

8th Current Issues in General Insurance Hotel Sea Princess, Mumbai 18th September 2025

Underwriting in the Age of Big Data & AI

Ashok Shetty
VP, Head Advanced Data Science, Swiss Re





The question!



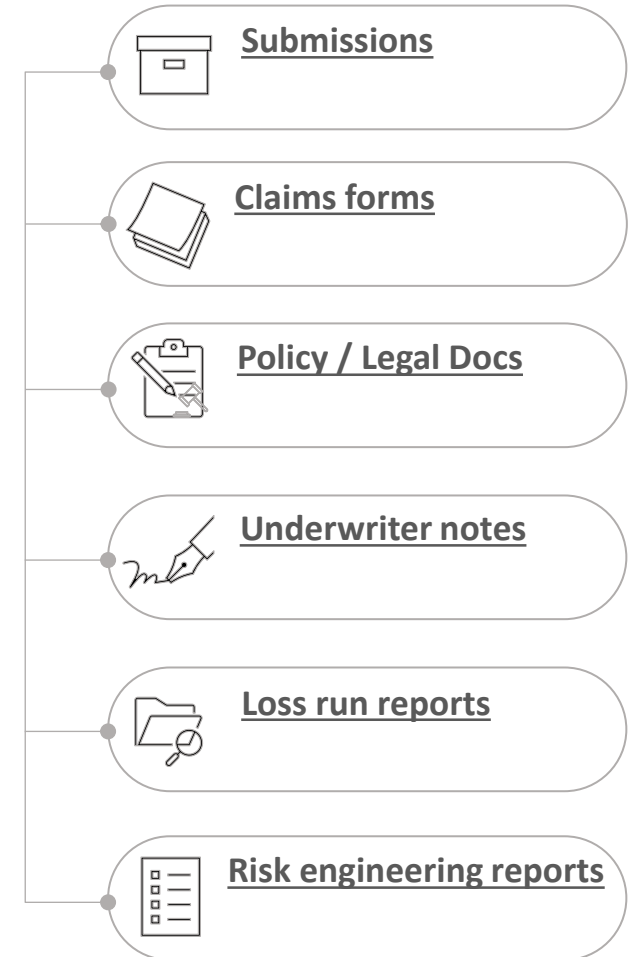
“If all actuaries/underwriters have access to the same tools, models, and technology, how would you set yourself apart and create a competitive advantage in your domain?”

The answer: Data

“The differentiator isn’t the model itself. It’s the data you use, how you shape it, and how you ask the right questions of it.”

