How Jumio

Minimizes Demographic Bias

in its AI Algorithms

jumio.

Background

More and more identity verification providers are leveraging artificial intelligence (AI) to determine if an identification document (ID) is authentic, to compare the selfie of a customer with the photo on their ID 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 greatly reduces the time it takes to verify an ID or an online identity which, in turn, helps Jumio's customers reduce their abandonment rates and increase new account conversions.

Here are the basic steps:

DATA → LABELING → MODELING → VALIDATION → MONITORING
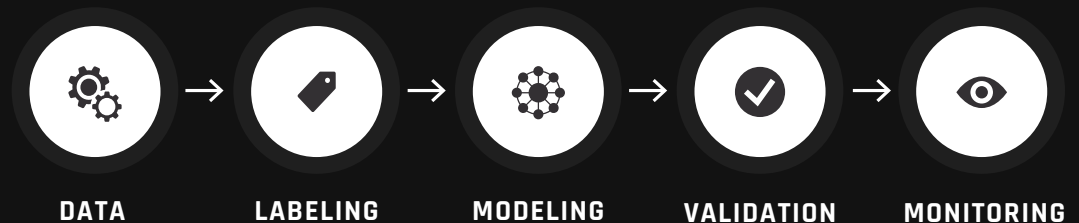
Data

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 IDs 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 the context of identity verification, labeling is how the IDs are tagged. If the photo of the ID has been manipulated, 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, the labels should reflect those characteristics. 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.

Modeling

Models are the algorithms in AI that replace the human decision-making process. Model development and selection is the process of researching, designing, training, evaluating and choosing a model. There may be many competing concerns when selecting a model, such as performance, 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 was 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

Once models are put into production, things can (and will) go wrong, so it's imperative to monitor the performance of models in order to detect anomalies and appropriately course-correct. Without careful monitoring, you can end up with unexpected output.

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 processed over 1 billion transactions and leverages massive and global data sets 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. For example, a face detection model that is trained on a large data set of faces from a single ethnicity will most likely fail to detect faces from another ethnicity. Jumio's accuracy is nearly equal across geographies, as the model is trained on well-distributed global Jumio data with different genders, ages and ethnicities.

REAL-WORLD PRODUCTION DATA

Jumio trains its AI models on real-world production data, not purchased data sets. A data set of images of documents under perfect lighting conditions with high-resolution cameras is not representative of the ID images that are captured in the real world. Not surprisingly, AI models built on unrealistic sample data will struggle with IDs that contain blur or glare or were captured in dim lighting. Consequently, those AI algorithms will be less robust and susceptible to more bias when the environmental conditions are not optimal, as systems that cannot handle images from low-quality cameras discriminate against people from less-privileged socioeconomic classes. This is why Jumio's algorithms are built on real-world production data that contains real-world imperfections. As a result, our AI models are more robust and less susceptible to demographic bias.

QUALITY CONTROL & GOVERNANCE

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 highly experienced verification experts to properly tag ID images, and we have built in quality controls to maximize accuracy. Many identity verification solution providers do not have experienced agents and simply purchase sample data sets that 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 data sets. Jumio has verification agents 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 demographic bias.

The AI-Powered Identity Verification Leader

BIG DATA

Jumio has processed over 1 billion transactions, 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 IDs since 2015. This experience and our global scale enable us to more accurately recognize fraudulent IDs in seconds.

jumio®

Learn more at jumio.com