

Three lectures on AI and its implications for actuarial (and/or financial) professions

Arthur Charpentier

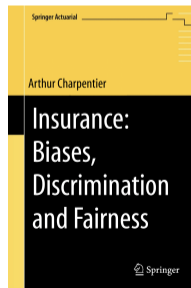
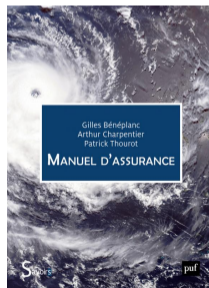
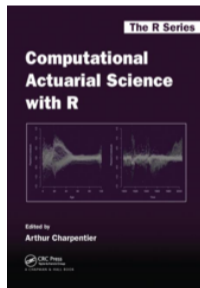
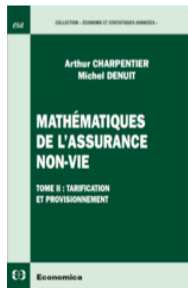
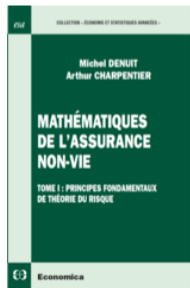
Université du Québec à Montréal & 京都大学
SCOR Foundation for Science

清华大学, April 2026

Bio (short)

Arthur Charpentier Professor at Université du Québec à Montréal (Canada)
Visiting professor at 京都大学 (Japan, 日本)

- [Charpentier and Denuit, 2004] *Mathematics of Non-Life Insurance* [FR],
- [Charpentier, 2014] *Computational Actuarial Science with R*,
- [Bénéplanc et al., 2022] *Handbook of Insurance* [FR],
- [Charpentier, 2024] *Insurance: Biases, Discrimination and Fairness*.



Agenda

Lecture 1: What insurance is (really) about?

Lecture 2 : What AI actually is, and where it enters insurance?

Lecture 3 : From risk pooling to risk personalization

Disclaimer: “AI” is a standard umbrella term, for actuaries and finance professionals,
→ here, **AI is mostly about turning data into decisions (under uncertainty)**

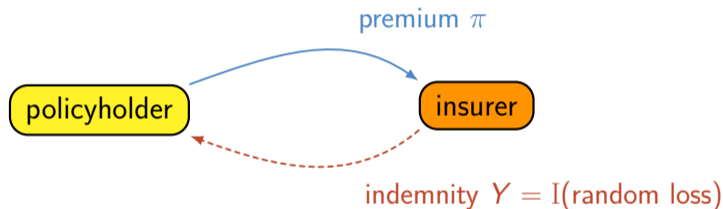
Lecture 1 — Why insurance exists?

Agenda

- Individuals face large losses that are rare, uneven, and hard to self-finance
- Insurance **transfers risk** from one balance sheet to another
- The contract converts uncertain future losses into a more certain premium today
- Insurance also **shares risk** across many policyholders through pooling
- This creates value even if no one can predict individual outcomes perfectly
- Insurance is both a financial technology and a **social institution**, [Borch, 1990]
- It is about compensation ex post, but also incentives ex ante
- AI matters only after this basic logic is understood clearly

Insurance contracts as risk transfer

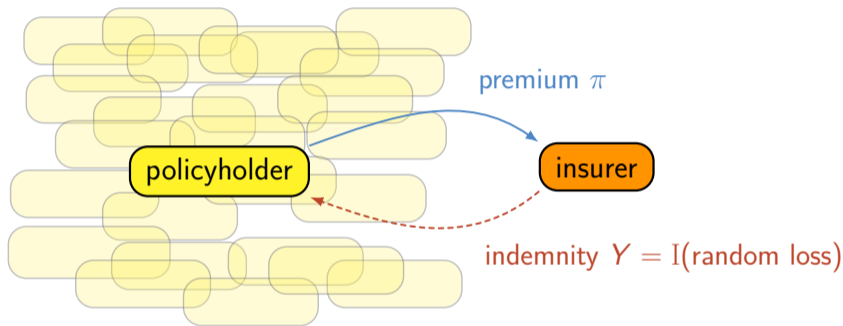
- Insurance **transfers risk** from one balance sheet to another, [Borch, 1990]



- Premium π is determined **ex ante**, whereas indemnity is paid **ex post**.
- Because of this inverted production cycle, insurance fundamentally depends on commitment and trust.

“Insurance is the contribution of the many to the misfortunes of the few”

- Insurance also **shares risk** across many policyholders through pooling, [Feng, 2023]



- Actuarial science aims to assess a **fair and adequate premium contribution** according to explicit **pricing principles** $\leftarrow \pi = \psi(\mathcal{L}_{w_Y})$
- Because ex ante / ex post, premium setting is **inherently predictive**.

Insurance contracts: risk transfer and risk sharing

- A contract defines covered events, exclusions, timing, limits, and obligations
- The policyholder **transfers part of a loss to the insurer in exchange for a premium** (fixed for one year of coverage)
- The insurer can do this because many risks are pooled together, [Borch, 1990]
- Pooling works best when losses are imperfectly correlated across people and firms
- **Insurance is not only about prediction; it is also about commitment and trust**
- The contract allocates information, incentives, and legal responsibilities
- Deductibles, limits, and exclusions are economic design choices, not only legal clauses ← “optimal” $I(\cdot)$ function
- Every contract reflects a trade-off between generosity, affordability, and simplicity

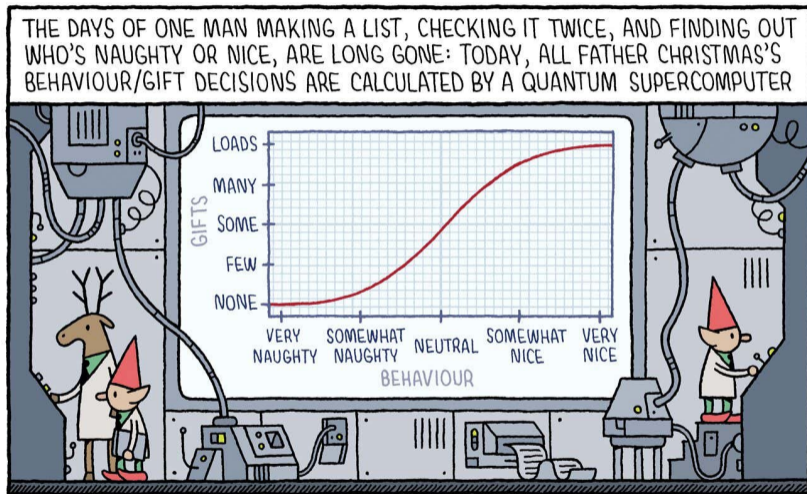
Where data and models enter insurance

- Data and models are not the whole business (in insurance); they support specific decisions, [Kaas et al., 2008]
 - Pricing asks: how much should the premium be for this risk? $\leftarrow \psi(\cdot) ?$
 - Underwriting asks: should we accept, reject, or modify this exposure?
 - Claims asks: did the event occur, what is covered, and how much should be paid?
 - Reinsurance asks: which risks should the primary insurer keep or transfer?
 - Capital management asks: how much uncertainty can the institution absorb?
- Some tasks use structured data, others use text, images, audio, or expert judgment
- **Models are therefore embedded inside multiple workflows**
- A good model is useless if the surrounding decision process is badly designed

Insurance pricing

- Pricing is often presented as technical, but it is partly normative, [Charpentier, 2024]
- One objective is **actuarial fairness**: expected premium aligned with expected risk
“pure premium principle”, $\psi(\cdot) = \mathbb{E}_{\mathbb{P}}[\cdot]$, or possibly $\psi(\cdot) = \mathbb{E}_{\mathbb{P}_X}[\cdot]$
- Another objective is commercial viability in a competitive market
- Another objective is affordability and social acceptability
- Features used in pricing (X) are **never purely neutral; they reflect choices**
- Granularity can improve prediction but may weaken pooling and solidarity
- Highly individualized pricing can make some customers cheaper and others uninsurable
- **Better pricing is not always equivalent to better welfare**
- When is precision desirable, and when is it excessive?

