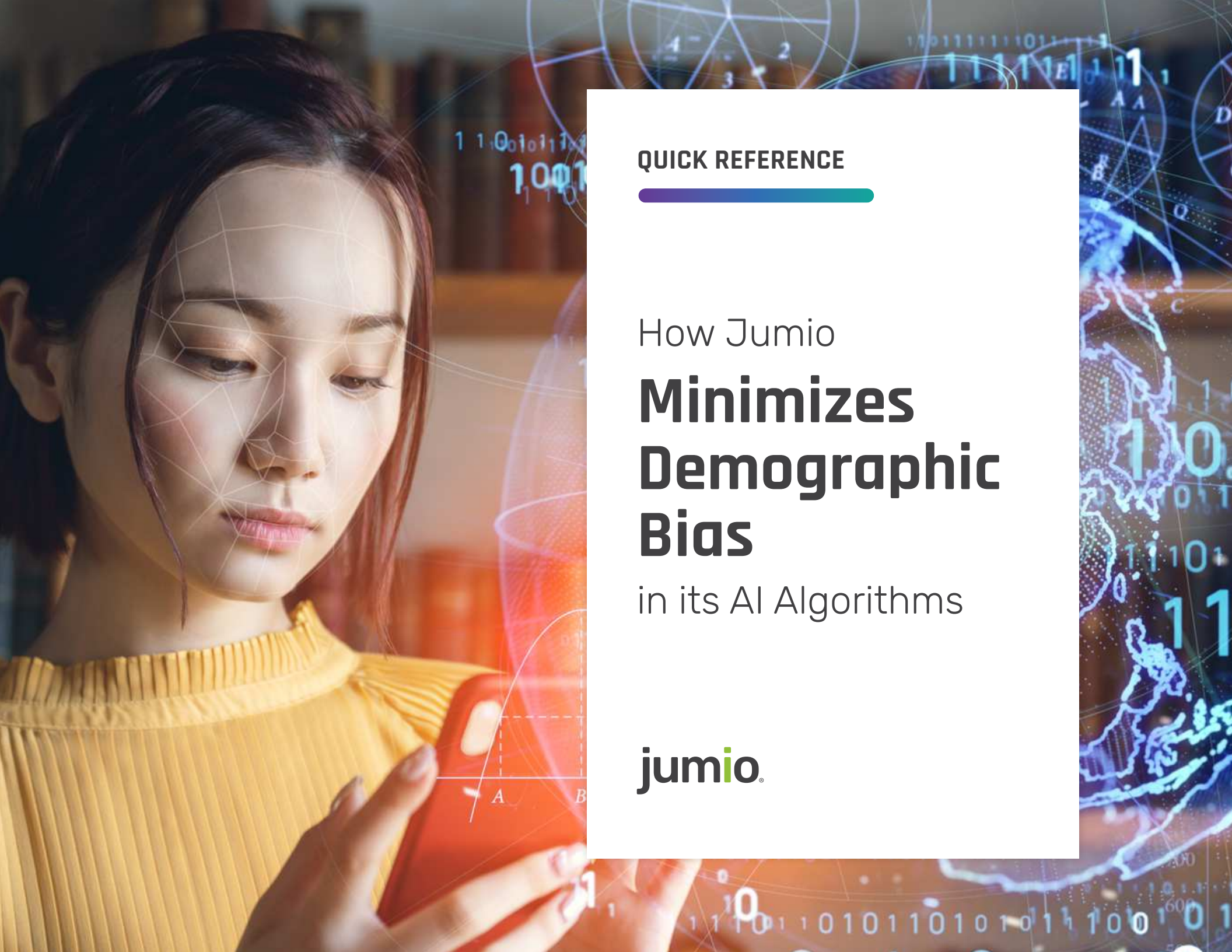


QUICK REFERENCE

How Jumio

Minimizes Demographic Bias

in its AI Algorithms

jumio

By 2022, more than

95%

of RFPs for document-centric identity proofing will contain clear requirements regarding minimizing demographic bias, an increase from fewer than 15% today.

Gartner

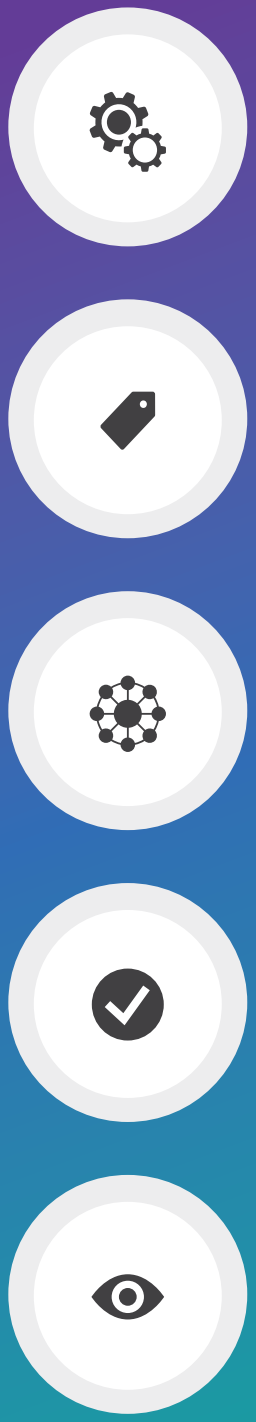
Background

More and more identity verification providers are leveraging AI to determine if an ID document is authentic, to compare the selfie of a customer with the photo in their identity document and even to determine if the person creating a new online account is physically present. When AI is being used to this extent, it's not surprising that there will be concerns about possible bias in the facial recognition process, especially when economic decisions are dependent upon the accuracy and reliability of those algorithms.

