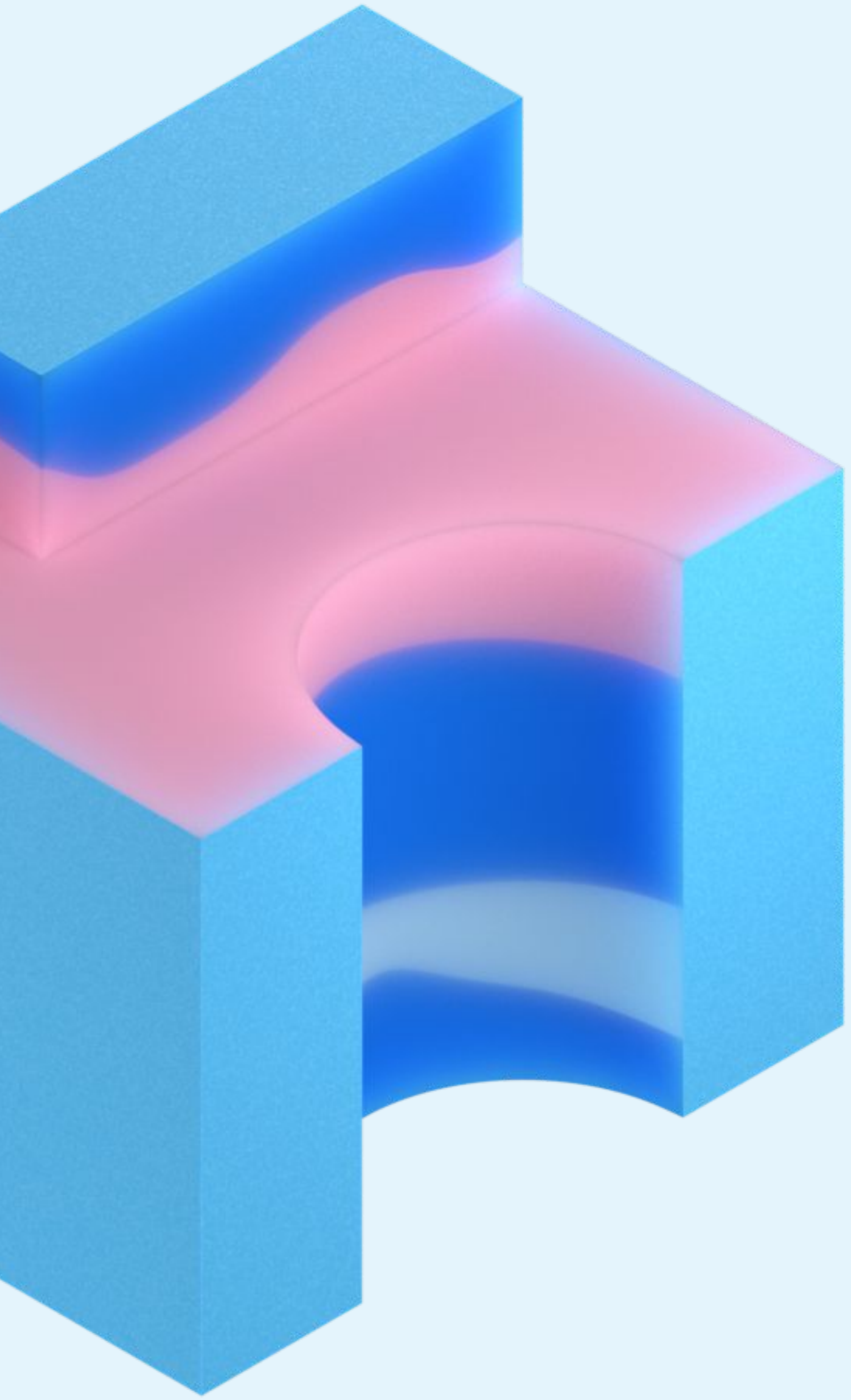
Machine learning and fairness in commercial insurance

