

ARTIFICIAL INTELLIGENCE GOVERNANCE PRINCIPLES: TOWARDS ETHICAL AND TRUSTWORTHY ARTIFICIAL INTELLIGENCE IN THE EUROPEAN INSURANCE SECTOR

A report from EIOPA's Consultative Expert Group
on Digital Ethics in insurance

This document represents the views of the members of EIOPA's Consultative Expert Group on Digital Ethics in insurance. The members named in this document support the overall AI governance framework put forward in this report, although they do not necessarily agree with every single statement in the document. EIOPA has created and supported the work of this stakeholder group, but the views included in this report do not necessarily represent the position of EIOPA.

PDF ISBN 978-92-9473-303-0 doi:10.2854/49874 EI-08-21-137-EN-N

Luxembourg: Publications Office of the European Union, 2021

© EIOPA, 2021

Reproduction is authorised provided the source is acknowledged.

For any use or reproduction of photos or other material that is not under the EIOPA copyright, permission must be sought directly from the copyright holders.

ARTIFICIAL INTELLIGENCE GOVERNANCE PRINCIPLES: TOWARDS ETHICAL AND TRUSTWORTHY ARTIFICIAL INTELLIGENCE IN THE EUROPEAN INSURANCE SECTOR

A report from EIOPA's Consultative Expert Group
on Digital Ethics in insurance

CONTENTS

I. FOREWORD	4
II. EXECUTIVE SUMMARY	6
III. GOVERNANCE PRINCIPLES FOR AN ETHICAL AND TRUSTWORTHY AI IN THE EUROPEAN INSURANCE SECTOR	8
IV. INTRODUCTION	9
1. The use of AI in the insurance sector	9
2. The importance of digital ethics in insurance	10
3. Approach taken, definitions and scope	14
V. AI USE CASE IMPACT ASSESSMENT	17
1. Assessing the impact of an AI use case to determine the relevant governance measures	17
2. AI use case impact assessment framework	18
VI. FAIRNESS AND NON-DISCRIMINATION	21
1. Fairness and non-discrimination in insurance	22
2. Fairness and non-discrimination in specific AI use cases in insurance	32
VII. TRANSPARENCY AND EXPLAINABILITY	40
1. Transparency and explainability in the insurance sector	41
2. Transparency and explainability in specific AI use cases in insurance	46
VIII. HUMAN OVERSIGHT	48
1. Human oversight in the insurance sector	49
2. Human oversight in specific AI use cases in insurance	53
IX. DATA GOVERNANCE AND RECORD KEEPING	56
1. Data governance and record keeping in insurance	57
2. Data governance and record keeping in specific AI use cases in insurance	60
X. ROBUSTNESS AND PERFORMANCE	62
1. Robustness and performance in insurance	63
2. Robustness and performance in specific AI use cases in insurance	65
XI. CONCLUSION	67

ANNEX 1 – BENEFITS AND ETHICAL CHALLENGES OF AI USE CASES ACROSS THE INSURANCE VALUE CHAIN	69
ANNEX 2 – AI USE IMPACT ASSESSMENT FRAMEWORK	74
ANNEX 3 – NON DISCRIMINATION REGULATORY FRAMEWORK IN INSURANCE	79
ANNEX 4 – LIST OF ACRONYMS	81
ANNEX 5 – MEMBERS OF EIOPA'S CONSULTATIVE EXPERT GROUP ON DIGITAL ETHICS IN INSURANCE	82
GLOSSARY	84
REFERENCES	86

I. FOREWORD

Due to technological advances digitalised data and its use play an increasingly important role in our societies. The amount of digital data doubles in short intervals, it is collected from different sources and formats and its manipulation gets more efficient. Data scientists invent novel ways of drawing better conclusions from the data. Technology is finally making Artificial Intelligence (AI) into a relevant tool to improve our societies.

