Fairness of predictive models: an application to insurance market

Arthur CHARPENTIER

SCOR
FOUNDATION FOR SCIENCE

FUNDED PROJECTS

EVENT - ACTUARIAL

Workshop in Paris (May 15th, 2025): Confidence and Fairness: Scientific Foundations in AI and Risk

Toon Calders
(University of Antwerp, Antwerp, Belgium)

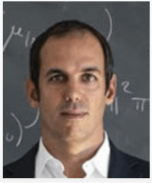
Arthur Charpentier
(UQAM, Montréal, Canada)

Evgenii Chzhen
(CNRS - Université Paris-Saclay, Paris, France)

François Hu
(Milliman, Paris, France)

Aurélie Lemmens
(Rotterdam School of Management, Erasmus University, Rotterdam)

Michele Loi
(Politecnico di Milano, Milano, Italy)



Jean Michel Loubés
(INRIA, Toulouse, France)

Antoine Ly
(SCOR, Chief Data Science Officer, Paris)

Isabel Valera
(Saarland University, Saarbrücken, Germany)

Fairness of predictive models
an application to insurance markets

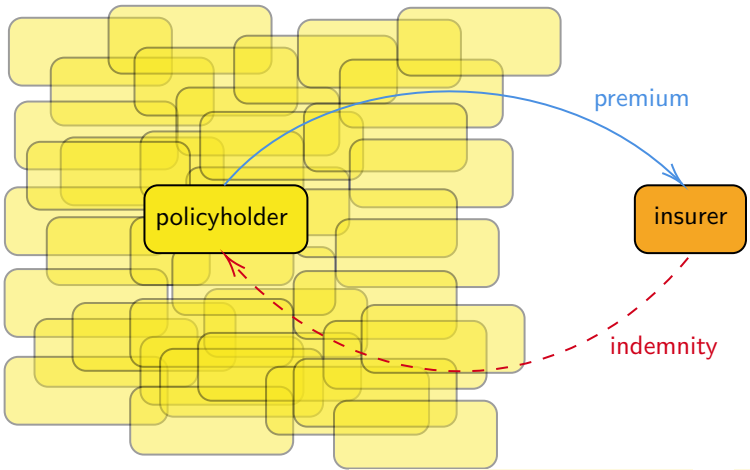
Arthur Charpentier
with Marie-Pier Côté, Olivier Côté Agathe Fernandes-Machado
Ewen Gallic, François Hu, Marouane Il-Idrissi & Philipp Ratz

Confidence and Fairness: Scientific Foundations in AI and Risk, May 2025

SCOR
FOUNDATION FOR SCIENCE

Confidence and Fairness

"*Insurance is the contribution of the many to the misfortune of the few*"



According to Arrow (1963), **actuarially fair premiums** = **expected losses**

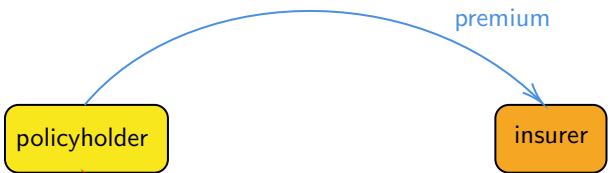
Confidence and Fairness

- There is no **law of one price** in insurance, see Froot et al. (1995)

→ with different models and different portfolio, we can have two different premiums

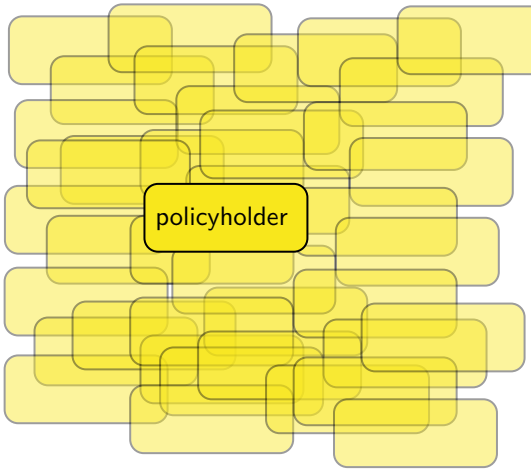
"*Insurance is the contribution of the many to the misfortune of the few*"

- Insurance is a **risk transfer** (from a policyholder to an insurance company)



- **Actuarial Fairness** is about the balance of this **risk transfer**

Confidence and Fairness



"*Insurance is the contribution of the many to the misfortune of the few*"

- As discussed in Charpentier (2025), insurance is also a **risk sharing** (among policyholders)
- **Fairness** (and **equity**) have to do with risk sharing and cross-subsidies within risk classes
- what is **expected losses**? $\mathbb{E}(Y)$ or $\mathbb{E}(Y \mid \boldsymbol{X})$? what is $\boldsymbol{X}$?

Confidence and Fairness

- Insurance is a **risk transfer**, see Ferrario and Loi (2022) on trust and confidence.

  Insurance is a promise, trust and confidence are key
- Insurance is a **risk sharing**, related to **cake-cutting** problems, see also Favier and Calders (2024).

"*Without trust and the mechanisms that build and maintain it, the insurance market becomes unsustainable*," Rozar and Eaton (2021), see also Guiso (2021)

Multiple publications on insurance, confidence, and personalization

Charpentier (2024a) Insurance: biases, discrimination and fairness. Springer Verlag

Charpentier and Vamparys (2025) Artificial intelligence and personalization of insurance: Failure or delayed ignition? *Big Data & Society*

Barry and Charpentier (2023a) Melting contestation: insurance fairness and machine learning. *Ethics and Information Technology*

Discrimination and Insurance

"**Discrimination** *is the act, practice, or an instance of separating or distinguishing categorically rather than individually*," Merriam-Webster (2022).

"*To be an* **actuary** *is to be a specialist in generalization, and actuaries engage in a form of decision making that is sometimes called actuarial. Actuaries guide insurance companies in making decisions about* **large categories that have the effect of attributing to the entire category certain characteristics that are probabilistically indicated by membership in the category, but that still may not be possessed by a particular member of the category**," Schauer (2006).