Insurance pricing



(source [Gauld, 2021])

TOM GAULD for NEW SCIENTIST

Why pooling works: the law of large numbers and the CLT

- Consider a portfolio of n policyholders with annual losses Y_1, \dots, Y_n .
- The aggregate loss is $S_n = \sum_{i=1}^n Y_i$ and $\bar{Y}_n = \frac{S_n}{n}$ is the average loss per policyholder.
- If losses are independent and identically distributed, with $\mathbb{E}[Y_i] = \mu$, $\text{Var}(Y_i) = \sigma^2$, then the **law of large numbers** says $\bar{Y}_n \xrightarrow{a.s.} \mu$, as $n \rightarrow \infty$.
- So **with a larger portfolio, the average loss becomes more predictable**.
- This is the basic statistical intuition behind mutualization, [Kaas et al., 2008].
- The **central limit theorem** goes further: $\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1)$.
- Equivalently, $S_n \approx \mathcal{N}(n\mu, (\sqrt{n}\sigma)^2)$, and $\bar{Y}_n \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$.
- Hence the standard deviation of the *average* loss is of order $1/\sqrt{n}$.
- This is why larger portfolios are statistically more stable.

Why pooling works: the law of large numbers and the CLT

- But this logic weakens sharply when losses are dependent, exposed to common shocks, or highly concentrated.
- The previous argument relies on **diversification**, [Berliner, 1982].
- Diversification requires many risks that are not too strongly dependent.
- In motor insurance, many small independent claims roughly fit this logic.
- In catastrophe insurance, cyber insurance, or AI-related losses, dependence can dominate diversification, and portfolios may no longer behave like “many small unrelated risks,” [McNeil et al., 2015]
- A **common shock** can generate many claims at once.
- And **contagion** can trigger sequential losses through propagation effects.

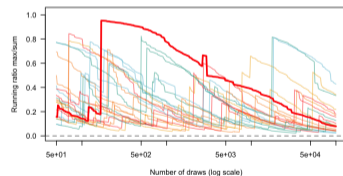
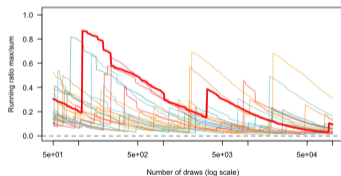
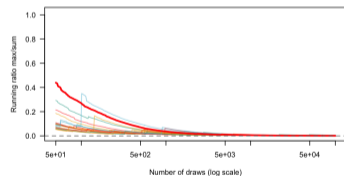
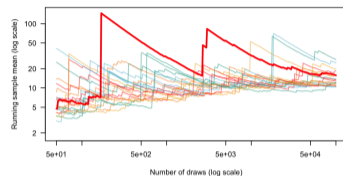
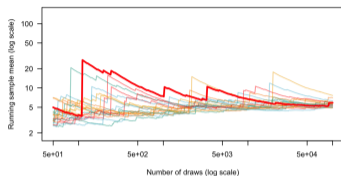
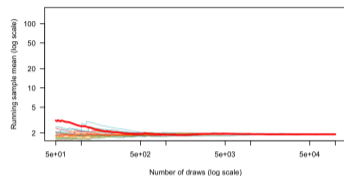
Why pooling works: the law of large numbers and the CLT

- In that case, the variance does not grow slowly enough for averaging to stabilize losses.
- This is why insurers care so much about accumulation, reinsurance, exclusions, and capital.
- The law of large numbers is not a magical theorem of insurability.
- It is a theorem that works under conditions — and those conditions matter.

$$\text{Var}(S_n) = n\sigma^2 + \sum_{i \neq j} \text{Cov}(Y_i, Y_j)$$

- Correlation is what turns a diversifiable portfolio risk into an **accumulation risk**.
- In many lines of business, variance might be infinite...

Law of Large Numbers with Large Risks



On top, $n \mapsto \frac{y_1 + \dots + y_n}{n}$ and below $n \mapsto \frac{\max\{y_1, \dots, y_n\}}{y_1 + \dots + y_n}$, with (finite mean, finite variance), (finite mean, infinite variance) and (infinite mean, infinite variance)

Behavior-based insurance: pricing, incentives, and continuous nudging

- Traditional insurance prices risk mostly from **ex ante characteristics** and **ex post claims**, [Cummins, 2002, Lemaire et al., 2016].
- Telematics and pay-as-you-drive products move insurance toward ongoing behavioral observation.
- Premiums can now depend on mileage, speed, braking, acceleration, time of day, routes, or phone use.
- This looks more “fair” because pricing becomes closer to measured exposure and observed conduct.
- But the premium is no longer only a price: it also becomes a signal, a reminder, and a **nudge**, [Parasuraman and Riley, 1997].
- Insurance then starts to shape behavior continuously, not only compensate losses after the fact.

Behavior-based insurance: pricing, incentives, and continuous nudging

- The difficulty is that behavior is never purely individual: it also reflects roads, jobs, schedules, and neighborhoods.
- A late-night trip may reflect risky driving, but it may also reflect shift work or family constraints.
- A hard brake may reflect driving style, but it may also reflect traffic density or road design.
- So **behavior-based insurance mixes actuarial classification, surveillance, and governance** .

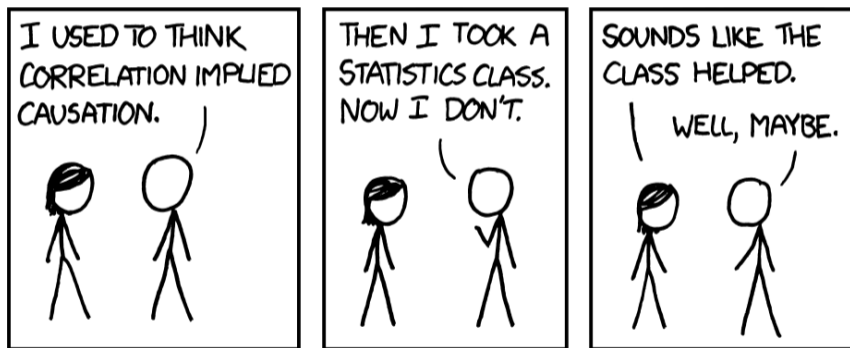
What makes a good rating variable?

- A useful rating variable is not only predictive, [Finger, 2006, Charpentier, 2024].
 - Actuarially, it should improve **accuracy**, create **homogeneous classes**, remain **credible**, and be **reasonably stable over time**.
 - Operationally, it should be objective, easy to collect, cheap to verify, and not overly ambiguous.
 - Socially, it raises questions of privacy, controllability, affordability, and acceptability.
 - Legally, it should not generate unfair discrimination.
- This is why a variable is never “just a variable”: it is also a choice about what differences may count.
- **Correlation can be sufficient for prediction, but not sufficient for legitimacy**
- A variable can predict risk without causing risk.
- A variable can also be a proxy for a protected or socially sensitive attribute.

What makes a good rating variable?

- So the key question is not only “does it work?” but also “what does it stand for, and can we defend its use?”

$$\mu(\mathbf{x}) = \mathbb{E}[Y \mid \mathbf{X} = \mathbf{x}] \quad (\text{predictive relevance does not imply causality})$$



(source [Munroe, 2014])

Insurance underwriting

*“use of a red line around the questionable areas on territorial maps centrally located in the Underwriting Division for ease of reference by all Underwriting personnel [...] mark off certain areas * * * to denote a lack of interest in business arising in these areas In New York these are called K.O. areas meaning knock-out areas; in Boston they are called redline districts. Same thing – don’t write the business,” [Kerner, 1968]*

- **Underwriting is the gatekeeping function of insurance**
- It determines eligibility, conditions, exclusions, limits, and sometimes monitoring
- Historically, underwriting mixed rules, experience, judgment, and medical or technical evidence, [Charpentier, 2024]
- It is where asymmetries of information become operationally important
- Underwriters try to detect adverse selection before the contract is signed
- They also manage concentration, ambiguity, and model uncertainty

Insurance underwriting

- Not all relevant information is numerical; much is textual, contextual, and uncertain
- This is one reason why AI may matter a lot in underwriting
- Should underwriting optimize prediction or protect insurability?
- Underwriting is where asymmetric information becomes operational.
- Before the contract is signed, the applicant may know more than the insurer.
- If the insurer charges a single average price to heterogeneous risks, low-risk applicants may leave.
- The remaining pool then becomes worse on average: this is the classical logic of **adverse selection**.
- Risk selection and premium segmentation are therefore partly defensive responses to **hidden heterogeneity**.
- Underwriting is not only about estimating expected claims; it is also about deciding whom to admit into the pool.

