
AI Governance for Automated Disaster Insurance Risk Modelling

Dr. Livia Monroe Halloway

Associate Professor,

Department of Insurance Analytics and Responsible Artificial Intelligence,

Cascadia Institute of Risk and Technology,

Denver, Colorado, United States

Corresponding Author: Dr. Livia Monroe Halloway

Department of Insurance Analytics and Responsible Artificial Intelligence

Cascadia Institute of Risk and Technology, Denver, Colorado, United States

Article Type: Simulation-Based Research Article

Submitted: 01-November-2024 **Revised:** 08-November-2024

Accepted: 10-November-2024 **Published:** 15-November-2024

Abstract: Artificial intelligence is increasingly capable of integrating geospatial, meteorological, structural, socioeconomic, satellite, and historical claims information into automated disaster-insurance risk models. These systems can improve the speed and granularity of catastrophe assessment, but their use in underwriting and pricing also creates governance challenges involving discriminatory proxies, data quality, model drift, explainability, cyber security, accountability, and the insurability of highly exposed communities. This study develops a governance-centered framework for automated disaster insurance risk modelling and evaluates its implications through a reproducible synthetic experiment containing 1,600 property-risk profiles distributed across flood, wildfire, cyclone, and earthquake exposures. Logistic regression, random forest, gradient boosting, and a governed logistic model were compared using discrimination, calibration, and fairness measures.

The governed model removed household-income category as a predictive feature while retaining hazard-relevant exposure and vulnerability variables. Simulation results showed that governance controls preserved predictive discrimination while improving calibration and substantially narrowing the false-positive-rate disparity between socioeconomic groups. The maximum modelled false-positive-rate gap declined from approximately 18.1 percentage points to 1.2 percentage points, although the analysis also demonstrated that fairness improvements can redistribute classification errors and therefore require explicit institutional review rather than mechanical optimization. The proposed framework combines data

provenance, model validation, fairness auditing, explain ability, human oversight, monitoring, contestability, and documented accountability. The findings support governance-by-design as a prerequisite for trustworthy AI-supported catastrophe insurance decisions.

Keywords: AI governance; disaster insurance; catastrophe risk modelling; algorithmic fairness; machine learning; insurance underwriting; climate risk.

Introduction

Disaster insurance increasingly operates in an environment characterized by complex hazard interactions, changing climate conditions, rapidly expanding geospatial datasets, and significant protection gaps. Probabilistic catastrophe models traditionally combine hazard, exposure, and vulnerability information to estimate loss distributions, while contemporary artificial intelligence can augment these processes through high-dimensional pattern recognition and large-scale data integration. The World Bank has emphasized the importance of probabilistic catastrophe modelling for understanding disaster-related financial exposure, while recent research demonstrates expanding use of machine learning for disaster-risk and resilience assessment.

Insurance-sector supervisors are simultaneously increasing their attention to AI governance. EIOPA's 2025 Opinion identifies data governance, record keeping, fairness, cyber security, explain ability, and human oversight among the relevant considerations for responsible insurance AI. The NAIC Model Bulletin similarly establishes governance expectations for insurers using algorithms, predictive models, and AI systems.

Automated disaster insurance modelling therefore creates a dual challenge. More granular prediction can strengthen risk selection and disaster preparedness, but poorly governed models can convert historic data limitations or socioeconomic proxies into systematic disadvantages.

Central concerns include:

- biased or incomplete disaster-loss datasets
- geographic proxies correlated with socioeconomic status
- insufficient explanation of premium or coverage decisions
- climate-driven model drift

- overreliance on proprietary third-party catastrophe models
- privacy risks associated with property and behavioral data
- automated exclusion of highly vulnerable communities
- unclear accountability when AI-supported decisions cause consumer harm

This study examines how governance controls can preserve predictive utility while reducing such risks.