- [Most] "*actuaries cannot think of individuals except as members of groups*" claimed Brilmayer et al. (1979). Each individual is assigned the same value as all other members of the **group** to which it is assigned. See also Mowbray (1921) or Bailey and Simon (1960), or more recently Board (2005) and Finger (2006).
- Nevertheless, insurers also consider **personalization**, Barry and Charpentier (2020) or Lemmens et al. (2024) (on personalization)

AI = data + models

General context, **AI** is **(big) data** + **(black box) models**

First problem with "big data" has to do with **proxy discrimination**

"*OK, let's not use race, but should we use zip code, which of course is a proxy for race in our segregated society?*," O'Neil (2016).

It is possible to predict the "race" based on non-protected variables, e.g. names and geolocation, see "Bayesian Improved Surname Geocoding (BISG)", Elliott et al. (2009), Imai and Khanna (2016)

y	urban	age	race
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮

y	urban	age	zip	lastname	model	credit
⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Proxy problems are more complicated, and related to **collected information**, see Antonio and Charpentier (2017).

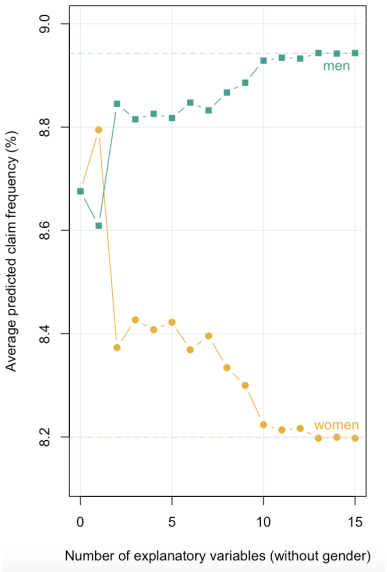
AI, Proxy and Generalization

On a French motor dataset, average claim frequencies are 8.94% (men) and 8.20% (women).

Consider some logistic regression to estimate annual claim frequency, on k explanatory variables **excluding gender**.

	men	women
$k = 0$	8.68%	8.68%
$k = 2$	8.85%	8.37%
$k = 8$	8.87%	8.33%
$k = 15$	8.94%	8.20%
empirical	8.94%	8.20%

Models simply tend to reproduce what was observed in the data, *clausula rebus sic stantibus* ("with things thus standing") $\longrightarrow$ **generalization**



AI, Proxy and Generalization

Multiple publications on insurance, discrimination and data
(without discussing further statistical learning and models)

Charpentier and Suire (2025) The insurance market in the era of digital transitions: Relationships between insurers, big tech, and insurtechs. *Society Of Actuaries Research Institute*

Côté and Charpentier (2025) Moral Maze. *The Actuary*

Charpentier (2025) Les paradoxes de la segmentation et de la discrimination en assurance. *Risques*.

Barry and Charpentier (2023b) Y-a-t-il une discrimination contre les pauvres ? *Risques*

Barry and Charpentier (2024) Partage des données, à qui profite le crime ? *Risques*

ML, Interpretability and Calibration

- Actuaries were using econometric techniques (Generalized Linear Models) then started using **machine learning**, see Charpentier et al. (2018)
- Problems of **transparency** and **interpretability**, see e.g., Delcaillau et al. (2022) and Goethals et al. (2023)
- Some concepts in actuarial science are **over-rated**, e.g., accuracy

  Some concepts in actuarial science are **under-rated**, e.g., **calibration**

"*When we speak of the 'probability of death', the exact meaning of this expression can be defined in the following way only. We must not think of an individual, but of a certain class as a whole* [...], *The phrase 'probability of death', when it refers to a single person, has no meaning for us at all*," von Mises (1928, 1939), see also **single-case interpretation of probability** in Reichenbach (1971).

ML, Interpretability and Calibration

As in Denuit et al. (2021), consider claims (annual) frequency in motor insurance,

	$\widehat{m}^{\text{glm}}$	$\widehat{m}^{\text{gam}}$	$\widehat{m}^{\text{gbm}}$	$\widehat{m}^{\text{rf}}$
average $\widehat{m}(\mathbf{x})$'s	0.0875	0.0877	0.0875	0.0886
10% quantile	0.0295	0.0288	0.0508	0.0006
90% quantile	0.1589	0.1612	0.1321	0.4073

First desirable property, **global balance**, $\mathbb{E}(Y) = \mathbb{E}(\widehat{m}(\mathbf{X}))$

Another desirable property, **local balance**,

average loss

average premium

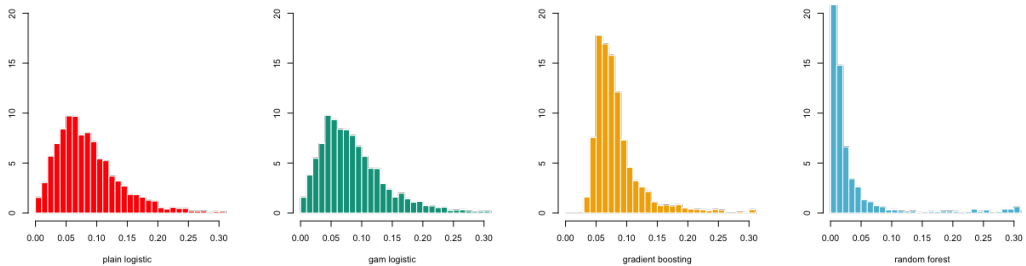
$$\mathbb{E}(Y \mid \widehat{m}(\mathbf{X})) = \mathbb{E}(\widehat{m}(\mathbf{X}) \mid \widehat{m}(\mathbf{X})) = \widehat{m}(\mathbf{X})$$

Calibration curve (for probabilistic scores) is defined as

$$g : \begin{cases} [0,1] \to [0,1] \\ p \mapsto g(p) := \mathbb{E}[Y \mid \widehat{m}(\mathbf{X}) = p] \end{cases}$$

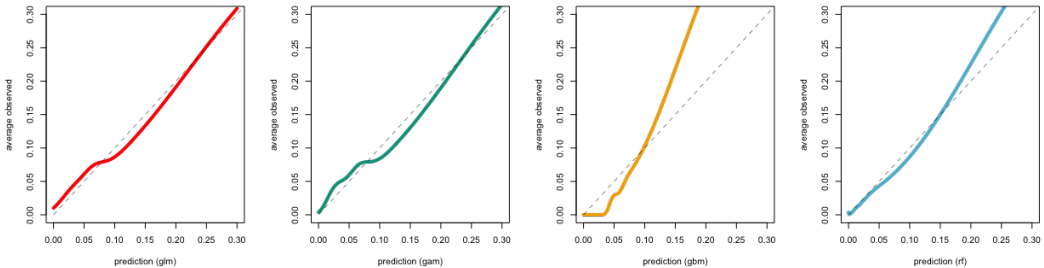
ML, Interpretability and Calibration

Distribution of $\widehat{m}(\mathbf{x}_i)$'s, for a plain logistic regression, GAM additive model, gradient boosting (sequential learning) and random forest.