Insurance Data Science Conference
London, UK

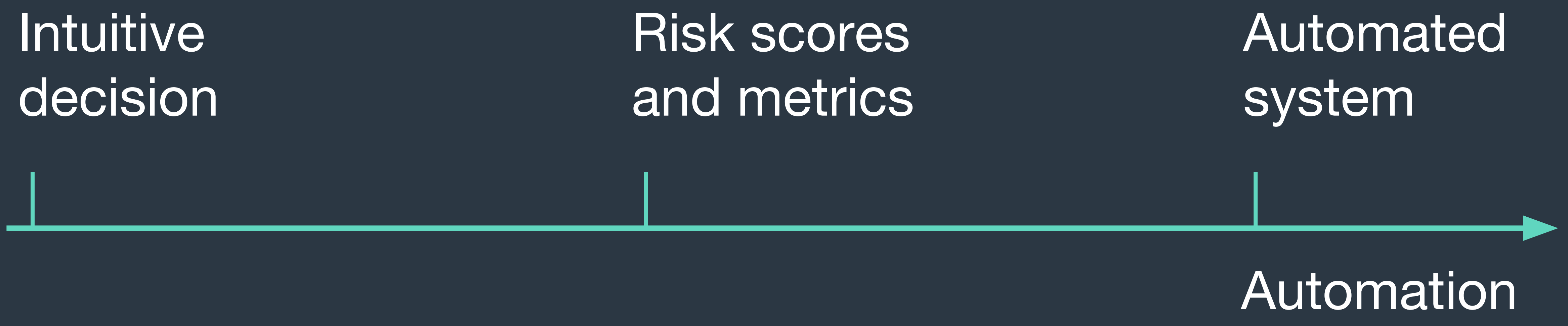
July 2018

Cytora



Cytora uses artificial intelligence to improve risk targeting, selection, and pricing for commercial insurance

Decisions that personally affect us are increasingly data-driven

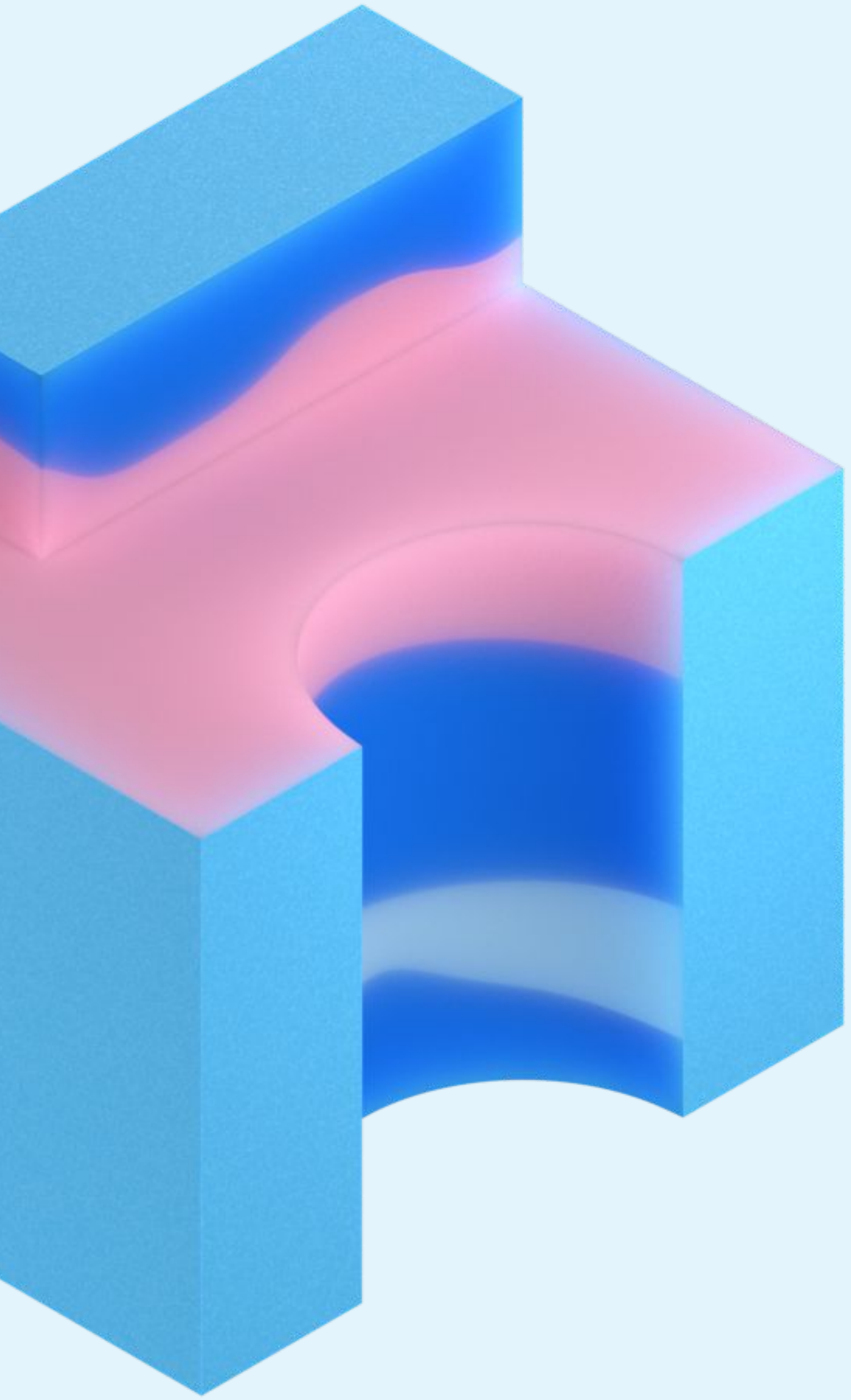


- Credit scoring
- Insurance
- Airport security
- Crowd monitoring
- Reoffending rates
- Advertising

