

DRAFT - 11/4/2022

Examining Bias in the Model Development Lifecycle

Nov. 10, 2022

Mark Sayre, FSA, CERA

Andy Reagan, Data Science at MassMutual

Presenters



Mark Sayre, FSA, CERA
Former Actuary and
Current Law Student



Andy Reagan, PhD
Data Scientist

Algorithms are Eating the World

“TikTok Reads Your Mind” - [The New York Times](#)

“Algorithms are better than people in predicting recidivism, study says” - [UC Berkeley News](#)

“Netflix’s New Algorithm Offers Optimal Recommendation Lists for Users with Finite Time Budget” - [infoQ](#)

If the 2010’s saw “software eating the world”, now we know that it’s really the algorithms embedded in that software. From what you see online from your friends, the news and TV on your home page, or how harshly you get sentenced in court, algorithms touch every part of our digital life and beyond.

Life insurance data often represents **a rich asset** from which to **discover patterns** with machine learning algorithms.

Insurance Too

DJIA 0.26% ▼ S&P 500 0.57% ▼ Nasdaq 0.90% ▼ U.S. 10 Yr 4.043% ▼ Crude Oil

WSJ

SUBSCRIBE SIGN IN

WSJ NEWS EXCLUSIVE | FINANCE

Allstate and Hancock Seek Careful Drivers With Healthy Diets

Partnership between insurers will reward safe drivers with potentially cheaper life insurance



Allstate and Hancock are teaming up on the theory that safe drivers live longer.

PHOTO: GERRY BROOME/ASSOCIATED PRESS

CNN

It's renters insurance season – Lemonade makes it easier than ever to get covered



Like a bike helmet, [renters insurance](#) is one of those things that you don't think you actually need until you really need it — and by then it's too late.

And just like how you put on a bike helmet even if you're

coverager

SUBSCRIBE

coverager

Avi Ben-Hutta
Editor at Coverager
May 8, 2019

Health IQ

Health IQ Raises \$55M

According to The Wall Street Journal, "the maker of an online quiz that prods marathoners and weightlifters about their exercise habits," aka Health IQ, has raised \$55m at a ~\$450m valuation. Investors that participated in the round include Aquiline Capital Partners, Greenspring Associates and Andreessen Horowitz.

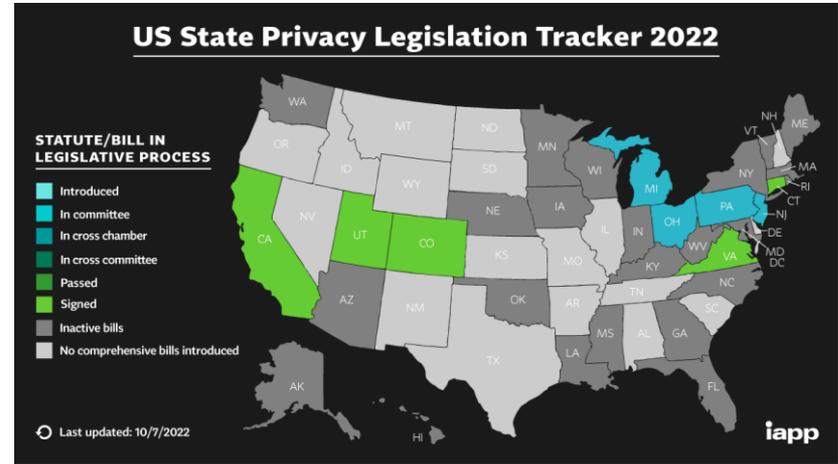
Predictive modeling has applications along the **complete insurance cycle**: in targeted marketing, new business underwriting, fraud detection, actuarial rate setting, and claims adjudication, among others.

Regulation - It Starts with Data

WIRED

California Unanimously Passes Historic Privacy Bill

The law will give Californians more control over the data that companies collect on them than ever before.



Regulation - It Starts with Data



The New York Times



PLAY THE CROSSWORD

G.D.P.R., a New Privacy Law, Makes Europe World's Leading Tech Watchdog

Give this article



28



Art. 22 GDPR

Automated individual decision-making, including profiling

1. The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.

Source: gdpr.eu

Regulation of A.I. is Rapidly Evolving



Proposed Algorithmic Accountability Act Targets Bias in Artificial Intelligence

JUNE 2019 | COMMENTARY



New Illinois Statute Among the First to Address AI-Aided Job Recruiting

Bloomberg Law WELCOME
BROWSE LOGIN

Daily Labor Report ®

✉️ f in t

NYC Targets Artificial Intelligence Bias in Hiring Under New Law

BY ERIN MULVANEY
Dec. 10, 2021, 5:41 AM



SUBSCRIBE

WILL KNIGHT BUSINESS APR 21, 2021 2:29 PM

Europe's Proposed Limits on AI Would Have Global Consequences

The EU released draft laws that would regulate facial recognition and uses of algorithms. If it passes, the policy will impact companies in the US and China.

CHINA BRIEFING

Q

EN

China's Sweeping Recommendation Algorithm Regulations in Effect from March 1

January 6, 2022 Posted by China Briefing Written by Arendse Huld Reading Time: 10 minutes

Insurance Too

2019

[New York Circular Letter No. 1 \(2019\)](#)

2020

[NAIC Adopts AI Principles](#)

2021

[CT DOI Big Data Notice and Certification](#)

[NAIC Forms Big Data and AI Working Group](#)

[Colorado Enacts S.B. 21-169](#)

2022

[CA DOI Bulletin 2022-5 on AI and Bias](#)

What Does Bias Mean?

3 Different Approaches

Disparate Impact

Does the model produce disproportionately negative impacts for certain populations?