Insurance underwriting

- In practice, much relevant information is incomplete, strategic, textual, contextual, or unverifiable.
- This makes underwriting a prediction problem, a **screening problem**, and a **contract-design problem** at the same time.
- Better prediction can reduce some selection problems, but it can also increase exclusion.
- This creates a central tension: should underwriting optimize prediction, or preserve insurability?

$$\begin{cases} \mathbb{E}_{\mathbb{P}}[Y \mid \mathbf{X} = \mathbf{x}, \text{apply} = 1] \neq \mathbb{E}_{\mathbb{P}}[Y \mid \mathbf{X} = \mathbf{x}] \\ \mathbb{E}_{\mathbb{P}}[Y \mid \mathbf{X} = \mathbf{x}, \text{apply} = 1] =: \mathbb{E}_{\mathbb{Q}}[Y \mid \mathbf{X} = \mathbf{x}] \end{cases}$$

New probability measure \mathbb{Q} captures underwriting selection bias (against \mathbb{P})

- Classic issue in life insurance, where policyholders are a selected population and thus follow mortality patterns that differ from those of the general population.

Insurers model their customers, not the market

- Insurers usually observe their portfolio, not the whole market.
- The training data come from applicants who applied, risks that were accepted, and contracts that were actually written.
- Rejected applicants, discouraged customers, and uninsured people are usually missing from the data.
- Claims are therefore observed on an already selected population.
- In other words, underwriting data are not a neutral sample from the population of all risks.
- Marketing strategy, distribution channels, underwriting rules, and past pricing choices all shape the observed data.
- If the firm changes its acceptance rule, it also changes the future data-generating process.
- A model may therefore look very good on the insurer's own portfolio while learning little about the wider market.

Insurers model their customers, not the market

- This is one reason why **underwriting models can create feedback loops and exclusion dynamics**.
- So model validation in underwriting is never only statistical: it is also institutional and strategic.

$$\begin{cases} \mathbb{P}(\mathbf{X} \mid \text{insured} = 1) \neq \mathbb{P}(\mathbf{X} \mid \text{market}) \\ \mathbb{P}(Y \mid \mathbf{X}, \text{insured} = 1) \neq \mathbb{P}(Y \mid \mathbf{X}) \end{cases}$$

Claims management

- Claims are the (ex post) “*moment of truth*” for insurance, when uncertainty is unveiled \leftarrow not only $y = Y(\omega)$ (realization) but also $\mathcal{L}aw(Y)$.
- A claim is simultaneously a financial event, a legal event, and a human event
- The insurer must verify the facts, interpret the contract, and estimate the loss
- Speed matters, but accuracy and consistency also matter
- Claims operations combine triage, fraud detection, reserving, negotiation, and service
- This is a natural domain for text analysis, image analysis, and **workflow automation**
- But claims handling also raises **empathy, accountability, and explainability issues**
- A bad claims experience can destroy trust even if the model is statistically good

AI in claims management: speed, evidence, and new fraud risks

- Claims management is one of the most operational uses of AI in insurance.
- In motor insurance, computer vision can analyze **accident photos** taken from several angles.
- From 3 to 5 pictures, the insurer may detect damaged parts, estimate severity, and flag whether the vehicle is likely repairable or close to total loss.
- The system can also generate a first estimate of repair cost, expected labor time, or the need for expert review.
- More broadly, AI is used for first notice of loss, document extraction, claim-note summarization, routing, and triage.
- **NLP tools** can read emails, invoices, medical reports, police reports, and policy documents.
- Anomaly-detection tools can identify **suspicious patterns**, inconsistencies, duplicate claims, or unusual claim trajectories.

AI in claims management: speed, evidence, and new fraud risks

- This can make claims handling faster, cheaper, and more standardized.
- But the reverse side is that fraud also becomes easier: fake invoices, fake narratives, fake voices, and fake images can now be produced at scale.
- A synthetic image of scratches or damage on a car may look convincing enough to fool a human reviewer, especially under time pressure.
- So claims management becomes not only a prediction problem, but also a problem of **evidence, provenance, and authenticity** .
- Nothing new here... but AI can make it faster.
- If seeing is no longer believing, how should insurers verify loss?

Reinsurance

- Reinsurance is “*insurance for insurers.*”
- It allows primary insurers to transfer part of their accumulated risk exposures.
- It supports **solvency, capacity, volatility management, and catastrophe absorption**
- It is especially important when losses are large, correlated, or difficult to diversify
- Reinsurance also reveals that insurance is a layered financial system
- The industry is therefore connected through contracts, models, and capital flows
- It also matters because model errors can become systemic rather than local

What is an “actuary”?

- Actuaries are involved in all those tasks.
- As claimed in [Schauer, 2006],

“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], [Bailey and Simon, 1960] or [Finger, 2006]).
- Categorization in underwriting, pricing, but also claims management.

Insurance pricing is ex ante: uncertainty never disappears

- Before moving to AI, a couple of slides on **uncertainty** .

“An actuary is a professional with advanced mathematical skills who deals with the measurement and management of risk and uncertainty,” [Wikipedia].

- An insurance premium is paid at the beginning of the coverage period, typically for one year.
- The insurer must therefore set a price *before* knowing whether a claim will occur.
- In that sense, **insurance pricing is fundamentally an ex ante exercise** .
- The actuary does not try to identify with near-certainty who will have an accident next year.
- The goal is more modest and more meaningful: estimate collective probabilities and expected costs well enough to fund losses through pooling.
- Better data and better models can refine the estimate of accident frequency or expected loss.

Insurance pricing is ex ante: uncertainty never disappears

- But **they do not remove uncertainty at the individual level**.
- Even with extremely rich data, losses remain intrinsically random, contingent, and context-dependent.
- There is also **epistemic uncertainty**: we never know the true model, the true parameters, or all future conditions with certainty.
- So prediction in insurance is not about certainty; it is about pricing an uncertain future in a defensible way.
- This is why **insurance relies on probabilities**, expected values, margins, and pooling rather than deterministic forecasts.
- If we could predict each individual loss almost perfectly, we would no longer be mutualizing risk; we would be individualizing cost.
- Better models improve pricing; they do not eliminate uncertainty.

Aleatory and epistemic uncertainty in insurance

- Not all uncertainty in insurance has the same nature.
- **Aleatory uncertainty** is the intrinsic variability of losses.
 - Even if we knew the model perfectly, some policyholders would still have accidents and others would not.
 - This is the randomness that insurance is designed to pool.
- **Epistemic uncertainty** comes from limited knowledge.
 - We do not know the true model, the true parameters, the true future environment, or all relevant covariates.
- Better data and better models can reduce epistemic uncertainty.
- But they cannot eliminate aleatory uncertainty.
- In practice, insurers face both at once.

Aleatory and epistemic uncertainty in insurance

- This is why **actuarial work is not only about prediction** : it is also about margins, robustness, validation, and prudence.
 - In stable settings, epistemic uncertainty may shrink with experience and portfolio size.
 - In changing environments, it may remain large or even grow again.
 - Insurance can reduce uncertainty financially, but it cannot make the future deterministic.
-
- Insurance has always been a business of prediction, AI might be a legitimate technique.
 - But it might be a new regime of prediction — more granular, more adaptive, and potentially more opaque.

Lecture 2 — AI is an umbrella term

- “AI” is not one thing and should not be taught as one thing, [Mitchell, 2019, Marcus and Davis, 2019]
- Some systems **predict** from data using statistical learning
- Some systems **recommend** or **optimize** decisions over time
- Some systems **generate** text, images, audio, code, or synthetic documents
- Some systems are mostly rule-based but marketed as AI
- Different AI families require different data, validation, and governance
- Different AI families also create different legal, ethical, and operational risks
- For actuarial audiences, the key move is from models to socio-technical systems
- Data, objective, model, human oversight, and feedback loop matter together

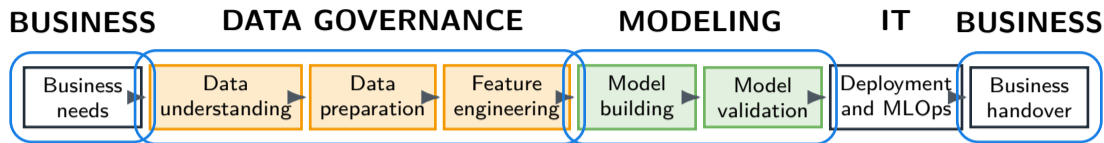
A practical taxonomy for these lectures

- **Predictive AI** : classification, regression, scoring, anomaly detection
- **Dynamic AI** : bandits, reinforcement learning, sequential decision systems
- **Generative AI** : large language models, image generators, voice cloning, summarization
- Computer vision: damage assessment, document inspection, image-based fraud signals
- NLP: claims notes, medical reports, emails, policy documents, customer interactions
- Rule engines remain important because insurance requires auditable decision logic
- Many real systems are hybrids rather than pure model categories
- The word “AI” hides this heterogeneity and can confuse governance debates
- “ **Different AIs, different promises, different harms** ”