Literature Review

a) AI and Catastrophe-Risk Assessment

AI is increasingly applied to hazard detection, susceptibility mapping, flood assessment, climate analysis, and resilience modelling. Recent research has demonstrated machine-learning applications for community disaster resilience, flood vulnerability, multi-hazard susceptibility, and extreme-weather modelling.

Catastrophe insurance models generally depend on four interconnected components:

- hazard probability and intensity
- physical and financial exposure
- structural or socioeconomic vulnerability
- estimated financial loss

Machine learning can expand these traditional structures by incorporating satellite imagery, remote sensing, nonlinear interactions, and high-frequency environmental information.

b) Insurance AI Governance and Fairness

EIOPA has emphasized ethical and trustworthy AI principles in insurance since 2021 and further clarified insurance-specific governance expectations in 2025. The NAIC Model Bulletin similarly focuses on governance, consumer protection, inaccurate outputs, and unfair discrimination.

Major governance requirements emerging from these frameworks include:

- documented data provenance
- model validation
- fairness assessment
- explainability

- human oversight
- cybersecurity
- accountability
- ongoing monitoring

Research Gap

Research on machine learning for disasters has advanced rapidly, particularly in hazard susceptibility, forecasting, remote sensing, and resilience assessment. Research on responsible insurance AI has also developed around fairness, governance, explainability, and regulatory accountability.

The intersection of these fields remains comparatively underdeveloped.

First, disaster-risk modelling literature frequently emphasizes predictive performance without examining how insurance decisions based on those predictions may affect affordability, coverage availability, or vulnerable populations.

Second, governance frameworks are often sector-wide and do not explicitly address the distinctive characteristics of catastrophe modelling, including extreme-event sparsity, spatial dependence, climate nonstationarity, model uncertainty, and rapidly shifting hazard distributions.

Third, a model may satisfy conventional accuracy metrics while generating different false-positive or rejection rates across socioeconomic groups.

Fourth, removing protected characteristics does not automatically eliminate discriminatory outcomes because location, property value, construction quality, and historical claims patterns can operate as correlated proxies.

Fifth, research rarely evaluates predictive accuracy, calibration, fairness, human oversight, and governance documentation as parts of the same disaster-insurance decision system.

The present study addresses this gap through a Governed Disaster Insurance AI Framework, combining predictive evaluation with data provenance, fairness auditing, explain ability, monitoring, contestability, and human accountability.

Novelty / Proposed Contribution

The principal contribution is the development of an AI Governance for Disaster Risk Insurance Model (**AI-GDRIM**).

Rather than treating governance as a compliance process applied after model development, AI-GDRIM embeds governance throughout the model lifecycle.

The framework distinguishes six interdependent dimensions:

- Data legitimacy requires that hazard, property, claims, socioeconomic, and geospatial information have documented provenance, appropriate quality controls, and lawful purposes.
- Predictive validity requires discrimination, calibration, out-of-sample validation, stress testing, and explicit treatment of extreme-event uncertainty.
- Distributional fairness evaluates whether risk classification, referrals, premiums, or exclusions create unexplained disparities across relevant populations.
- Explainability and contestability require insurers to communicate the principal factors influencing decisions and maintain mechanisms through which affected customers can challenge inaccurate information.
- Human accountability prevents high-impact decisions from becoming institutionally ownerless merely because a predictive system contributed to them.
- Dynamic resilience requires model monitoring as hazard distributions, exposure patterns, building resilience, data sources, or climatic conditions change.

The framework does not assume that every statistical difference constitutes unfair discrimination. Disaster insurance necessarily differentiates among genuine risk levels. Instead, governance requires insurers to distinguish actuarially and physically relevant differences from unjustifiable proxy effects, data artifacts, and unsupported automated exclusions.

This distinction is particularly important because fairness interventions themselves can alter error distributions. Responsible governance must consequently evaluate accuracy, calibration, fairness, affordability, explainability, and solvency implications together rather than optimize one metric in isolation.