Insurance has been a heavy user of data from practically early days of its existence. The collection of data, even when available, has been expensive. Analysis of this data has been expensive too and often inaccurate. Instead of an as exact as possible knowledge of insured persons and physical objects insurers have had to live with crude indicators of the risk inherent in each case. The emergence of Big Data (BD) and AI are changing this, making it possible to have more exact knowledge and changing the ways insurers interact with policyholders

Insurance has also through all of its existence dealt with ethical problems. Fair treatment of the insured pool and each policyholder has created problems that have been solved with varying degrees of success. Developments with BD and AI are not creating new challenges in this area. Instead, they are offering possibilities to deal with some in a better way but also exacerbating other. Current ethical issues in insurance are also acute only partly due to changes in BD/AI. Maybe even more often topical issues in this area result from changes in our societies, i.e., from changing thoughts on what a good life is and how individuals should be treated.

Ethics is about good life. There have been different efforts to formalise ethics, i.e., to create a framework to determine in an undisputed manner what is ethical and what is not. This has proved to be impossible. Therefore there cannot be an algorithmic way to integrate ethics into the use of data in a way that always reaches correct solutions. This report approaches ethical issues in a more down-to-earth manner. Ethics is thought to mean approaches that are fair based on international and national recommendations, standards and treaties, and of course legislation. Our understanding is that this represents what most people would understand as ethical.

Insurance exists in many forms. One dividing line is between (mandatory) social insurance and private insurance. This report concentrates on private insurance. Possible issues on BD/AI in social insurance would need a separate analysis.

Ethical challenges in insurance result from separate interests of the main stakeholders of insurance activity. We can identify three key players:

- an individual seeking insurance cover or being insured,
- the pool of insured risks, and
- the insurer who manages the pool.

Usually the individual in this case is looking for suitable cover at a price that is as low as possible. The pool is a group of risks, independent enough that allows for risk sharing among the group utilising the Law of Large numbers or one of its softer forms. In the interest of the pool there should be certainty that none of its members is taking inappropriate advantage of the pool.

In many cases there are legal, contractual or informal ways of returning a certain part of the profit of the insurer to the pool and its insured even in situations where the insurer is a profit-making entity.

The requirements of insurability and the conflicting interests of these three stakeholders create situations with ethical dilemmas. In many cases this is related to the fair treatment of an individual when the interests of the pool and the insurer are taken into account. One can ask to what extent the legitimate interests of one of these players can be limited in order to honour the legitimate interests of the other two.

In our work we have looked at the challenges to fairness with the emergence of new technologies. Fairness is especially threatened with the treatment of individuals in more or less vulnerable situations. We have outlined tools in transparency and explainability to help identifying areas where fairness is threatened. And we have suggestions on how the governance of the use of AI should be organised to safeguard sound use of AI.

The scope of our work was ambitious. Analysing how BD/AI influences insurance's many processes and interactions with policyholders was a significant challenge that we took eagerly knowing that compromises in the number of analysed cases would be required. Some readers may wish that our report had covered specific forms of insurance in greater detail and provided more specific guidance for them. We believe that, while not covering every possible case, our report provides the tools for individuals and organisations to reflect on the ethical challenges of BD/AI in insurance and apply BD/AI techniques in a trustworthy manner. Should this require additional specialist knowledge, market participants (consumer associations, insurers and national supervisors) may want to work together in their respective markets to address those specific forms of insurance

The Chairs of the GDE: Esko Kivisaari, Lutz Wilhelmy and Pedro Écija Serrano

II. EXECUTIVE SUMMARY

The advent of new technologies such as artificial intelligence (hereinafter “AI”), cloud computing or the internet of things (hereinafter IoT), coupled with the increasing availability of data in today’s digital society and economy, are enabling opportunities for future growth and development in the insurance sector. In order to capitalize on the opportunities offered by digitalisation and leveraging on their experience on data analytics processes, in recent years several European insurance undertakings and intermediaries (hereinafter insurance firms) have embarked in ambitious digital transformation projects where AI plays a pivotal role. As shown by EIOPA’s Big Data Analytics thematic review in motor and health insurance, in 2018, already 31% of the participating European insurance firms were using AI and another 24% were at a “proof of concept” stage.