Where AI enters the insurance value chain

- Marketing and distribution: lead scoring, personalization, chat interfaces
- Pricing: richer features, nonlinear interactions, more granular segmentation
- Underwriting: document reading, medical text extraction, anomaly flags, decision support
- Claims: triage, image-based severity estimation, document summarization, workflow routing
- Fraud: pattern detection, network analysis, synthetic evidence detection
- Customer service: assistants for explanation, status updates, and document generation
- Operations and compliance: internal search, summarization, policy checking, reporting
- Risk prevention: nudges, monitoring, alerts, and behavior-based intervention
- Reinsurance and capital: scenario analysis, exposure aggregation, portfolio monitoring

The Life Cycle of an AI Project



Data engineering

Model validation

- The process starts with a business objective: what decision should be improved?
- **Data work** translates that objective into usable information through collection, cleaning, preparation, and feature engineering.
- **Modeling** then turns data into predictions or scores, but only after proper validation.
- IT and MLOps make the model operational through deployment, monitoring, maintenance, and governance.
- The final test is business handover: a model matters only if it is used, trusted, and embedded in real decisions.

The most important conceptual shift (categories → personalization)

- Insurance used to rely heavily on **coarse categories and limited observability** .
- AI plus new data infrastructures can make individuals far more legible.
- This shifts insurance from broad pooling toward **granular classification** .
- It also shifts power from ex post compensation toward ex ante prediction and intervention.
- A telematics-based product does not merely price risk; it can reshape behavior.
- A health app does not merely observe risk; it can become part of risk management.
- An LLM does not merely summarize documents; it changes administrative scale and speed.
- The real issue is therefore not just better forecasting but different institutional logics.
- This is why AI in insurance is an **economic and political topic** ,
not only a technical one .

Parallel implications for finance professions

- Insurance and finance share core tasks: risk classification, pricing, monitoring, allocation.
- Credit scoring and insurance underwriting are close cousins conceptually.
- Portfolio management and pricing both depend on prediction under uncertainty.
- Fraud, AML, compliance, and document-heavy work are natural targets for AI automation.
- In both sectors, **data asymmetries can create market power and exclusion risks** .
- In both sectors, **personalization can increase efficiency but also opacity** .
- In both sectors, **explainability becomes more difficult when workflows become modular and hybrid** .
- In both sectors, **the profession moves from manual analysis toward model supervision** .

Machine learning works by generalization

- In **supervised learning**, we observe pairs (\mathbf{x}_i, y_i) : features and outcomes.
- We do not want a model that only reproduces the past.
- We want a model that **predicts well on new observations**, [Denuit et al., 2020, Denuit et al., 2019].
- This is why the data are usually split at random into a training sample and a validation (or test) sample.
- On the training sample, we fit several candidate models.
- On the validation sample, we compare how well those models predict outcomes that were not used for fitting.
- We then keep the model with the best out-of-sample predictive performance.
- The core idea is **generalization**: learning a pattern that extends beyond the fitted sample.

Machine learning works by generalization

- This also explains why overfitting is dangerous: a model can fit the training data extremely well and still predict poorly on new data.
- In short: machine learning is not mainly about explaining the past, but about predicting unseen cases.

$$\left\{ \begin{array}{l} \hat{R}_{\text{train}}(m) = \frac{1}{n_T} \sum_{i \in \mathcal{D}_T} \ell(m(\mathbf{x}_i), y_i), \quad \longrightarrow \hat{m} \in \underset{m \in \mathcal{M}}{\text{argmin}} \{ \hat{R}_{\text{train}}(m) \} \\ \hat{R}_{\text{test}}(\hat{m}) = \frac{1}{n_V} \sum_{i \in \mathcal{D}_V} \ell(\hat{m}(\mathbf{x}_i), y_i) \end{array} \right.$$

- Idea: select models using **out-of-sample performance**, not in-sample fit.
- This corresponds to **accuracy** in machine learning.

accuracy is overrated

What explainability can and cannot do

- See Sherlock Holmes and Dr. John Watson discussion

When I looked at him he had finished reading the note, and his eyes had assumed the vacant, lack-lustre expression which showed mental abstraction.

“How in the world did you deduce that?” I asked.

“Deduce what?” said he, petulantly.

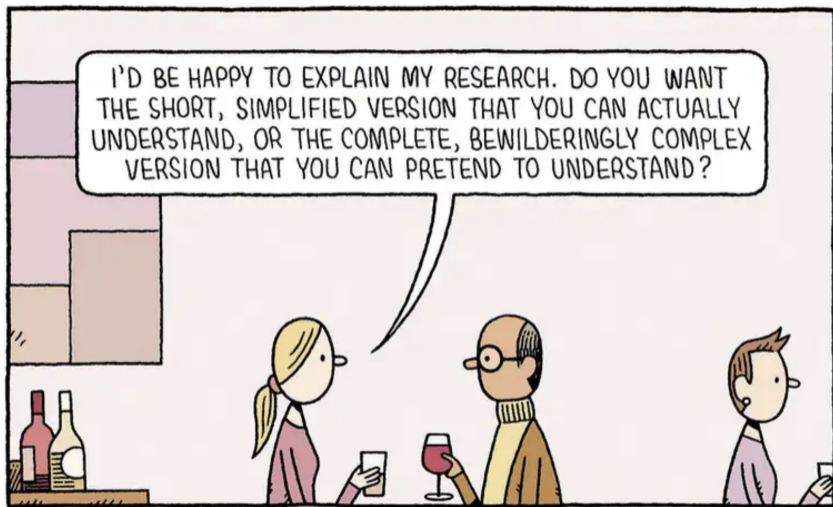
“Why, that he was a retired sergeant of Marines.”

“I have no time for trifles,” he answered, brusquely; then with a smile, *“Excuse my rudeness. You broke the thread of my thoughts; but perhaps it is as well. So you actually were not able to see that that man was a sergeant of Marines?”*

“No, indeed.”

“It was easier to know it than to explain why I knew it. If you were asked to prove that two and two made four, you might find some difficulty, and yet you are quite sure of the fact... [Conan-Doyle, 1887]

What explainability can and cannot do



(source [Gauld, 2021])

TOM GAULD for NEW SCIENTIST

What explainability can and cannot do

- The rhetoric of insurance exclusion – numbers, objectivity and statistics – forms what Brian Glenn calls

“the myth of the actuary,” “a powerful rhetorical situation in which decisions appear to be based on objectively determined criteria when they are also largely based on subjective ones” or “the subjective nature of a seemingly objective process.” “Virtually every aspect of the insurance industry is predicated on stories first and then numbers,”

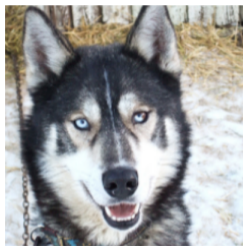
See [Glenn, 2000, Glenn, 2003]

- Importance of **interpretation** and **explainability** of models
- Some models have a high accuracy... for wrong reasons...

What explainability can and cannot do

- E.g., classifiers, $y \in \{0, 1\}$
- why a prediction of $\hat{y} = 1$?

“On a collection of additional 60 images, the classifier predicts “Wolf” if there is snow (or light background at the bottom), and “Husky” otherwise, regardless of animal color, position, pose, etc.”
[Ribeiro et al., 2016]



(a) Husky classified as wolf



(b) Explanation

Figure 11: Raw data and explanation of a bad model’s prediction in the “Husky vs Wolf” task.

| | Before | After |
|-----------------------------|--------------|--------------|
| Trusted the bad model | 10 out of 27 | 3 out of 27 |
| Snow as a potential feature | 12 out of 27 | 25 out of 27 |

Table 2: “Husky vs Wolf” experiment results.

Machine learning works by generalization

- Explainability is not one single thing, [Raji et al., 2020, Molnar, 2025].
- A regulator, an actuary, a customer, and a model developer do not ask for the same explanation.
- Sometimes we want a **global explanation** : how does the model behave overall?
- Sometimes we want a **local explanation** : why did this individual receive this score or decision?
- Sometimes we want a **counterfactual explanation** : what would have to change for the decision to change?
- But an explanation of a prediction is not automatically an explanation of the world.
- A variable can be important for a model without being a genuine cause of the outcome.
- A local explanation can describe the score without telling us whether the decision is fair, robust, or legally defensible.