Related to **Equity**

Unfair Discrimination

Does the model exhibit sufficient correlation to relevant risks based on generally accepted actuarial principles?

Related to **Actuarial Justification**

Proxy Discrimination

Are the correlations in the model a result of the inputs functioning as proxies for protected classes?

Related to **Fairness**

Putting it All Together

“Bias is everywhere in AI”
- UNESCO Caribbean AI Policy Roadmap

Framing and Planning

- Define business problem and success metrics
- Establish cross-functional team
- Determine if AI should be used and how
- Identify source data
- Evaluate risks

Development

- Extract and explore data
- Feature selection and engineering
- Determine modeling approach(es)
- Train model(s)
- Develop explainability tools

Testing

- Outline testing strategy
- Identify and prepare testing set
- Perform testing
- Document results and approval

Deployment

- Deploy in production
- Publish guidelines for model use
- Train end users
- Provide transparency via explainability tools
- Adopt controls and security measures

Management

- Monitor model performance
- Determine frequency and scope of reviews
- Measure against business objectives and metrics
- Modify or discontinue model if needed

Framing & Planning

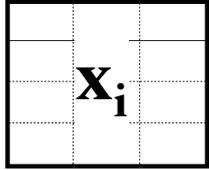
Potential Sources of Bias

- Homogenous team
- Team members' awareness of bias and mitigation techniques
- Team structure
- Narrow framing of problem
- Data access or quality

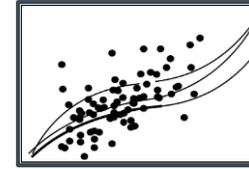
Mitigation Techniques

- Diverse, cross-functional teams
- Training
- Engage outside experts

Development of model



Pre-Processing
Adjust Training Data



In-Processing
Model Fitting

- Re-weight cases or (under/over) sample groups
- Change (some) labels
- Suppress sensitive and correlated features
- Learn a fair representation mapping
- Select semantically fair features
- Use actuarial judgement

- Embed fairness constraint within regularization term
- Train separate models for each sensitive category; combine output
- Use causal model or 'fair-aware' learning algorithms

Framing and Planning

Development

Testing

Deployment

Management

Development of model: practical steps

1. From universe of inputs, **exclude** protected classes and **known proxies** for protected classes
2. **Rational basis** for inclusion
3. Statistical **actuarial justification** test:
GAMs

1. **Choose linear, interpretable** models
2. **Black box models:** require model **explainability** tooling
3. **Document** key decisions and inputs.



Testing: Prediction & Assessment



- Scores should balance FP/FN
- Different thresholds for different groups might be necessary
- Assess fairness of model decisions
- Evaluate system as a whole by conducting fairness assessment on outcome distribution
- Does our model performance align to the end user's expectations
- Are we able to understand and explain the model's decision making?
- What data elements did we ultimately use in the development process, and is this consistent with what was expected?

Statistical study of disparate impact

*If protected class is **not** observed, but available externally, perform an external audit. Example: race identification.*

Framing and Planning

Development

Testing

Deployment

Management

Deploy

This phase involves the first time the model begins to have actual impact on individual customers, applicants, and users

1. **Data Missingness** should be handled in the applicant's favor, except where negative action is actuarially sound
2. Final results should be **explainable and transparent**, while protecting trade secrets
3. Individual decisions should be **consistent and repeatable**
4. Egregious **outliers** should be investigated

Framing and Planning

Development

Testing

Deployment

Management

Deploy & Explain

Output:

- Model-level performance metrics
- Individual predictions for individual

-- Example --

Applicant: John Snow **Score:** 85

-10	feature 1	+10	feature 2
- 5	feature 4	+ 5	feature 3
-15	feature 5		

100 (max score) - **30** + **15**

Framing and Planning

Development

Testing

Deployment

Management

Managing Post-Deployment

Potential Sources of Bias

- Changes in data sourcing or landscape
- Model and Concept Drift
- Changes in frequency and structure of data inputs
- End user adoption
- Complacency

Mitigation Techniques

- Model monitoring
- Ongoing training
- End user feedback
- Adopt a model review plan
 - Regular reviews
 - Ad-hoc reviews

Monitor

- Incoming data labels
- Distribution of requests
- Distribution of predictions
- Quality of predictions



Intervene

- Correct Data Problem
- Retrain/recalibrate model

Framing and Planning

Development

Testing

Deployment

Management

Model Monitoring

Monitor

- Incoming data labels
- Distribution of requests
- Distribution of predictions
- Quality of predictions



Intervene

- Correct Data Problem
- Retrain/recalibrate model

Framing and Planning

Development

Testing

Deployment

Management

Other Challenges

- Two alternative scenarios regarding access to data:
 - Data available through an external source only (race).
 - Data generally unavailable (sexual orientation).
- [Others]

Reading List

1. European Data Protection Board - Guidelines on Automated Decision Making under the GDPR, <https://ec.europa.eu/newsroom/article29/items/612053>
2. Society of Actuaries - Avoiding Unfair Bias in Insurance Applications of AI Models, <https://www.soa.org/4a36e6/globalassets/assets/files/resources/research-report/2022/avoid-unfair-bias-ai.pdf>
3. Ian Foster et al, *Data Science Methods and Tools for Research and Practice*, Ch. 11 - Bias and Fairness, <https://textbook.coleridgeinitiative.org/chap-bias.html>
4. U.S. House Committee on Financial Services, *Diversity and Inclusion: Holding America's Largest Insurance Companies Accountable*, https://financialservices.house.gov/uploadedfiles/d.i_insurance_report_092022.pdf