Several studies indicate that the Covid-19 pandemic has accelerated the adoption of AI across all sectors of the economy, which would reinforce the trend in insurance towards increasingly data-driven business models throughout the insurance value chain. More specifically, AI systems are increasingly used by insurance firms to process new and old datasets to underwrite risks and price insurance products, launch targeted marketing campaigns or to offer enhanced products and services to consumers (e.g. usage-based insurance products), using mobile phone applications or chat bots conveniently accessible on a 24/7 basis from any location. The benefits arising from AI in terms of prediction accuracy, automation, new products and services or cost reduction are remarkable. However, there are also growing concerns amongst stakeholders about the impact that the increasing adoption of AI could have on the financial inclusion of groups of protected classes or vulnerable consumers or on our society as a whole.

There is already a comprehensive legislative framework underpinning the activity of insurance firms, which is also applicable to the use of AI within their organisations. This is, particularly, the case of the Solvency II Directive, the Insurance Distribution Directive (IDD), the General Data Protection Regulation (GDPR) and the upcoming e-privacy Directive (ePD). For example, Article 41 (1) of Solvency II Directive requires “insurance and reinsurance undertakings to have in place an effective system of governance which provides for sound and prudent management of the business.” Existing legislation should indeed form the basis of any AI governance framework, but the different pieces of legislation need to be applied in a systematic manner and require unpacking to assist organisations understand what they mean in the context of AI. Furthermore, an ethical use of data and digital technologies implies a more extensive approach than merely complying with legal provisions and needs to take into consideration the provision of public good to society as part of the corporate social responsibility of firms.

Against this background, several initiatives have proliferated in recent years at international, European and national level aiming to promote an ethical and trustworthy AI in our society. Leveraging on these cross-sectorial initiatives, in particular on the Ethics Guidelines for Trustworthy AI developed by the European Commission’s High Level Expert Group on AI (hereinafter AI HLEG), EIOPA’s Consultative Expert Group on Digital Ethics (hereinafter GDE) has developed six AI governance principles to promote an

ethical and trustworthy Artificial Intelligence in the European insurance sector. The principles developed by EIOPA's multidisciplinary stakeholder group take into account the specificities of the insurance sector and lay down the key governance pillars for ethical and trustworthy AI in insurance.

The high-level principles are accompanied by additional guidance for insurance firms on how to implement them in practice throughout the AI system's lifecycle. For example, in order to implement the principle of proportionality, the report develops an AI use case impact assessment which could help insurance firms understand the potential outcome of AI use cases and subsequently, determine in a proportionate manner the "mix" of governance measures necessary to implement ethical and trustworthy AI systems within their organisations. Taking into account the large variety of different AI use cases in insurance, several of the recommendations included in this report would apply only to those use cases that have a higher impact on consumers and/or insurance firms. However, the fact that a specific AI use case does not require, for instance, a high level of explainability, does not imply a low level of control over the data and technologies used, since all applicable regulations must be respected at all times.

With regards to the use of AI in insurance pricing and underwriting, the report includes guidance on how to assess the appropriateness and necessity of rating factors, noting that correlation does not imply causation. Insurance firms should also avoid certain types of price and claims optimisation practices such as those aiming to maximise consumer's "willingness to pay" or "willingness to accept". From a transparency and explainability perspective, consumers should be provided with counterfactual explanations, i.e. they should be informed about the main rating factors that affect their premium to promote trust and enable them to adopt informed decisions. Concerning the use of AI for fraud detection purposes, adequate human oversight is key in so far fraud always needs to be proved by the insurance firm, and for these practices it may not be possible to provide very detailed explanations to avoid compromising insurance firm's legitimate interest to fight against fraud.

