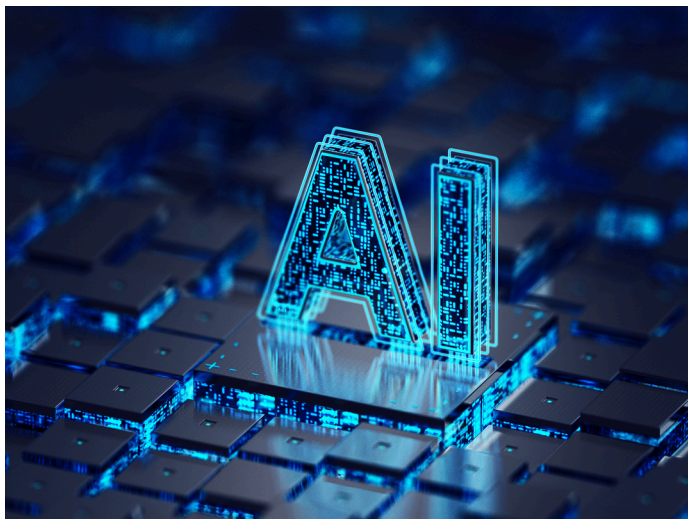
Team

03.

Modeling Risks of Artificial Intelligence Through Incident Report Analysis

How do we quantify legal and financial risks arising from artificial intelligence failures?

Artificial intelligence is increasingly used in high stakes legal, professional, and commercial contexts.



AI is being adopted more widely, yet failures such as hallucinated outputs, data misuse, and procedural errors have resulted in growing litigation and regulatory scrutiny. These risks are hard to insure due to limited loss data and evolving legal standards. As a result, insurers often rely on exclusions rather than structured risk transfer, limiting responsible AI adoption.

This project is relevant to a wide range of stakeholders:

Insurers/Reinsurers	Underwrite and manage AI risks better
AI Developers & Vendors	Clearer insights into legal risks and governance expectations
Orgs Deploying AI	Guidance on high-risk use cases and mitigation
Courts/Regulators	Evidence-based understanding of emerging AI misconduct
Policyholders/Public	Safer AI deployment and higher accountability

A collaborative project with:

University of Toronto

York University

Munich RE

Next page

Methods

1. Data Collection

Curated 2 complementary AI-related legal case and incident datasets.

2. Exploratory Analysis

Examined trends in time, geography, parties, and penalties.

3. Risk Categorization

Applied NLP techniques, including LLM-based semantic embeddings, to categorize hallucination risks.

4. Sanction Prediction

Built classification models to predict sanction likelihood.

5. Penalty Estimation

Developed regression models to predict penalty severity.

6. Expected Loss Framework

Integrated frequency and severity into an expected loss model.

Findings

Heterogeneous Risk

AI legal risk varies widely across use cases, jurisdictions, and hallucination types.

Rising Frequency

Cases are increasing due to broader AI adoption and stricter regulatory enforcement.

Jurisdictional Impact

Legal outcomes and penalty severity differ significantly by region.

Heavy-Tailed Losses

A small number of high-severity cases drive most portfolio risk.

Failure and Context Driven

Risk is concentrated in high-stakes AI applications and technical/expert-level failures.