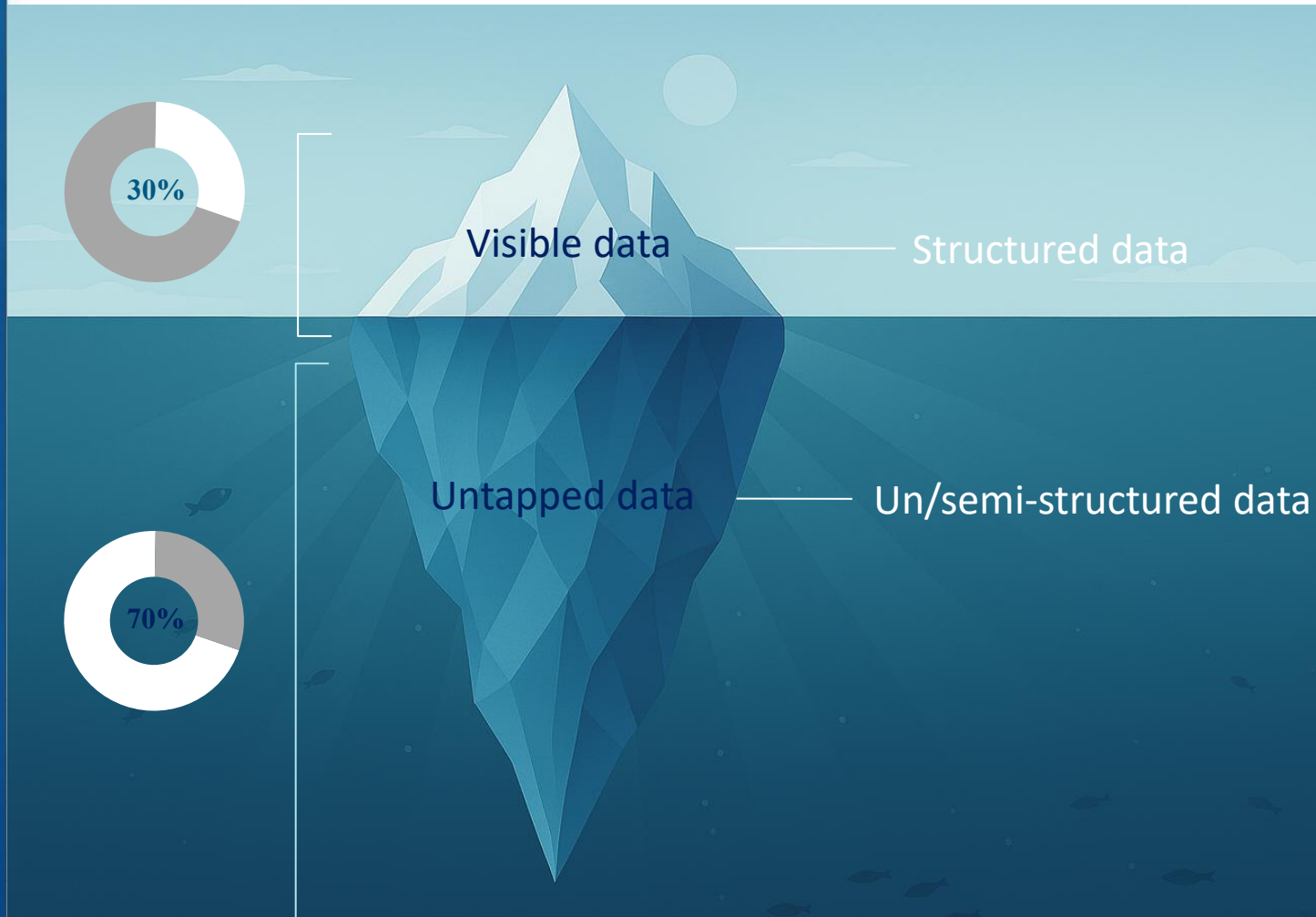
But we already have data...!

Insurers sit on massive datasets.



1. Are We Really Utilizing Our Data?

Insurers sit on massive datasets, but...~70% of enterprise data is locked away in the Semi/Unstructured format.



