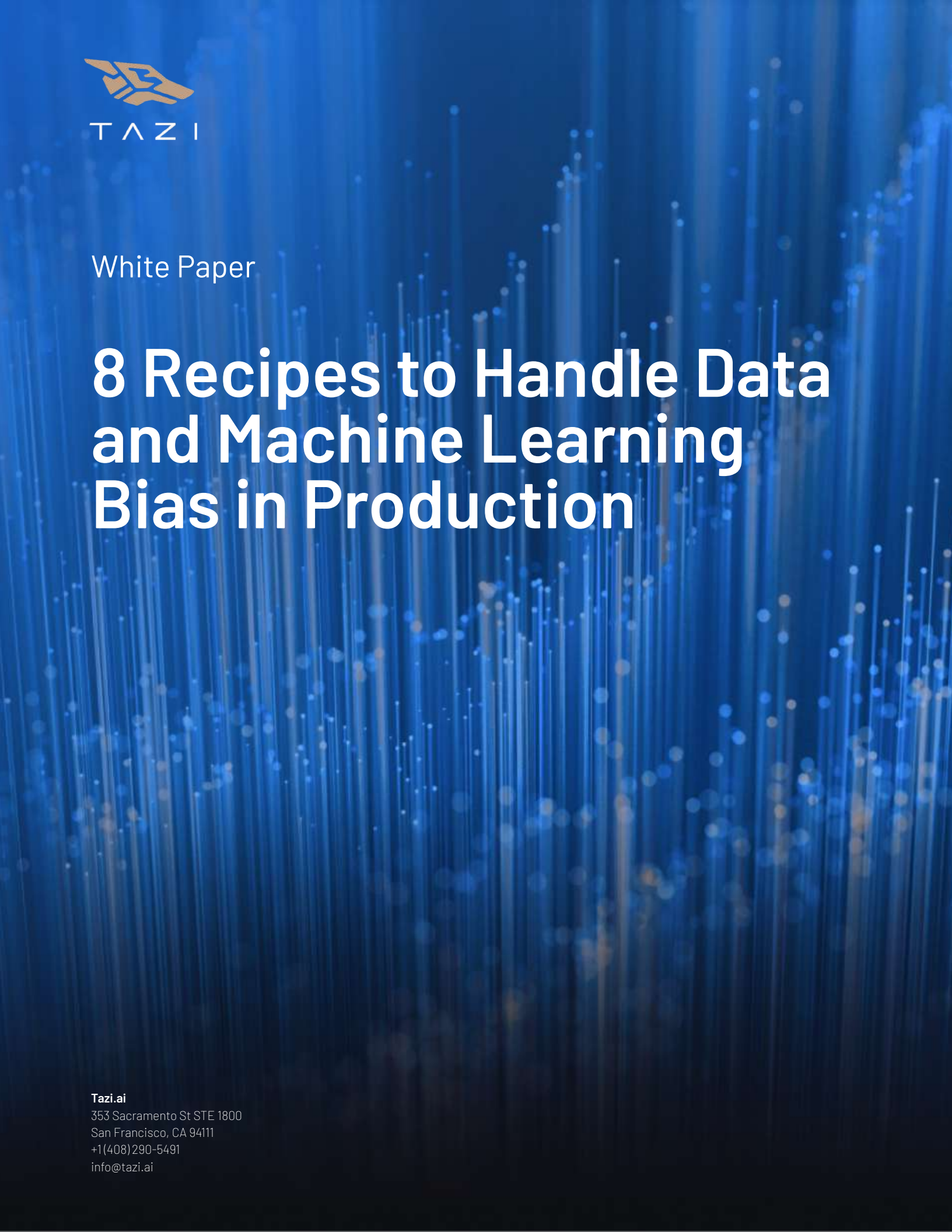

TAZI

White Paper

# 8 Recipes to Handle Data and Machine Learning Bias in Production

**Tazi.ai**
353 Sacramento St STE 1800
San Francisco, CA 94111
+1(408)290-5491
info@tazi.ai

## SUMMARY

Bias, whether conscious or unconscious, can lead to discrimination and unfairness. Detecting and addressing bias in datasets and machine learning (ML) models is crucial. Accountability and accessibility play key roles in handling bias, with audit logs and user-friendly interfaces aiding in the process. Data-related bias can be detected by re-sampling and considering feature relevances. Correlated features can be identified through clustering or building ML models for each sensitive feature. Model-related bias can be addressed by assigning class weights and using model explanations. Surrogate models and streamlined actions facilitate prompt resolution of bias. Diverse teams and AI-enabled processes enhance bias detection and prevention.



**TAZI.AI**
353 Sacramento St STE 1800
San Francisco, CA 94111
+1(408)290-5491
info@tazi.ai

## INTRODUCTION

Bias refers to a conscious or unconscious preference towards a particular group, often to the exclusion of and unfairness towards others. For people with certain racial, ethnic, gender, religious groups, and disabilities, bias results in discrimination and systemic barriers to opportunities and success. Data created in a biased world is inherently biased. Creating and deploying machine learning (ML) models always come with a significant risk of bias. ML solution environments should provide human-usable explanations to detect and remedy bias.

Accountability and accessibility are essential in handling bias. Accountability is needed to make sure that whoever notices bias does something about it. The accessibility of ML systems in production allows bias to be handled in a timely manner. Accountability can be partially addressed by audit logs. Lowering the entry barrier to ML by automatization and easy-to-use UI/UX can help with accessibility.

