Plausible Risk of Structural Racism from the Widespread Use of AI

Rajeshwarie VS and Vivek Gupta

Any views and ideas expressed in the essays are the author's alone and may not reflect the views and ideas of the Society of Actuaries, the Society of Actuaries Research Institute, Society of Actuaries members, nor the author's employer.

INTRODUCTION

Human intelligence[1] is the mental ability to learn and adapt to new situations, to understand and deal with abstract concepts, communication, reasoning, problem solving, and memory formation. In the past two decades, artificial intelligence (AI) has emerged as a powerful tool to fill in the gaps of human intelligence – speed and capacity. With its ability to process extremely large amounts of data in exceedingly short time spans, AI first began taking over repetitive and human energy/resource intensive tasks and evolved from automation to prediction and classification. The National Artificial Intelligence Act 2020 defines artificial intelligence as a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments. Both definitions of human and artificial intelligence are not very different except for perhaps communication, problem-solving and adapting to new situations which are still unique human strengths. AI follows a human-like evolutionary path of learning and growing, first being taught/trained to look for patterns and then to start spotting them at a much larger scale and at phenomenal speed yet unmatched by the human brain. Humans learn their biases through prejudices, stereotypes and rhetoric that have and may still prevail in society, and not always from scientific data or study. To counteract this, States across the globe have enacted laws to protect vulnerable minorities from discrimination. These laws categorically prohibit use of inherent traits to deny/discriminate people in any offerings unless they prove differentiation is fair and based on actual scientific data. These laws are made with the social goal to have a fair and equitable society where protected classes are not left marginalized. When AI is fed with data that carries the imprint of human biases, there is a great chance that these biases will enter models, as advanced algorithms can find new proxies to confirm these data biases and create more unexplainable models that strengthen these biases.

The nature of insurance business is "discriminatory" in the sense that it differentiates policyholders based on information about them and classifies them into "good" and "bad" risks. Such risk classification in order to assess and charge premiums is inherent to insurance business and is justifiable. But to what extent can one discriminate is a key question. Laws exist that prevent using certain information as rating factors or as a basis for providing/rejecting coverage. Gender, race, pre-existing health conditions and genetic information cannot be used to deny coverage or charge different premiums.

[1] Human Intelligence - an overview | ScienceDirect Topics

While direct discrimination is prohibited by laws, there are still ways by which scrupulous parties can unknowingly discriminate against protected groups by creating surrogate variables or rules in their processes thus undermining the objective of legislation. Two prominent types in which it can occur are:

> Structural racism[2] is defined as social, economic, and cultural differences between different groups of people that have developed over time leaving patterns entrenched in society. So, although an insurance pricing model may not be built with race as one of the rating factors, the results of the pricing exercise may still be to the advantage of some groups and/or a disadvantage to some others.
>
> Disparate Impact[3]: Policies, processes, rules, or other systems that appear to be neutral, result in a disproportionate impact on the protected classes. An example would be a fraud score which uses zip code as one of the parameters as this may cause a higher cost to poor minority neighborhoods. Zip codes have no direct impact on fraud score of an individual and are most easily available information. But zip codes are proxies to income, ethnicity and race and therefore it is possible that using a seemingly neutral factor such as a zip code in an algorithm to assign a fraud score, may produce unfair outcomes to some socio-economic groups which could then result in higher borrowing costs or insurance premiums.

In the rest of this paper, we look at the possible biases that may creep in if an insurance company uses AI extensively in its customer-facing processes largely due to rating factors or socio-economic indicators that may unintentionally work as proxies for protected factors in AI algorithms. These algorithms are often proprietary and hence black boxes which makes it difficult to assess the prediction logic. We conclude with a review of nascent fields of regulation and proactive measures that are developing to use AI ethically.

AI IN INSURANCE: THE PROBLEM OF PROXIES AND BLACK BOXES

AI is revolutionizing insurance in a number of ways with the ability to process vast amounts of data and generate insights for pricing and risk evaluation, targeted marketing, claims handling and settlement, streamlining manual and repeatable tasks, thus speeding up the insurance process for all parties involved as well as reducing the cost of delivering services. By leveraging AI, insurance companies can create and customize insurance products as per the need of the insured and price it better in commiseration to risk. With data ingestion platforms, a lot of data entry tasks can be reduced for underwriters and claim handlers thus freeing up their time to focus more on value added services of risk evaluation and claim adjudication. As a logical next step, insurance companies might leverage AI to make decisions without human intervention or with very limited human oversight. There is a good possibility that it may have negative impacts on the way insurance companies treat the insured. A number of recent studies have pointed to such negative impacts and below are some examples.