Submissions



Claims forms



Policy / legal Docs



Underwriter notes



Loss run reports

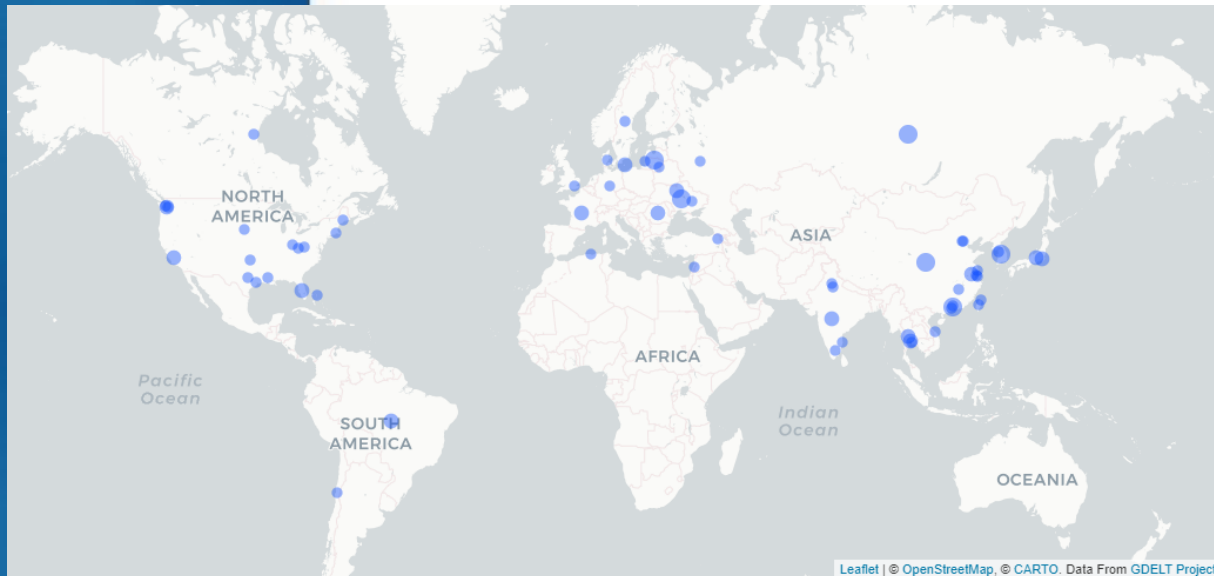


Risk engineering reports

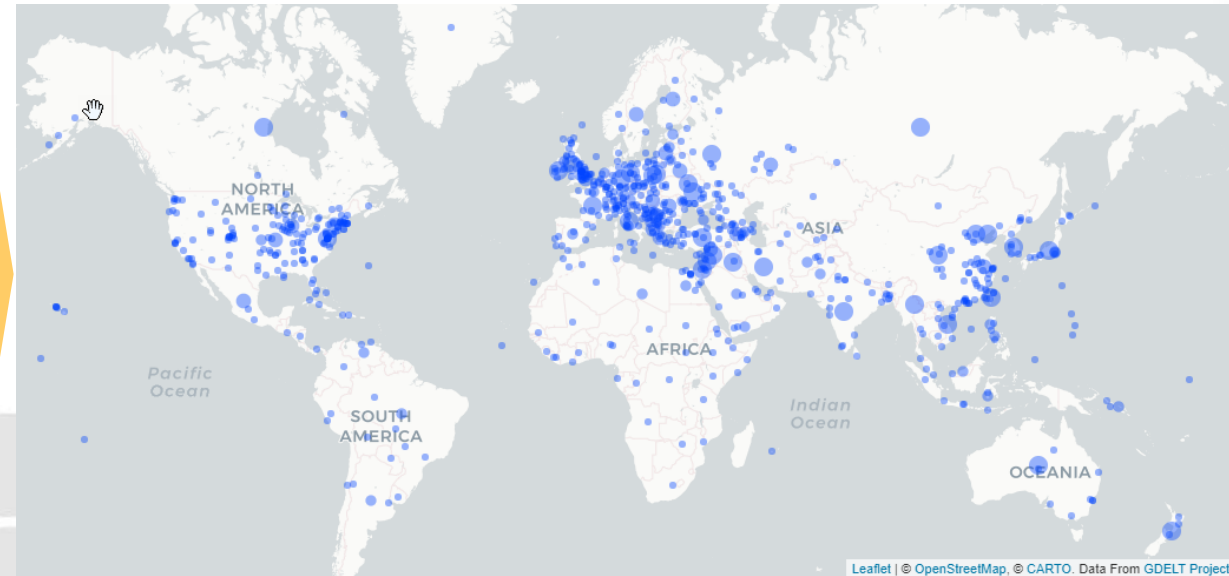
2. Is that data enough?

- Re/insurers have a blind spot or very limited amount of information over numerous disasters/events that take place across the world. Some of such disasters may possess direct or indirect threat to their business
- Also, past is not always a true representative of future. Hence, we also need to identify the unseen drivers and their trends to quantify various risks pools

E.g. - Man-made disasters experienced by an insurer during a specific period



Man-made disasters the world observed during that period.



How are Re/Insurers responding?



Claims

Pricing

Portfolio Mgt.

Risk Mgt.

Underwriting ➤

1990s – Early 2000s ➤

Mid-2000s ➤

2010–2017 ➤

2017 onward



Submission,
Claims docs [...]