Finally, the report is based on the state-of-the-art of AI at the time of its publication. The GDE acknowledges that AI is an evolving technology with an ever-increasing number of applications and where extensive research is on-going. This is, particularly, the case in the area of transparency and explainability, as well as in the area of active fairness seeking to develop fairness and non-discrimination metrics to assess the outcomes of AI systems. As these areas of application and research evolve, the recommendations included in this report may also need to be revised in due course.

III. GOVERNANCE PRINCIPLES FOR AN ETHICAL AND TRUSTWORTHY AI IN THE EUROPEAN INSURANCE SECTOR

Principle of proportionality: Insurance firms should conduct an AI use case impact assessment in order to determine the governance measures required for a specific AI use case. The AI use case impact assessment and the governance measures should be proportionate to the potential impact of a specific AI use case on consumers and/or insurance firms. Insurance firms should then assess the combination of measures put in place in order to ensure an ethical and trustworthy use of AI.

Principle of fairness and non-discrimination: insurance firms should adhere to principles of fairness and non-discrimination when using AI. They should take into account the outcomes of AI systems, while balancing the interests of all the stakeholders involved. As part of their corporate social responsibility insurance firms should also take into account financial inclusion issues and consider ways to avoid reinforcing existing inequalities, especially for products that are socially beneficial. This includes assessing and developing measures to mitigate the impact of rating factors such as credit scores and avoiding the use of certain types of price and claims optimisation practices like those aiming to maximise consumers' "willingness to pay" or "willingness to accept". Fair use of data means ensuring that it is fit for purpose and respect the principle of human autonomy by developing AI systems that support consumers in their decision-making process. Insurance firms should make reasonable efforts to monitor and mitigate biases from data and AI systems. This may include using more explainable algorithms or developing fairness and non-discrimination metrics in high-impact AI applications. Insurance firms should develop their approach to fairness and keep records on the measures put in place to ensure fairness and non-discrimination.

Principle of transparency and explainability: Insurance firms should adapt the types of explanations to specific AI use cases and to the recipient stakeholders. Insurance firms should strive to use explainable AI models, in particular in high-impact AI use cases, although, in certain cases, they may combine model explainability with other governance measures insofar as they ensure the accountability of firms, including enabling access to adequate redress mechanisms. Explanations should be meaningful and easy to understand in order to help stakeholders make informed decisions.

Insurance firms should transparently communicate the data used in AI models to consumers and ensure that they are aware that they are interacting with an AI system, and its limitations.

Principle of Human Oversight: Insurance firms should establish adequate levels of human oversight throughout the AI system's life cycle. The organisational structure of insurance firms should assign and document clear roles and responsibilities for the staff involved in AI processes, fully embedded in their governance system. The roles and responsibilities of staff members may vary from one AI use case to another. It is also important that insurance firms assess the impact of AI on the work of employees and provide staff with adequate training.

Principle of data governance of record keeping: The provisions included in national and European data protection laws (e.g. GDPR) should be the basis for the implementation of sound data governance throughout the AI system lifecycle adapted to specific AI use cases. Insurance firms should ensure that data used in AI systems is accurate, complete and appropriate and they should apply the same data governance standards regardless of whether data is obtained from internal or external sources. Data should be stored in a safe and secured environment and, in particular for high-impact use cases, insurance firms should keep appropriate records of the data management processes and modelling methodologies in order to enable their traceability and auditability.

Principle of Robustness and Performance: Insurance firms should use robust AI systems, both when developed in-house or outsourced to third parties, taking into account their intended use and the potential to cause harm. AI systems should be fit for purpose and their performance should be assessed and monitored on an on-going basis, including the development of relevant performance metrics. It is important that the calibration, validation and reproducibility of AI systems is done on a sound manner that ensure that the AI systems outcomes are stable overtime and/or of a steady nature. AI systems should be deployed in resilient and secured IT infrastructures, including against cyber-attacks.