Machine learning works by generalization

- In insurance and finance, explainability is therefore not only technical; it is also operational, legal, and institutional.
- Explaining a model is not the same as explaining reality. If $\hat{y} = m(\mathbf{x})$
- **explainability usually describes m , not the mechanism generating y**
- Model explanation, local/global explanation, and limits of interpretability.

Predicting is not understanding — and understanding is not predicting

- A model may predict very well while telling us little about the underlying mechanism.
- Conversely, a structural or causal model may clarify mechanisms and still have modest predictive performance.
 - Prediction asks: who is likely to claim, default, lapse, or commit fraud?
 - Understanding asks: what generates that behavior, and what would happen if we intervened?
- These are different scientific goals.
- In practice, **insurance often uses prediction** for pricing, underwriting, triage, and fraud detection.
- But prevention, regulation, and public policy often require **understanding** rather than mere forecasting.
- This is why predictive accuracy cannot be the only criterion of model quality.

Predicting is not understanding — and understanding is not predicting

- A highly accurate model may still be useless for intervention.
- A less predictive model may still be more **useful for governance** if it identifies plausible mechanisms.
- Prediction \neq understanding, i.e.,

$$\min_m \{ \mathbb{E}[\ell(m(\mathbf{X}), Y)] \} \neq \text{what changes } Y \text{ when } \mathbf{X} \text{ is changed?}$$

- A predictive model answers “what is likely?” not necessarily “why?” or “what should we change?”
- “Why?” is usually an important question \leftarrow **causal inference**

Conditional is not causal

- One of the central confusions in **AI (ML) is to read association as causation**, [Pearl, 2009, Imbens and Rubin, 2015].
- A model may estimate how outcomes vary with observed covariates.
- But this does not tell us what would happen under an intervention.
- In other words, $\mathbb{E}[Y \mid \mathbf{X} = \mathbf{x}]$ is not the same object as the effect of setting $\mathbf{X} = \mathbf{x}$ (or possibly a subset of features) \rightarrow $\text{do}(\cdot)$ operator, see $\mathbb{E}(Y \mid \text{do}(\mathbf{x}))$.
- This distinction matters everywhere: pricing, underwriting, prevention, public policy, and fairness analysis.
- **If a variable is predictive, it may still be a proxy, a marker, or a consequence rather than a cause**.
- Large datasets do not automatically solve this problem.
- More data can improve prediction while leaving causal ambiguity untouched.
- This is why **counterfactual reasoning matters**.

Conditional is not causal

- The key question becomes: compared with what alternative state of the world?

$$\mathbb{E}[Y | \mathbf{X} = \mathbf{x}] \neq \mathbb{E}(Y | \text{do}(\mathbf{X} = \mathbf{x}))$$

- $\mathbb{E}(Y | \text{do}(\mathbf{X} = \mathbf{x}))$ denotes the average outcome under an intervention that forces \mathbf{X} to be equal to \mathbf{x} (coined as “intervention”).
- It can be read as an average potential outcome: what would happen, on average, if we actively set $\mathbf{X} = \mathbf{x}$.
- $\mathbb{E}[Y | \mathbf{X} = \mathbf{x}]$ is the average outcome among units naturally observed with $\mathbf{X} = \mathbf{x}$.
- Conditional relationships are useful for prediction; causal effects are needed for intervention.

Observing is not experimenting

- **Most insurance and financial data are observational, not experimental**, [Charpentier, 2024].
- Firms observe customers, applicants, contracts, claims, and defaults after many selection mechanisms have already operated.
- In observational data, the treated and untreated groups are often not comparable.
- This is why **naive comparisons can be badly misleading**.
- A difference in outcomes may reflect treatment, but it may also reflect selection, background risk, or omitted variables.
- Randomized experiments remain the cleanest benchmark when they are feasible.
- When they are not feasible, we look for approximate designs: matching, propensity scores, difference-in-differences, regression discontinuity, natural experiments.
- This matters for telematics, prevention programs, fraud alerts, customer nudges, model-driven interventions, etc.

Observing is not experimenting

- AI does not remove the need for identification.
- Better prediction does not eliminate the **need for careful causal design**.

$$\tau = \mathbb{E}[Y(1) - Y(0)] \text{ but } \mathbb{E}[Y | T = 1] - \mathbb{E}[Y | T = 0]$$

- Observational comparisons are easy to compute and hard to interpret.

Bias often comes from the data-generating process

- Bias does not come only from the algorithm, [Sculley et al., 2015, Noble, 2018].
- It may arise much earlier: sampling, observation, measurement, labeling, portfolio construction, or survival in the data.
 - Sampling bias appears when the observed sample is not representative of the intended population.
 - Survival bias appears when we only observe the cases that remain visible, accepted, or successful.
 - Measurement bias appears when outcomes or covariates are poorly measured or recorded differently across groups.
- **Spurious correlation** appears when variables move together for reasons that are not substantively meaningful.
- In insurance, many datasets are shaped by past underwriting rules, past pricing, and past claims handling.

Bias often comes from the data-generating process

- This means that the data already contain **institutional history**.
- A model can therefore be **perfectly fitted to biased data**.
- So the real question is not only “is the model biased?” but also “how was the dataset produced?”

$$\mathbb{P}(\mathbf{X}, Y \mid \text{observed}) \neq \mathbb{P}(\mathbf{X}, Y \mid \text{target population})$$

- Garbage in, garbage out is too weak: the deeper issue is often selective visibility.
- Classical problem is Simpson's paradox (omitted variable interpretation problem).

I admit that lectures on AI is an opportunity to get back to standard questions on statistical modeling...

Bias often comes from the data-generating process

| | Total | Men | Women | Proportions |
|-------|------------------|------------------------|----------------------|-------------|
| Total | 5233/12763 ~ 41% | 3714/8442 ~ 44% | 1512/4321 ~ 35% | 66%-34% |
| Top 6 | 1745/4526 ~ 39% | 1198/2691 ~ 45% | 557/1835 ~ 30% | 59%-41% |
| A | 597/933 ~ 64% | 512/825 ~ 62% | 89/108 ~ 82% | 88%-12% |
| B | 369/585 ~ 63% | 353/560 ~ 63% | 17/ 25 ~ 68% | 96%- 4% |
| C | 321/918 ~ 35% | 120/325 ~ 37% | 202/593 ~ 34% | 35%-65% |
| D | 269/792 ~ 34% | 138/417 ~ 33% | 131/375 ~ 35% | 53%-47% |
| E | 146/584 ~ 25% | 53/191 ~ 28% | 94/393 ~ 24% | 33%-67% |
| F | 43/714 ~ 6% | 22/373 ~ 6% | 24/341 ~ 7% | 52%-48% |

Data from [Bickel et al., 1975]. Formalized as follows: S is the (binary) genre, Y the admission decision, and X the program (category),

Bias often comes from the data-generating process

$$\mathbb{P}[\text{admission} \mid S = \text{men}] \geq \mathbb{P}[\text{admission} \mid S = \text{women}]$$

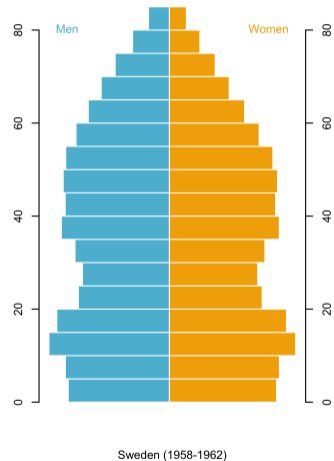
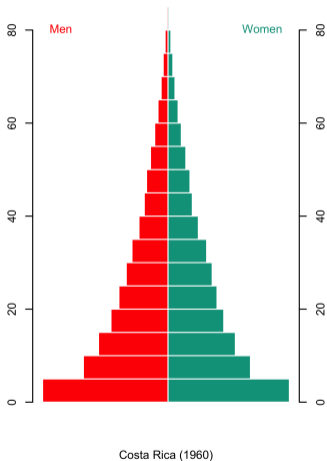
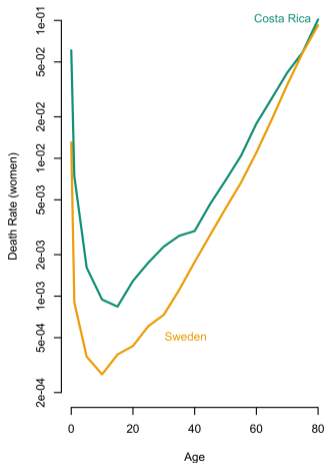
overall admission

$$\mathbb{P}[\text{admission} \mid X = x, S = \text{men}] \leq \mathbb{P}[\text{admission} \mid X = x, S = \text{women}], \forall x.$$

conditional on program

“the bias in the aggregated data stems not from any pattern of discrimination on the part of admissions committees, which seems quite fair on the whole, but apparently from prior screening at earlier levels of the educational system. Women are shunted by their socialization and education toward fields of graduate study that are generally more crowded, less productive of completed degrees, and less well funded, and that frequently offer poorer professional employment prospects,” [Bickel et al., 1975]

Bias often comes from the data-generating process



Overall mortality rate for women, **8.12‰** in Costa Rica, against **9.29‰** in Sweden.