Defining fairness

Cytora

Fairness means treating individuals from different groups equally



Fairness through unawareness is <u>not sufficient</u> to guarantee equal treatment for individuals in protected groups

Careful consideration is necessary when designing decision systems

Data

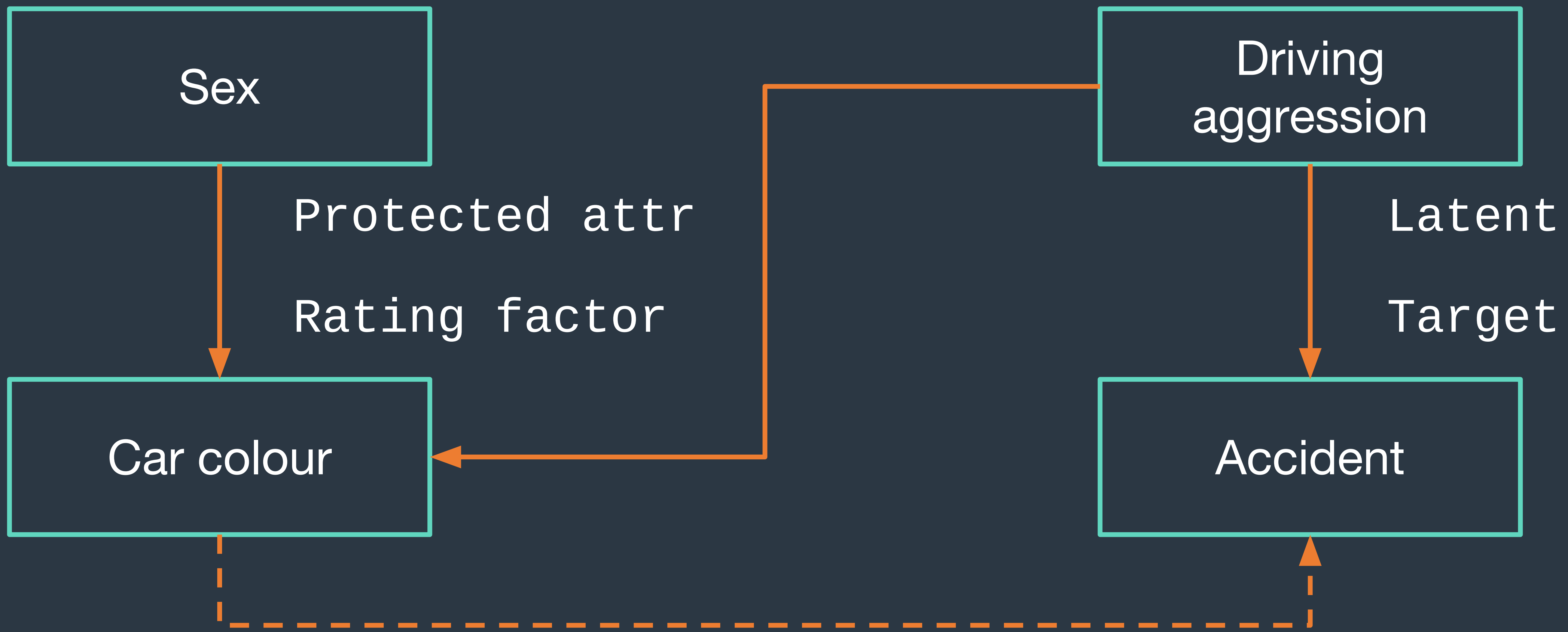
- Inherent data biases
- Reasoned vetting of variables
- True measures of latent risk
- Measure the protected attribute

