Granularity, Mutualization, and AI Risk in Insurance

Arthur Charpentier
Université du Québec à Montréal & 京都大学

2026 International Workshop on Risk and Insurance, 서울
(AI Revolution and Cyber Risks Session)

Granularity and Mutualization

- Turning heterogeneity into risk classes
- Information asymmetry: private information and selection
- Information granularity: reducing epistemic uncertainty?
- Natural catastrophe maps
- Actuarial Fairness

AI Risk in Insurance

- From AI for insurance to insurance for AI
- Common models, common failures
- Prioritizing AI risks
- LLMs: plausible outputs, costly verification
- Proof, traceability, and insurability
- Vibe coding

Granularity and Mutualization

Turning heterogeneity into risk classes

- Variance decomposition, for some latent risk factor Θ,

$$\mathrm{Var}[Y] = \underbrace{\mathbb{E}[\mathrm{Var}[Y|\Theta]]}_{\text{within}} + \underbrace{\mathrm{Var}[\mathbb{E}[Y|\Theta]]}_{\text{between}}.$$

- In practice, insurers observe only a vector of covariates $\boldsymbol{X}$,

$$\mathrm{Var}[Y] = \underbrace{\mathbb{E}[\mathrm{Var}[Y|\boldsymbol{X}]]}_{\to \text{ insurer}} + \underbrace{\mathrm{Var}[\mathbb{E}[Y|\boldsymbol{X}]]}_{\to \text{ policyholder}}.$$

- The residual uncertainty decomposes as

$$\mathbb{E}[\mathrm{Var}[Y|\boldsymbol{X}]] = \underbrace{\mathbb{E}[\mathrm{Var}[Y|\Theta]]}_{\text{irreducible risk}} + \underbrace{\mathbb{E}[\mathrm{Var}[\mathbb{E}[Y|\Theta]|\boldsymbol{X}]]}_{\textbf{epistemic uncertainty}}.$$

- [De Pril and Dhaene, 1996] interpret this distinction as the difference between random solidarity and subsidiary solidarity, while [Charpentier and Denuit, 2004, Charpentier and Denuit, 2005, Antonio and Valdez, 2012] show how observable heterogeneity is translated into tariff classes.

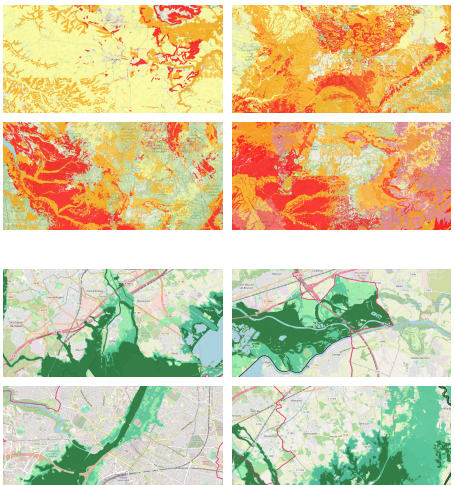
Information asymmetry: private information and selection

- Insurance markets are vulnerable to asymmetric information : when policyholders know more about their risk than insurers, pooling contracts may attract high-risk types and destabilize the market [Akerlof, 1970, Rothschild and Stiglitz, 1976].
- Empirical tests of asymmetric information ask whether coverage choices still predict losses after controlling for observable risk characteristics [Chiappori and Salanié, 2000].
- Genetic information made this problem especially visible: private knowledge about future health risks can undermine mutualization even before any claim occurs [Hoy and Polborn, 2000, Lehtonen and Liukko, 2011].
- Risk-classification bans can protect individuals against reclassification risk, but they also change participation, cross-subsidies, and market equilibrium [Dionne and Rothschild, 2014].
- The classical concern is therefore not too much information, but unequal access to information.

Information granularity: reducing epistemic uncertainty?

