracy of the consumer or business partner privacy notice.
- □ Review the human oversight and data decision making processes.

Inputs Needed

- Problem statement
- Entities and variables
- Rules and logic
- Hypothesis
- Desired solution outcome
- Development and staging environments
- Relevant development skills
- Third-party and/or proprietary technologies as needed

Outputs Delivered

- Documented solution scope including known risks for biases
- Design documentation package including technical visual aids
- Proof of concept or prototype solution
- Iteration plan

Considerations

- Has the data, the data analytics, or technology changed from the design phase?
- Assess whether there are filters, tools, or scoring that are used to alter the results?
- Assess whether the technology and data are being used as planned/designed? Have you confirmed this through design and data records (design, training, and testing records should be maintained and applied at this stage)?
- Does the data produce the results requested or are there unintended or unpredictable results (e.g., exploiting children, disparate treatment of persons based on a protected classification, or does not achieve the advertising goal)?
- Is certain sensitive data used that would require additional testing and controls?
- Has the organization deploying the AI been transparent about its capabilities, use, and results?
- Has the output been reviewed by a human, an unrelated third party, or an auditor?
- Have the data or technology designers changed?
- Has the organization been transparent with consumers and customers about the data use and that an algorithm is being used?

Bias Questions/Risk Level Assessment

- Has the project team aligned on roles and responsibilities?
- Has the project team aligned on the problem statement, entities/variables, rules/logic, desired outcome, proposed tech stack, and development framework?
- Has the project team aligned on how their technical strengths and weaknesses may contribute to being biased towards one technology over another and associated pros and cons of that technology?
- Has the development team worked to mitigate all known biases in the technical procedures?

Phase: Development, Tuning, and Testing

Primary Stakeholders: Architect
Secondary Stakeholders: Requestor and consumer

An AI-powered product is a process, not a one-time build out.

What Happens in This Stage

An algorithm is never done. An AI-powered product is a process, not a one-time build out. The process to make a great AI-powered product is one of ongoing iteration. As such, the development and tuning stage involves trial and error, and it does not end the day the product ships. The industry and environment are always changing and input data sets are always adjusting, which means an algorithm left alone is likely to become increasingly irrelevant. We need to continually evolve just to stay in place, by developing newer more refined versions of a model in an endless quest to maximize the quality of the predictions a model is making.

The testing phase bridges quality assurance and marketing. Responsible AI algorithm development involves a rigorous structured testing process that measures the quality of an algorithm's predictions across an array of metrics, and then feeds results back into the next phase of an algorithm's development. Successful tests also lend themselves extremely well to effective marketing of an AI-powered product once launched. The key to minimizing bias is to ensure the testing design is made by a team that understands representativeness, and continually questions assumptions.

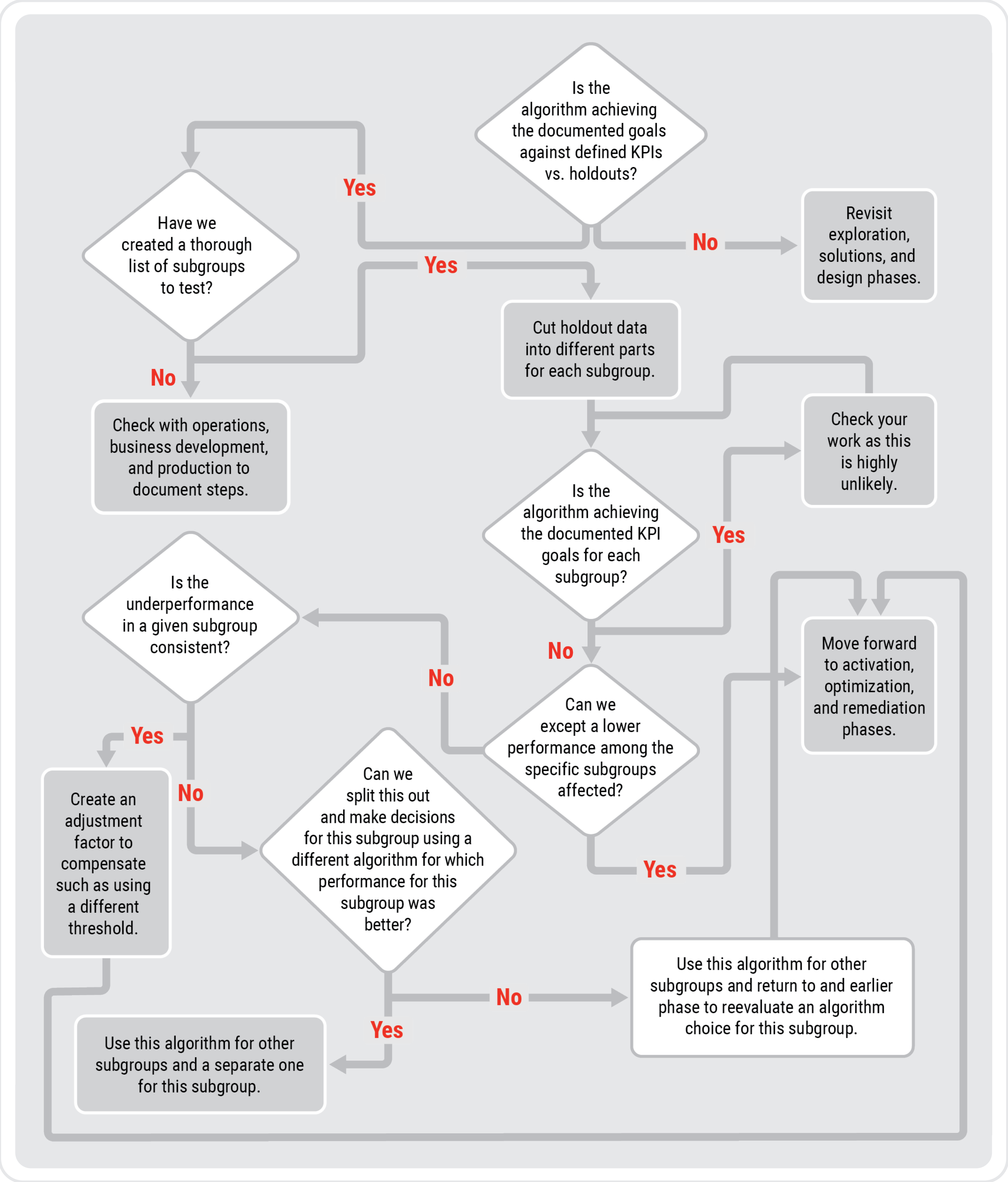
Checklist

Key steps during this process include:

- □ Test multiple versions of a model that try different combinations of inputs, to see which is most effective.
- □ Develop testing metrics and methodology that will be used to determine the success of an algorithm.
- □ Early testing is generally done "in the lab."
  - □ Initially use a "holdout" from the training data set (i.e., a % of the initial training data set that is randomly withheld from the training process, and then used after an algorithm has been developed to determine if the algorithm is performing well).
  - □ Then use a separate data set pulled and stored from the outside world, to determine if the algorithm is performing as well outside the training data set as it is inside.