Manual
Risk
Extraction

powered by ➤



Decision

Statistical NLP + ML Algorithms

- Rise of statistical machine learning with more data + computing power.
- Techniques: Hidden Markov Models (HMMs), Conditional Random Fields (CRFs), Maximum Entropy models.
- Enabled POS tagging, Named Entity Recognition (NER), information retrieval at scale.

Limitation: Needed lots of labeled data, still shallow understanding.

Word Embeddings + Deep Learning Models

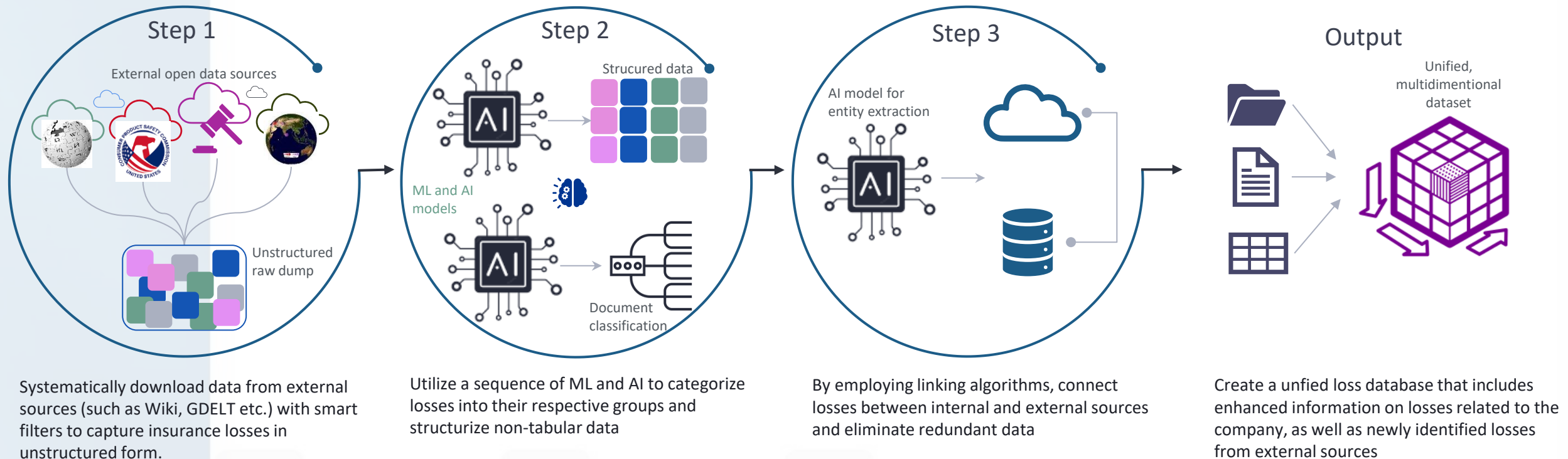
- Shift from handcrafted features → representation learning.
- Key breakthroughs:
- Word2Vec (2013) ; GloVe (2014)
- Recurrent Neural Networks (RNNs), LSTMs, GRUs for sequences.