- Granular information reduces epistemic uncertainty by making latent heterogeneity partially observable [Bühlmann and Gisler, 2005].
- More variables, finer spatial resolution, telematics, genetic data, and behavioural traces all make it easier to replace pooled uncertainty by classified heterogeneity [Eling et al., 2022].
- But finer classification does not only improve prediction: it also redistributes premiums, changes participation, and may weaken the social meaning of insurance [Lehtonen and Liukko, 2011].
- Removing protected variables is not enough, because granular covariates may reconstruct protected or socially sensitive information [Lindholm et al., 2022a, Lindholm et al., 2022b, Charpentier, 2024].
- Granularity also raises inclusive-insurance issues: some predictive information may be accurate but should fade over time, as in debates on criminal records, rehabilitation, and the right to be forgotten [Charpentier and Schraub, 2026].

Natural catastrophe maps: pricing granularity and withdrawal



- Vulnerability maps turn spatial heterogeneity into observable and increasingly priceable risk classes [Dalbarade et al., 2026].
- Even when natural catastrophe coverage is formally pooled, access to insurance may depend on the underlying household contract.
- Granular pricing can attract low-risk households, while high-risk areas face higher premiums, stricter conditions, or localized non-quoting.
- The key regulatory object is therefore not only the average premium, but also local availability, insurer presence, and effective withdrawal.

Source: https://www.georisques.gouv.fr/cartes-interactives/

Actuarial fairness: fairness relative to information

- Actuarial fairness is usually defined by the idea that equal risks should pay equal premiums [Charpentier, 2024].
- But "equal risks" depends on the information set: full profile, tariff class, score, group, or protected attribute.
- Coarser information gives weaker notions of fairness: global balance, calibration by score, calibration by group, or full risk-based pricing [Charpentier et al., 2026a].
- [Charpentier et al., 2026b] frame the resulting tension as a trilemma between actuarial adequacy ("actuarial fairness")), solidarity ("cross-subsidies"), and causal legitimacy.

AI Risk in Insurance

From AI for insurance to insurance for AI

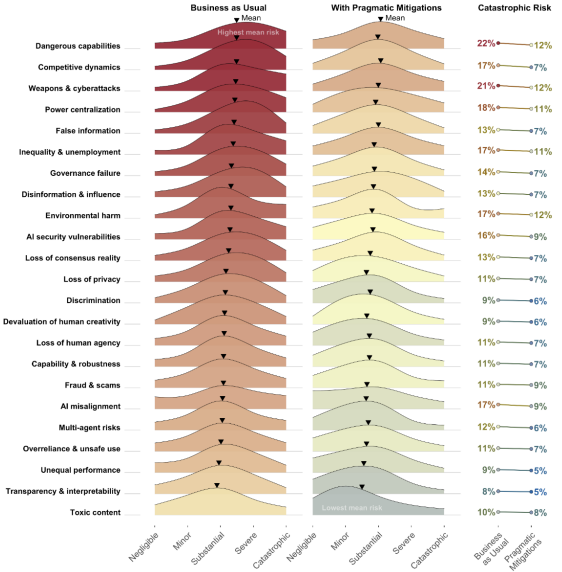
- AI is not only a tool used by insurers; it is also becoming an insured exposure embedded in firms, products, services, and professional decisions [Szpruch et al., 2025].
- Many AI losses will first appear through familiar insurance lines: cyber, technology E&O, professional liability, product liability, D&O, and regulatory risk.
- What is new is the combination of opacity, scale, speed, model updates, and difficult ex-post attribution.
- The actuarial question is therefore not simply whether AI is risky, but whether the risk can be bounded, observed, priced, monitored, and evidenced.

Common models, common failures

- AI creates accumulation risk : portfolios may look diversified by sector or geography while depending on the same models, APIs, clouds, datasets, or agent frameworks [Charpentier, 2026a].
- A model update, data poisoning event, prompt-injection vulnerability, retrieval failure, or compromised API can affect many insureds at once.
- This weakens the usual law-of-large-numbers intuition: losses may be synchronized by shared technological infrastructure .
- Underwriting therefore requires a map of the AI supply chain: provider, model version, cloud, API, data source, retrieval system, tools, agents, and change-management process.