- □ Later testing should be done “in the wild,” equating as much as possible what decision-making scenarios will be like once the algorithm is deployed and in use in a live situation.
  - □ This is critical especially for programmatic advertising, as there are so many other factors that could lead to bias in the algorithm results once deployed in the real world. Some examples:
    - □ Filtered out inventory for reasons unrelated to the algorithm that may have inherent biases, such as price floors above what you want to pay (you may be eliminating inventory that over-indexes with certain consumer populations), or categories of content deemed inappropriate for a client.
    - □ Accessing a non-representative set of supply sources. Perhaps there is selection bias in the choice of publisher partners, such as to keep inventory costs low, or due to a particular client’s whitelist.
    - □ Being outbid. Even if an algorithm predicts certain bid requests are valuable, some portion of these may be unavailable because they are valuable to other bidders too that are willing to bid more. This might lead to results being worse in the wild than they are in the lab, because in the lab we may be assuming we can win any auction we want.
  - □ The point of moving testing to the wild is not to try to eliminate all biases in the availability of inventory, but rather to help ensure you know before deploying an algorithm how it will perform in the wild, adjusting for whatever biases do exist in the wild, which is often very different to how it performs in the lab.
- □ Because algorithm development is a virtuous cycle, it’s critical to continually feed testing results back into the design stage, to continually develop newer and better performing versions of an algorithm. You should update algorithms as often as weekly or at the very least quarterly.

Assessing Bias in AI Decision Tree



Is the algorithm achieving the documented goals against defined KPIs vs. holdouts?

No

Revisit exploration, solutions, and design phases.

Yes

Have we created a thorough list of subgroups to test?

No

Check with operations, business development, and production to document steps.

Yes

Cut holdout data into different parts for each subgroup.

Is the algorithm achieving the documented KPI goals for each subgroup?

Yes

Check your work as this is highly unlikely.

No

Can we except a lower performance among the specific subgroups affected?

Yes

Move forward to activation, optimization, and remediation phases.

No

Is the underperformance in a given subgroup consistent?

Yes

Create an adjustment factor to compensate such as using a different threshold.

No

Can we split this out and make decisions for this subgroup using a different algorithm for which performance for this subgroup was better?

Yes

Use this algorithm for other subgroups and a separate one for this subgroup.

No

Use this algorithm for other subgroups and return to and earlier phase to reevaluate an algorithm choice for this subgroup.

Inputs Needed

- Training data set holdout as described above.
- Captured data set of data points with full input variables from the real world, for use in lab testing (described above). Must include known dependent variables, for use in determining the accuracy of algorithm prediction.
- Ability to run a test that includes real-world filters and biases that are representative of what an algorithm will face once deployed in the real world, for in the wild testing. An example in programmatic advertising would be a list of auctions won, not bid requests seen.
- A list of subgroups that are important to test: A critical step in minimizing bias is to test an algorithm not only on an overall basis, but also on a subgroup basis. An algorithm may be great at predicting the quality of an applicant for a computer science job position, but it could be inaccurate for one minority of applicants without affecting the overall score dramatically enough for anyone to notice, so you need to document each subgroup you want to test. This concept was discussed as intersectionality in the awareness and discovery stage earlier in this document. For a marketing use case, these subgroups for instance could be:
    - Demographics
    - Usage intensity: regular users vs. infrequent users
    - Lifecycle stage: new or prospective users vs. repeat users
    - Short tail inventory sources vs. long-tail inventory sources
    - Early adopters vs. later adopters (especially important for launching new products, as early optimization algorithms can optimize to early adopters that are not representative of those customers that are needed to “cross the chasm” to mainstream)
    - Before and after a database match: Database matches can lead to significant data loss and data bias, especially for probabilistic matches (such as cross-platform identity graphs). If some scenarios are occurring after a database match, and some before, it’s important to understand how the algorithm is performing for each.

Outputs Delivered

The critical step in determining the output of this stage is designing a set of algorithm tests that gives a well-rounded view into an algorithm’s effectiveness while showing the representativeness of algorithm predictions.

- Overall success criteria: accuracy, precision, recall, area under the curve, etc.
- Success criteria for each subgroup defined in the **Inputs** stage above.
- Comparative process that finds any outliers in performance for any subgroup.
- For traditional metrics such as accuracy, precision, and recall, best practice is to test algorithms at various thresholds, to understand the trade-off between scale and accuracy and hopefully find an inflection point that blends both.

Considerations

A mistake many companies make is measuring algorithms on an overall basis only. Algorithms can be highly ineffective for specific subgroups or minorities, but if those subgroups or minorities are a small percentage of scenarios, it may not affect overall algorithm effectiveness enough to warrant further consideration.

  - As an example, imagine a bidding algorithm trying to predict app download or in-app purchases for a mobile freemium game. Many games like these depend on "whales," a select few players that make a high percentage of in-app purchases. Measuring the algorithm overall will primarily report the effectiveness of the algorithm at predicting the actions of non-whales since non-whales make up most of the consumer population.
  - Thus, it's critical to document subgroups, particularly highly desired ones, even if they are relatively small, and then test algorithms at the subgroup level to ensure the algorithm is effective not just overall, but also for critical subgroups. Refer to the section earlier on intersectionality discussed in the awareness and discovery stage to revisit the importance of this approach across stages of algorithm development.

- Another big error companies make is mistaking in the lab test results for in the wild results. There are countless hurdles an algorithm may face in the real world which don't exist in the lab. It's critical to see how an algorithm performs in a scenario that shadows as much as possible what it'll face upon deployment. Examples are provided above in the **Checklist** section.
- Get involved and stay involved. **Constant human involvement with AI is crucial.** Question all assumptions and compare human decisions to model decisions, digging into any differences or patterns you can find. As a marketer, make sure not to commit too early to a "set-and-forget" automation use case for AI, and instead periodically ensure the algorithm is working the way you want.

Bias Questions/Risk Level Assessment

The key theme at this stage is that bias moves from being a theoretical issue to being a practical issue. As such, minimizing bias at this stage is less about the risk of specific types of bias and the underlying causes thereof, and more about rigorous testing to determine if some bias exists, and the real-world quantitative implications of that bias. If that testing process uncovers that bias exists, and you can isolate the implications of that bias, then you move to diagnose that bias, at which point you dive back into the theoretical. But first you must objectively determine if bias exists. The following question list can help you get there:

- Have you created a thorough list of subgroups to test? To do this, engage with the operations team to document all the sometimes forgotten or overlooked steps that occur between client request and execution, to ensure you're set to test every single potential source of difference between in the lab test results as compared to in the wild test results. The business development team may need to be consulted as well, as they may need to engage with partners to understand steps that may be affecting the process from client request to execution. Many examples of these types of real-world considerations have been suggested above in the **Inputs** and **Considerations** sections, but your business is bound to have unique ones. Have you thought of *everything*?
- Is the algorithm achieving the documented goals according to the KPI defined in earlier stages when testing against the testing data set?