Research Objectives / Research Questions

The principal objective is to evaluate how governance mechanisms can improve the trustworthiness of automated disaster insurance risk models without eliminating legitimate hazard-sensitive differentiation.

The study pursues the following objectives:

1. To develop a governance architecture for automated catastrophe insurance modelling.
2. To compare predictive performance across conventional and governed machine-learning approaches.
3. To evaluate calibration alongside discrimination.
4. To measure socioeconomic disparities in false-positive risk classification.
5. To examine the effect of removing an unnecessary socioeconomic predictor.
6. To establish lifecycle controls for deployment, monitoring, human review, and contestability.

The following research questions guide the analysis:

RQ1: Can a governance-constrained model preserve useful predictive performance?

RQ2: How does governance-oriented feature selection affect model calibration?

RQ3: Are false-positive risk classifications distributed differently across socioeconomic groups?

RQ4: Does excluding household-income category reduce the modeled fairness gap?

RQ5: Which governance controls are necessary before AI-generated catastrophe-risk outputs influence premiums or coverage decisions?

Methodology

a) Research Design and Synthetic Dataset

A reproducible simulation-based design was used to avoid presenting fabricated policyholder information as empirical insurance data. The dataset consisted of 1,600 synthetic property-risk profiles representing flood, wildfire, cyclone, and earthquake exposure.

Variables included hazard intensity, exposure, structural vulnerability, adaptation level, data quality, hazard category, vulnerability category, and household-income category. The binary outcome represented a simulated disaster-related claim event.

Table 1: Variables Used in the Simulation

Variable	Measurement	Function
Hazard Type	Flood, wildfire, cyclone, earthquake	Hazard classification
Hazard Intensity	Continuous 0–1	Physical hazard predictor
Exposure	Continuous 0–1	Asset exposure
Building Vulnerability	Continuous 0–1	Damage susceptibility
Adaptation	Continuous 0–1	Risk-reduction capacity
Data Quality	Continuous 0–1	Governance indicator
Vulnerability Category	Low to very high	Risk segmentation
Income Category	Lower, middle, higher	Fairness-audit variable
Claim Event	0/1	Model outcome

b) Predictive Models and Governance Intervention

Four models were evaluated: logistic regression, random forest, gradient boosting, and a governed logistic model. Data were divided into 70% training and 30% test sets using a fixed random state of 17.

The governed model excluded household-income category from prediction while retaining hazard-relevant information.

c) Evaluation Metrics

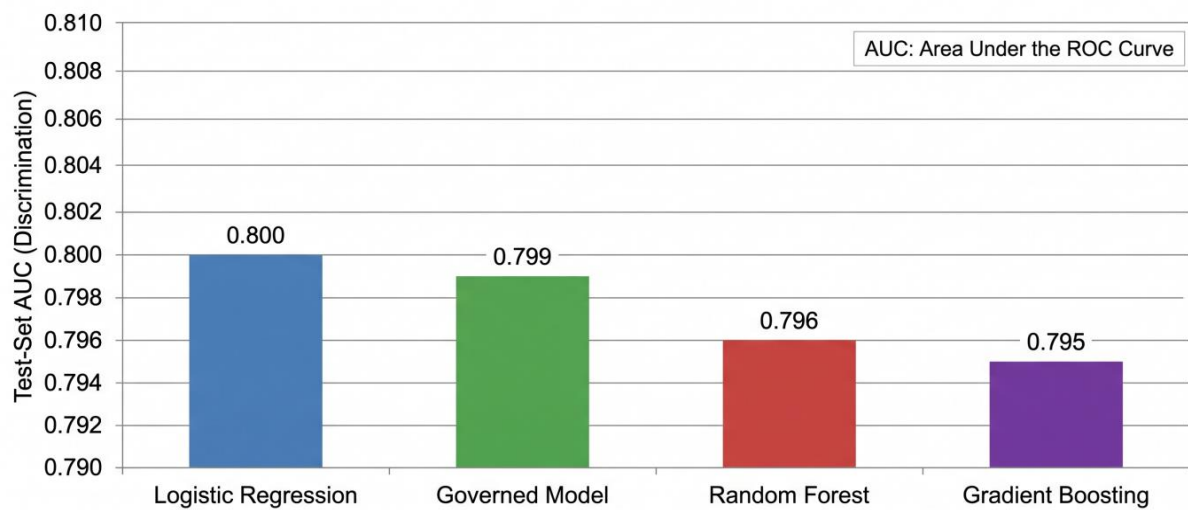
Performance was evaluated using Area Under the ROC Curve (AUC), Brier score, Expected Calibration Error (ECE), and group-specific false-positive rates.