Predictable Outcomes

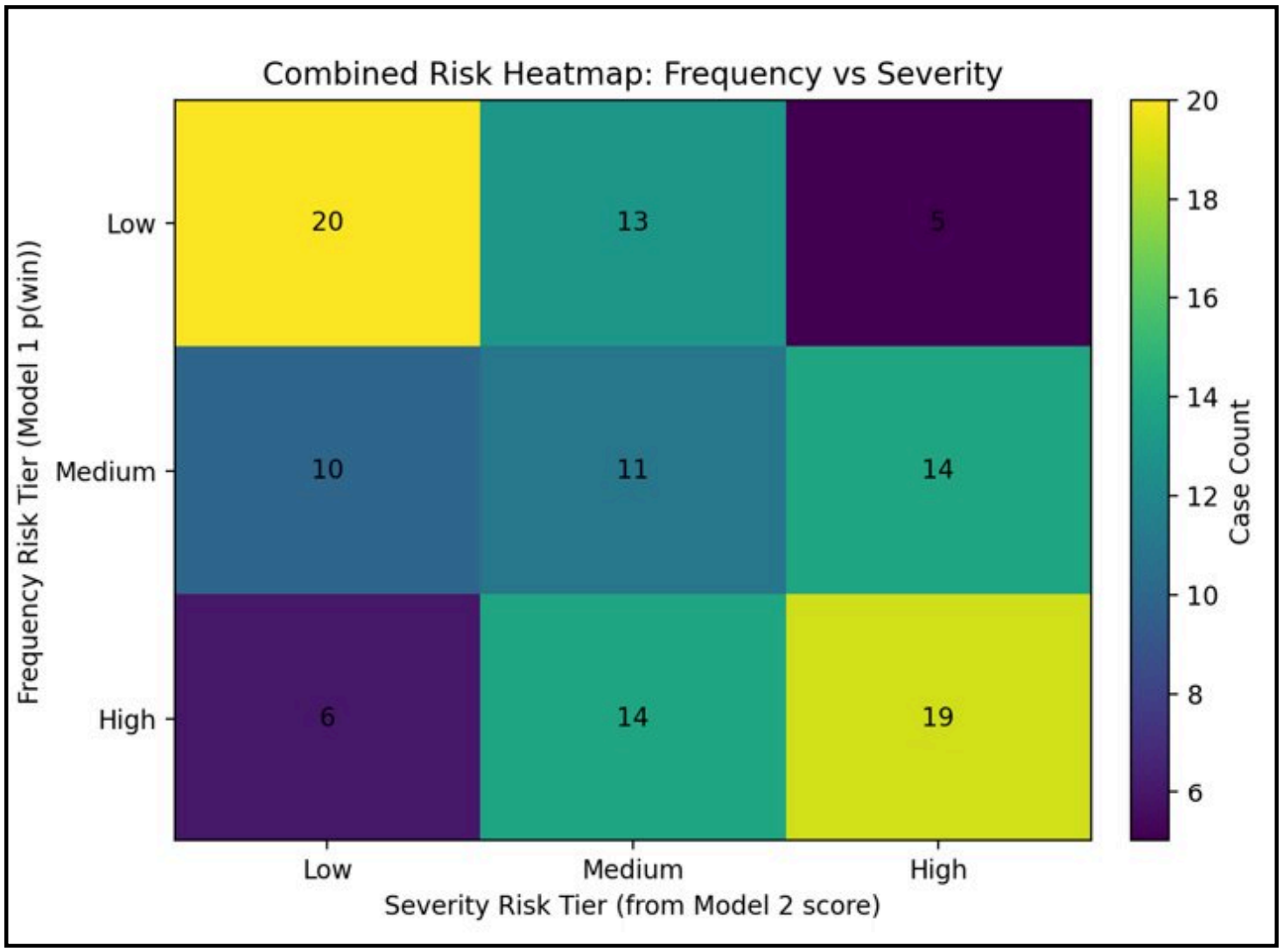
Sanction likelihood and penalty severity can be systematically modeled.

Portfolio Implications

Tail risks highlight the need for targeted limits, reinsurance, and capital allocation strategies.

Recommendation

The team proposes a 2-stage actuarial-style framework **combining sanction probability and severity modeling to estimate expected loss.** This enables insurers to move from blanket exclusions to risk-based underwriting and pricing of AI-related liabilities.



Limitations

Publicly reported cases likely understate true AI-related risk due to confidential settlements and undisclosed resolutions. Additionally, rapidly evolving regulations and legal standards limit the long-term generalizability of historical findings.

Team 3

HAIVE 10

Risk and Insurance Studies Centre

Team Members



Hantian Chen
Actuarial Science
University of Toronto



Robert Okine
Master in Financial Insurance
University of Toronto



Stefan Jafry
Data Science
York University



Industry Mentor:
Yuanyuan Li, PhD
Senior Research Scientist , Munich Re



Academic Mentor:
Christopher Blier-Wong, PhD
Assistant Professor, University of Toronto