- Is it possible to cut the testing data set by each subgroup defined above? In other words, if your team decides it needs to test whales against less frequent users, are you able to create a testing data set of just whales separate from a testing data set of just less avid users?
- Is the algorithm achieving the documented goals according to the KPI defined in earlier stages for each one of the subgroups defined above? Here's a hint: The answer is probably no. If it looks like yes, check your work.
- If/when you find that certain subgroups are scoring worse than others, then ask yourself does it matter? No algorithm will ever be perfect and bias-free. It pays to be practical. Can you live with less success with certain subgroups? For instance, if your only priority is to find more whales, then maybe if an algorithm has higher recall but lower precision with whales as compared to overall, this is something you can live with.
- If you determine a certain subgroup's performance is below an acceptable level, then how consistent is the underperformance? Does the subgroup always perform the same relative to the overall testing group? Bias is much easier to solve for or adjust if it's consistent.
- If a subgroup's performance is consistently poor, is it possible to adjust the results for this subject and adjust for any bias? For example, perhaps you can use a higher threshold for inventory from some publishers, and lower thresholds for inventory from others.
- If a subgroup's performance is underperforming but it is not consistent and is unpredictable, are there versions of the algorithm that don't have this issue? You have likely iterated with multiple versions of the model. Maybe certain versions of the model work better in certain circumstances, and other versions of the model work better in others, and that can be fine. Maybe some use a variable that has bias in it, and others don't use that variable.
- If a subgroup's performance is underperforming but it is not consistent and is unpredictable, and you don't have any versions of the model that perform acceptably, then are you dealing with an unknown bias that needs to be diagnosed and minimized? In this case, you need to take a step back and return to an earlier stage of algorithm development. Revisit the different types of bias that could lead to this subgroup's underperformance, and reinvestigate if the risk factors associated with each of those types of bias exist in your algorithm process.
- If you do need to revisit earlier stages of algorithm development, can you isolate specific subgroups that might need their own algorithms? It's possible that the work you've done to this point is extremely useful in many scenarios, and you should use it. You may need to parallel path a separate algorithm process for specific scenarios that don't work with this algorithm.
- If you do have all subgroups working at an acceptable level, however you got there (whether by using multiple algorithms, or different thresholds, or accepting underperformance with certain groups), then pat yourself on the back and move forward to full implementation. Never forget, however, that algorithms left alone become less useful. Stay involved and continually improve.

Phase: Activation, Optimization, and Remediation

Primary Stakeholders: Architect and legal and compliance
Secondary Stakeholders: Requestor and consumer

What Happens in This Stage

At this stage you obtain the results of the algorithm and are optimizing its benefits, but your anti-bias testing responsibilities are just starting. Algorithmic output testing is never complete. Rather testing your technology, data set, putting in place risk assessments and mitigation systems, reviewing your data logs and results to assess algorithmic bias, building inhuman oversight, testing the human oversight for biases, and compiling adequate documentation for everything should be an ongoing requirement. What is critical in this read-out phase is to detect and test to minimize unfair and unintended outcomes. The Federal Trade Commission and the European Union have issued guidance requiring businesses that use algorithms to use an algorithm in a manner that does more good than harm, to be transparent about the data uses and outcomes to consumers, and have a high level of robustness, security, and accuracy in any algorithm. Specifically, FTC Commissioner Slaughter noted that algorithms can produce harmful results by “faulty outputs and a failure to test.”[7]

What to avoid at this phase. You need to determine whether the results and the outcomes are producing biased or unexpected consequences. Some of the outcomes would have been explored in the Awareness and Discovery phase, but new, unforeseen ones can be discovered in this phase. While intentions may be to grow share, the result has the potential to cause discrimination or unintended biases. Health care companies may use data to advertise less expensive healthcare benefits to a certain population. In segmenting the data set, if the requestor specs a population with longer life expectancy, fewer illnesses, and higher incomes, the results could unintentionally target advertisements to more limited groups and communities.

Legal and compliance would ask when seeing the test results. Are you depriving a group of participants of a benefit that is contrary to regulations, internal policies, or anti-bias standards? The requestor should assess if this business is losing potential revenue by applying these biases? Also, from a reputational standpoint, how would the public consider these product offerings, meaning others were deprived of the offer? As an example, the data protection officer viewing an algorithm’s use of video testing of content and advertising could raise that the outputs show that this testing could discriminate against persons with disabilities.

The builder and the product lawyer may notice that their algorithm data used for job advertisements unintentionally produced results showing that they were advertising only to younger, less experienced individuals, although the client specifically asked not to produce this result and this result could raise federal and state employment law issues. The data protection officer may also notice that in reviewing the documentation of another AI or algorithm that the builder and requestor changed the data inputs in the development phase and those changes were not included in the privacy notice, so the consumer was not notified how his or her data was being used.

[7] Yale Information Society Project, Y*ale Journal of Law, 7 Technology*, Vol. 23, August 2021

Regulators globally and in the U.S. are paying more attention to unfair or discriminatory outcomes of AI. To reduce this regulatory risk, the remediation, testing, and optimization phase should be consistent and constant.

> ***The requestor, builder, and legal and compliance participants should establish a cadence of review and testing procedures that documents the outcome and results of the AI—as part of an AI governance process.***

What do you want to do here? The requestor, builder, and legal and compliance participants should establish a cadence of review and testing procedures that documents the outcome and results of the AI—as part of an AI governance process. This documentation process should include testing the original goal against the results and testing the data, human oversight and intervention, and the interpretation of the results against bias indicators. Any changes in design, in data, or in maintenance that occur to the algorithm should be tested as well.

These steps can help the platform continue to deliver upon the expectations of the stakeholders. Remember we're building intelligent systems, so ongoing governance is to be expected. Luckily the AI development community has begun to craft tools and approaches to support this process.

Some tools focus on specific tests, while other tools are broader and look at the entire process. Here are a few areas where you can already get support:

- **Model Robustness -** Adversarial approaches to help determine when your models and platforms are at risk either from manipulation or privacy issues.
- **Explainability -** Approaches to help explain to stakeholders what your model is doing and how. These can be distributed as fact sheets to all stakeholders for tracking and continued understanding as the models evolve.
- **Transparency -** Clarity on which features are being used for which decisions can help to root out bias in models. This transparency should extend to the ultimate consumer.
- **Fairness -** Understanding when groups are positioned at a systemic disadvantage will help teams create fairer models.

Organizations are developing resources that view AI governance as a critical step in developing a trusted AI future. IBM Research, for example, has focused on these types of tools for years, while many startups and new players are focusing on developing feature sets that support monitoring models and understanding how they evolve.

Checklist

- □ Review the data logs and outcome results.
- □ Confirm the AI performed according to your established metrics.
- □ Test the data for intended and unintended consequences, including negative regulatory consequences.
- □ Commission third-party or independent audit testing of the above.
- □ Confirm security practices governing sensitive data such as encryption, aggregation, and deidentification.
- □ Confirm security practices governing all data.
- □ Confirm any vendor data or technology inputs and outputs.
- □ Establish a testing documentation process.
- □ Establish a change process and document it.
- □ Establish a feedback system and process.
- □ Confirm the accuracy of a consumer and business partner privacy notice.
- □ Review the human oversight and data decision making processes.