	$\widehat{m}^{\text{glm}}$	$\widehat{m}^{\text{gam}}$	$\widehat{m}^{\text{gbm}}$	$\widehat{m}^{\text{rf}}$
average $\widehat{m}(\mathbf{x})$'s	0.0875	0.0877	0.0875	0.0886
10% quantile	0.0295	0.0288	0.0508	0.0006
90% quantile	0.1589	0.1612	0.1321	0.4073

ML, Interpretability and Calibration



Evolution of the (estimated) calibration curve, $p \mapsto \mathbb{E}[Y|\widehat{m}(\mathbf{X}) = p]$
where $\widehat{m}(\mathbf{x})$ is a predicted claims (annual) frequency, for a plain logistic regression, GAM additive model, gradient boosting (sequential learning) and random forest.

Even if **calibration** is an important property in insurance pricing (see also Baumann and Loi (2023)), there are no guarantees for most machine learning models.

ML, Interpretability and Calibration

Multiple publications on insurance and "predictive analytics"

Aas et al. (2025) Insurance analytics: prediction, explainability and fairness. *Annals of Actuarial Science*

Fernandes Machado et al. (2024c) Probabilistic scores of classifiers, calibration is not enough. *arXiv:2408.03421*

Fernandes Machado et al. (2024b) Post-calibration techniques: Balancing calibration and score distribution alignment. *Thirty-Eighth Annual Conference on Neural Information Processing Systems (NeurIPS)*

Fernandes Machado et al. (2024a) From uncertainty to precision: Enhancing binary classifier performance through calibration. *arXiv:2402.07790*

Il Idrissi et al. (2025) Unveil Sources of Uncertainty: Feature Contribution to Conformal Prediction Intervals. *Forthcoming*

Spectrum of Fair Predictors

"*What is unique about insurance is that* **even statistical discrimination which by definition is absent of any malicious intentions, poses significant moral and legal challenges**. *Why? Because on the one hand, policy makers would like insurers to treat their insureds equally, without discriminating based on race, gender, age, or other characteristics, even if it makes statistical sense to discriminate* (...) *On the other hand,* **at the core of insurance business lies discrimination** *between risky and non-risky insureds. But riskiness often statistically correlates with the same characteristics policy makers would like to prohibit insurers from taking into account.* " Avraham (2017)

"*Machine learning won't give you anything like gender neutrality 'for free' that you didn't explicitly ask for*," Kearns and Roth (2019)

Spectrum of Fair Predictors

demographic parity $\longleftrightarrow \quad \widehat{Y} \perp\!\!\!\perp S \quad \rightarrow \quad \mathbb{E}[\,\widehat{Y} \mid S = \mathrm{A}\,] \overset{?}{=} \mathbb{E}[\,\widehat{Y} \mid S = \mathrm{B}\,]$

sensitive sensitive score $\widehat{y}$

(see Hertweck et al. (2021))

or a stronger condition, $(\,\widehat{Y} \mid S = \mathrm{A}\,) \overset{?}{\underset{\mathcal{L}}{=}} (\,\widehat{Y} \mid S = \mathrm{B}\,)$

equalized odds $\longleftrightarrow \quad \widehat{Y} \perp\!\!\!\perp S \mid Y \quad \rightarrow \quad \mathbb{E}[\,\widehat{Y} \mid Y = y, S = \mathrm{A}\,] \overset{?}{=} \mathbb{E}[\,\widehat{Y} \mid Y = y, S = \mathrm{B}\,]$

outcome y score $\widehat{y}$

(see Heidari et al. (2019))

calibration $\longleftrightarrow \quad Y \perp\!\!\!\perp S \mid \widehat{Y} \quad \rightarrow \quad \mathbb{E}[\,Y \mid \widehat{Y} = u, S = \mathrm{A}\,] \overset{?}{=} \mathbb{E}[\,Y \mid \widehat{Y} = u, S = \mathrm{B}\,]$

outcome y score $\widehat{y}$

(see Baumann and Loi (2023))

Spectrum of Fair Predictors

Multiple publications on definitions of "fair price" for insurance premiums

Côté et al. (2024) A fair price to pay: exploiting causal graphs for fairness in insurance. *Journal of Risk and Insurance*.
Côté et al. (2025b) Selection bias in insurance: Why portfolio-specific fairness fails to extend market-wide. *SSRN*
Côté et al. (2025a) A scalable toolbox for exposing indirect discrimination in insurance rates. *Submitted*
Moriah et al. (2024) Measuring and mitigating biases in motor insurance pricing. *European Actuarial Journal*
Charpentier (2024b) Quantifying fairness and discrimination in predictive models *in* Machine Learning for Econometrics and Related Topics. Springer

Two applications and two focuses

- moving from **group fairness** to **individual fairness**
- **mitigation** in the context of **demographic parity**

Interpretability and Counterfactuals

- Regulatory perspective, **group fairness** (discussed previously)
- Policyholders perspective, **individual fairness**

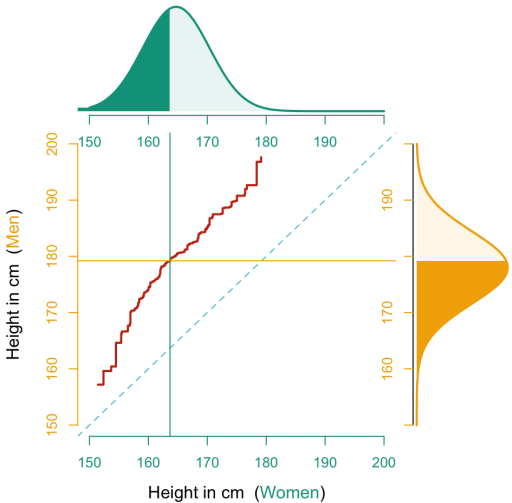
A decision satisfies individual fairness if "*had the protected attributes (e.g., race) of the individual been different, other things being equal, the decision would have remained the same*."