Limitation: still sequential, struggled with long contexts

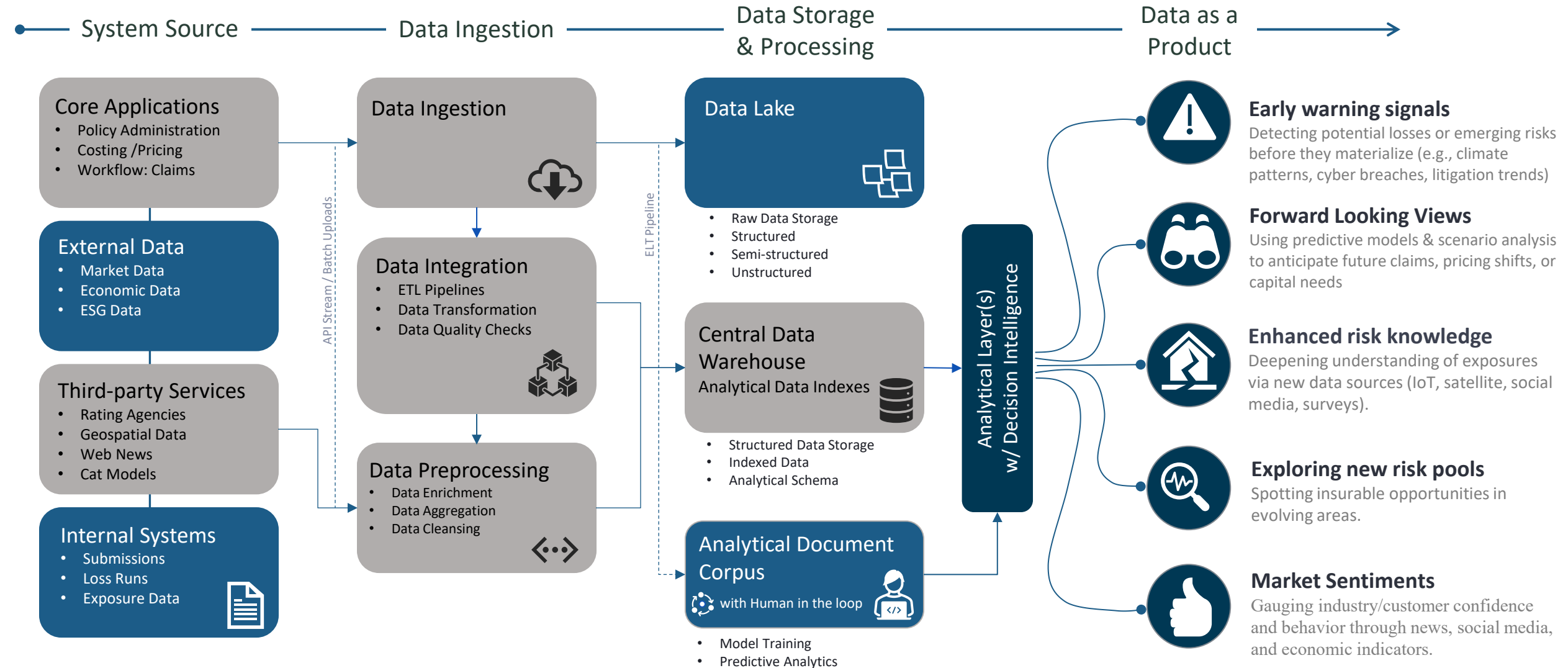
Transformers: The Revolution

- Game-changer: “Attention is All You Need” (Vaswani et al., 2017).
- Transformers enabled parallel processing of sequences, capturing long-range dependencies.
- Led to BERT (2018), GPT (2018), XLNet, RoBERTa, etc.
- Large Language Models (LLMs): GPT-3 (2020), ChatGPT (2022), GPT-4/5, Claude, Gemini.