A number of rating factors are used to underwrite and price risks. For auto insurance, traditionally, vehicle and driver related information have always been the rating factors, for example, vehicle make and model, vehicle and driver age, history of previous claims/accidents, mileage, and zip code. A number of states can use non-driving related factors such as gender, marital status, home ownership details, address or zip-code, education level, past insurance purchase history. While it may seem that some of these factors are obvious choices for determining driver behavior and the risk the drivers represent, some of these factors can act as

---

[2] Definition of structural racism - NCI Dictionary of Cancer Terms - NCI

[3] Defining Fairness and Equity in AI-enabled Fraud Detection | Voyatek

proxies for others that are explicitly banned. A number of studies[4] point to drivers from predominantly Black/African American[5] U.S. neighborhoods being charged significantly higher premiums than those from white neighborhoods, whites who live in rented houses being charged lower premiums than Black homeowners, and Black drivers with better credit scores having higher premiums than white drivers with poorer credit scores. None of the algorithms that generate these rates have race as an input. Another more common example is exploiting the correlation between smoking and gender, while gender cannot be used as a rating factor, health and life insurance requires smoking status, and justifiably so!

While the above two examples illustrate the problem of proxies, there is also the problem of AI algorithms being black boxes, where it is not always obvious how the input data is being used to generate insights. An algorithm[6] used for predicting complex health care needs assigns a risk score to each individual based on past data of health care costs, and patients with risk scores above a threshold are automatically enrolled in the care program. The algorithm takes insurance claims data – age, sex, insurance type, diagnosis, procedures, and costs but explicitly excludes race. In predicting future healthcare costs based on the data, the algorithm performs consistently and predicts more or less equal costs for Blacks and whites irrespective of risk score. But the results showed that while referring people to support/care programs, a lower number of Blacks than whites at the same risk level were automatically selected. A study[7] on the results generated by the algorithm highlights that at the same level of health, Blacks tend to spend less on healthcare than whites and the kinds of expenses the two groups incurred were also different. Blacks spent more on emergency hospital visits while whites spent more on inpatient surgical procedures and specialist fees. While the algorithm was accurate in predicting future care costs per person, when it came to identifying care needs, using this predicted variable as an indicator of who needs care was failing to produce similar and equitable results for all groups.

DATA

AI is being used in claims settlement for initiating, assessing as well as approving claims, to quicken the process and offer customized responses to policyholders. Algorithms are widely used to file and screen claims and often involve processes such as checks to assess if forms have been filled correctly, images of damage are uploaded with required resolution, and screening basic information for possibility of fraud, thus saving many manhours. There have been studies that show that these practices have been discriminatory[8] with many instances of policyholders belonging to racial minorities being asked for additional information or documentation thus leading to a longer and more complicated settlement process. While no algorithms may have been built with the explicit aim to produce such delays, differences in levels of awareness, knowledge, digital literacy, access to uninterrupted internet, speed or resolution of devices that policyholders use may all impact the quality of the inputs they provide resulting in the longer delays, rejections, and requirement for additional validation. All these factors are known to vary with economic circumstances which vary widely between different ethnic groups and segments of society.

On the underwriting side, in order to classify and evaluate the risk represented by a prospective or existing policyholder, data is vital. In earlier times, this data was mostly available from the policyholders themselves and from the insurer's own experience. With advancement in AI, a large number of companies collect data and process it from various sources – shopping trends, credit scores, health parameters, travel, and food