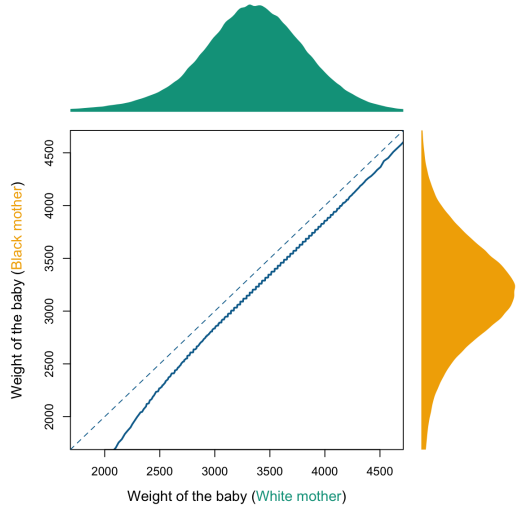
- also named **counterfactual fairness** in Kusner et al. (2017), and should be related to classical causal inference problem, (conditional) average treatment effect (the **treatment** being the sensitive attribute),

  "*other things being equal*" ? **ceteris paribus** ? See "revolving variable" in Kilbertus et al. (2017).

- two groups, distribution of **height** (cm) with women and men (in the U.S.)
- two groups, distribution of **baby weight** (g) with non-black and black mothers, delivering babies (in the U.S.)

Interpretability and Counterfactuals





Interpretability and Counterfactuals

Interpretability and explainability are important whem dealing with discrimination, Goethals et al. (2023, 2024). See also Karimi et al. (2020) and Tokpo and Calders (2023) for counterfactual explanations. Von Kügelgen et al. (2022) Majumdar et al. (2025) used also a causal framework to discuss fairness issues.

As De Lara et al. (2024), we used techniques based on **Optimal Transport**

Charpentier et al. (2023a) Optimal transport for counterfactual estimation: A method for causal inference. *in* Optimal Transport Statistics for Economics and Related Topics, pages 45–89. Springer

Fernandes Machado et al. (2025b) Sequential conditional transport on probabilistic graphs for interpretable counterfactual fairness. *39th Annual AAAI Conference on Artificial Intelligence*.

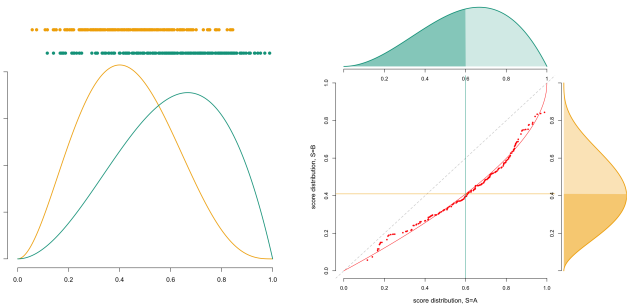
Fernandes Machado et al. (2025a) Optimal transport on categorical data for counterfactuals using compositional data and Dirichlet transport. *34th International Joint International Conference on Artificial Intelligence (IJCAI)*

We used **causal graphs** to derive an interpretable transport map.

Mitigation

In insurance, we usually observe differences among protected groups, e.g. Sauce et al. (2023) and Moriah et al. (2024). Can we **mitigate** such discrimination?

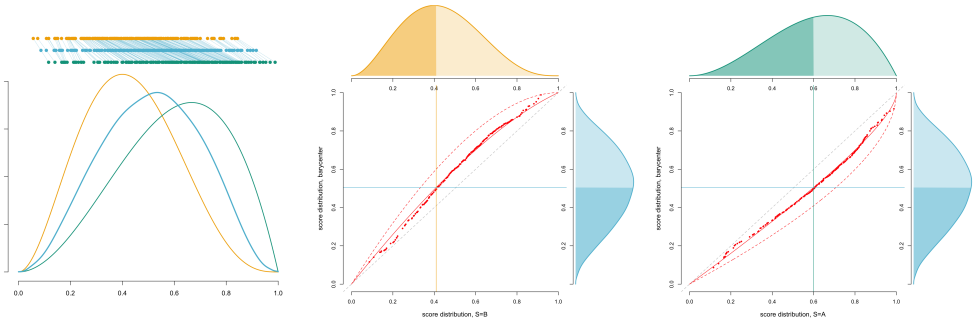
- we can consider **in processing** approaches, as Zafar et al. (2017a,b, 2019) (add a fairness penalty in the learning optimization problem)
- we can consider **post processing** approaches.



- individual in group A with a score $\widehat{y}(\mathrm{A}) = 60\%$ corresponding to quantile α (here 0.5)
- in group B, the same quantile α corresponds to $\widehat{y}(\mathrm{B}) = 40\%$

Mitigation

One can consider **Wasserstein barycenter** (and the associated transport mapping) to distort unfair predictors, and **mitigate** discrimination of models.



Strong mathematical guarantees, discussed in Le Gouic and Loubes (2017), and then Chzhen et al. (2020a,b), Chzhen and Schreuder (2022), Gaucher et al. (2023), Denis et al. (2024), Krco et al. (2023).

Mitigation

Multiple publications on the use of **Wasserstein barycenter** to mitigate discrimination, including discussions about multiple sensitive attributes.

Fernandes Machado et al. (2025c) Equipy: Sequential fairness using optimal transport in Python. *ArXiv:2503.09866*

Hu et al. (2023b) Parametric fairness with statistical guarantees. *ArXiv:2310.20508*

Charpentier et al. (2023b) Mitigating discrimination in insurance with wasserstein barycenters. bias. *3rd Workshop on Bias and Fairness in AI, International Workshop of ECML-PKDD*.