How AI is Used to Make Verification Decisions

Before exploring how to reduce demographic bias, it's important to understand how AI is used in the identity verification process. At Jumio, we use AI to create deep learning models in order to improve identity verification accuracy, identify fraudulent IDs and make the verification process faster. AI has already been productionalized to reduce the time it takes to verify an ID document or an online identity which, in turn, is helping Jumio's customers reduce their abandonment rates and increase new account conversions.

Here are the basic steps:



Data

Machine learning depends heavily on data — without data, it is impossible for an AI algorithm to learn. For identity verification, that starts with a database of ID documents and corroborating selfies. Ideally, the database contains a large and representative sample of IDs, including driver's licenses, passports and ID cards from a large sample of countries and territories.

Labeling

In most AI projects, classifying and labeling data sets takes a fair amount of time, especially with enough accuracy and granularity to meet the expectations of the market. In the context of identity verification, labeling is how the ID documents are tagged. If the photo of the ID has been manipulated, then the document will be tagged as fraudulent with photo manipulation. If the picture of the ID has excessive glare, blur or was captured in poor lighting, then the labels should reflect those characteristics.

Modeling

Model development and selection is the process of researching, designing, training, evaluating and choosing a model. There may be many competing concerns when performing model selection beyond model performance, such as complexity, maintainability and available resources.

Validation

Model validation is the process where a trained model is evaluated with a testing data set. The testing data set is a separate portion of the same data set from which the training set is derived. The purpose of validation is to find and optimize the best model with the highest predictive power in the real world (i.e., in order to generalize results beyond the training set).

Monitoring

Without careful monitoring, your machine learning module can return unexpected output. Once models are put into production things can (and will) go wrong, so it's imperative to monitor the performance of our models in production in order to detect anomalies and appropriately course-correct.

How Jumio Minimizes Demographic Bias

Bias can creep into algorithms in several ways. AI systems learn to make decisions based on training data, which can include biased human decisions or reflect historical or social inequities, even if sensitive variables such as gender, race or sexual orientation are removed. Here are some of the steps that Jumio is taking to minimize demographic bias in our AI algorithms.

LARGE & REPRESENTATIVE DATA SETS

Jumio has verified over 400 million identity verifications and leverages massive and global datasets to help create and refine our AI models. This volume helps reduce bias, but size isn't everything. To minimize bias, it's critical to have a database that is also diverse and representative. Think about a face detection model that is trained on a large dataset of faces from a single ethnicity. It will most likely fail to detect faces from another ethnicity.

PRODUCTION & REALISTIC DATA

Jumio trains its AI models on real-world production data, not purchased data sets. A dataset of images of documents under perfect lighting conditions with high-resolution cameras is not representative of ID images that are captured in the real world.

Not surprisingly, AI models built on unrealistic models will struggle with IDs that contain blur or glare or were captured in dim lighting. Consequently, these AI algorithms will be less robust and susceptible to more bias when the environmental conditions are not optimal. This is why Jumio's algorithms are built on the back of real-world production data which contain real-world imperfections. As a result, our AI models are more robust and less susceptible to demographic bias.

QUALITY CONTROL & GOVERNANCE

With AI, garbage in results in garbage out. If the wrong labels are used when tagging individual identity verification transactions, the AI models will bake that information into the algorithms which will make the models less accurate. Jumio relies on our experienced verification experts to properly tag ID images and we have built in the quality controls which govern how ID images are properly tagged. Many identity verification solution providers that are new to the space do not have the luxury of experienced agents who have tagged historical transactions and have purchased off-the-shelf datasets which have been tagged improperly and without these controls in place.

Jumio's Embedded Diversity

Reducing bias is also about the people who are developing the AI algorithms and tagging the datasets. Jumio employs a diverse team of verification agents (located around the globe) and a diverse team of AI engineers from a variety of nationalities, genders, ethnicities, professional experiences and academic backgrounds. This diversity helps us examine problems from different perspectives, which also helps reduce some demographic bias.

The AI-Powered Identity Verification Leader

BIG DATA

Jumio has verified over 400 million identities, which means we have deep experience reviewing large volumes of government-issued IDs from across the globe. This is key to training robust algorithms and accurately extracting and structuring the data, as well as spotting manipulation.

INFORMED AI

To better inform the algorithms, there needs to be a continuous feedback loop to incorporate the judgment of our verification experts. Jumio leverages verification experts to review and refine our algorithms which speeds up the learning curve.

AI ENGINEERS

Jumio created AI Labs to boost the development of machine learning models, develop best practices and to continue as a platform of innovation.

EXPERIENCE

Jumio has been developing our AI models by intelligently tagging valid and invalid ID documents over the last five years. This experience and our global scale enable us to more accurately recognize fraudulent IDs in seconds.

jumio®

Learn more at jumio.com