Inputs Needed

- A clear objective for your testing and documentation and what you're trying to achieve.
- An AI anti-bias monitoring procedure.
- Confirm your project's core concept details:
  - Did you reach your target audience? Did you exclude certain segments caused by data, people, or machine bias?
  - What was your outcome and did it have unintended bias?
  - Did you achieve the anti-bias goals that you started with?
- Did you satisfy the regulatory and compliance obligations?
- Confirm your privacy notice provides the data subject with an accurate description of how data is collected, used, shared, transferred, sold, stored, and deleted in the AI.
- Establish what you want the third-party auditor to test and what standards the third party should use.

Outputs Delivered

- Data test report
- Algorithm test report
- Outcome and result report
- Cyclical reporting on project lifecycle
- Third-party audit
- Report as a comparison to documented algorithmic records
- Algorithmic documentation process review certification
- Validity of consumer or customer notice (risk scores, what data was used in the model)

Considerations

Testing AI should be an ongoing process to minimize human error. Too often, AI is deployed without testing, when with testing the unintended bias or harm could be mitigated. Predeployment testing coupled with regular monitoring, testing, and audits can minimize disparate outcomes. Regardless of company size, including third-party audits, preparing testing and results documentation, and having checklists in place is critical to your AI and your company's success.

A recent study found racial bias in AI that was intended to provide access to health care for high-risk patients. The company used lower health care costs to predict increased care needs, when Black patients had higher health care costs typically than white patients. The embedded bias in the AI reduced the number of Black patients offered this health care by 50%. If the outcome of this AI was tested, the bias could have been mitigated.[8]

To minimize such bias, consider asking the following questions:

- Has the data, the data analytics, or technology changed from the design phase?
- Are there are filters, tools, or scoring that are used to alter the results?
- Are the technology and data being used as planned/designed? Have you confirmed this through design and data records (design, training, and testing records should be maintained and applied at this stage)?
- Does the data produce the results requested or are there unintended or unpredictable results (e.g., exploiting children, exclusion and disparate treatment of persons based on a protected classification, or does not achieve the advertising goal)?

[8] Source: Yale Information Society Project, *Yale Journal of Law, 7 Technology*, Vol. 23, August 2021

For years financial institutions' credit and loan advertising has been regulated to prevent historical aggregation and cultural bias. Financial institutions and their marketers often are prohibited from using data such as zip codes, precise geolocation, and city as key data points for a credit advertising audience pool. Why? The use of the data in advertising had been found to affect minority populations' access to credit, as certain zip codes and cities were removed from the applicant pool. Rigorous testing of the outcomes by researchers could have identified this anomaly.

Some questions to ask in this scenario:

- Has the output been reviewed by a human, an unrelated third party, or an auditor?
- Have the data or technology designers changed?
- What is the accuracy, precision, recall, area under the curve, etc.?
- Is there test success criteria for each subgroup? Is there a comparative process that finds any outliers in performance for any subgroup?
- Is your decision fair?
- Did you deny consumers something of value?

Finally, as technology drives our advertising world, we need to consider the potential for unwanted and unintended bias with the use of technology. Increasingly regulators are calling on companies to be transparent and clear about how they use consumers' and prospective employees' and employees' data. Testing your company's AI notices and disclosures should be part of your testing process. Consider:

- Has the organization been transparent with consumers and customers about the data use and that an algorithm is being used?
- Is certain sensitive data used that would require additional testing and controls?
- Has the organization deploying the AI been transparent about the AI's capabilities, use, and results?

Bias Questions/Risk Level Assessment

- Are there input controls to prevent bias?
- What is the accuracy, precision, recall, area under the curve, etc.?
- Is there a monitoring and testing plan to prevent bias?
- Are there test success criteria for each subgroup? Is there a comparative process that finds any outliers in performance for any subgroup?
- Is the outcome or the impact fair?
- Did you deny a consumer something of value? For example, who is paying a higher price or not receiving a benefit?
- Did your data model account for biases?
- How are you auditing your results?
- Have you used a third-party bias tool or assessment?

Conclusion

The current industry-related discourse on bias and AI is heavily grounded in the ideology that AI introduces something that could cause negative results. However, the reality is that bias is a human concept that, when not appropriately addressed, is transmitted into machines and amplified across systems.

Business, product, and design leaders need to understand that bias can emerge due to many factors, including but not limited to the design of the algorithm or the unintended or unanticipated use or decisions relating to the way data is coded and collected, selected, or used to train the algorithm.

> ***Building bias mitigation strategies into our product and campaign life cycles are essential to businesses operating toward privacy, data minimization, equity, and equality.***

Building bias mitigation strategies into our product and campaign life cycles are essential to businesses operating toward privacy, data minimization, equity, and equality. These steps need not be disruptive. The entirety of this guide is intended as a starting point for how organizations craft their own policies and approaches. It's important to note that a checklist, though useful, is not enough. As your organization begins its exploration consider these actions:

Team Collaboration

- Establish who's ultimately accountable for the AI project and its use, as well as key team members' project roles and responsibilities.
- Arrange exploratory meetings with an AI builder and legal and compliance team to:
  - Determine whether AI is appropriate and fit for your purpose, or if its limitations preclude its use.
  - Establish the legal basis for processing personal data and perform a privacy impact assessment.

Core Business Requirements

- Align on the problem statement, entities and variables, solution rules, key metrics, and hypothesis.
- Establish the core details:
  - Who is this project for?
  - What are you studying? Not only the question and problem, but what kind of data that you are proposing to analyze and leverage.
  - Why are you doing this analysis?

- When are you doing this analysis? Is the data or the analysis timebound, or is it continuously refreshed?

- Document your strategy around biases at this stage:
  - Consider if there's unconscious or confirmation bias in the problem you're trying to solve (i.e., "How bad is driving for you?")
  - Determine if there's bias in the human-driven process you're trying to replace or improve upon that you need to adjust for.
  - Consider the intersectionality and diversity of your potential data set.
  - Consider the potentially harmful effects of the output.
- Propose materials and associated risks for technical biases.
- Build proof of concept solution(s).

Business and Technical Assessments

- Analyze third-party technology and output to avoid liability for third party's own biases.
- Review data logs and outcome results.
- Confirm security practices governing all data including sensitive data such as encryption, aggregation, and deidentification.
- Confirm accuracy of consumer and business partner privacy notice.
- Confirm any vendor data or technology inputs/outputs.
- Review human oversight and data decision making processes.
- Review data logs and outcome results for intended and unintended consequences.
- Evaluate if the AI performed per your established metrics.
- Observe PoC solution and deconstruct unforeseen biases.
- Conclude future of solution post PoC.
- Establish a change process and document it.
- Establish a feedback system and process.