From Using AI to Insuring AI

- So far, we have asked how AI changes insurance.
- We should also ask the reverse question: **how can insurance deal with AI?** , [Hufeld et al., 2016]
- AI is already embedded in products, workflows, advice, customer service, pricing, compliance, and operations.
- This means that AI risk is often already present inside existing insurance portfolios.
- In many cases, it appears as a form of **silent coverage** : neither clearly included nor clearly excluded.
- The key difficulty is that AI risk is not a single risk.
- It combines operational risk, liability risk, cyber risk, model risk, governance risk, and reputational risk.
- It is also dynamic: models change, APIs change, prompts change, retrieval pipelines change, users adapt.

From Using AI to Insuring AI

- So the question is not simply “can AI be insured?”
- The deeper question is: **under what conditions can AI-related losses become insurable risks?**
- Insurance works best when losses are sufficiently heterogeneous and not perfectly correlated,.
- AI often creates the opposite structure: common providers, common models, common updates, common vulnerabilities [Gai, 2013].
- A single defect may generate not one claim, but thousands of similar claims at once.
- This is why insurers worry less about one very large loss than about many simultaneous losses.
- In that sense, **AI risk can become a systemic risk** .
- Historical claims data are often scarce, unstable, or misleading.
- The environment is **non-stationary** : performance today does not guarantee performance tomorrow.

From Using AI to Insuring AI

- Many losses may also be **latent** : the damage appears only months after deployment.
- **Liability is often diffuse** : developer, deployer, user, vendor, integrator, or downstream operator?
- Insurability therefore depends not only on frequency and severity, but also on dependence and attribution.

Insurance as a Technology of Governance

- Insurance does not only pay after the loss.
- **Insurance also creates incentives before the loss** .
- For AI, this may be the most important role of insurance.
- A serious insurer would want logs, documentation, versioning, monitoring, incident response, and audit trails.
- It would also care about red teaming, testing, fallback procedures, human escalation, and change management.
- The hidden cost of AI is often not generation, but verification.
- If nobody pays for proof ex ante, everyone may pay for the loss ex post.
- This is why internal auditing matters.
- Real auditing is not cosmetic: it is slow, methodical, documented, and revisable.
- In other words, insuring AI may require building an **infrastructure of proof** .

What Would an AI Insurance Contract Need?

- Need for a clear definition of the insured event.
 - What exactly counts as a loss: third-party damage, compliance failure, bad advice, operational disruption, harmful output?
- Need for a precise technical perimeter.
 - Which model, which version, which provider, which retrieval system, which prompts, which tools?
- Need for a change regime.
 - When does a model update, a fine-tune, or a new agent architecture change the insured risk?
- Duties of the insured: monitoring, logging, evidence retention, periodic testing, incident reporting.
- Audit clauses and rights of review.

What Would an AI Insurance Contract Need?

- Limits (or sublimits) for common dependencies, common providers, or systemic events.
- Possibly: usage-based pricing, bonus-malus mechanisms, or parametric triggers.
- Need to stress that pricing would depend on technical and governance evidence.

Insuring AI: New Risk or Old Insurance Questions in New Form?

- Some aspects are genuinely new: speed, scale, software monoculture, continuous updates, agentic workflows.
- **But many questions are familiar to insurance theory .**
- Correlation and accumulation are classic problems of insurability.
- Limited history is not new: catastrophe models also rely on a priori modeling and expert judgment.
- Deep uncertainty is not new: insurance has long distinguished aleatory and epistemic uncertainty.
- Information asymmetry is not new: the insured usually knows more than the insurer.
- Moral hazard is not new: once protected, actors may underinvest in care or verification.
- So the real challenge is not to abandon actuarial thinking.

Insuring AI: New Risk or Old Insurance Questions in New Form?

- It is to **reconnect actuarial science, contract design, model governance, and technical auditing**.
- This is why insuring AI is a natural bridge toward broader questions of accountability and regulation.

- The next question is therefore not only whether AI changes insurance, but how AI is concretely used in insurance.
- This means moving from general debates to operational questions: use cases, decision points, value creation, and governance.

Lecture 3 — From probabilistic actuarial models to machine learning

- **Actuarial science was historically built on probability and statistics**.
- The core objects are random losses, conditional expectations, frequencies, severities, and portfolio distributions.
- This framework makes uncertainty explicit: we can discuss means, variances, tail events, confidence, calibration, and solvency.
- In that tradition, a pricing or reserving model starts from a probabilistic representation of risk.
- Machine learning is often organized differently in practice.
- Instead of specifying a full stochastic model, we often search for a function m that predicts outcomes well from observed covariates, and a loss ℓ .
- In supervised learning, the goal is typically to approximate a regression function or a classifier from data.
- In ensemble methods, we often aggregate several such functions rather than postulate one transparent probabilistic mechanism.

Lecture 3 — From probabilistic actuarial models to machine learning

- This is powerful for prediction, but it can make uncertainty harder to interpret and communicate.
- So the contrast is not statistics versus AI.
- The contrast is often: **explicit probabilistic modeling of uncertainty versus empirical function approximation for prediction**.
- This distinction matters in insurance, because insurers do not only need scores; they also need prices, reserves, capital, and explanations.

Actuarial view: $Y \sim \mathbb{P}_\theta(\cdot | \mathbf{X} = \mathbf{x})$, and $m_\theta(\mathbf{x}) = \mathbb{E}[Y | \mathbf{X} = \mathbf{x}]$

$$\text{ML view: } \hat{m} = \underset{m \in \mathcal{M}}{\operatorname{argmin}} \left\{ \frac{1}{n} \sum_{i=1}^n \ell(m(x_i), y_i) \right\}$$

- Nuance: some machine-learning methods are probabilistic; but many practical workflows focus on predictive performance rather than explicit stochastic structure.

Personalization, an epistemological perspective

- Without a probabilistic model, how can we make sure that $\hat{m}(\mathbf{x}) = 12.345\%$ is a good approximation of $\mathbb{P}[Y = 1 \mid \mathbf{X} = \mathbf{x}]$?
- Related to **probabilistic interpretation for single case event** (no LLN)

“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, e.g., ‘all insured men forty-one years old living in a given country and not engaged in certain dangerous occupations’. A probability of death is attached to the class of men or to another class that can be defined in a similar way. The phrase ‘probability of death’, when it refers to a single person, has no meaning for us at all,”

[von Mises, 1928, von Mises, 1939]

- This has not much to do with “accuracy”, but it has to do with **“calibration”**.

Personalization, an epistemological perspective

As explained in [Van Calster et al., 2019], "*among patients with an estimated risk of 20%, we expect 20 in 100 to have or to develop the event,*"

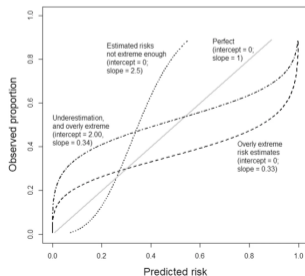
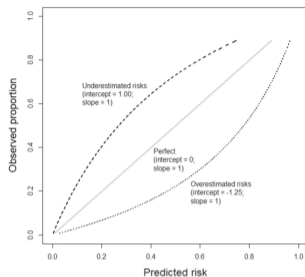
- If 40 out of 100 in this group are found to have the disease, the risk is underestimated
- If we observe that in this group, 10 out of 100 have the disease, we have overestimated the risk.

