

AI for Good vs. AI made evil: why ethical AI should be a top priority for insurers



AI for Good vs. AI made evil: why ethical AI should be a top priority for insurers

Insurance is lagging behind on AI

The use of AI has quickly expanded over the past few years across all sectors, driven by proven substantial financial benefits for organizations. The report [“Smart Money, How to drive AI at scale to transform the financial services customer experience”](#) from CapGemini Research Institute pinpoints that 13% of insurers have witnessed a reduction in cost of operations and 11% an increase in revenue per customer after deploying AI.

Nonetheless, the adoption rate of AI greatly varies across sectors, with insurance lagging behind other fields: only 6% of insurers have been able to deploy AI at scale - according to the same report, based on a survey carried out in March-April 2020 - vs life science at 27%, retailers at 21% or even government at 9%.

The conundrum of AI turning evil and how Transparent AI can defeat that

Two main factors can explain the lag of the Insurance sector in using AI at scale:

- One is the relative ‘lack of incentives’: the overall stability of the insurance competitive landscape, coupled with formerly relatively advantageous interest rates that allowed for attractive financial investments for decades have been comfortable conditions up until pretty recently. Yet, this has come to an end with rising competitive pressure - from GAFAs and insurtechs - and low interest rates, putting pressure on insurers’ operational bottom lines. (see position paper 1)

- The second is the perceived risks of AI: traditional AI approaches are usually blackbox and entail uncontrollable and unacceptable biases, generating issues both in terms of process and output.
 - **Blackbox models:** in the insurance space specifically, traditional AI cannot be used in production for core processes that require utmost transparency, such as pricing. Uncontrollable blackbox AI can expose carriers to risks of adverse selection, with significant financial impact if AI is misused in pricing decisions.
 - **Uncontrollable and unacceptable biases:** there are many examples of AI-entailed biases. For example, AI-driven facial recognition technologies can misidentify nonwhite or female faces ; AI-driven hiring software can reinforce gender and racial prejudice. Vendors such as IBM, halted the sale of such technologies. Another illustration of this risk is the gender-bias controversy around Apple's credit card in November 2019. Even though gender did not appear on the credit card applicant forms, the algorithms reconstructed this information by associating variables such as type of spending, or retailers visited, and applied a higher risk factor to a large share of women on these grounds - without data scientists or bank rating modelers knowing or noticing it.

Such use cases, where traditional blackbox AI doesn't cut it, are numerous in the insurance sector. This is where Transparent AI comes into play.

The notion of Transparency applied to AI is actually at the core of governing principles for ethical AI. "Ethics guidelines for trustworthy AI" from the European Commission established a list of 7 principles for governing the ethical issues arising from AI system use. Principle number 4 is "Transparency". AI systems should be based upon the principle of explainability, and should encompass transparency and communication of the elements involved, namely the data, the system and business models.

Scrutiny and pressure from regulatory bodies around AI transparency are increasing, to control the risk to society, in terms of fairness of treatment, and the potential cost of adverse selection for insurers. In the insurance sector, transparency and auditability cannot be compromised with.

The US National Association of Insurance Commissioners (NAIC) has formed a special committee focused on race and insurance, and adopted guiding principles that demand AI be fair, ethical, accountable, and safe. Insurers and rating and advisory organizations "*should be responsible for the creation, implementation, and impact of any AI system, even if the impact is unintended*" according to NAIC's Innovation and Technology Task Force.

In other words, AI should be at the service of, and controlled by humans; not served by, and controlling humans.

As such, transparent AI is a privileged way to counter the risks and unwanted side effects of traditional AI. In other words, an ethical AI defeating an AI that could go wrong - making transparent AI a prerequisite for insurance companies to be able to deploy AI at scale.

Just so we are clear, Transparent AI is much more than explainable AI (XAI)

Christopher Woolard, Executive Director of Strategy and Competition at the Financial Conduct Authority, says "*Algorithmic decision-making needs to be 'explainable'*". "*But what level does that explainability need to be? Explainable to an informed expert, to the CEO of the*

firm or to the consumer themselves? It's possible to 'build in' an explanation by using a more interpretable algorithm in the first place, but this may dull the predictive edge of the technology. So what takes precedence - the accuracy of the prediction or the ability to explain it?" from "FCA to explore the "explainability" of AI in consumer finance," July 2019.

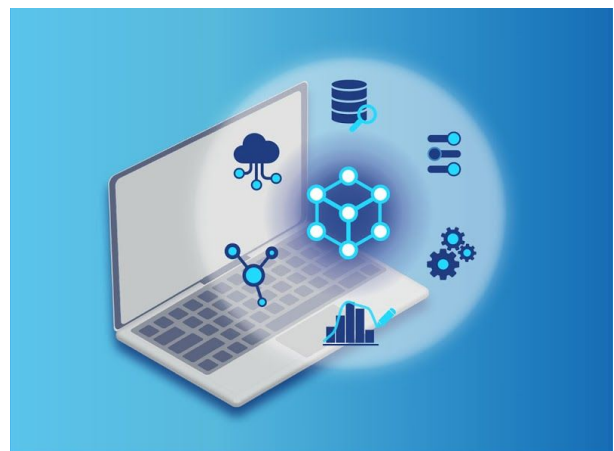
So, can we then settle for an AI that is explainable?

The subtlety that is hidden here comes from the difference between *explainability* and *transparency*.

An *explainable* AI can indeed be explained, but only in hindsight. It could be compared to an inverted train of thought, where you start from the end and arrive at the beginning, to retrace the path that was chosen by the algorithm. *Explainable* AI allows to understand and justify the decision that was made by the algorithm, but only once the decision has been made and the result is determined.

It does not provide a transparent understanding of the train of thought which leads to the decision that is made and leaves many questions unanswered: what choice will the algorithm make in a different situation? For what reasons? Using what variables? In exactly which way? As an example, correlations between variables that are computed by algorithms (typically what happened in the Apple credit card gender bias case) are almost uncontrollable.

In other words, only the impact of the decision can be understood but not the process that was followed to get there. This is precisely what *Transparent* AI addresses. On top of allowing to fully explain the output of a model, it allows to modify, amend, influence the computation and modeling process in itself. As such, AI *transparency* goes much further than *explainable* AI, combining both the automation power and performance of AI with the reliability, auditability - in a nutshell the transparency - of traditional manual models.



Transparent AI is fair AI: "AI for good"

At the end of the day, the mission of insurers is to support customers who are facing hardships through a solidarity-based model. If blackbox AI is incompatible with fairness of treatment, transparent AI, on the contrary, allows insurers to deliver greater value to their customers and to ensure they get fair and equal treatment.

Regulators are increasingly chasing opaque pricing practices that result in unfair and unequal consumers treatment. The UK Financial Conduct Authority (FCA) recently slammed insurers for "complex and opaque" pricing for car and home insurance policies, proposing radical reforms designed to boost competition, deliver fair value and increase trust in the sector. If its proposals are accepted, customers will save an estimated £3.7bn over 10 years.