Testing Requirements

- Establish a testing documentation process.
- Develop testing metrics and methodology that will be used to determine success of an algorithm.
- Early test in the lab:
  - Initially use a holdout from the training data set (i.e., a % of the initial training data set that is randomly withheld from the training process, and then used after an algorithm has been developed to determine if the algorithm is performing well).
  - Then use a separate data set pulled and stored from the outside world to determine if the algorithm is performing as well outside the training data set as it is inside.
- Continually feed testing results back into the design stage to develop newer and better-performing versions of an algorithm. Update algorithms as often as weekly or at the very least quarterly.
- Test multiple versions of a model that try different combinations of inputs to see which is most effective.
- Establish some data set testing of intended and unintended consequences.
- Test in the wild, equating as much as possible what decision-making scenarios will be like once the algorithm is in a real situation. The point of moving testing to the wild is not to try to eliminate all biases, but rather to help ensure you know before deploying an algorithm how it will perform in the wild, adjusting for whatever biases exist, which is often very different to how it performs in the lab.

Standard Terminologies and Definitions

Bias and the multitudes of ways it can affect our processes, products, and outcomes we seek is complex and varied. We can employ methodologies to reduce bias across the project or product lifecycle.

Whether you are a requestor, builder, end-user, consumer, or in legal and compliance, it is essential to your role in helping to reduce bias that you understand what biases are and how they can occur. Understanding the fundamental concepts and terms can help you identify the risk of unintended outcomes in the work you’re supporting.

Here are definitions of the most common cognitive biases that can affect teams and the development process. Please note that while this list is lengthy, it is not exhaustive.

Learn these terms and continue to explore related terms to deepen your knowledge of bias and how it can impact consumers, brands, and campaign performance.

- Algorithmic Bias
- Algorithmic Bias and Machine Learning
- Anchoring Bias
- Artificial Intelligence
- Availability Heuristic Bias
- Bandwagon Bias
- Bias
- Bias Blind Spot
- Bias Variance Tradeoff
- Confirmation Bias
- Data Bias
- Fairness
- Halo Effect
- Impartiality
- In-Group/Out-Group Bias
- Intersectionality
- Not Invented Here Bias
- Selection Bias
- Self-Interest Bias
- Status Quo Bias
- Unconscious Bias
- User Interaction Bias

Algorithmic Bias

Algorithmic bias represents systematic and repeatable flaws in a machine system that produce unfair consequences, such as privileging one arbitrary group of users over another.

In Depth

Algorithmic bias can be identified in many different platforms and can have repercussions varying from unintentional privacy violations to augmenting social prejudices of race, gender, sexuality, and ethnicity.

Our growing reliance on algorithms can displace human responsibility for the outcomes we develop them to achieve. Bias can enter algorithmic systems due to pre-existing cultural, social, or organizational expectations, from technical limitations, or when used in unanticipated contexts or by audiences not considered in their initial design.

Algorithmic bias can lead to other acute types of bias. Popularity and evaluation biases might emerge when manipulative methods are in place or when inappropriate or disproportionate benchmarks are employed in model evaluation. You could also encounter emergent bias when data analysis leads to algorithms that feed viewpoints from a previous data set or position (or ranking) bias where top-ranked items might affect the algorithm's ability to learn.

Why It's Important

As we seek to expand the ability of algorithms to automate, optimize, and enhance our ability to engage with our audiences, we need to concern ourselves with how unanticipated output and manipulation of data can affect the human experience. These impacts will be amplified by the algorithms we lean on.

Algorithmic Bias and Machine Learning Fairness

Algorithmic bias is the appearance of systematic and repeatable errors in a system that produce unfair outcomes, such as disadvantaging one group of users under the advantage provided to others.

In Depth

Bias is inherent to any decision-making process. The common misconception is that AI can eliminate bias universally. However, the more accurate statement is that implementing AI solutions and algorithmic procedures can offer deeper transparency into a decision-making process, enabling solution architects and stakeholders to minimize or mitigate undesired biases. It is also essential to understand the source of algorithmic bias assumed in the algorithms, where data can be the primary source of the issue.

Why It's Important

Considering the amount of data, decisions, and interactions generated in digital advertising, AI is a natural companion to the evolution and growth of our industry. With this great tool comes the responsibility of understanding how we can introduce biases into our algorithms.

Anchoring Bias

An anchoring bias affects an individual or group's decisions when influenced by a particular reference point, expectation or anchor. Once the anchor is established, every consideration after that is influenced by that anchor, and might differ significantly from a decision not anchored in that belief.

Another way to think of anchoring bias is that fixed expectations equal outcome influenced design.

In Depth

People rely too heavily on the very first piece of information they acquire, which can severely affect the decision they make. This reaction to available data is a common way an anchoring bias can be applied to a system. If the system is built according to the anchoring bias, then all results will be directed towards the set expectation. A system will not learn from the data but rather interpret it into the intended outcome.

Typically, anchoring effects emerge from an individual's beliefs, but studies have shown that anchoring bias can appear in a group setting as well. In this scenario, a possible cause could be the discriminatory way information is transmitted, processed, and aggregated based on each individual's anchored knowledge and belief. Anchoring has been widely studied and identified as a tough bias to mitigate. When a belief system is anchored, the individual has difficulty removing the expectation and tends to reframe future inputs to meet the anchored outcome.

Why It's Important

In marketing, we all seek to deliver an outcome that is in line with the brand's needs, but sometimes the intended result can become the anchor for the decisions we make in building a system, application, or campaign. The consciousness of the impacts of anchoring can help us mitigate it in our planning processes.

Artificial Intelligence

The academic realm of computer science defines artificial intelligence (AI) research as the study of intelligent agents which could be any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals. On a technical level, AI can be characterized as a system's ability to correctly interpret external data, to learn from it, and to use those learnings to achieve specific goals and tasks through flexible adaptation.

In Depth

AI is driven by intended utility functions or goals that can be defined or induced. If AI is programmed for reinforcement learning, goals can be implicitly induced by rewarding some types of behavior or punishing others. Other methods include evolutionary goals by using a fitness function to mutate and preferentially replicate high-scoring systems. Others can define goals implicit in their training data called 'nearest neighbor' systems. AI often revolves around algorithms built on top of each other.

Merriam-Webster defines *algorithm* as a finite sequence of well-defined, computer-implementable instructions to solve a class of problems or to perform a computation. Algorithms are used as specifications for performing calculations, data processing, automated reasoning, and other tasks commonly found within the domain of AI and ML.

Why It's Important

This is the umbrella of edge technologies that will continue to evolve and drive transformative change across the enterprise.

Availability Heuristic Bias

Availability bias is the assumption that the most available and published data is the most applicable and required. If there's a lot of information, for example, news, about a topic, you might be biased to believe it's more important than other observable data. This can affect our interpretation of data, outcomes, or the potential impact of strategies. You might also encounter the terms *availability heuristic* and *availability bias* that mean the same thing.

