MedOrion

How to avoid Bias in **Healthcare Insurance AI**

know
why.
know
how.

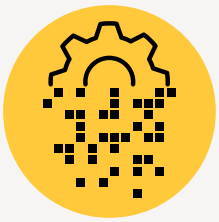
Intro

Artificial intelligence is now used in almost every industry, harnessing the proliferation of data to streamline services, offer new business insights, and increase revenue. The healthcare industry have been early adopters of AI technology with clinician decision support systems integrated into numerous areas of clinical practice. Today payers and providers are looking to AI to create new healthcare pathways, improve member experience, and improve overall outcomes.

Today many healthcare systems are grappling with the challenge of creating fair and unbiased access to services. As AI plays an increased role in healthcare provision, it is important to address potential 'flaws' in the systems, most notably the potential for AI algorithms to become biased against specific subgroups within a target audience. Join us to learn more about how AI can become biased and what practical measures can be implemented to combat this challenge.

84%

Of business executives think AI will help them gain a competitive edge



75%

Of executives expect AI adoption to transform their company in the next 3 years

$327.5B

Global AI market in 2021

Potential of AI

Artificial intelligence is changing the face of businesses around the globe. The increasing availability of digital data streams has enabled the development of AI-based systems meeting a myriad of needs. In testament to the accelerated growth possible by leveraging AI-based systems, in 2021 the global AI market stood at $US 327.5 billion.

Business leaders are taking notice with over **84% of business executives** reporting that they think that AI will help them gain a competitive edge in their niche. Harvard Business Review found that 75% of executives in companies that have adopted AI believed that using AI technology would substantially **transform their companies within the next three years.**

As this versatile technology has the potential to impact many processes, executives are looking to AI to enhance their business in a variety of ways. One **survey** found that 35% of executives anticipated that AI would help them make better business decisions, while 36% were looking to optimize their internal business operations. A further 30% looked to AI to optimize their marketing and sales process.

Within the healthcare space, AI adoption is currently 16%, which accounts for both clinical applications and other adjacent processes. The potential for growth in AI healthcare applications is vast, but as the healthcare systems scale the use of AI, the potential for bias also grows.

30%

Of executives think AI will help them optimize their marketing and sales process

36%

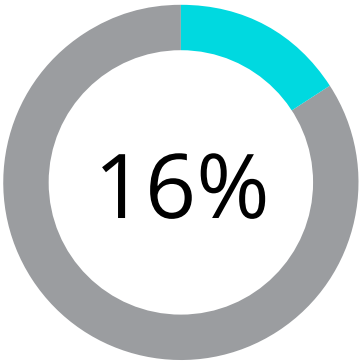
Of executives think AI will help them optimize their internal business operations

35%

Of executives think AI will help them make better business decisions

Types of AI in Healthcare

AI in healthcare falls into one of three broad categories:



Adoption of AI in the healthcare industry

Process automation

Allows healthcare systems to streamline basic processes that might otherwise have required excessive manpower. This is particularly impactful in supporting the revenue cycle with services such as claims processing, payment processing, maintaining EHR and compliance management.

Cognitive insight

Algorithms detect patterns in data and interpret their meaning. This has seen extensive implementation in the healthcare space with image recognition allowing for AI-based clinician decision support in areas such as radiology and cardiology. Cognitive insight applications also allow systems to predict customer responses, 'learning' based on past responses.

Cognitive engagement

Enables healthcare systems to offer personalized, responsive communications through automation. Chatbots are the most frequently used application in the wider world. In the healthcare space, cognitive engagement is used to create individualized member journeys and customized care plans.

How AI Can be Biased

Bias in AI is defined as results that are systematically prejudiced due to faulty assumptions. By definition AI systems are created by humans and trained using specific data sets. Bias can creep into the process either as a result of human error or incorrect correlations made when analyzing the data set.

The most common sources of bias in AI are:

Background bias

Biases present in the background knowledge used to create the system can be replicated. This means that societal biases can be repeated in the AI algorithm. For example, in the case of NLP language based approaches, if the background data used to train the language recognition associates European-American names with positive sentiments and African-American names with negative sentiments then the algorithm will become biased.

Outcome bias

It is not only training data derived from human judgement that can become biased. Even systems based on objective facts can become biased when that objective data may reflect societal bias. For example, non-adherence to treatment regimes might be factual data, however if African-American's were more likely to be out of work (and so lack the funds to purchase repeat prescriptions) the societal descrimination against African-Americans is impacting outcome.