High-level solution design



The new data strategy

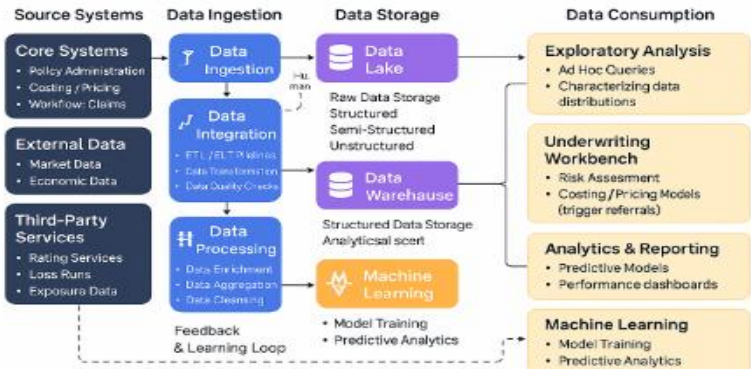


GenAI-Powered Solutions



Underwriting Risk Advisor

End-to-End Data Architecture for Underwriting Risk Advisor



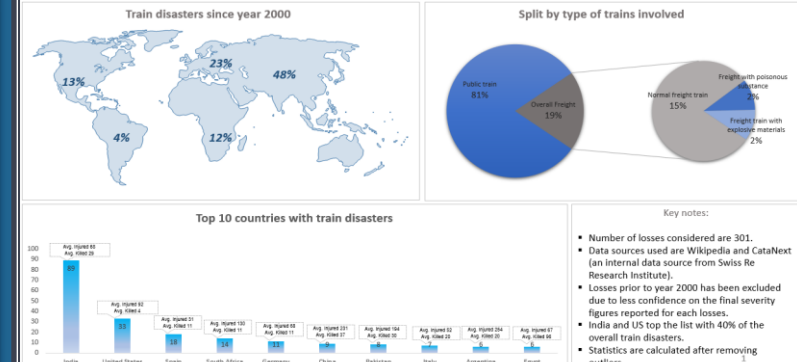
Submission Risk Classifiers

Data - Insight - Action



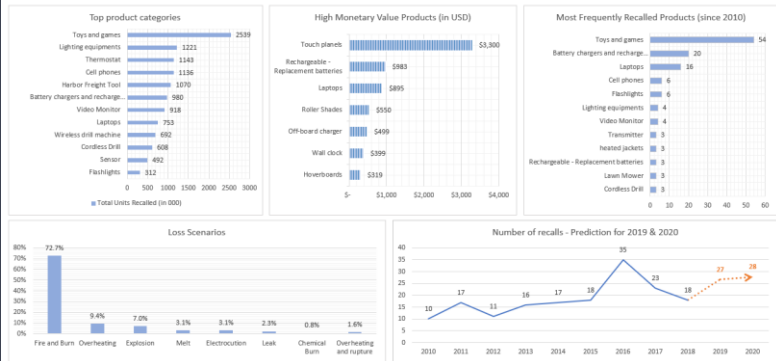
Risk Radar

Executive Summary – Train Accidents



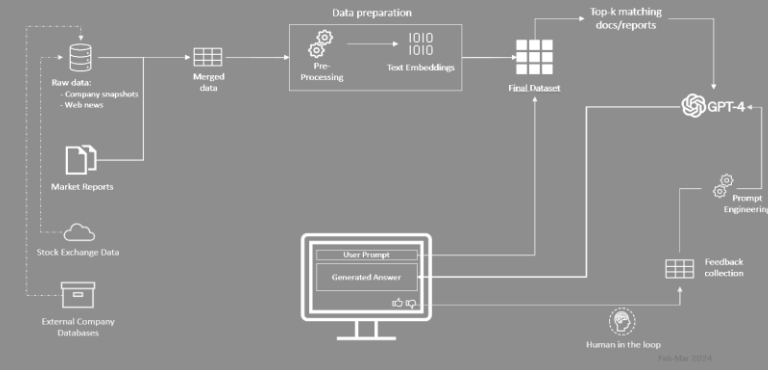
Claims IQ

Diagnostics & Predictions



Smart Data Ingestion

behind the scenes...



Risk Data Enrichment

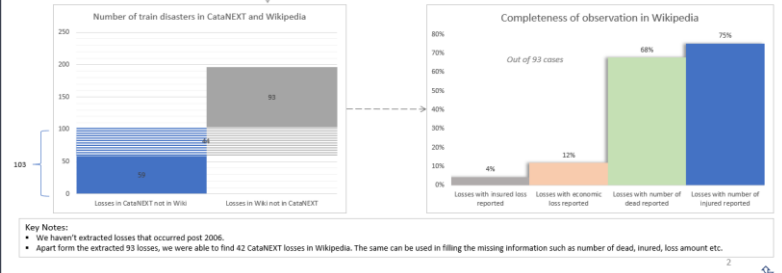
Summary

Types of major losses extracted from Wikipedia (Man-made disasters):

