



Reinventing data quality: how AI is transforming pensions and insurance

Machine Learning: securing actuarial integrity and trust through smarter data transformation

Actuarial and financial models are built upon one **fundamental assumption: that the underlying data is reliable. We call this the data challenge behind every model. Even the most advanced model is only as reliable as its data – especially in pensions and insurance. Over time, organizations have accumulated millions of records on policies, benefits, claims etc. collected from various systems, mergers, and migrations, creating a complex mix of formats, codes, and inconsistencies.**

Regulatory expectations are rising. Frameworks like Solvency II, IFRS 17, and the WTP demand transparent, traceable, and trustworthy data so every decision can be explained and aligned with long-term commitments to policyholders and participants.

Manual checks can't keep up – fixing errors after they appear is too late, adding cost and delay to reporting. Artificial Intelligence (AI) is emerging not as a replacement for actuaries, but as an intelligent assistant. Its role is to secure data integrity before it reaches the model. By learning from historical patterns, AI can detect, explain, and even correct anomalies in real time – shifting data control from rule-based checks to intelligent assurance¹.

D. Rahmanifard MSc AAG (left works as a senior actuary at Actuarial Reporting, Assumptions & Data at a.s.r.

R. Cats MSc AAG works as a senior financial risk manager at Asset Risk Modelling at a.s.r.



FROM GOVERNANCE TO INTELLIGENCE – THE RISE OF AI-DRIVEN DATA QUALITY

Data quality management traditionally relied on rule-based checks – *for example, date of birth < date of retirement or sum of contributions = total premium* – to ensure accuracy, completeness, and consistency. These rules formed the backbone of traditional governance.

Static rules can't keep up with today's complexity. Pension and insurance data now come from payrolls, medical systems, and investment platforms – far too diverse for fixed checks to catch every anomaly. AI changes this. Instead of following preset rules, machine learning learns what 'normal' looks like and flags deviations automatically, adapting as data and patterns evolve.

A mature AI-driven framework typically includes:

1. **Data profiling:** incoming datasets are scanned for structure, missing values, and outliers.
2. **Pattern learning:** unsupervised models detect unusual behaviour without relying on preset limits.
3. **Semantic mapping:** natural-language tools link technical fields to business meaning.
4. **Automated correction:** AI proposes or applies fixes, sometimes retrieving accurate data automatically.
5. **Continuous feedback:** human validation refines the model, improving accuracy over time.

Together, these capabilities turn data quality from a compliance task into an intelligent, self-learning process – giving actuaries cleaner data and the confidence to focus on insight instead of error checking.

PUTTING AI TO WORK – FROM PENSION DATA TO ACTUARIAL MODELS

Imagine a large pension administrator managing millions of participant records – with recurring issues like inconsistent statuses, duplicates, and missing partner details. Despite regular reconciliations, legacy databases and data marts make these errors hard to trace and correct.

To resolve this, the organization can introduce an AI-driven data-quality layer between its administration systems and actuarial models. The organization will go through the following phases:

Phase 1 – Profiling and discovery: historical datasets are scanned for missing fields, uncommon transitions, mapping inconsistencies, and structural anomalies.



Phase 2 – Learning normal behaviour: by analysing several years of historical data, the system will learn typical transition paths (such as active → deferred → pensioner) and flagged anomalies like reactivated retirees or contribution gaps.

Phase 3 – Prediction and correction: using machine learning, the system learns patterns to predict missing data and estimate likely values. It also detects contribution gaps from upload errors. Each prediction includes an explainability score, helping data stewards verify the reasoning.

AI-driven data quality tools integrated into actuarial pipelines – in R, Python, or commercial platforms – automatically check data before each model run. If results fall outside set limits, the model pauses, alerts are triggered, and dashboards display the affected records, all fully logged for audit.

The outcome: fewer manual fixes, faster reporting, and greater trust in model inputs. Each run gets a data-quality score, ensuring reliability, traceability, and speed.

In this way, AI becomes part of the actuarial ecosystem – scaling human expertise across millions of records while strengthening transparency and governance.

CONQUERING THE INHERITED CHALLENGES

Legacy systems are a major barrier to modern data management. Pension and insurance providers often rely on old mainframes or custom tools that don't connect easily with new analytics platforms and replacing them is costly and risky.

AI offers a smarter alternative: it adds intelligence around existing systems. Middleware² extracts data, applies machine-learning checks and corrections, and returns clean outputs – without changing the legacy core.

Through APIs³ and data connectors⁴, AI can work in near real time across diverse systems. This 'non-invasive modernization' boosts data reliability while reducing disruption and cost – enabling organizations to upgrade data quality without a full IT overhaul.

GOVERNANCE, ETHICS, AND THE HUMAN IN THE LOOP

Regulators require pension and insurance providers to base their calculations on accurate, complete, and traceable data. AI-driven quality systems support this with automated audit trails, explainable dashboards, and full data-lineage tracking – logging every anomaly, correction, and decision for transparent audits.

Automation, however, doesn't replace human judgment – it strengthens it. AI spots anomalies but can't always read context, such as whether a surge in lump-sum withdrawals reflects an error or a real policy change. Human experts review and refine these findings, feeding insights back into the system. This 'human-in-the-loop' model blends machine efficiency with professional accountability.

Ethical governance is crucial. Because pension and insurance data are highly sensitive, AI must meet strict standards of privacy, fairness, and bias control. Above all, transparency ensures users understand how and why decisions are made – building trust in AI as a responsible partner in actuarial work.

FUTURE OUTLOOK – TOWARDS PREDICTIVE GOVERNANCE

As AI matures, data quality management is shifting from reactive to predictive. Instead of fixing issues after they occur, systems can now forecast where and when they're likely to arise. Using metadata such as system logs, transaction volumes, and user activity, predictive governance anticipates risks – for example, flagging a likely rise in missing partner data after a system migration so checks can be done early.

At the same time, self-healing data pipelines are emerging. Using reinforcement learning, they test and correct data automatically, improving with every cycle. Combined with blockchain-based lineage tracking, they could soon provide fully auditable, tamper-proof data environments.

For actuaries, this means fewer surprises during model reviews. The question shifts from '*Is our data correct?*' to '*How confident are we it will stay correct?*'

Predictive governance transforms data quality from a defensive duty into a strategic advantage.

CONCLUSION – THE SILENT GUARDIAN OF ACTUARIAL RELIABILITY

In actuarial work, models get most of the attention – their design, assumptions, and outputs are rigorously tested. Yet every model depends on one thing: its data. AI is becoming the quiet guardian of that foundation. By continuously learning, monitoring, and improving data quality, it lets actuaries focus on insight and strategy instead of data cleaning. Machine learning can operate within strict governance, transparency, and ethical standards, turning data quality from a periodic check into a continuous, intelligent process. Ultimately, securing data integrity is more than a technical task – it's a commitment to accuracy and trust. Every pension payment, insurance claim, and solvency figure depends on it. In this way, AI not only improves efficiency but also protects the trust that underpins the entire financial system. ■

1 – <https://actuary.org/article/model-behavior-applications-of-artificial-intelligence-in-actuarial-science-2/>
https://actuaries.org/app/uploads/2025/05/AITF2024_C2_Data-and-control-cycle_DRAFT.pdf

2 – software that acts as a bridge between different systems, applications, or databases – helping them communicate and share data smoothly

3 – API (Application Programming Interface) is a digital gateway that lets two systems communicate directly.

4 – Data Connector is the practical link that uses an API to move or sync data between systems