Perceptive bias

Many AI systems seek to replicate human judgements. However, human judgements are frequently either consciously or unconsciously biased. Therefore any training data coming from human judgement is likely to contain social biases.

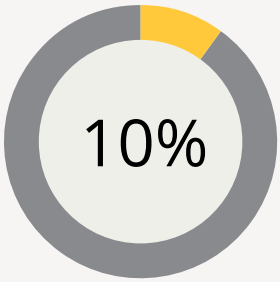
Availability bias

Machine learning algorithms are built to recognize clear frequently repeated patterns. Therefore, data that does not fit into these patterns may be overlooked by these AI systems.

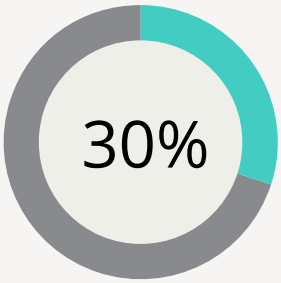
Why Bias in AI Matters

As the use of AI increases globally, maintaining trustworthy AI systems is becoming increasingly important. In a recent **survey of AI professionals**, 31% of respondents believed that the most significant problem to tackle in AI/ML today is the social impacts caused by bias in data and models. While only 10% of companies had already implemented a solution to ensure fairness, a further 30% of companies planned to implement steps to mitigate AI bias in the next 12 months.

Correcting bias in AI is essential to ensure fair access for all people. In 2020, the **Organisation for Economic Co-operation and Development (OECD)** adopted principles for the use of AI which have been adopted by 42 countries, including the United States. These principles are intended to ensure that AI respects the rule of law, human rights, democratic values and diversity, and includes appropriate safeguards to ensure a fair and just society. Any system that allows bias to continue puts these values at risk.



Of companies currently have a plan in place to mitigate bias in AI



Of companies intended to implement a plan to mitigate bias in AI in the next year.

Best Practice for AI in Healthcare Insurance

The National Association of Insurance Commissioners (NAIC) adopted guiding principles on artificial intelligence that were established to inform and articulate general expectations for businesses, professionals, and stakeholders across the insurance industry as they implement AI tools to facilitate operations.

Members deserve to have peace of mind that any AI systems used in their healthcare experiences are fair and free from bias. **Research** has shown that while members appreciate the many positive impacts that AI can have on healthcare, they remain concerned that AI tools might reinforce existing biases. It is up to insurers to earn the trust of their members.

The NAIC AI principles state that AI used in healthcare insurance should be:

- Fair and ethical
- Accountable
- Compliant
- Transparent
- Secure/ Safe / Robust

How to Mitigate Bias in Healthcare AI - a Practical Example

Awareness and testing are the keys to mitigating bias within any AI system. It requires first understanding where bias could be created and how it would impact outcomes. Secondly, multiple layers of outcome testing and review can act as a 'safety net' for the AI. We look at MedOrion's AI-based Health Behavior Management as an example of industry best practice.

How does the system work?

MedOrion's proprietary AI platform generates individual level Health Behavior insights and automatically applies them in all aspects of health plans interactions with their members, from risk adjustment through all levels of risk management and member engagement, leading to optimal health plans, improved member experience and overall better member care.

MedOrion's software models and measures members' health behaviors and stores them within the Health Behavior Management platform infrastructure. These insights are used to uncover certain trends, plan future products and even execute individualized member communications to encourage specific actions taking, all according to the objectives and priorities set by the health plan in real time. MedOrion designs a decision model for every health behavior objective (like medication adherence). Within that model are a number of decision factors (like cost, fear of side effects) and the relevance of each factor is calculated for every member.

The decision factors within the Health Behavior Management are scalable across populations and derived from validated models from behavioral science and psychology that have been shown to be consistent across different populations.

Where could bias creep in?

As we outlined above, any AI system is susceptible to some degree of bias. In the case of MedOrion, machine learning classification methods are used to 'fill in the gaps' if there are missing or "dirty" data elements for some members, to create 100% population coverage for the decision factors' relevance. Although the AI avoids using variables related to ethnicity, race, religion or sexual identity, some of the data elements might be correlated with them. Therefore, fairness tests must validate that no discrimination leading bias takes place.