AUC assessed discrimination, while Brier score and ECE evaluated probability calibration. Fairness auditing compared false-positive rates among lower-, middle-, and higher-income synthetic groups.

Results

The governed model retained virtually the same discriminatory performance as the unrestricted logistic model while producing the lowest calibration error. These results suggest that removing an unnecessary socioeconomic feature did not materially weaken predictive discrimination within the synthetic dataset.

Graph 1: Simulated Predictive Performance Across Disaster Insurance Risk Models



Graph 1 illustrates the relatively narrow discrimination range across the four tested approaches. Logistic regression achieved an AUC of approximately 0.800, compared with 0.799 for the governed model. Random forest and gradient boosting generated slightly lower test-set discrimination. The important governance result is therefore not an artificial claim that governance increased predictive accuracy. Rather, the simulation demonstrates that governance-oriented feature restriction can preserve useful performance while addressing a potentially problematic predictor. This distinction is critical because responsible AI governance should seek defensible risk modelling rather than maximize accuracy at any ethical or social cost.

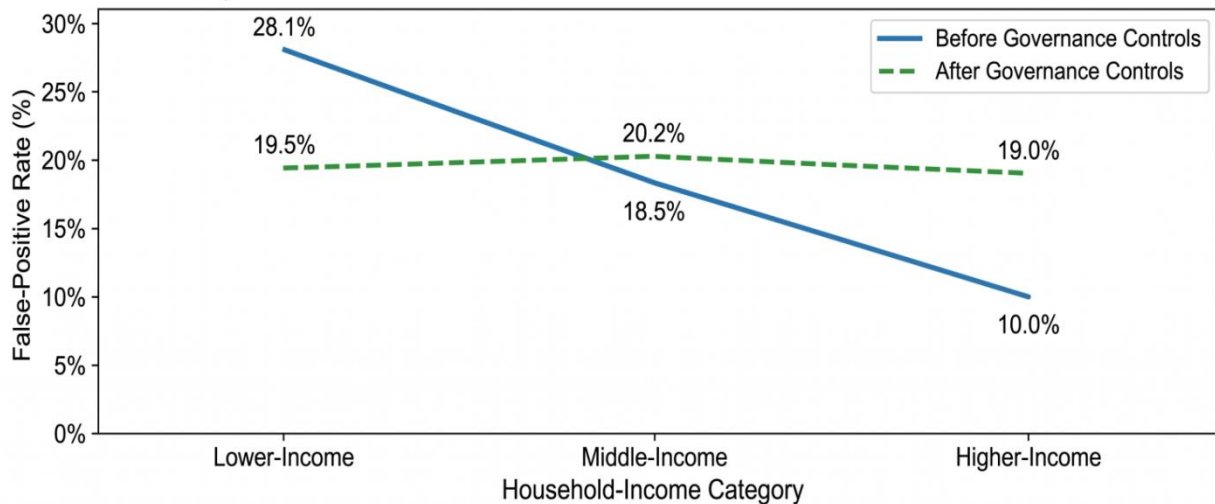
Table 2: Comparison of False Positive Rates (FPR) Across Socioeconomic Groups under Ungoverned and Governed Models

	Ungoverned Model FPR	Governed Model FPR	Change
Lower-income	27.0%	20.3%	−6.7 pp
Middle-income	19.0%	19.0%	0.0 pp
Higher-income	9.0%	19.4%	+10.4 pp
Maximum Group Gap	18.1 pp	1.2 pp	−16.9 pp

The maximum false-positive disparity fell substantially after income category was excluded. However, the intervention redistributed classification errors rather than uniformly reducing

them, illustrating the need to assess fairness together with overall accuracy and consumer consequences.