IV. INTRODUCTION

1. THE USE OF AI IN THE INSURANCE SECTOR

The increasing availability of data in today’s digital society combined with increasingly powerful data storing and processing technologies like cloud computing or AI are very relevant developments for the insurance sector, given that data analytics has always been at the core of its business model. Mathematical and statistical methods have historically been used in insurance to process personal and non-personal data in order to underwrite risks and price insurance policies, to quantify losses and to pay customer’s claims or to identify and prevent insurance fraud.

In recent years, the European insurance sector has embarked on a digital transformation process where AI

plays a pivotal role, given its potential to increase the efficiency of operational processes and reduce costs. EIOPA’s thematic review on the use of Big Data Analytics in motor and health insurance¹ showed that, in 2018, already 31% of participating European insurance undertakings were using AI and another 24% were at a “proof of concept” stage. The adoption of AI in the financial services sector reportedly has accelerated during the Covid-19 pandemic.²³ As shown in the non-comprehensive table below (further developed in Annex 1),⁴ there are multiple AI use case applications across all of the stages of the insurance value chain.

1 EIOPA (2019)
 2 KPMG (2021)
 3 Mckinsey (2020)
 4 Developed by the GDE based on work from TECHNGI research group, Loughborough University, www.techngi.uk

Figure 1 – Examples of AI use cases across the insurance value chain

Product design and development	Pricing and underwriting	Sales and distribution	Customer service	Loss Prevention	Claims management
<ul style="list-style-type: none"> Historical customer and survey data analysis to inform new products Predictive modelling of disease development patterns Novel products, e.g. parametric and usage-based insurance 	<ul style="list-style-type: none"> Enhanced risk assessments combining traditional and new data sources (including IoT data) Price optimisation: micro-segment / personalised pricing based on non-risk individual behavioural data (e.g. to estimate price elasticity, lifetime value and propensity to churn) and market competition analysis 	<ul style="list-style-type: none"> Digital marketing techniques based on the dynamic analysis of online search behaviour Virtual Assistant and Chatbots that utilise Natural Language Processing (NLP) and insurance ontologies to support communication Proactive customer communication, nudging and cross-selling of related services (“next-best action”) based on consumer data from Customer Relationship Management (CRM) systems 	<ul style="list-style-type: none"> Call centre sentiment analysis, route cause analysis, dynamic scripting and agent allocation Customer self-service through multiple channels using NLP, voice recognition, insurance ontology maps and chatbots Robotic Process Automation (RPA) including Optical Character Recognition (OCR) to extract information from documents (e.g. FNOL, email with questions complaints etc.) and route them to the correct department 	<ul style="list-style-type: none"> Provide diagnostic advice and coaching based on AI analytics from health and automotive big data, e.g. suggest exercise and driving behaviour changes 	<ul style="list-style-type: none"> Enhanced fraud analytics: claims scoring, anomaly detection, social network analytics and behavioral modelling Loss reserving: use of AI to estimate the value losses, in particular for high-frequency claims AI image recognition to estimate repair costs in household property insurance, business premises and automotive Automated segmentation of claims by type and complexity and automated invoice verification and payment process

Source: EIOPA Consultative Expert Group on Digital Ethics in insurance

AI presents numerous opportunities both for the insurance industry as well as for consumers, in particular from the perspective of prediction accuracy, automation, new products and services and cost reduction. More particularly, in the area of product design and development, the use of AI systems to process large amounts of real-time IoT data collected from health wearables devices, car telematics or connected homes allow insurance firms to offer new tailored products and services to the consumer needs and demands. For example, in the case of the so-called usage-based insurance products, the premium paid by consumers partly depends on their lifestyles or driving behaviour. This gives consumers access to a wider choice of suitable services including a number of loss prevention/risk-mitigation services such as driving behaviour coaching or suggestions to adopt healthier lifestyles, which can eventually lead to less car accidents or health problems.

As far as pricing and underwriting are concerned, the processing of traditional and new data sources by increasingly accurate AI systems allows insurance firms to more efficiently underwrite risks and price policies that more closely reflect the risk and characteristics of the individual. Some insurance firms also use price optimisation practices, where a number of non-risk related factors are used to estimate consumer's price elasticity, lifetime value estimation or propensity to churn in combination with market competition analysis.