Hu et al. (2023a) Fairness in multi-task learning via wasserstein barycenters. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases – ECML PKDD*

Hu et al. (2024) A sequentially fair mechanism for multiple sensitive attributes. *38th Annual AAAI Conference on Artificial Intelligence*.

Mitigation?

- Supreme Court Justice Harry Blackmun stated, in 1978,

  "*In order to get beyond racism, we must first take account of race. There is no other way. And* **in order to treat some persons equally, we must treat them differently**," Knowlton (1978), cited in Lippert-Rasmussen (2020)

  To counteract **disparate impact**, intentional **disparate treatment** is necessary

  See philosophical discussions about **affirmative action**, e.g., Rubenfeld (1997); Pojman (1998); Anderson (2004)

- In 2007, John G. Roberts of the U.S. Supreme Court submits

  "*The way to stop discrimination on the basis of race is to* **stop discriminating on the basis of race**," Sabbagh (2007) and Turner (2015)

  corresponds to the "colorblind" approach

  Rejects any form of **disparate treatment**, even for corrective purposes, and reproduction of historical inequalities will lead to **disparate impact**

Conclusion?

- Thanks to this three-year funding by the SCOR Foundation for Science, a lot of work has been done,
- 18 months gone, 18 months to go
- Amazing team of collaborators, postdoctoral fellows, students and interns, in **Montréal** (Canada)

This was only the introduction to the workshop, the most interesting part is yet to come, we just had a glimpse of the main themes that will be discussed all during the day...

Conclusion?

To follow our work, see our newsletter, every 6 months...

https://foundation.scor.com/funded-projects/fairness-predictive-models-application-insurance-markets

SCOR-UQAM Fairness and Insurance Project, Newsletter #1 2024

SCOR-UQAM NEWSLETTER #1 (October 2023-March 2024)

Arthur Charpentier (UQAM; charpentier.arthur@uqam.ca)

In the era of big data and hardly interpretable machine learning techniques, ensuring fairness and non-discrimination in predictions poses a significant challenge. At the core of the insurance industry lies the differentiation between high-risk and low-risk policyholders. However, with insurers gathering increasingly extensive data, distinguishing between justified pricing based on risk and unfair discrimination based on sensitive factors becomes more intricate. As Kranzberg's law suggests, *"technology is neither inherently good nor bad; nor is it neutral."* Hence, it's imperative to accurately measure discrimination to prevent its adverse effects. This project aims to address biases within automatic artificial intelligence algorithms utilized for determining optimal pricing in individual policies. Its goals are to mitigate or eliminate biases that could lead to inequities or discriminatory practices based on factors such as gender, race, religion, or origin in the coverage provided by insurers or reinsurers to policyholders. This subject holds both theoretical significance, as it involves enhancing control over the "black box" nature of AI-based models, and practical importance, by reducing the risks associated with discrimination and inequality. In light of the increasing influence of social networks and the consequential reputational challenges faced by insurers and reinsurers, this issue is particularly relevant and timely.

People involved

Several people are involved in the project.

Students (interns)

Suzie Grondin (alumni: MSc ENSAE), came for six months to work on dynamic pricing games, with reinforcement learning. She also worked on equity. ▶ ◀

See https://freakonometrics.hypotheses.org/

Students (PhD)

Two PhD students started in the Fall 2023, on fairness and discrimination, in insurance markets.

Agathe Fernandes (alumni: MSc IMT Atlantique & Eurécia), started a PhD at UQAM, supervised by **Arthur Charpentier** and **Ewen Gallic** (AMSE, France), on fairness and calibration of predictive models.

Olivier Côté (alumni: MSc Laval) started a PhD at Université Laval (Québec), supervised by **Arthur Charpentier** and **Marie-Pier Côté** (Université Laval), on fairness, discrimination and causal models.

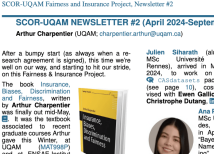
Postdoctoral Fellows

François Hu (alumni: PhD ENSAE Institut Polytechnique), finished a postdoctoral fellowship, supervised by **Arthur Charpentier** and **Manuel Morales** (Université de Montréal). He will go back to Paris, to join the R&D team of Milliman.

Two postdoctoral fellowships will be offered, starting during the Summer 2023.

Other People

Ewen Gallic, maître de conférences at Aix-Marseille School of Economics, is invited for one year at UQAM, to work with Arthur Charpentier.



SCOR-UQAM Fairness and Insurance Project, Newsletter #2 2024

SCOR-UQAM NEWSLETTER #2 (April 2024-September 2024)

Arthur Charpentier (UQAM; charpentier.arthur@uqam.ca)

After a bumpy start (as always when a research agreement is signed), this time we're well on our way, and starting to hit our stride, on this Fairness & Insurance Project.

The book Insurance, Biases, Discrimination and Fairness, written by **Arthur Charpentier** was finally out mid-May, . It was the textbook associated to recent graduate courses Arthur gave this Winter, at UQAM (MATH69) and at ENSAE-Institut Polytechnique de Paris, in France , and at the Warsaw School of Economics this Fall .

ISBN 9780331497827

People involved

Several new people are involved in the project, including several interns!

Students (interns)

Noé Bosc-Haddad (alumni: MSc Centre-Supélec), arrived in March 2024, to work on "dependence measures and RKHS" (reproducing kernel Hilbert space).

Florent Crouzet (alumni: MSc Ecole Polytechnique, France), arrived in April 2024, to work on "transfer learning" and applications in climate modeling. This work was presented in Leuven in June 2024.

Julien Siharath (alumni: MSc Université de Rennes), arrived in March 2024, to work on the CASdatasets package (see page 10), cosupervised with **Ewen Gallic** and **Christophe Dutang**.

Ana Patron Piñeros (alumni: MSc Universidad de Los Andes, Colombia), arrived in April 2024, to work on "Bayesian Improved First Name Surname Geocoding" (BIFSG) model, cosupervised with **Agathe Fernandes Machado**.

Cassandra Mussard (alumni: MSc ENSEEIHT & INSA Toulouse, France), arrived in August 2024, to work on "Gini Pseudo-Metric Spaces" (applied to k-means and k-nearest neighbors).