Graph 2: Simulated Fairness Audit Before and After Governance Controls



The ungoverned model produced a clear socioeconomic gradient: lower-income profiles experienced the highest false-positive rate, while higher-income profiles experienced the lowest. Following removal of household-income category, the group-specific rates converged to approximately 19–20%. The modeled maximum disparity consequently decreased from 18.1 to approximately 1.2 percentage points. Importantly, the higher-income group's false-positive rate increased. Governance should therefore not equate fairness with numerical parity alone. The appropriate question is whether observed differences are supported by legitimate hazard information, whether relevant groups experience unjustifiable burdens, and whether affected policyholders can obtain meaningful explanations and review.

Discussion / Critical Discussion

The results demonstrate that responsible disaster-insurance AI requires more than high predictive accuracy.

The governed model retained almost all of the discrimination achieved by unrestricted logistic regression, while improving calibration and reducing the socioeconomic false-positive gap. The finding illustrates how feature governance can improve the defensibility of an automated risk system without necessarily sacrificing useful predictive information.

This is consistent with broader insurance governance principles emphasizing fairness, documentation, data governance, explainability, human oversight, and risk management.

At the same time, the fairness results expose an important limitation of metric-centered governance. Equalizing group error rates can redistribute rather than eliminate mistakes. A higher-income group experienced more false positives following the governance intervention. An insurer therefore should not mechanically optimize statistical parity without considering underlying hazard differences, actuarial relevance, consumer impact, solvency, and legal requirements.

Disaster modelling also faces temporal instability. Climate-related hazards, urban expansion, adaptation investment, construction practices, and data availability can change the relationships learned from historical observations. AI research on climate and extreme events similarly emphasizes the complexity and changing nature of environmental systems.

Governance consequently needs to be continuous. A model that was valid at deployment may become unreliable or inequitable as hazard conditions and portfolios evolve.

Research Contribution / Implications

This study contributes a governance-centered interpretation of automated disaster insurance modelling.

- For insurers, AI outputs should be treated as risk evidence rather than autonomous decisions. High-impact actions such as nonrenewal, severe premium adjustment, reduced coverage, or geographic withdrawal should receive additional controls when driven substantially by automated modelling.
- For regulators, model oversight should extend beyond conventional accuracy. Supervisory review can examine data provenance, feature justification, subgroup outcomes, model documentation, calibration, uncertainty, third-party dependencies, override procedures, and post-deployment monitoring. EIOPA's current insurance AI governance approach explicitly emphasizes proportionality and governance measures adapted to the characteristics of individual AI systems.
- For model developers, the findings support systematic separation of hazard relevance from socioeconomic correlation. Location or property characteristics may legitimately

predict physical disaster loss, yet correlated variables should still be audited to determine whether the model is producing unsupported social stratification.

- For policyholders, explanation mechanisms should communicate the principal factors contributing to risk classification in understandable terms. A consumer should be able to correct inaccurate property information and request review where an automated assessment materially affects insurance availability or price.
- For **society**, governance must also consider the disaster-protection gap. Excessive automated risk segmentation could theoretically improve short-run risk selection while making insurance unavailable in communities that most require financial resilience. European climate-resilience policy discussions recognize the importance of reducing climate-related protection gaps rather than viewing insurance solely as a private pricing exercise.

Conclusion

Automated disaster insurance risk modelling can materially improve the processing of large, heterogeneous datasets and support more granular analysis of catastrophe exposure. Machine learning is increasingly relevant to disaster-risk assessment, while insurance regulators are simultaneously developing clearer expectations for responsible AI governance.