In Depth

An availability heuristic is considered a cognitive shortcut and occurs when an individual or group relies on immediate, recallable information related to a specific topic, concept, methodology, or decision within their own or collective memory. Pre-processed information, typically more recently acquired, is weighted more heavily than any new information for consideration, causing the individual or group to form biased opinions.

The availability heuristic can provide access to quick and immediate actions, for example, when assessing an outcome based on the observable world around you. When we continue to observe information, it becomes more available in our minds, and the more likely we will be to believe our interpretation of that data. While recollection isn't necessarily flawed, reinforced information tends to stand out for us more than other information, guiding us to make decisions informed by the loudest amount of information, regardless of whether it's correct.

Why It's Important

Within an industry with a healthy number of opinions and media coverage, availability bias might affect your overall marketing or advertising decisions based on what is currently trending in the media. Closer to the organization, you might allow a previously concluded campaign to bias your approach to a new campaign, missing an opportunity to drive better performance or insights.

Bandwagon Bias

Bandwagon bias is a kind of groupthink. It's a cognitive bias that encourages us to accept something because other people think it. As an example, bandwagon bias can make us believe something achievable is improbable because others have failed before us. Bandwagon bias is also known as the bandwagon effect.

In Depth

Bandwagon bias, or the bandwagon effect, is the tendency for people to adopt behaviors, beliefs, or attitudes because others are also doing it. It is a cognitive bias where the opinions of a group or their behaviors can alter due to actions and beliefs rallying. Bandwagon bias is attributed to a psychological phenomenon in which the adoption of opinions and ideas grows based on the number of individuals who have declared the same beliefs or ideas. As the adoption of these beliefs develops, others "hop on the bandwagon," often ignoring any underlying data or evidence to the contrary.

Bandwagon effects can occur because individuals prefer to conform or be accepted by the larger group—this conformity allows them to fit in, even if they initially held different beliefs. Within the context of business, leaders must be aware of the bandwagon effects and position ideas and concepts as things to evaluate and understand instead of theories to adopt.

The reverse bandwagon effect is a cognitive bias that causes people to avoid doing something because they believe others are doing it.

Why It's Important

In creating solutions, the bandwagon effect can greatly affect outcomes both for the business and the consumer of the solution. When we don't catch the effects early, waste and misdirection can derail a project. Businesses should adopt approaches that allow individuals to voice independent opinions to diffuse bandwagon bias.

Bias

Bias, in general, is any disproportionate weight either in favor of or against an idea, thing, or group. Most often, this influence applies in a close-minded, prejudicial, or unfair way. Humans can learn biases or inherit them from experiences, or they can be specific to the conditions an individual experiences. In machine learning, these biases can be transferred from the individual to the machine, potentially amplifying the effects.

The term bias is often used interchangeably with similar concepts like prejudice, partiality, and discrimination.

In Depth

Bias is a broad and varied topic that applies to the way humans think, and how cultures form and emerge in accepted social norms. Bias most often has a negative connotation, but in some cases, bias is essential for a system to operate, providing a designed overcompensation for an intended outcome.

Within AI and machine learning systems, bias usually arises as algorithmic bias or data bias, where the algorithm's training uses biased data. These interpretations are often inherited from humans either defining the outcomes, manipulating the data, or architecting and programming the algorithms.

Algorithmic bias happens when a system repeats and amplifies biased learning and creates unfair outcomes, for example, privileging one group of users over another. These outcomes arise from inputs of designers, architects, or possibly unintended use. Often, decisions related to the way the data is interpreted, coded, collected, or modeled for training can cause biases.

Bias is born out of human decisions and cognitive biases. Any efforts to stem that bias from algorithms will require tools and a healthy process for identification.

Why It's Important

Biases left unchecked can cause cascading effects across the data lifecycle, potentially amplifying the bias through outcomes in organizational interpretation and future strategies and can cause consumer social or economic effects. Continuing to use this data or practice can spread the biased complications to other processes.

Bias Blind Spot

A bias blind spot is recognizing the impact of biases on the judgment of others while failing to see the impact of biases on your own judgment. Blind spots tend to arise before, during, and after the development of a model or strategy—making them difficult to detect.

No individual or team is immune to the effects of bias blind spots.

In Depth

A bias blind spot can generate unintended consequences, potentially amplifying the biases. They can be caused by other biases present in the process or through an individual's need to positively view themselves and their impact on the process. Because of the negative connotation of bias, we tend to want to believe that we are rational, accurate and our approach is free of bias—a blind spot.

When made aware of various biases acting on our opinion, judgments, or assessments, research has shown that we cannot control them. This adds to the bias blind spot in that even if an individual is informed that they are biased, they cannot modify their perception.

Why It's Important

When we fail to explore our own biases, or generally overlook them because we believe as a group we're doing the right thing, we create blind spots in our process. Blind spots can bind themselves into our processes, platforms, and interpretations, greatly skewing outcomes across the lifecycle.

Bias Variance Tradeoff

The bias-variance tradeoff is the tension between the error introduced by the bias and the variance created by submitting more expansive training data. Other ways you might encounter it are as bias-variance dilemma or bias-variance problem.

In Depth

If a model is too simple and has very few parameters, it may have high bias and low variance. By contrast, if a model has a large number of parameters, it will have high variance and low bias. Look to find the right/good balance without overfitting and underfitting the data.

This tradeoff in complexity is why there is a tradeoff between bias and variance. The bias-variance decomposition is a way of examining an algorithm's assumed generalization error involving a particular problem as a sum of bias, variance, and the irreducible error, resulting from noise in the problem itself.

Why It's Important

An algorithm cannot be more complex and less complex at the same time. It is this tradeoff in variance and bias that helps derive the best possible outcome for any model actions.

Confirmation Bias

Confirmation bias is an individual or group's tendency to interpret new data as confirmation of one's existing beliefs. This bias appears when we select the data or information that supports our opinions or beliefs, resulting in an echo chamber, while disregarding any additional data or views that offer contradictory outcomes. You might also see confirmation bias called myside bias or confirmatory bias.

In Depth

Confirmation bias often appears when individuals or groups are seeking a specific desired outcome. This can be further amplified for emotionally charged issues or deeply entrenched beliefs. The tug on our cognitive functions for the desired outcome can skew our interpretation of data, causing us to focus on the results that meet our desired outcome, while disregarding or discrediting data that supports an alternative point of view.

This type of bias can arise in situations like biased search for information, biased interpretation of information, biased memory recall—which are all related to our cognitive need to be confident in our personal beliefs. There are four effects to consider:

- **Attitude polarization:** When parties have the same data or evidence available but the debate becomes more extreme.
- **Belief perseverance:** When a biased belief persists despite the evidence being proven false.
- **Irrational primacy effect:** When we refer back to earlier evidence even though new data is available.
- **Illusory correlation:** A false perceived correlation between two events.

Why It's Important

Suppose we as marketers lean too heavily on our own beliefs and expectations, for example, on a specific campaign outcome. In that case, we may fail to realize some new insight or opportunities because they do not conform to our expectations. Today, we may be doing this as the signals and data that we consume continuously change around us, missing opportunities to evolve with our consumers and their needs and wants.

