

# Promoting Responsible Artificial Intelligence in Insurance

**Benno Keller**

Special Advisor, Digital and Innovation, The Geneva Association

*The use of artificial intelligence in insurance has the potential to yield economic and societal benefits that go beyond insurers and their customers by improving risk pooling and enhancing risk reduction, mitigation and prevention. In order to foster the adoption of AI systems and realise these benefits, insurers need to earn the trust of customers by using new technologies responsibly. In order to promote key principles for responsible AI such as transparency and explainability, fairness, safety, accountability and privacy, insurers should establish internal guidelines and policies, adopt an appropriate governance structure to address related risks, and develop and roll out comprehensive training programmes for employees and agents.*

## Artificial intelligence in insurance

Since the end of the 'AI Winter', the period from the 1970s to the end of the 1990s characterised by setbacks and disappointment, artificial intelligence (AI) has made remarkable progress. Today, many insurers are rolling out intelligent systems that automate routine tasks or assist human decision-making along the entire insurance value chain. Such systems combine new types of learning algorithms with the analysis of data from new sources, such as online media and the Internet of Things (IoT).

For example, progress in natural language processing allows intelligent systems to 'talk' and interact with humans. Insurers are increasingly using conversational agents (e.g. chatbots) that can identify and respond to complex customer queries and are available 24/7. Furthermore, intelligent systems can 'view' and recognise objects in pictures and extract related information. Such computer vision allows insurers to automate manual and cognitive routine tasks, e.g. to extract data from written documents and pictures for use in the underwriting or claims process.

Intelligent systems excel in detecting patterns and correlations in complex data in ways that humans find difficult or even impossible. Such identified patterns are the basis for analytical tasks such as classification, regression and clustering that play an important role in the insurance business model. Compared to traditional modelling approaches used in insurance (such as generalised linear models), intelligent systems have the potential to provide much more accurate predictions, because they can learn complex non-linear relationships between variables.

Further progress in learning algorithms will enable these systems to autonomously make standardised decisions under human supervision in a growing number of areas.

## Benefits of AI in insurance

In order to foster the adoption of AI systems and realise the benefits, insurers need to earn the trust of their customers by using the new technology responsibly. AI can also help insurers to enhance their role in risk reduction, mitigation and prevention (see Figure 1).

**Figure 1: Socio-economic benefits of AI in insurance**

### Expand the scope of risk pooling

- Extend insurance cover to new and previously uninsured customer segments by facilitating access to personalised products (e.g. life insurance for individuals with pre-existing conditions)
- Expand the range of risks for which insurance cover is available through improved risk insights (e.g. cyber risks)

### Reduce the cost of risk pooling

- Offer more cost-efficient insurance through the automation of specific tasks, better risk assessments and reduction of moral hazard and adverse selection

### Mitigate and prevent risks

- Yield novel risk insights that help mitigate and prevent risks
- Enable early warning systems that make it possible to reduce losses

Source: The Geneva Association

## Principles for Responsible AI

Over the past few years, an intense and ongoing debate has emerged on what the responsible use of AI entails, and various governmental and non-governmental actors have issued guidelines for its ethical use. An analysis of these guidelines reveals a global convergence towards five core principles for responsible AI: (1) *transparency and explainability*, (2) *fairness*, (3) *safety*, (4) *accountability* and (5) *privacy*. Considerable uncertainty remains, however, regarding how ethical principles and guidelines should be implemented in a specific context.

In insurance, these core principles have long played an important role. As a matter of fact, various laws and regulations—including insurance law, privacy and data protection laws, anti-discrimination laws and supervisory requirements—govern the fair, transparent and accountable behaviour of insurers as well as the protection of privacy.

Nevertheless, the use of AI raises some intricate questions for insurers. Which trade-offs arise with the implementation of core principles? How can insurers foster and demonstrate adherence to such principles? What changes, if any, are necessary to existing governance mechanisms and risk management

frameworks for this purpose? Insurance regulators too are increasingly considering these questions.

The Geneva Association's report, *Promoting Responsible Artificial Intelligence in Insurance*, analyses important ethical guidelines for the use of AI and explores how to implement them in the context of insurance, taking into account current developments in computer science on fair machine learning. A definitive answer to these questions will remain elusive. The Geneva Association's report, however, aims to contribute to identifying and discussing key trade-offs that arise when implementing core principles for responsible AI in insurance. While all of the above-mentioned are critical, the report focuses specifically on the principles of 1) transparency and explainability and 2) fairness, because they raise particularly complex issues in insurance.

## Transparency and explainability

Transparency and explainability are important for building trust with customers and other stakeholders. Some guidelines mention the importance of transparency and explainability to enable individuals to seek redress against decisions affecting them. Providing an explanation is particularly important when



a decision has a significant impact on the affected individual. Therefore, the degree to which explainability is needed highly depends on the context and severity of the consequences of an erroneous or otherwise inaccurate output. Interpretability of algorithmic outcomes is also an indispensable prerequisite for the assessment of the performance of AI systems and their continuous improvement and thus for sound data science.