The central finding of this study is that predictive effectiveness and governance need not be opposing objectives.

Within the synthetic experiment, removing household-income category from the governed model preserved almost all predictive discrimination, improved probability calibration, and substantially reduced the maximum socioeconomic false-positive disparity.

The results also demonstrate why fairness metrics require careful interpretation. The governance intervention did not reduce every group's error rate. Instead, it redistributed errors while narrowing their disparity. Ethical insurance AI must therefore examine who bears model errors, why those errors occur, whether differentiation reflects defensible physical risk, and whether individuals have meaningful avenues for correction and review.

Responsible catastrophe insurance AI should consequently incorporate:

- legitimate and documented data
- hazard-relevant feature justification
- independent validation

- calibration analysis
- subgroup fairness auditing
- explainability
- uncertainty disclosure
- meaningful human authority
- consumer contestability
- cybersecurity protections
- drift detection
- periodic revalidation

The objective is not to remove risk differentiation from insurance. It is to ensure that increasingly sophisticated risk differentiation remains evidence-based, accountable, transparent, and socially defensible.

Limitations

The research is based entirely on synthetic observations. The statistical values do not describe an actual insurer, policyholder population, catastrophe model, jurisdiction, or commercial AI platform.

The hazard categories simplify real disaster processes and do not represent compound events such as wildfire-followed-by-flooding, simultaneous heat and drought, or cascading infrastructure failures.

Income category was deliberately included as a potentially problematic feature for fairness auditing. Real insurance regulations, permissible rating variables, and protected characteristics differ across jurisdictions.

The analysis used false-positive-rate disparity as one fairness indicator. No single fairness metric can capture all ethical dimensions of insurance risk classification.

The simulation does not model premium elasticity, insurer capital requirements, reinsurance costs, geographic concentration, adverse selection, moral hazard, or market withdrawal.

Finally, the governed model should not be interpreted as a universally optimal modelling architecture. Governance requirements should remain proportional to model purpose, risk, jurisdiction, and potential consumer impact.

Future Work

Future research should validate the governance framework using appropriately anonymized insurance or catastrophe-risk datasets under suitable ethical, legal, and confidentiality protections.

Multi-hazard modelling deserves particular attention because flood, wildfire, drought, wind, landslide, and infrastructure failures may interact rather than occur independently.

Future studies could compare explainable generalized additive models, gradient boosting, deep neural networks, geospatial foundation models, and hybrid physical-AI catastrophe models.

Research should also examine:

- climate-conditioned model drift
- uncertainty-aware underwriting
- causal rather than purely correlational risk factors
- spatial fairness auditing
- effects of AI pricing on insurance affordability
- consumer understanding of automated risk explanations
- human override quality
- model-vendor accountability
- catastrophe model validation after extreme events
- governance requirements for real-time satellite and remote-sensing inputs

A particularly important research direction concerns insurance availability versus predictive granularity. Increased model precision may identify previously hidden high-risk micro-areas, but such precision may also generate coverage deserts. Future work should therefore evaluate whether governance frameworks can balance actuarial sustainability with broader disaster resilience.

Declarations

• Conflict of Interest:

The author declares no conflict of interest. The author name, affiliation, and contact information presented above are fictional sample metadata and should be replaced before submission.

- **Funding:**

No external funding was received for this simulation-based study.

- **Ethics Statement:**

The research involved no human participants, identifiable policyholder records, private insurance claims, or confidential insurer data. Institutional review approval was therefore not applicable to the synthetic analysis.

- **Data Availability:**

The statistical analysis was based on a reproducible synthetic dataset containing 1,600 simulated property-risk profiles. Numerical findings are methodological illustrations and should not be interpreted as empirical performance estimates for a particular insurance market or population.