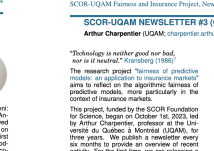
Several internships are offered for next summer ("Bias Globalink Research Internships"), "Interpretability of black box predictive models" , "combining decision trees and regression models" , and "causal models for machine learning" .

Miscellaneous (prices & rewards)

Franklin Feukam Kouhoue, who was a student at ENSAE, and wrote his MSc thesis on "Interpretabilité des modèles d'actuariat en tarification", supervised by Arthur Charpentier and Laurence Barry, obtained the **Prix des Sciences du Risque** from the **Optimind-Accenture Foundation**, in France, . A nice portrait was published, .

See https://freakonometrics.hypotheses.org/

Funding of grants and fellowships is coherent with amounts asked by the student association Support Our Science, in Canada .



SCOR-UQAM Fairness and Insurance Project, Newsletter #3 2025

SCOR-UQAM NEWSLETTER #3 (October 2024-March 2025)

Arthur Charpentier (UQAM; charpentier.arthur@uqam.ca)

"Technology is neither good nor bad, nor is it neutral." Kranzberg (1986)

The research project "fairness of predictive models: an application to insurance markets" aims to reflect on the algorithmic fairness of predictive models, more particularly in the context of insurance markets.

This project, funded by the SCOR Foundation for Science, began on October 1st, 2023, led by Arthur Charpentier, professor at the Université du Québec à Montréal (UQAM), for three years. We publish a newsletter every six months to provide an overview of recent activity. For the first time, we are releasing a version in French initiative #1 .

So here we are, halfway through the project, eighteen months after its launch. Following the publication of the book **Insurance, Biases, Discrimination and Fairness** (in English, presented in Newsletter #2), activity has slowed down a bit. As winter set in, the interns left, and we welcomed our daily activities, progressing in the writing of articles and presenting our work (notably at NeurIPS and AAAI, page 18). We also organized a very successful conference in Guanajuato, Mexico (see page 16), with another one planned for mid-May, this time in Paris (we will discuss it on page 21, and in the next newsletter).

Here's a quick overview of our activities over the past six months. We revisited competitive insurance markets, addressing issues of fairness (Selection bias in insurance: why portfolio-specific fairness fails to extend market-wide), algorithmic collusion (beyond human intervention: algorithmic collusion through multi-agent learning strategies),

See https://freakonometrics.hypotheses.org/

Melvin Kranzberg is an American historian who set out six laws of technology in a "Presidential Address" in the journal *Technology and Culture*, which he founded. This is the first law.

interactions with digital players (The insurance market in the era of digital transitions), The role of government vs. private sector provision of insurance, as well as an article in the newspaper *Le Monde*, Wildfires in California and natural disaster insurance). We also examine approaches related to fairness, with a short article Insurance analytics: prediction, explainability and fairness providing a brief state-of-the-art review, a dissemination article Moral maze: ethics and discrimination in machine learning aimed at actuaries, as well as A comment on the proposed automobile insurance rating and underwriting supervision guidance for regulators in Ontario. Let us mention a more legal article (The existence of gender bias through the use of algorithms in decision-making processes in liability and insurance law) and more mathematical articles, including Optimal transport on compositional data for conterfactuals and Equipy, sequential fairness using optimal transport are presented.

Finally, several articles revisit learning models: Post-calibration techniques: balancing calibration and score distribution alignment gets back on recalibration, Data augmentation with variational autoencoder for imbalanced dataset on the difficulty of predicting imbalanced data, k-nearest neighbors and k-means in Gini prametric spaces on handling outliers, Hierarchical decomposition of black-box models with dependent inputs on formal challenges in explainability approaches (when inputs are correlated), Functional central limit theorems for epidemic models with varying infectivity and waning immunity on epidemiology issues (variable infectivity and waning immunity), and a short article on learning algorithms and AI in economics (Machine learning and economics).

In addition to academic publications, magazines such as *The Actuary* (page 12), *Risques* (13-14), *The European Actuary* or *l'Actuariel* from the French Institute of Actuaries (15) review our work related to this project.

Toon Calders
(University of Antwerp, Antwerp, Belgium)

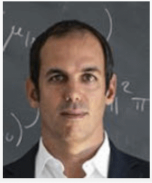
Arthur Charpentier
(UQAM, Montréal, Canada)

Evgenii Chzhen
(CNRS - Université Paris-Saclay, Paris, France)

François Hu
(Milliman, Paris, France)

Aurélie Lemmens
(Rotterdam School of Management, Erasmus University, Rotterdam)

Michele Loi
(Politecnico di Milano, Milano, Italy)



Jean Michel Loubés
(INRIA, Toulouse, France)

Antoine Ly
(SCOR, Chief Data Science Officer, Paris)

Isabel Valera
(Saarland University, Saarbrücken, Germany)

References

Aas, K., Charpentier, A., Huang, F., and Richman, R. (2025). Insurance analytics: prediction, explainability and fairness. *Annals of Actuarial Science*.

Anderson, T. H. (2004). *The pursuit of fairness: A history of affirmative action*. Oxford University Press.

Antonio, K. and Charpentier, A. (2017). La tarification par genre en assurance, corrélation ou causalité ? *Risques*, (110).

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *The American Economic Review*, 53(5):941–973.

Avraham, R. (2017). Discrimination and insurance. In Lippert-Rasmussen, K., editor, *Handbook of the Ethics of Discrimination*, pages 335–347. Routledge.

Bailey, R. A. and Simon, L. J. (1960). Two studies in automobile insurance ratemaking. *ASTIN Bulletin: The Journal of the IAA*, 1(4):192–217.

Barry, L. and Charpentier, A. (2020). Personalization as a promise: Can Big Data change the practice of insurance? . *Big Data & Society*.

Barry, L. and Charpentier, A. (2023a). Melting contestation: insurance fairness and machine learning. *Ethics and Information Technology*, 25(4):49.

References

Barry, L. and Charpentier, A. (2023b). Y-a-t-il une discrimination contre les pauvres ? *Risques*.

Barry, L. and Charpentier, A. (2024). Partage des données, à qui profite le crime. *Risques*.