Data Bias

A data bias is any bias that might be present in a data set or prescribed upon that data set based on the views of the human interpreting or building models. Often the data biases rely on human interpretations of cultural, historical, temporal, or aggregation-based views.

In Depth

Data biases can be difficult to identify because they are often ingrained in either the processes and procedures for organizing data or our strategies for understanding the data. Here are a few examples of how data can contain bias:

- Cultural bias might be data that lacks diversity based on characteristics like gender, race, age, education, geography, language, or value systems like religion.
- Historical bias occurs when data is based on past occurrences and has no reflection upon the current, real-world observations.
- Aggregation bias happens when an incorrect conclusion is made about a subgroup due to false assumptions about the overall population.
- Temporal bias occurs when using data where the population or behaviors change over time.

Why It's Important

Our marketing and advertising performance is heavily dependent on data, but biases that exist within the data or our interpretation of it can greatly influence outcomes.

Fairness

Within the practice of machine learning, an algorithm is considered fair (or has fairness) when the intended outcomes do not depend upon or correlate with independent variables, especially those of a sensitive nature. Attributes that might cause an imbalance in fairness could include gender, ethnicity, sexual orientation, income, or other characteristics that could cause a group to be unequally considered, despite features that lack relevance to the outcome.

Fairness is sometimes also related to the terms equality, equity, nondiscriminatory, nonpartisan, and group fairness.

In Depth

Fairness is applied to machine learning when observing the ways a model might perform differently for distinct groups or individuals inside a set of data. An algorithm is seen as unfair when its outcomes provide some benefit, value, or advantage to a protected group (or individual) within the data, while an unprotected group receives none.

In an ideal situation, an algorithm would be customized to an individual and fairness-related criteria would be based on the algorithm providing similar outcomes for individuals that share similar characteristics. Many countries and industries seek to address fairness in terms of demographic groupings like race, gender, or socioeconomic status. This relates to how advertisers typically attempt to group audiences into groups based on similar traits so that we can deliver relevant messaging.

Fairness is not a universal solution. There is a level of modulation to be considered depending on composure of the source data, the attributes present, and the intended outcomes. Apply a fairness lens to the use of data and algorithms, even from campaign to campaign.

Why It’s Important

Whenever systems are making decisions (for example, which ads to expose to a consumer, how much money to optimize towards a certain portion of a segment, or the appropriate characteristics on which to model a new audience), it is crucial to evaluate the fairness of those uses.

Halo Effect

The halo effect occurs when positive opinions of one facet of a person, company, brand, or product bias the opinion of other aspects positively. You might also encounter the terms halo error, halo bias, or halo effect bias.

In Depth

A cognitive bias, the halo effect, can hinder the ability to accept a belief, data, or outcome biased with the idea by a baseless view on what is good or bad. This unconscious behavior surfaces when one positive metric causes you to conclude the entire campaign is successful.

The halo effect is commonly observed in marketing and advertising. Sometimes it is a tactic for product promotion or a methodology within attribution. For example, auto brands often use high-end concept cars for their halo effect on other models within the same make.

The opposite of the halo effect is the horn effect.

Why It's Important

While the halo effect can positively influence outcomes within marketing tactics, biased interpretations of data can also cause systemic failure in optimization or future strategy planning.

Impartiality

Impartiality is the opposite of bias and occurs in a system where decisions are based on objective criteria. Impartiality minimizes the existence of characteristics that might be considered biased and focuses on only the elements that are important to the decision.

Impartiality is also referred to as fair-mindedness or evenhandedness.

In Depth

If you are partial to one thing, you invite bias into a system. Impartiality in contrast does not consider any characteristics that might lead to a biased decision. Instead, an impartial system will focus on features that are present across the data set and evaluate them without respect to any characteristics like gender, ethnicity, or social standing. As an example, a churn prediction model might look at historical consumption patterns and willingness to renew a subscription, but will not consider economic situations based on location, age, or ethnicity.

Why It's Important

The goal should be systems that are impartial and objective, however achieving this at scale is difficult. Strive to be impartial by considering all the sources of data and decisions that can influence a system.

In-Group/Out-Group Bias

In-group bias is a social psychology pattern where a group supports its own members' beliefs, ideals, and actions over out-group members. In-group biases influence decisions and can be amplified within an algorithm to a broader out-group.

You might also encounter the terms in-group favoritism, intergroup bias, or in-group preference.

In Depth

In-group bias is believed to be a cultural phenomenon often arising out of groups that live or work together and develop a shared sense of interpretation of the world around them. These might include moral or cultural norms evolving within the group. In-group bias could also include trait-based effects where an individual might have an in-group bias for the needs of their own gender, ethnicity, or generation.

Why It's Important

Group-based cognitive biases can be amplified by machine learning algorithms and are harder to detect because of their wider-ranging control of the group's views.

Intersectionality

Intersectionality is the interconnected nature of characterizations such as race, class, and gender as they apply to a given individual or group. These intersections can create systemic overlapping and interdependent systems of discrimination or disadvantage.

Intersectionality has also been referred to as intersectional bias.

In Depth

Intersectionality focuses on how various aspects of an individual's identity can create different modes of discrimination and privilege when combined or overlapping with others. Interconnected signals affecting bias could include gender, race, age, sexual orientation, religion, ethnicity, disability, and the outcomes of their intersection might be empowering or oppressive.

These signals could be employed in algorithms where their intersection causes disparate impacts between groups—causing adverse outcomes.

Why It's Important

While many systems and campaigns are designed to limit the number of interacting signals, there is a possibility that two adjacent systems like bid-optimization and creative optimization could be introducing different characteristics that might have intersecting effects. Data sets and a lack of diverse engineers could also lead to AI systems that cause adverse bias.

Not Invented Here Bias

A not invented here (NIH) bias is typically a group-think effect that avoids information, data, or outcomes from outside sources. NIH can result from an unwillingness to accept competitive groups' data or outcomes because their results do not align with the group's requirements.

You might also encounter the terms not invented here syndrome or NIH.

In Depth

Not invented here bias can occur for several reasons. Some examples include:

- Building a solution instead of partnering to avoid partnership fees.
- A lack of interest in valuing the work or research of others because they might show differing outcomes.
- Concerns around IP and patent infringement.
- A type of tribalism that avoids others efforts because they're external to a group.

Why It's Important

Within advertising and marketing, often organizations will dismiss results, methodologies, or applications because they're external to the organization's way of working or are merely competitive. The consequences can be far reaching from unnecessary loss of revenue (due to re-engineering an application for example) to widely inconsistent metrics and expectations.

Selection Bias

Selection bias can be introduced when selecting individuals, groups, or data for analysis that does not allow for appropriate randomization. Improper methods can mean the obtained sample does not represent the intended population for the study.

You might also encounter the terms selection effect or sampling bias.