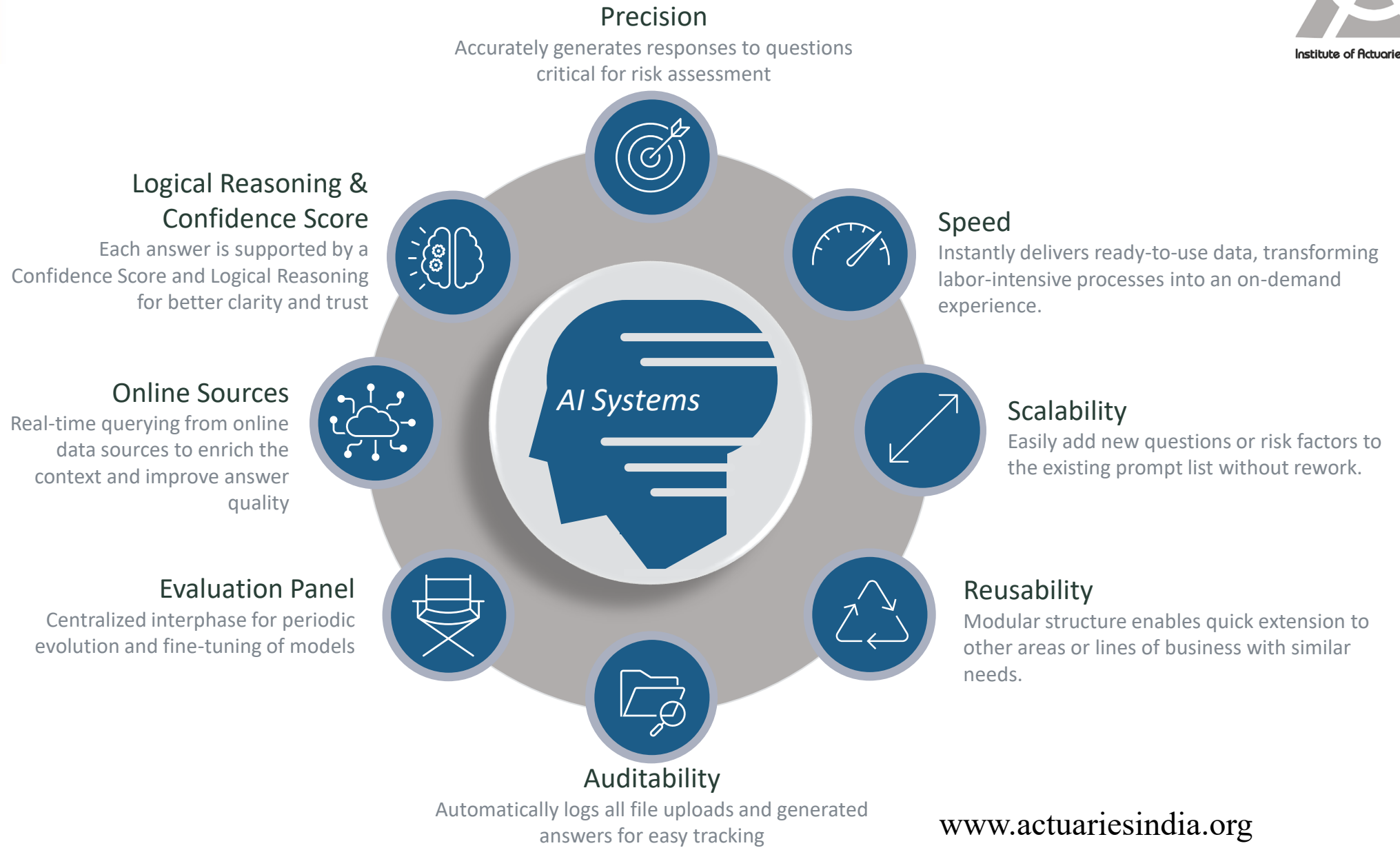
1. Train Accidents
2. Premises Disaster
3. Utility Catastrophe
4. Construction Disaster
5. Food Industry Catastrophe

Thresholds of CataNEXT:

- A loss needs to meet any one of the below criteria (from year 2016) in order to get recorded in CataNEXT database –
1. >=20 Killed
 2. >=50 Injured
 3. >=100 million USD economic loss



TOP FEATURES OF AI SYSTEMS



To Summarize



GenAI and data science are redefining what it means to be an actuary or an underwriters empowering us to unlock deeper insights, manage emerging risks, and drive innovation.

Thank you!



Ashok Shetty
Head Advanced Data Science
Swiss Re