Providing meaningful explanations is a challenge, as some 'black box' algorithms are by nature complex—the price to pay for better accuracy—and therefore difficult to interpret and explain. In recent years, considerable efforts have been undertaken in computer science to overcome the challenges of interpreting and explaining 'black box' algorithms. *Reverse engineering* approaches, for example, consist of building interpretable algorithmic surrogates,<sup>1</sup> a recent technique which needs to be better understood. *Design approaches* rely on imposing certain constraints on the predictions.

When it is not possible to provide an explanation of the role of different variables in a given decision, other approaches, such as certification of the AI system by an independent body, may be used to foster trust. Nevertheless, the implementation of interpretable models should be encouraged, in particular if their outcomes have a significant impact on customers. When used for risk selection and pricing, trust in AI systems can be fostered by using data sources that are related to the insured risk in a way which is intuitively understandable to customers. The benefits of overly complex models may not always justify their reduced interpretability.

## Fairness

How to ensure fairness is particularly intricate as there are several mutually exclusive definitions of a fair decision. Fairness is usually associated with many different values such as freedom, dignity, autonomy, privacy, non-discrimination, equality and diversity, among others. These values often need to be interpreted in context, including the cultural context. It is therefore impossible to provide a universal standard of fairness.

Decisions should be fair in the sense that they do not discriminate and disadvantage individuals or groups of individuals. To eliminate or minimise unfair bias in AI-driven decisions is thus a core requirement in ethical guidelines.

Humans are not free from bias, and the use of AI may actually enhance fairness of decisions in some circumstances. However, if deployed at scale, even a minimal bias in an AI system can affect large numbers of individuals. For data scientists, the challenge is therefore to identify, measure and mitigate potential bias that could put certain groups at a systematic disadvantage.

In insurance, the following fairness concepts are of particular relevance:

*Actuarial fairness* demands that similar risks are treated in similar ways, so that the premium an individual pays corresponds to the actual risk.

*Non-discrimination* implies that the premium an individual pays is not based on factors that the individual cannot influence, in particular if these factors relate to socially protected groups. Non-discrimination is related to the notions of group fairness—the goal of groups defined by protected attributes (such as gender, ethnicity, sexual orientation, etc.) to receive similar treatments or outcomes to those without protected attributes and avoid disparate treatment from decisions that are (partly) based on a sensitive attribute. Group fairness complements the notion of disparate impact, i.e. outcomes that disproportionately hurt or benefit people with certain sensitive attributes.

*Solidarity or mutualisation* is often mentioned as a key feature of insurance. Increasing individualisation of insurance, driven by AI systems and data analytics, may disadvantage certain groups, for example by charging unaffordable premiums or denying cover altogether.

The relative importance of different fairness concepts varies between different types of insurance and jurisdictions. For example, in many jurisdictions, solidarity is considered an essential feature in health insurance.

To identify, measure and mitigate potential bias, insurers need to identify context-specific fairness definitions for each use of AI. As fairness is not a new concept in insurance, existing frameworks and norms (such as actuarial ethics) should be leveraged.

---

1 Algorithmic surrogates can be understood as approximation models that mimic the output of the algorithm and are easily interpretable.

## Recommendations

In order to promote the responsible use of AI within their organisations, insurers should consider the following three actions:

### **1) Establish internal guidelines and policies for the use of AI**

Internal guidelines and policies play an important role in raising awareness of the benefit–risk trade-offs in the use of AI in insurance. Insurers should therefore develop and adopt relevant guidelines and policies that include principles for dealing with issues of transparency and explainability and of fairness. In particular, guidelines should help to clarify how the benefits and risks of using AI should be assessed on a case-by-case basis. Actuaries, risk managers, data scientists and data protection officers should closely cooperate in the development and implementation of such guidelines and policies.

In doing so, insurers may adopt a risk-based approach to the governance of AI, implying a special focus on the uses of AI systems that may have a significant impact on individuals. The significance of impact refers to the consequences decisions have on affected individuals and depends on the specific circumstance of the use of AI.

### **2) Adopt appropriate governance structures**

The responsible use of AI in insurance requires governance structures that hold decision makers—individuals, committees and departments—accountable. Decision makers should have the necessary competencies, skills and expertise, and organisations should have in place effective processes, including triggers and escalation.

With many different governance models possible, each with its own advantages and disadvantages, the optimal organisational model will depend on a company’s structure and culture.

### **3) Develop and roll out internal training programmes**

Finally, ensuring a responsible use of AI requires awareness of related benefits and risks across different functions and managerial levels. In order to raise awareness, insurers should consider developing and rolling out comprehensive training programmes on the benefits and risks of AI and their respective internal guidelines and policies. Such training programmes would ideally target employees at different management levels and from various internal functions involved in decision-making, including agents and other customer-facing employees.