For example, if the cost factors have been identified as one of the barriers to pharmacy adherence, then the AI might use median income to classify which members are more likely to be persuaded by cost-based messaging. A fairness requirement in this case would be to make sure that the cost factor did not 'take over' the other decision factors and create a bias for a specific population. If people who are immigrants are more likely to have a lower income, then a fairness test would make sure that all immigrants were not disproportionately persuaded by addressing a cost decision factor.

What is 'fairness' in AI?

The idea of 'fairness' in AI refers to the various attempts at bias correction. Results are considered fair if they are independent of given variables, especially those considered sensitive, such as traits that should not be correlated with the outcome (i.e. gender, ethnicity, sexual orientation, disability, etc).

Bias detection and mitigation

Bias prevention

Before a new deployment of the Health Behavior Management AI platform to a new population or decision objective, there are two stages of bias mitigation.

1. Customer sync

The MedOrion data scientists collect input from the customers about possible data elements which might be considered as discriminating and any features known to be correlated with them, such as specific geographic areas that might be predominantly associated with one ethinic group. These features will be considered to be removed from the Health Behavior Management AI scoring mechanism.

2. Behavioral scientist review

MedOrion's Health Behavior Management AI platform is heavily based on validated behavioral science theories. The behavioral science team reviews the raw member data to see what if any elements might cause bias. Together with the features suggested in the customer sync, these data elements will be updated in the testing suite (detailed in the next sections) in order to validate that no bias exists for them.

The default data elements (when provided) which would be classified as possibly affecting fairness include:

- Race
- Religion
- Ethnicity
- National Origin
- Marital Status
- Socioeconomic status
- Age
- Gender
- Education (as a proxy to health literacy) - e.g. college vs high school

Any issues detected at this stage are considered to be removed from the Health Behavior Management AI scoring mechanism.

Bias detection suite

Before deployment, MedOrion runs an extensive bias testing suite to test it's AI configurations. The Bias Testing Suite is based on industry standards and covers the acceptable fairness criteria for classification problems. As MedOrion's Health Behavior Management platform training and deployment is streamlined in a structured pipeline, the bias testing suite is part of the validation and testing phase. An update is not deployed to a production environment until it has been tested in case the suite fails.

A variety of different tests are run comparing predicted results with actual results.

They include:

Statistical parity (STP)
Verifies that all unprivileged subgroups have the same portion of positive predictions as the privileged group (the groups which are considered not to suffer from unfairness).

Equal opportunity (TPR)
Verifies that the classifier has equal True Positive Rate for each subgroup.

Predictive Parity (PPV)
Verifies that the model has equal Positive Predictive Value for each subgroup.

Predictive Equality (FPR)
Verifies that classifiers have equal False Positive Rate for each subgroup.

Accuracy Equality (ACC)
Verifies that the models have the same Accuracy for each subgroup, i.e the portion of errors the model makes within the privileged group is the same as the portion of errors the model makes within the unprivileged one.

Bias Mitigation

In the case that any of the tests should fail, then immediate mitigation steps are taken to adjust the model so that fairness criterias are met again.

Findings of any bias detected will first be presented to the customer to determine whether fairness criteria were disrupted. MedOrion recommends that the following mitigation steps be taken for each step in the data processing flow:

1. Pre-processing - removing the unwanted correlations using techniques such as: Disparate impact remover, Reweighting and Resampling.
2. In-processing - optimizing classifiers not only to reduce classification error but also to minimize a fairness metric.
3. Post-processing - model output modification so that predictions and miss-predictions among subgroups in the protected population are more similar. Known techniques include: Reject Option based Classification Pivot and Cutoff manipulation.

All the stages listed above take place in the pre-production environment. The bias detection testing suite takes place after the model testing and validation. Even after all bias detection testing has been successfully completed, a behavioral scientist performs a final review to ensure that everything is in order before deployment.

Conclusion

AI is here to stay. With the increase in AI applications, the need to maintain fairness and mitigate bias is only increasing. The challenge of beating bias in AI can be met with three distinct stages; identifying problematic areas before deployment, rigorous testing to detect bias within the system, and immediate response if any bias is detected.

MedOrion is proud to be leading the way within the healthcare space improving member outcomes through AI-powered member Health Behavior Management. We remain committed to mitigating bias at every stage of our development process.