---

[4] Study Points to Rate Bias in U.S. Auto Insurance Industry (investopedia.com)
[5] For brevity throughout the remainder of this essay, the author uses "Black" to represent "Black/African American."
[6] For minorities, biased AI algorithms can damage almost every part of life | SOAS
[7] Dissecting racial bias in an algorithm used to manage the health of populations | Science
[8] State Farm accused of covert racial discrimination in claims processing | WGLT

preferences are all accessible to insurers to deliver tailor-made products at accurate pricing. While we may believe that this data collected independently and through disjoint sources is unbiased and reflects lifestyle and behaviors of users as is, we need to understand that whatever insights it can provide would be limited by differences[9] in access, knowledge, awareness, ease/comfort with technology among different users or groups of users. It is well known and acknowledged that race and many socio-economic factors are correlated. For instance, in health care, poverty differences, lack of awareness to seek care, difference in access to transportation and health facilities, other competing demands like jobs or childcare, and doctor-patient relationships result in different levels in use of health care. Most of these factors would similarly impact adoption and optimal use of new technology and therefore any data collected through these will have an under-representation of some groups. While this data is used as such by AI/machine learning algorithms, these patterns are further amplified. Any prediction logic built with the majority, white-collared, educated and tech-savvy individual in mind would work well only for this segment of the population and may result in unfair results to any others that might not meet even one of these criteria.

REMEDIAL AND NEXT STEPS

The legal and regulatory landscapes have been constantly evolving to help counteract any inequality or discrimination that is perpetuated by technology. Beginning with the Fair Housing Act to counteract redlining, there are now many laws[10] such as Health Insurance Portability and Accountability Act of 1996, Genetic information non-discrimination act 2008, and Patient Protection and Affordable Care Act 2010 that apply to insurers to prevent any unjustified discrimination. More recently, states[11] in the U.S. are coming up with legislation specifically to avert unfair discrimination resulting from the extensive use of AI. These acts aim to ensure that an insurer's use of external consumer data sources and predictive models are augmented by a risk assessment and management framework that has to be documented and submitted to the regulators. Such frameworks require, on an ongoing basis, insurers to

- disclose external information sources and the algorithms or predictive modes they use,
- explain the manner in which both the data and the predictive models are used,
- assess where the use of these data sources and algorithms can cause unfair discrimination based on gender, race, color, ethnicity, disability, sexual orientation, and other prohibited factors.
- provide a reasonable time frame to remedy any such discriminatory impact of an algorithm.

Apart from the legal and regulatory aspects to prevent unfair discrimination, insurers can proactively work on the data and models to ensure the same. Companies could choose to work with the input data, the models themselves, the outputs or how they use these outputs in a manner to prevent any disparate impact.

- Data: A good understanding of the data that is used and how it has been collected is key to assessing the limitations and biases inherent in it and is a first step in eliminating biased outputs. The next step could be pre-processing input data, or post-processing outputs to mitigate the effects of biases that the data contains.[12]
- Models: It is important to understand what happens in AI models – their prediction or classification logic. Identifying the right relationship between the predicted variable and various

[9] The data divide | Ada Lovelace Institute
[10] repository.law.umich.edu/cgi/viewcontent.cgi?article=1163&context=law_econ_current
[11] 2021a_169_signed.pdf (colorado.gov)
[12] How insurers can mitigate the discrimination risks posed by AI - UNSW BusinessThink

predictors and testing this logic is necessary and is also a key area for regulators. Data insufficiencies or limitations can be corrected in the model training phase.

- Output: There is an emerging body of work on 'fairness criteria'[13] that can be used alongside AI models. Such criteria can determine if the output produced is fair to all racial/ethnic and other protected classes. Companies can adjust outputs to remove the effect of these biases.
- Using the output: In some situations, because of very limited availability of data and access to technology, some sections of consumers may be heavily underrepresented in model input data. In such cases, it is essential that companies review the applicability of model outputs to such groups. While it may work for some, it may not work for all!

CONCLUSION

In some areas such as criminal sentencing and facial recognition in public spaces, there is a demand to ban use of AI. In insurance, although it is still an area that people are suspicious about with the recent spate of lawsuits and literature around unfair discrimination, It is also widely acknowledged that there are benefits of AI in hastening up the various tasks along the insurance value chain that translate to savings in cost and resources. The alarm that is being raised is not merely resistance to change and should not be dismissed as voices against adoption to new technology. Companies and practitioners of AI need to acknowledge that the risks exist, and these concerns are legitimate as a first step to working on a solution to provide equitable and fair insurance offerings to all.

* * * * *

Rajeshwarie VS is a P&C actuary based in India. She can be reached at r.eshwari.894@gmail.com.

Vivek Gupta is an insurance consultant based in India. He can be reached to hutchie@gmail.com.

[13] Deloitte_Trustworthy AI_Fairness_Whitepaper_Dec2021.pdf