Baumann, J. and Loi, M. (2023). Fairness and risk: an ethical argument for a group fairness definition insurers can use. *Philosophy & Technology*, 36(3):45.

Board, A. S. (2005). Risk classification (for all practice areas). *Actuarial Standard of Practice (ASOP)*, 12.

Brilmayer, L., Hekeler, R. W., Laycock, D., and Sullivan, T. A. (1979). Sex discrimination in employer-sponsored insurance plans: A legal and demographic analysis. *University of Chicago Law Review*, 47:505.

Charpentier, A. (2024a). *Insurance: biases, discrimination and fairness*. Springer Verlag.

Charpentier, A. (2024b). Quantifying fairness and discrimination in predictive models. In *Machine Learning for Econometrics and Related Topics*, pages 37–77. Springer.

Charpentier, A. (2025). Les paradoxes de la segmentation et de la discrimination en assurance. *Risques*, 142.

Charpentier, A., Flachaire, E., and Gallic, E. (2023a). Optimal transport for counterfactual estimation: A method for causal inference. In *Optimal Transport Statistics for Economics and Related Topics*, pages 45–89. Springer.

References

Charpentier, A., Flachaire, E., and Ly, A. (2018). Econometrics and machine learning. *Economie et Statistique*, 505(1):147–169.

Charpentier, A., Hu, F., and Ratz, P. (2023b). Mitigating discrimination in insurance with wasserstein barycenters. bias. In *3rd Workshop on Bias and Fairness in AI, International Workshop of ECML PKDD*.

Charpentier, A. and Suire, R. (2025). The insurance market in the era of digital transitions: Relationships between insurers, big tech, and insurtechs. *Society Of Actuaries Research Insitute*.

Charpentier, A. and Vamparys, X. (2025). Artificial intelligence and personalization of insurance: Failure or delayed ignition? *Big Data & Society*, 12(1):20539517241291817.

Chzhen, E., Denis, C., Hebiri, M., Oneto, L., and Pontil, M. (2020a). Fair regression via plug-in estimator and recalibration with statistical guarantees. *Advances in Neural Information Processing Systems*, 33:19137–19148.

Chzhen, E., Denis, C., Hebiri, M., Oneto, L., and Pontil, M. (2020b). Fair regression with wasserstein barycenters. *Advances in Neural Information Processing Systems*, 33:7321–7331.

Chzhen, E. and Schreuder, N. (2022). A minimax framework for quantifying risk-fairness trade-off in regression. *The Annals of Statistics*, 50(4):2416–2442.

Côté, M. and Charpentier, A. (2025). Moral Maze. *The Actuary*.

References

Côté, O., Côté, M.-P., and Charpentier, A. (2024). A fair price to pay: exploiting causal graphs for fairness in insurance. *Journal of Risk and Insurance*.

Côté, O., Côté, M.-P., and Charpentier, A. (2025a). A scalable toolbox for exposing indirect discrimination in insurance rates. *Forthcoming*.

Côté, O., Côté, M.-P., and Charpentier, A. (2025b). Selection bias in insurance: Why portfolio-specific fairness fails to extend market-wide. *SSRN*.

De Lara, L., González-Sanz, A., Asher, N., Risser, L., and Loubes, J.-M. (2024). Transport-based counterfactual models. *Journal of Machine Learning Research*, 25(136):1–59.

Delcaillau, D., Ly, A., Papp, A., and Vermet, F. (2022). Model transparency and interpretability: Survey and application to the insurance industry. *European Actuarial Journal*, 12(2):443–484.

Denis, C., Elie, R., Hebiri, M., and Hu, F. (2024). Fairness guarantees in multi-class classification with demographic parity. *Journal of Machine Learning Research*, 25(130):1–46.

Denuit, M., Charpentier, A., and Trufin, J. (2021). Autocalibration and tweedie-dominance for insurance pricing with machine learning. *Insurance: Mathematics & Economics*.

Elliott, M. N., Morrison, P. A., Fremont, A., McCaffrey, D. F., Pantoja, P., and Lurie, N. (2009). Using the census bureau's surname list to improve estimates of race/ethnicity and associated disparities. *Health Services and Outcomes Research Methodology*, 9:69–83.

References

Favier, M. and Calders, T. (2024). Cherry on the cake: Fairness is not an optimization problem. *arXiv:2406.16606*.

Fernandes Machado, A., Charpentier, A., Flachaire, E., Gallic, E., and Hu, F. (2024a). From uncertainty to precision: Enhancing binary classifier performance through calibration. *arXiv preprint arXiv:2402.07790*.

Fernandes Machado, A., Charpentier, A., Flachaire, E., Gallic, E., and Hu, F. (2024b). Post-calibration techniques: Balancing calibration and score distribution alignment. *NeurIPS BDU Workshop*.

Fernandes Machado, A., Charpentier, A., Flachaire, E., Gallic, E., and Hu, F. (2024c). Probabilistic scores of classifiers, calibration is not enough. *arXiv preprint arXiv:2408.03421*.

Fernandes Machado, A., Charpentier, A., and Gallic, E. (2025a). Optimal transport on categorical data for counterfactuals using compositional data and dirichlet transport. *arXiv*, 2501.15549.

Fernandes Machado, A., Charpentier, A., and Gallic, E. (2025b). Sequential conditional transport on probabilistic graphs for interpretable counterfactual fairness. *39th Annual AAAI Conference on Artificial Intelligence*.

Fernandes Machado, A., Grondin, S., Ratz, P., Charpentier, A., and Hu, F. (2025c). Equipy: Sequential fairness using optimal transport in Python. *ArXiv:2503.09866*.

References

Ferrario, A. and Loi, M. (2022). How explainability contributes to trust in ai. In *Proceedings of the 2022 ACM conference on fairness, accountability, and transparency*, pages 1457–1466.

Finger, R. J. (2006). Risk classification. In Bass, I., Basson, S., Bashline, D., Chanzit, L., Gillam, W., and Lotkowski, E., editors, *Foundations of Casualty Actuarial Science*, pages 287–341. Casualty Actuarial Society.

Froot, K. A., Kim, M., and Rogoff, K. S. (1995). The law of one price over 700 years. *National Bureau of Economic Research Cambridge*.