In Depth

Sampling bias is a systematic error that arises when non-random samples of a population or data set cause some members to be less likely to be included than others. These types of bias can occur when we self-select data, pre-screen it for specific characteristics or apply limiting filters, for example, the amount of time a subject has been part of the data collection. Here are a few key areas where we need to consider the effects of selection bias.

Single or selective data sources can be problematic as they might not reflect the population. A limited view can lead to selecting individuals that best represent the problem but not necessarily the actual audience. The methodology for sampling is also essential. Probability sampling will eliminate voluntary response bias and guard against under-coverage bias. Increasing the sample size reduces any errors, allowing for greater population representation, but increased size will not solve discrimination in the methodology.

In addition to the selection bias, measurement bias can occur when we mismatch and measure against select features across the sample. When these features are similar but have misconstrued relationships, they can cause unexpected outcomes. Similarly, omission bias can be present when a variable is left out of the model. This might further instill biased results without full consideration of all the possible inputs.

Why It's Important

Data is the cornerstone of any work we will do with AI. Still, our cognitive biases and approaches to acquiring and using that data can manifest in ill-conceived models and unintended outcomes. A healthy approach to sample data acquisition can ensure that your results reflect the reality you're attempting to address.

Self-Interest Bias

When we observe positive events and successes through our character or actions but determine that negative results occurred because of external factors, this is called a self-interest bias.

You might also encounter this as self-serving bias.

In-Depth

Self-interest bias is a cognitive or perceptual bias that is twisted by any individual or group's need to maintain a positive or overly favorable self-image by valuing successful outcomes as originating with the individual or group. By contrast, poor results are observed as arising from outside of the individual or group's control.

In marketing, this happens when reviewing campaign results where the positive performance is seen as a direct result of actions taken by the agency or brand. Adverse outcomes are caused by the consumer or outside forces and are not directly related to activities or circumstances controlled by the brand or agency.

Why It's Important

An organization or individual creating a solution grounded in a self-interest-based strategy might find themselves in a loop where they're continually looking to confirm (confirmation bias) their own self-serving biases.

Status Quo Bias

When we refer to the current situation as the norm and deviation from that norm is considered a failure, we employ the emotion-based status quo bias.

In Depth

The status quo bias is a cognitive bias that involves individuals or groups preferring that conditions stay as they are. This bias can affect human behavior and decisions that might influence how a system is encoded to make decisions. If all decisions are focused on maintaining a status quo, then the system could be blinded from other possibilities.

While status quo bias refers to our desire to maintain the current state, it does not mean that we won’t act against things that might affect the status quo. It is not inaction but the action to maintain the current state.

Why It’s Important

Generally speaking, it’s easy to look at the current state of success and drive all strategies and outcomes towards maintaining that state. If we only focus on maintaining the status quo, we might miss the opportunity to enhance it, or miss something that could be a threat.

Unconscious Bias

An unconscious bias is when prejudice or unsupported judgment favors a thing, person, or group over others, grounded by an experience or learned associations. An unconscious bias is often undetectable by the party employing it because it has become ingrained in their conceptualization of a situation.

Unconscious bias is often referred to as implicit bias.

In Depth

Unconscious biases can be shaped with experience and learned connections between distinct groups and social characteristics, including race, gender, and income. An unconscious bias can also be an aspect of inherent social cognition: the phenomenon that beliefs, attitudes, and stereotypes are employed before conscious intention. Individuals will likely be unaware they hold an unconscious bias, and their pre-existing perceptions can influence their choices and behaviors without their knowledge. By contrast, an explicit or conscious bias is intentional.

Unconscious biases are thought to be the outcome of connections acquired through prior experiences. They could be activated by the situation and work before a person's conscious intent.

Why It's Important

Decisions are cultivated and enriched by the observations and learnings that we make as marketers and advertisers. They can be difficult to identify and have far reaching consequences. When unconscious biases find their way into our cognitive processes the results can creep into the strategies we establish and the systems we build.

User Interaction Bias

When the user is present in the equation, the functionality, interface, and the user can all be sources of bias in the system. The evaluation of user interaction bias should consider how user input, output, and feedback loops are designed, presented, and managed.

In Depth

User interaction bias can produce new or updated data, if the functionality, interface, or the users themselves are biased, then new data added will contain further bias and continue to scale. Beyond the specifics of an application or interface, here are a few broad areas of consideration where interaction bias might surface:

- Social bias can occur when people’s actions or content produce outcomes that can affect judgments.
- Presentation bias arises through the potential for differences in the presentation of an interface or content between two different groups, especially when feedback is sought to improve the system.
- Linking bias can misrepresent users’ actual behaviors by misusing network attributes about their related or shared actions.
- Behavioral biases can arise from user behavior across platforms, contexts, or their appearance in different data sets.
- Production bias can originate from user-generated content inequalities across a population, often affected by varied structural, lexical, semantic, and syntactic differences.

Why It’s Important

Most platforms and solutions revolve around the humans they are seeking to understand and reach. When their engagement with platforms is considered, we are opening the door for their interactions to influence how our AI might grow, optimize, and improve. Consider how bias might occur within this relationship and plan for its effects.

About IAB and the IAB Programmatic+Data Center

IAB empowers the media and marketing industries to thrive in the digital economy. Its membership comprises more than 650 leading media and technology companies that are responsible for selling, delivering, and optimizing digital advertising or marketing campaigns. The trade group fields critical research on interactive advertising, while also educating brands, agencies, and the wider business community on the importance of digital marketing. In affiliation with the IAB Tech Lab, it develops technical standards and best practices. Founded in 1996, IAB is headquartered in New York.

PROGRAMMATIC+
DATA CENTER

Founded to enhance existing IAB resources and to drive the "data agenda" for the digital media, marketing, and advertising industry, the Programmatic+Data Center (PDC) defines boundaries, reduces friction, and increases value along the data chain, for consumers, marketers, and the ecosystem that supports them. The PDC's work is to support the direct brand economy, drive accelerated digital marketing transformation through emerging technologies, advance programmatic growth while supporting media buying for emerging formats, lead industry consumer privacy and ethics initiatives, and define data transparency, quality, and identity to inform measurement and attribution within the supply chain. For more information or to get involved, please contact data@iab.com.

About the IAB AI Standards Working Group

Artificial Intelligence and Machine Learning business activities are used in a multitude of new and exciting ways impacting data-driven decision making. The IAB Programmatic + Data Center of Excellence has formed a working group to develop industry standards, guidelines, and best practices to ensure proper application of these techniques. The Working Group will define digital media industry approaches to business use cases and an advocacy plan to continually evolve and help guide future implementations of AI and ML.

The IAB AI Standards Working Group is open to IAB members. If you are interested in participating, please email datas@iab.com to join the working group.

Contact Information

Orchid Richardson
Senior Vice President, Programmatic+Data Center
orchid@iab.com

Angelina Eng
Vice President, Measurement and Attribution, IAB
angelina@iab.com

IAB Media Contacts
Kate Tumino/Britany Tibaldi
212-896-1252/347-487-6794
ktumino@kcsa.com/btibaldi@kcsa.com