We explain below how data-related bias and model-related bias can be detected and handled via systematic explanations of data and ML models. Data-related bias is defined as the bias that already exists in the dataset. For example, in a customer churn prediction use-case, 90% of the dataset could contain white customers, leading to race bias in the dataset. Model-related bias is defined as the bias that is produced within the model. In this case, since white people make up 90% of the population, the model which aims to minimize the error, would predict churn better for the whites, resulting in race bias in the model. Using this model to take actions to prevent churn would target the white population and under-serve the others.

## PROBLEM STATEMENT

### Data Bias Detection and Handling

The first and more common type of data-related bias happens when some variable values occur more frequently than others in a dataset (representation bias). For example, for a clinical trial, 90% of the participants could be males.

Recipe 1: The representation bias can be partially handled by re-sampling [1] the data to represent different groups equally. However, when there is less

**TAZI.AI**
353 Sacramento St STE 1800
San Francisco, CA 94111
+1(408)290-5491
info@tazi.ai

information and details for the underrepresented groups, the ML model may be learning them less.

Data-related bias also occurs when there are highly correlated variables with the target feature. Recipe 2: In order to detect bias according to certain sensitive features, the feature relevances, i.e. the correlation [2] of each column with respect to the target feature, can be calculated. The user can ignore the highly relevant sensitive features, such as gender or age, that might lead to bias.

Even when the sensitive and relevant features are removed, there may be other features correlated with those sensitive features. For example, zip code may be highly correlated with race, even if race is removed from model building, keeping the zip code may still cause biased models. Recipe 3: Clustering or grouping variables based on their correlations with each other may help detect and remove such correlated features. Another way to detect complex data bias is through the creation of an ML model for each sensitive feature. The features that contribute most to the prediction of the sensitive features should be ignored in the ML models.

**Machine Learning Model Bias Detection and Handling**

For model-related bias, consider both the inputs to the ML model and the output predictions of the model. When the dataset is unbalanced, sensitive features might be too relevant to the target feature and cause bias. Recipe 5: Some ML platforms assign automated class weights during model building to emphasize the underrepresented classes.

Machine learning model explanations also help with the detection and prevention of model-related bias. Recipe 6: There are local or global feature importances, providing information on how each feature's value affects the model outcome [3][4]. For example, if increased age results in lower credit score predictions, then the model has age-related bias. However, it is difficult to determine exactly where in the model the bias is. Recipe 7: Use easily interpreted surrogate model explanations, such as linear models or decision trees. Surrogate models approximate and explain the underlying ML model used for decision-making. They allow more granular detection of bias. A decision tree surrogate model contains automatically generated micro-segments of model prediction, each resembling a rule, e.g. "if the Agency_type is silver and gender is male then the customer will Churn."

**TAZI.AI**
353 Sacramento St STE 1800
San Francisco, CA 94111
+1(408)290-5491
info@tazi.ai

Recipe 8: When bias (or any other problem) is detected on an ML model, the ease and speed of actions determine how fast it is resolved. Ease of creating and sharing data and model explanations, model building, update, deployment, and monitoring determine whether a larger set of users (instead of just IT or data scientists) can take action.

## CONCLUSION

Systematic detection and prevention of bias in data and machine learning models is possible. Hiring users from diverse backgrounds and AI-enabling them, allow not only better detection and prevention of bias but also remedies when bias detection systems or ML models themselves fail or are even hacked [5].

**TAZI.AI**
353 Sacramento St STE 1800
San Francisco, CA 94111
+1 (408) 290-5491
info@tazi.ai

# ABOUT TAZI

Artificial intelligence (AI) is a source of both huge excitement and apprehension, transforming enterprise operations today. It is more intelligent as it unlocks new sources of value creation and becomes a critical driver of competitive advantage by helping companies achieve new levels of performance at greater scale, growth, and speed than ever before, making it the biggest commercial opportunity in today's fast-changing economy.

TAZI is a leading global Automated Machine Learning product/solutions provider with offices in San Francisco. TAZI is a Gartner Cool Vendor in Core AI Technologies (May 2019) and is considered as "The Next Generation of Automated Machine Learning" by Data Science Central.

## WHO WE ARE

Founded in 2015, TAZI has a single mission which is to help businesses directly benefit from Automated Machine Learning by using TAZI as a superpower, shaping the future of their organizations while realizing direct benefits like cost reduction, increasing efficiency, enhanced (dynamic) business insight, new business (uncovered), and business automation.

## WHAT WE OFFER

Through its understandable continuous machine learning from data and humans, TAZI is supporting companies in the banking, insurance, retail, and telco industries in making smarter, more intelligent business decisions.

**TAZI.AI**
353 Sacramento St STE 1800
San Francisco, CA 94111
+1(408)290-5491
info@tazi.ai

TAZI solutions are based on a most compelling architecture that combines the experiences of 23 patents granted in AI and real-time systems, proven at different global implementations.

Some unique differentiators of TAZI products are:

- Business users can automatically configure custom ML models based on their KPI and the available data. TAZI's Profiler accelerates this process through data understanding and automated cleaning, feature transformation, engineering, and selection capabilities.
- TAZI models learn continuously and are suitable for today's dynamic, real-time data environments.
- TAZI models are GDPR compliant (no black-box models). They provide an
- explanation in the business domain's terminology for every result they produce.
- TAZI supports multiple (heterogeneous) data sources, i.e.: external, batch, streaming, and others.
- TAZI can learn both from human domain experts and from data, which speeds up accuracy improvement.
- TAZI's hyperparameter optimization feature reduces human time spent on model configuration. TAZI products contain algorithms that are developed and coded to be lean, efficient, and scalable.

**TAZI.AI**
353 Sacramento St STE 1800
San Francisco, CA 94111
+1(408)290-5491
info@tazi.ai