Prioritizing AI risks: severity, vulnerability, responsibility

- [Saeri et al., 2026] use a Delphi study of 272 international experts to rank 24 AI risk domains by severity, likelihood, vulnerability, and responsibility.
- Under a business-as-usual scenario, experts judge 18 of the 24 domains to have more than a 10% probability of catastrophic outcomes by 2030.
- For insurance, the key message is the mismatch between vulnerability and control : users and affected stakeholders are most exposed, while developers, regulators, governments, and standards bodies are seen as most responsible.



LLMs: plausible outputs, costly verification

- LLM failures are not limited to hallucinations: they include retrieval errors, prompt injection, privacy leakage, biased outputs, insecure code, and incorrect tool use [Bender et al., 2021, Bommasani et al., 2021, Weidinger et al., 2021].
- Benchmarks evaluate controlled tasks, while insurance must deal with deployed systems, incentives, changing environments, and downstream losses [Bengio et al., 2025a, Bengio et al., 2025b].
- AI systems often shift the cost of reliability from production to verification: outputs are cheap, but checking them may be expensive.
- A human-in-the-loop is not a control unless the human has time, competence, authority, and incentives to challenge the model output [Elish, 2019].

Proof, traceability, and insurability

- Insurance can make proof financially relevant: coverage may depend on logs, audits, monitoring, incident reporting, and documented controls [National Institute of Standards and Technology, 2023, OECD, 2025].
- The insured system must be defined: use case, model, version, provider, data source, retrieval layer, tools, agents, and update regime.
- Governance must be operational, not symbolic: checklists do not prove that controls were active when the loss occurred [Raji et al., 2020, Mitchell et al., 2019].
- Conditional insurability follows: bounded, documented, monitored AI may be insurable; autonomous, opaque, unlogged AI may require exclusions, sublimits, or refusal.

From “code is law” to “vibe code is liability”

- [Lessig, 1999] famously argued that “code is law”. It means that software does not merely execute rules; it also shapes behavior by defining what users can do, what they cannot do, what is easy or costly, and what remains visible or hidden.
- Vibe coding turns software risk into provenance risk: after a loss, can the insured reconstruct the path from prompt to production [Charpentier, 2026b]?
- AI-generated code creates a software supply-chain exposure: prompts, model versions, generated diffs, tests, dependencies, permissions, and approvals become part of the risk record.
- The underwriting question is not only “Do you use AI?”, but “How can AI-generated code enter production?”
- Policy tools include approved-tool representations, retained logs, mandatory review, vulnerability scanning, SBOM updates, exclusions for unreviewed autonomous deployment, and aggregation clauses for common model or tool failures [Sculley et al., 2015, Pearce et al., 2021].

References

Akerlof, G. A. (1970).
The market for "Lemons": Quality uncertainty and the market mechanism.
The Quarterly Journal of Economics, 84(3):488–500.

Antonio, K. and Valdez, E. A. (2012).
Statistical concepts of a priori and a posteriori risk classification in insurance.
AStA Advances in Statistical Analysis, 96:187–224.

Bender, E. M., Gebru, T., McMillan-Major, A., and Shmitchell, S. (2021).
On the dangers of stochastic parrots: Can language models be too big?
In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pages 610–623.

Bengio, Y., Clare, S., Prunkl, C., et al. (2025a).
International AI safety report 2025: First key update: Capabilities and risk implications.
Technical report, International AI Safety Report.

Bengio, Y., Clare, S., Prunkl, C., et al. (2025b).
International AI safety report 2025: Second key update: Technical safeguards and risk management.
Technical report, International AI Safety Report.

References

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., et al. (2021).
On the opportunities and risks of foundation models.
arXiv, 2108.07258.

Bühlmann, H. and Gisler, A. (2005).
A Course in Credibility Theory and its Applications.
Universitext. Springer, Berlin, Heidelberg.

Charpentier, A. (2024).
Insurance, Biases, Discrimination and Fairness.
Springer.

Charpentier, A. (2026a).
Insuring AI: Common models, common failures, and the price of proof.
The Actuary, Forthcoming.

Charpentier, A. (2026b).
Vibe coding and the next silent cyber risk: Why provenance matters for casualty insurers.
Under review.