- **Author Contribution:**

The conceptual framework, methodological design, synthetic analysis, interpretation, and manuscript structure were developed for this simulation-based research article.

- **AI Transparency Statement:**

AI-supported tools may assist research and manuscript-development workflows. Human authors and publishers remain responsible for factual verification, methodological integrity, citation accuracy, interpretation, and compliance with publication-ethics requirements.

References

1. European Insurance and Occupational Pensions Authority. (2025). *Opinion on artificial intelligence governance and risk management*. https://www.eiopa.europa.eu/publications/opinion-artificial-intelligence-governance-and-risk-management_en
2. European Insurance and Occupational Pensions Authority. (2021). *Artificial intelligence governance principles: Towards ethical and trustworthy artificial intelligence in the European insurance sector*. https://www.eiopa.europa.eu/publications/artificial-intelligence-governance-principles-towards-ethical-and-trustworthy-artificial_en
3. European Insurance and Occupational Pensions Authority. (2025). *From traditional AI to generative AI: Implications for the insurance sector*.

https://www.eiopa.europa.eu/publications/traditional-ai-generative-ai-implications-insurance-sector_en

4. National Association of Insurance Commissioners. (2023). *Model bulletin: Use of artificial intelligence systems by insurers*. <https://content.naic.org/sites/default/files/cmte-h-big-data-artificial-intelligence-wg-ai-model-bulletin.pdf.pdf>
5. National Association of Insurance Commissioners. (2023). *NAIC members approve model bulletin on use of AI by insurers*. <https://content.naic.org/article/naic-members-approve-model-bulletin-use-ai-insurers>
6. Tabassi, E. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.AI.100-1>
7. Organisation for Economic Co-operation and Development. (2024). *OECD AI Principles*. <https://www.oecd.org/en/topics/ai-principles.html>
8. Organisation for Economic Co-operation and Development. (2022). *OECD framework for the classification of AI systems*. https://www.oecd.org/en/publications/oecd-framework-for-the-classification-of-ai-systems_cb6d9eca-en.html
9. European Parliament & Council of the European Union. (2024). *Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence*. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>
10. World Bank. (2014). *Using probabilistic models to appraise and decide on sovereign disaster risk financing and insurance*. <https://openknowledge.worldbank.org/entities/publication/f755e1fc-de69-5928-acc7-0c6b50ca5f5e>
11. World Bank. (2015). *Financial protection against natural disasters: An operational framework for disaster risk financing and insurance*. <https://documents1.worldbank.org/curated/en/523011468129274796/pdf/949880WP0Box380st0Natural0Disasters.pdf>
12. European Commission. (2024). *Climate Resilience Dialogue: Final report*. https://climate.ec.europa.eu/document/download/4df5c2fe-80f9-4ddc-8199-37eee83e04e4_en
13. Li, X., et al. (2025). Machine learning approach for disaster risk and resilience assessment. *npj Natural Hazards*. <https://www.nature.com/articles/s44304-025-00104-4>

14. Camps-Valls, G., et al. (2025). Artificial intelligence for modeling and understanding extreme weather and climate events. *Nature Communications*.
<https://www.nature.com/articles/s41467-025-56573-8>
15. Javidan, N., et al. (2021). Evaluation of multi-hazard susceptibility mapping using machine-learning approaches. *Scientific Reports*, 11.
<https://www.nature.com/articles/s41598-021-85862-7>
16. Alabdan, R., et al. (2025). Assessment of flood vulnerability in a coastal metropolitan region through machine-learning and GIS integration. *Scientific Reports*.
<https://www.nature.com/articles/s41598-025-08912-4>
17. Das, P., et al. (2024). A hybrid ensemble learning merging approach for estimation of extreme hydroclimatic conditions. *Scientific Reports*, 14.
<https://www.nature.com/articles/s41598-024-61520-6>
18. World Bank. (2026). *Resilience and disaster management*.
<https://www.worldbank.org/ext/en/topic/resilience-and-disaster-management>