The prediction $\hat{m}(\mathbf{X})$ of Y is a well-calibrated prediction if

20 out of 100 (proportion $y = 1$)

$$\mathbb{E}[Y \mid \hat{Y} = \hat{y}] = \hat{y}, \forall \hat{y}$$

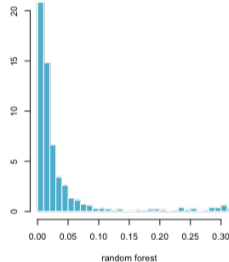
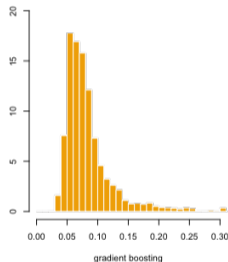
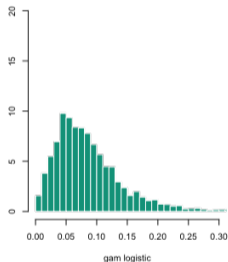
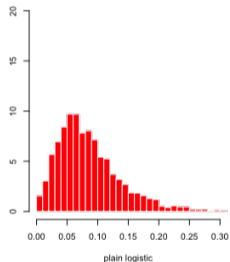
↑
estimate risk $\hat{y} = 20\%$



Personalization, an epistemological perspective

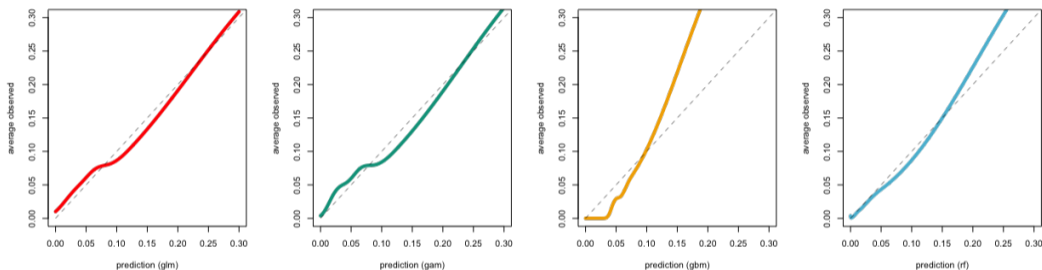
- A related challenging question is “would could be the true underlying distribution of the score” ?

Annual claim frequency, **GLM**, **GAM**, **Random Forest** and **Gradient Boosting**



| | $\hat{\pi}^{\text{glm}}$ | $\hat{\pi}^{\text{gam}}$ | $\hat{\pi}^{\text{gbm}}$ | $\hat{\pi}^{\text{rf}}$ |
|------------------------------------|--------------------------|--------------------------|--------------------------|-------------------------|
| average $\hat{\pi}(\mathbf{x})$'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 |

Personalization, an epistemological perspective



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

Privacy, surveillance, and telematics

- Telematics is one of the clearest examples of AI-related tension in insurance
- It promises better pricing, better incentives, and sometimes safer behavior
- It also raises deep issues of consent, monitoring, and boundary-setting
- The insurer may observe not only outcomes, but patterns of everyday behavior
- This moves insurance closer to continuous surveillance than traditional contracting
- A key distinction is ex post information versus ex ante and real-time information
- Another key distinction is voluntary data sharing versus de facto coercive adoption
- Would you trade privacy for a cheaper premium?

Generative AI and fraud

- Generative AI creates a two-sided race between detection and deception, [Darby and Karni, 1973]
- It can help summarize claims files, draft communications, and detect anomalies
- But it can also generate fake invoices, fake photos, fake voices, and fake narratives, [Anderson, 2020]
- Fraud may become more scalable, more persuasive, and cheaper to produce
- **The problem is not only accuracy but evidentiary trust**
- If synthetic content becomes easy, verification becomes more expensive and more central
- Claims management may therefore become a battle over authenticity
- If seeing is no longer believing, how should insurance verify loss?
- This topic also connects insurance to cybersecurity and platform governance

Discrimination and unfair exclusion

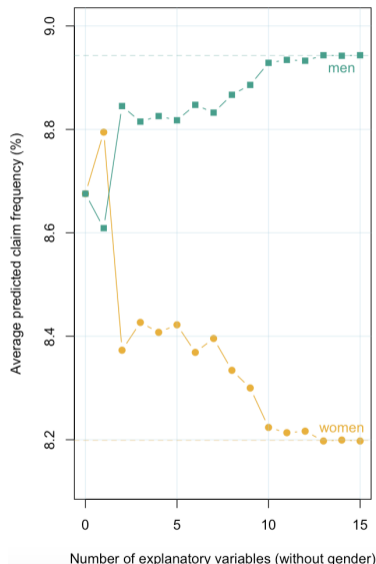
- Insurance has always classified people; AI changes the scale and subtlety of classification
- Bias does not require explicit use of protected variables
- Proxy variables, data quality problems, and feedback effects can create unfair outcomes
- **There is a difference between actuarial differentiation and socially unacceptable discrimination**
- That boundary is legal, ethical, and political rather than purely technical
- More variables do not automatically mean more fairness
- More predictive accuracy can coexist with more exclusion
- **Should insurers be allowed to use any variable that predicts risk?**
- Partial answer: removing a sensitive variable is never the good answer...

Discrimination and unfair exclusion

- 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 ← **generalization**



Discrimination and unfair exclusion

- Dealing with “discrimination” in insurance is challenging...

*“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]*

Several definitions of “fairness” or “non-discriminatory”

demographic parity $\rightarrow \mathbb{E}[\hat{Y} | S = A] \stackrel{?}{=} \mathbb{E}[\hat{Y} | S = B]$

sensitive (green arrow pointing to $S = A$)

sensitive (orange arrow pointing to $S = B$)

score \hat{y} (blue arrow pointing from $S = A$ to $S = B$)

equalized odds $\rightarrow \mathbb{E}[\hat{Y} | Y = y, S = A] \stackrel{?}{=} \mathbb{E}[\hat{Y} | Y = y, S = B], \forall y$

outcome y (orange arrow pointing to $Y = y$)

score \hat{y} (blue arrow pointing from $S = A$ to $S = B$)

calibration $\rightarrow \mathbb{E}[Y | \hat{Y} = u, S = A] \stackrel{?}{=} \mathbb{E}[Y | \hat{Y} = u, S = B], \forall u$

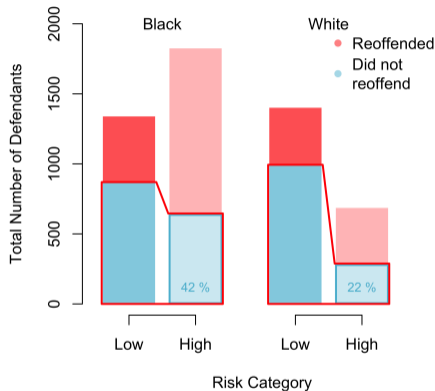
outcome y (orange arrow pointing to Y)

score \hat{y} (blue arrow pointing from $\hat{Y} = u$ to $\hat{Y} = u$)

Isn't it a problem to have several definitions?

From [Feller et al., 2016],

- for White people, among those who did not re-offend (y), 22% were wrongly classified (\hat{y}),
- for Black people, among those who did not re-offend, 42% were wrongly classified,
- **Problem**, since $42\% \gg 22\%$

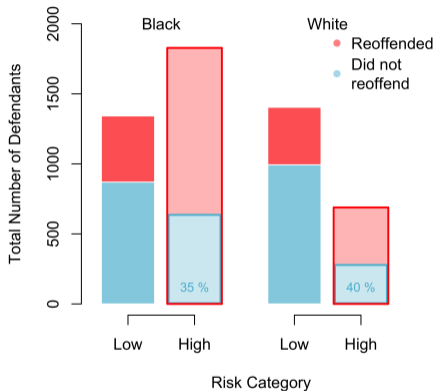


$$\mathbb{P}[\hat{Y} = \text{high} \mid Y = \text{no}, S = \text{black}] = 42\% \stackrel{?}{=} \mathbb{P}[\hat{Y} = \text{high} \mid Y = \text{no}, S = \text{white}] = 22\%$$

Isn't it a problem to have several definitions?