Gaucher, S., Schreuder, N., and Chzhen, E. (2023). Fair learning with wasserstein barycenters for non-decomposable performance measures. In *International Conference on Artificial Intelligence and Statistics*, pages 2436–2459. PMLR.

Goethals, S., Martens, D., and Calders, T. (2023). Explainability methods to detect and measure discrimination in machine learning models. In *CEUR workshop proceedings*, volume 3442, pages 1–5.

Goethals, S., Martens, D., and Calders, T. (2024). Precof: counterfactual explanations for fairness. *Machine Learning*, 113(5):3111–3142.

Guiso, L. (2021). Trust and insurance.

References

Heidari, H., Loi, M., Gummadi, K. P., and Krause, A. (2019). A moral framework for understanding fair ml through economic models of equality of opportunity. In *Proceedings of the conference on fairness, accountability, and transparency*, pages 181–190.

Hertweck, C., Heitz, C., and Loi, M. (2021). On the moral justification of statistical parity. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 747–757.

Hu, F., Ratz, P., and Charpentier, A. (2023a). Fairness in multi-task learning via wasserstein barycenters. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases – ECML PKDD*.

Hu, F., Ratz, P., and Charpentier, A. (2023b). Parametric fairness with statistical guarantees. *arXiv*, 2310.20508.

Hu, F., Ratz, P., and Charpentier, A. (2024). A sequentially fair mechanism for multiple sensitive attributes. *Annual AAAI Conference on Artificial Intelligence*.

Il Idrissi, M., Fernandes Machado, A., Gallic, E., and Charpentier, A. (2025). Unveil sources of uncertainty: Feature contribution to conformal prediction intervals. *forthcoming*.

Imai, K. and Khanna, K. (2016). Improving ecological inference by predicting individual ethnicity from voter registration records. *Political Analysis*, 24(2):263–272.

References

Karimi, A.-H., Barthe, G., Balle, B., and Valera, I. (2020). Model-agnostic counterfactual explanations for consequential decisions. In *International conference on artificial intelligence and statistics*, pages 895–905. PMLR.

Kearns, M. and Roth, A. (2019). *The ethical algorithm: The science of socially aware algorithm design*. Oxford University Press.

Kilbertus, N., Rojas Carulla, M., Parascandolo, G., Hardt, M., Janzing, D., and Schölkopf, B. (2017). Avoiding discrimination through causal reasoning. *Advances in neural information processing systems*, 30.

Knowlton, R. E. (1978). Regents of the university of california v. bakke. *Arkansas Law Review*, 32:499.

Krco, N., Laugel, T., Loubes, J.-M., and Detyniecki, M. (2023). When mitigating bias is unfair: A comprehensive study on the impact of bias mitigation algorithms. *arXiv:2302.07185*.

Kusner, M. J., Loftus, J., Russell, C., and Silva, R. (2017). Counterfactual fairness. In *Advances in Neural Information Processing Systems*, pages 4066–4076.

Le Gouic, T. and Loubes, J.-M. (2017). Existence and consistency of wasserstein barycenters. *Probability Theory and Related Fields*, 168:901–917.

References

Lemmens, A., Roos, J. M., Gabel, S., Ascarza, E., Bruno, H., Gordon, B. R., Israeli, A., Feit, E. M., Mela, C. F., and Netzer, O. (2024). Personalization and targeting: How to experiment, learn & optimize. *Learn & Optimize*.

Lippert-Rasmussen, K. (2020). *Making sense of affirmative action*. Oxford University Press.

Majumdar, A., Kanubala, D. D., Gupta, K., and Valera, I. (2025). A causal framework to measure and mitigate non-binary treatment discrimination. *arXiv:2503.22454*.

Merriam-Webster (2022). *Dictionary*.

Moriah, M., Vermet, F., and Charpentier, A. (2024). Measuring and mitigating biases in motor insurance pricing. *European Actuarial Journal*, 14(3):833–869.

Mowbray, A. (1921). Classification of risks as the basis of insurance rate making with special reference to workmen's compensation. *Proceedings of the Casualty Actuarial Society*.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Pojman, L. P. (1998). The case against affirmative action. *International Journal of Applied Philosophy*, 12(1):97–115.

Reichenbach, H. (1971). *The theory of probability*. University of California Press.

References

Rozar, T. and Eaton, R. (2021). Building trust in a digital world: Addressing modern threats to trust requires modern solutions. *The Actuary*.

Rubenfeld, J. (1997). Affirmative action. *Yale Law Journal*, 107:427.

Sabbagh, D. (2007). *Equality and transparency: A strategic perspective on affirmative action in American law*. Springer.

Sauce, M., Chancel, A., and Ly, A. (2023). Ai and ethics in insurance: a new solution to mitigate proxy discrimination in risk modeling. *arXiv:2307.13616*.

Schauer, F. (2006). *Profiles, probabilities, and stereotypes*. Harvard University Press.

Tokpo, E. K. and Calders, T. (2023). Model-based counterfactual generator for gender bias mitigation. *arXiv:2311.03186*.

Turner, R. (2015). The way to stop discrimination on the basis of race. *Stanford Journal of Civil Rights & Civil Liberties*, 11:45.

Von Kügelgen, J., Karimi, A.-H., Bhatt, U., Valera, I., Weller, A., and Schölkopf, B. (2022). On the fairness of causal algorithmic recourse. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 9584–9594.

von Mises, R. (1928). *Wahrscheinlichkeit Statistik und Wahrheit*. Springer.

References

von Mises, R. (1939). *Probability, statistics and truth*. Macmillan.

Zafar, M. B., Valera, I., Gomez Rodriguez, M., and Gummadi, K. P. (2017a). Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment. In *Proceedings of the 26th international conference on world wide web*, pages 1171–1180.

Zafar, M. B., Valera, I., Gomez-Rodriguez, M., and Gummadi, K. P. (2019). Fairness constraints: A flexible approach for fair classification. *Journal of Machine Learning Research*, 20(75):1–42.

Zafar, M. B., Valera, I., Rogriguez, M. G., and Gummadi, K. P. (2017b). Fairness constraints: Mechanisms for fair classification. In *Artificial intelligence and statistics*, pages 962–970. PMLR.