References

Charpentier, A., Côté, M.-P., Côté, O., and Fernandes Machado, A. (2026a).
Actuarial fairness as an information-level property: Loss ratios, autocalibration, and demographic parity.
Working paper.

Charpentier, A., Côté, O., and Côté, M.-P. (2026b).
Governing fairness in insurance pricing: The trilemma of actuarial adequacy, solidarity and causal legitimacy.
Under review.

Charpentier, A. and Denuit, M. (2004).
Mathématiques de l'assurance non-vie, Tome 1.
Economica.

Charpentier, A. and Denuit, M. (2005).
Mathématiques de l'assurance non-vie, Tome 2.
Economica.

Charpentier, A. and Schraub, D. (2026).
Criminal records, the right to be forgotten, and inclusive insurance: Risk, rehabilitation, and data governance in underwriting.
SSRN, 6170208.

References

Chiappori, P.-A. and Salanié, B. (2000).
Testing for asymmetric information in insurance markets.
Journal of Political Economy, 108(1):56–78.

Dalbarade, R., Charpentier, A., Barry, L., and Hillairet, C. (2026).
Granular pricing and effective withdrawal in climate-exposed household insurance in france.
Under review.

De Pril, N. and Dhaene, J. (1996).
Segmentering in verzekeringen.
DTEW Research Report 9648, pages 1–56.

Dionne, G. and Rothschild, C. (2014).
Economic effects of risk classification bans.
The Geneva Risk and Insurance Review, 39(2):184–221.

Eling, M., Nuessle, D., and Staubli, J. (2022).
The impact of artificial intelligence along the insurance value chain and on the insurability of risks.
The Geneva Papers on Risk and Insurance – Issues and Practice, 47:205–241.

References

Elish, M. C. (2019).
Moral crumple zones: Cautionary tales in human-robot interaction.
Engaging Science, Technology, and Society, 5:40–60.

Hoy, M. and Polborn, M. (2000).
The value of genetic information in the life insurance market.
Journal of Public Economics, 78(3):235–252.

Lehtonen, T.-K. and Liukko, J. (2011).
The forms and limits of insurance solidarity.
Journal of Business Ethics, 103(Supplement 1):33–44.

Lessig, L. (1999).
Code and Other Laws of Cyberspace.
Basic Books, New York.

Lindholm, M., Richman, R., Tsanakas, A., and Wüthrich, M. V. (2022a).
Discrimination-free insurance pricing.
ASTIN Bulletin, 52(1):55–89.

References

Lindholm, M., Richman, R., Tsanakas, A., and Wüthrich, M. V. (2022b).
A discussion of discrimination and fairness in insurance pricing.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., and Gebru, T. (2019).
Model cards for model reporting.
In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pages 220–229. ACM.

National Institute of Standards and Technology (2023).
Artificial intelligence risk management framework (AI RMF 1.0).
NIST AI 100-1, National Institute of Standards and Technology, Gaithersburg, MD.

OECD (2025).
Towards a common reporting framework for AI incidents.
OECD Artificial Intelligence Papers 34, OECD Publishing, Paris.

Pearce, H., Ahmad, B., Tan, B., Dolan-Gavitt, B., and Karri, R. (2021).
Asleep at the keyboard? assessing the security of GitHub Copilot's code contributions.

References

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*, pages 33–44. ACM.

Rothschild, M. and Stiglitz, J. (1976).
Equilibrium in competitive insurance markets: An essay on the economics of imperfect information.
The Quarterly Journal of Economics, 90(4):629–649.

Saeri, A. K., Graham, J., Noetel, M., Slattery, P., Thompson, N., et al. (2026).
Prioritization of risks from artificial intelligence: A delphi study of 272 international experts.

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.
In *Advances in Neural Information Processing Systems*, volume 28.

Szpruch, L., Orfanoudaki, A., Maple, C., Wicker, M., Bengio, Y., Lam, K.-Y., and Detyniecki, M. (2025).
Insuring AI: Incentivising safe and secure deployment of AI workflows.
SSRN working paper.

References

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., Glaese, M., Balle, B., Kasirzadeh, A., et al. (2021).
Ethical and social risks of harm from language models.
arXiv, 2112.04359.