From [Dieterich et al., 2016],

- for White people, among those who were classified as high risk (\hat{y}), 40% did not re-offend (y),
- for Black people, among those who were classified as high risk (\hat{y}), 35% did not re-offend (y),
- **No problem**, since $35 \approx 40\%$



$$\mathbb{P}[Y = \text{no} \mid \hat{Y} = \text{high}, S = \text{black}] = 35\% \stackrel{?}{=} \mathbb{P}[Y = \text{no} \mid \hat{Y} = \text{high}, S = \text{white}] = 40\%.$$

Isn't it a problem to have several definitions?

- 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,”* [?], cited in [?]
- To counteract **disparate impact**, intentional **disparate treatment** is necessary
- See philosophical discussions about **affirmative action**, e.g., [?, ?, ?]
- 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,”* [?] and [?]
- 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**

Granularity and the veil of ignorance

- **Insurance relies on partial ignorance about individual (future) losses**
- That ignorance makes pooling both possible and acceptable
- If AI removes too much of that ignorance, the logic of mutualization may weaken
- The premium starts to look like an individualized price rather than a shared contribution
- Extreme segmentation can make solidarity fragile even if it improves predictive accuracy
- This is why **granularity is not just a technical parameter but a social choice**
- One provocative claim: perfect risk classification could destroy part of what insurance is for

Governance and regulation

- Once AI is understood as a system, governance can no longer focus only on the model
- Questions include data provenance, validation, oversight, monitoring, and recourse
- Governance also includes vendor dependence, documentation, and auditability
- Human-in-the-loop is not enough if the human cannot meaningfully challenge the output
- Risk management frameworks matter because performance is not the only objective
- In finance and insurance, rights, fairness, and trust are load-bearing constraints
- **Regulation is becoming central**
- A good line: high-performance AI in a low-trust system is not a success
- Another good line: governance is part of model design, not an afterthought

What changes for actuarial and financial professions?

- The profession moves from hand-crafted calculation toward model supervision and judgment
- Less time may be spent on routine analysis and more on framing, validation, and challenge
- Professionals will need to evaluate data quality, model limits, and workflow design
- They will also need to **communicate uncertainty** to managers, regulators, and customers (ideally ex ante)
- The human comparative advantage may shift toward problem formulation and accountability
- Future actuaries may need to understand statistics, machine learning, and institutional design
- AI will not remove judgment; it will relocate judgment

Take away messages

- Insurance is about pooling uncertainty, not eliminating it.
- AI does not enter insurance from the outside: insurance has always been a business of prediction.
- But AI changes the regime of prediction: more granular, more adaptive, and more opaque.
- AI is not one thing: different systems create different promises, risks, and governance needs.
- An AI project should start from a business decision, not from a model.
- A model creates value only when it is embedded in data, workflows, IT, and real organizational use.
- Accuracy is overrated: focus on calibration, uncertainty, and fairness.
- Prediction is not understanding: conditional patterns are not causal effects.
- More personalization is not always progress: granularity can weaken pooling, solidarity, and insurability.
- Governance is part of model design, not an afterthought.

References

- Anderson, R. (2020).
Security Engineering: A Guide to Building Dependable Distributed Systems.
Wiley, Indianapolis, IN, 3 edition.
- 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.
- Bénéplanc, G., Charpentier, A., and Thourot, P. (2022).
Manuel d'assurance.
Presses Universitaires de France.

References

- Berliner, B. (1982).
Limits of Insurability of Risks.
Prentice-Hall, Englewood Cliffs, NJ.
- Bickel, P. J., Hammel, E. A., and O'Connell, J. W. (1975).
Sex bias in graduate admissions: Data from berkeley.
Science, 187(4175):398–404.
- Borch, K. H. (1990).
Economics of Insurance.
North-Holland, Amsterdam.
- 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.

References

- Charpentier, A. (2014).
Computational actuarial science with R.
CRC press.
- Charpentier, A. (2024).
Insurance: biases, discrimination and fairness.
Springer Verlag.
- Charpentier, A. and Denuit, M. (2004).
Mathématiques de l'assurance non-vie - Tarification et provisionnement (Tome 1).
Economica.
- Conan-Doyle, A. (1887).
A study in scarlet.
Ward Lock & Co's.
- Cummins, J. D., editor (2002).
Deregulating Property-Liability Insurance: Restoring Competition and Increasing Market Efficiency.
AEI-Brookings Joint Center for Regulatory Studies, Washington, DC.

References

- Darby, M. R. and Karni, E. (1973).
Free competition and the optimal amount of fraud.
The Journal of law and economics, 16(1):67–88.
- Denuit, M., Hainault, D., and Trufin, J. (2019).
Effective Statistical Learning Methods for Actuaries III (Neural Networks and Extensions).
Springer Verlag.
- Denuit, M., Hainault, D., and Trufin, J. (2020).
Effective Statistical Learning Methods for Actuaries II (Tree-based methods and Extensions).
Springer Verlag.
- Dieterich, W., Mendoza, C., and Brennan, T. (2016).
Compas risk scales: Demonstrating accuracy equity and predictive parity.
Northpointe Inc, 7(7.4):1.

References

- Feller, A., Pierson, E., Corbett-Davies, S., and Goel, S. (2016).
A computer program used for bail and sentencing decisions was labeled biased against blacks. it's actually not that clear.
The Washington Post, October 17.
- Feng, R. (2023).
Decentralized insurance.
Springer.
- 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.
- Gai, P. (2013).
Systemic Risk: The Dynamics of Modern Financial Systems.
Oxford University Press, Oxford.

References

- Gauld, T. (2021).
Two ways to explain scientific research.
New Scientist.
- Glenn, B. J. (2000).
The shifting rhetoric of insurance denial.
Law and Society Review, pages 779–808.
- Glenn, B. J. (2003).
Postmodernism: the basis of insurance.
Risk Management and Insurance Review, 6(2):131–143.
- Hufeld, F., Koijen, R. S. J., and Thimann, C., editors (2016).
The Economics, Regulation, and Systemic Risk of Insurance Markets.
Oxford University Press, Oxford.

References

- Imbens, G. W. and Rubin, D. B. (2015).
Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.
Cambridge University Press, Cambridge.
- Kaas, R., Goovaerts, M., Dhaene, J., and Denuit, M. (2008).
Modern actuarial risk theory: using R.
Springer.
- Kerner, O. (1968).
Report of The National Advisory Commission on Civil Disorder.
Bantam Books.
- Lemaire, J., Park, S. C., and Wang, K. C. (2016).
The use of annual mileage as a rating variable.
ASTIN Bulletin: The Journal of the IAA, 46(1):39–69.

References

- Marcus, G. and Davis, E. (2019).
Rebooting AI: Building Artificial Intelligence We Can Trust.
Pantheon Books, New York.
- McNeil, A. J., Frey, R., and Embrechts, P. (2015).
Quantitative Risk Management: Concepts, Techniques and Tools.
Princeton University Press, Princeton, revised edition edition.
- Mitchell, M. (2019).
Artificial Intelligence: A Guide for Thinking Humans.
Farrar, Straus and Giroux, New York.
- Molnar, C. (2025).
Interpretable Machine Learning: A Guide for Making Black Box Models Explainable.
Self-published, 3 edition.

References

- 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.
- Munroe, R. (2014).
What if?: Serious scientific answers to absurd hypothetical questions.
Houghton Mifflin Harcourt.
- Noble, S. U. (2018).
Algorithms of Oppression: How Search Engines Reinforce Racism.
NYU Press, New York.
- Parasuraman, R. and Riley, V. (1997).
Humans and automation: Use, misuse, disuse, abuse.
Human factors, 39(2):230–253.

References

- Pearl, J. (2009).
Causality: Models, Reasoning, and Inference.
Cambridge University Press, Cambridge, 2 edition.
- Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., and Barnes, P. (2020).
Closing the ai accountability gap: Defining an end-to-end framework for internal algorithmic auditing.
In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (ACM FAccT)*, pages 33–44.
- Ribeiro, M. T., Singh, S., and Guestrin, C. (2016).
" why should i trust you?" explaining the predictions of any classifier.
In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144.

References

- Schauer, F. (2006).
Profiles, probabilities, and stereotypes.
Harvard University Press.
- Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., and Dennison, D. (2015).
Hidden technical debt in machine learning systems.
Advances in neural information processing systems, 28.
- Van Calster, B., McLernon, D. J., Van Smeden, M., Wynants, L., and Steyerberg, E. W. (2019).
Calibration: the achilles heel of predictive analytics.
BMC medicine, 17(1):1–7.
- von Mises, R. (1928).
Wahrscheinlichkeit Statistik und Wahrheit.
Springer.

References

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