Modelling

- Quantify feature contributions
- Tune for fairness
- Bias in, bias out

Cytora

Protected attributes encoded in “harmless” rating factors

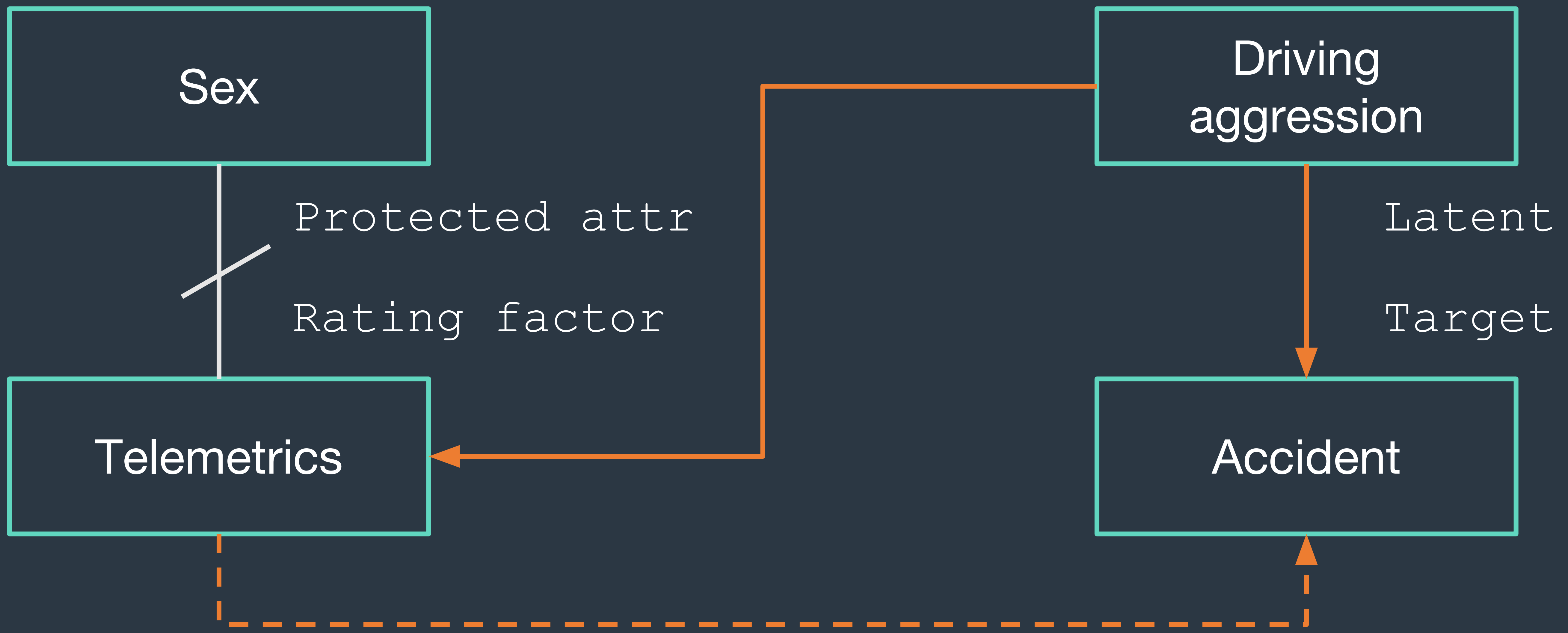


Kusner, Loftus, Russell, Silva (2018)
"Counterfactual Fairness"

Strategies for fairer pricing in insurance

Cytora

1. Observe relevant rating factors



2. Adjust premiums to optimise metrics of fairness

- Profit (accuracy)
- False positive rate, equal opportunity (FP/N)
- False negative rate (FN/P)
- Equalised odds (FPR & FNR)
- Equality of opportunity (FPR)
- Calibration (true probabilities)
- Demographic parity

Actual \ Predicted	No loss	Loss
No loss	TN	FP
Loss	FN	TP

Confusion matrix per protected group

Berk, Heidari, Jabbari, Kearns, Roth (2017) "Fairness in Criminal Justice Risk Assessments: The State of the Art"

Kleinberg, Mullainathan, Raghavan (2016) "Inherent Trade-Offs in the Fair Determination of Risk Scores

3. Design and train algorithms with fairness baked-in

- Structural models
  - Kilbertes, et. al. (2018) "Avoiding discrimination through causal reasoning"
  - Kusner, Loftus, Russell, Silva (2018) "Counterfactual Fairness"

- Penalised / constrained loss functions
  - Zafar, et. al. (2017) "Fairness Beyond Disparate Treatment & Disparate Impact"
  - Zhao, et. al. (2017) "Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints"

- Model inspection
  - Tan, Caruana, Hooker, Lou (2018) "Detecting Bias in Black-Box Models Using Transparent Model Distillation"

Example: Restaurant shutdown

Rating factor: Cuisine type = Krusty Burgers

Protected attribute: Shelbyville or Springfield resident

Solutions:
1. Observe management quality, menu, online reviews...
2. Geographic analysis of offered premium (adjust?)
3. Use fairness-calibrated algos

Summary

Data-driven modelling and machine learning can improve fairness by

1) Find better approximations of latent risk
2) Quantify effects on domain-specific fairness metrics
3) Calibrate decision making process to optimise fairness

Questions?

oliver@cytora.com

Learn more at cytora.com

Cytora