The increasing personalisation / micro-segmentation of pricing increases the profitability for insurance firms in a highly price competitive sector and enables consumers with lower risk profiles to benefit from lower premiums. Some high-risk consumers who previously had difficulties to access insurance, may now face less hurdles to get an insurance coverage. For example, young drivers installing telematics devices in their cars or consumers with Type 2 diabetes using health wearable bracelets and providing access to the data to the insurance firms reportedly are able to obtain insurance coverage at lower premiums than what they would have been able to obtain without these devices (if any access).

The sales and distribution area of the insurance value chain is one of the areas that encounters a greater number of AI use cases. Digital marketing techniques based on the real-time dynamic analysis of online behaviour can allow insurance firms to capture consumer's attention via tailored offers increasing sales via their websites, apps or other digital distribution channels. Virtual assistants and chatbots powered by Natural Language Processing (NLP) techniques, facilitate the consumer's purchase journey on the insurance website and can solve non-sensitive Q&As.

Customer retention AI systems based on consumer's data centralised in comprehensive Customer Relationship Management (CRM) systems allow insurance firms to send targeted and personalised marketing offers and increase the loyalty of their customer base. Moreover, in addition to Robotic Process Automation (RPA) techniques traditionally used to automate repetitive tasks in call centres, AI (NLP) can be used to develop more sophisticated non-repetitive tasks in the area of consumer service such as to call centre voice sentiment analysis, to route cause analysis and to dynamic scripting or agent allocation.

Finally, enhanced fraud analytics techniques powered by AI systems such as claims scoring, anomaly detection, social network analytics and behavioural modelling help insurance firms more efficiently to fight against fraudulent practices including those of organised crime, which eventually results in lower premiums for honest consumers. By flagging potentially fraudulent claims, investigators can focus on claims that are likely to be fraudulent and reduce the number of false positives and false negatives.

Insurance firms can also optimise the calculation of technical provisions with the use of AI systems to estimate the value of losses, in particular in lines of business with high-frequency claims, i.e. where there is a sufficiently large number of data points available to train the AI system. Moreover, in the area of claims management, insurance firms are developing valuation systems using image analytics to assess vehicle damage with the aim of replacing the need for an engineer's inspection and lowering the claims management costs. Several insurance firms already use AI to automate invoice verification and claims payments, in particular for low-value claims.

2. THE IMPORTANCE OF DIGITAL ETHICS IN INSURANCE

Addressing digital ethics for the insurance industry is a necessary task. The operation of the insurance market has important economic and welfare functions for the wider society and it can generate both positive and negative externalities. In terms of social inclusion, life, health and non-life insurance lines all play an important role. The advent of the digital economy has afforded many industries an unprecedented opportunity to utilise new technologies such as AI in order to process information on clients. This has prompted many international institutions and national governments to produce reports and White Papers on the ethical use of digital technologies, including

the EU, the OECD and a number of Member State governments.⁵

That said, given the centrality of the insurance industry to the life of EU citizens, there is a need for a bespoke approach on digital ethics as it pertains to the insurance profession. The digital economy introduces new classes of risk around the consumer and, in many instances, threatens an information asymmetry between the client and an insurance firm that favours the latter. From a regulatory perspective, there are risks around both competition, fairness and non-discrimination. The challenge for regulators and supervisors alike resides in allowing the European insurance sector to take advantage of the innovation offered by the digital economy, whilst, at the same time, protecting the interests of consumers and citizens.⁶

Wider developments in behavioural analytics and the power of *Big Tech* in this area represents a potentially disruptive environment with an unprecedented amount of personal data now available. With the *Internet of Things* still in its nascent phase, data on consumers/citizens will become ever more granular. This trend towards ever more personalised services would seem to call into question the very principle of mutualisation on which insurance and its social pact are founded. This is particularly the case in the area of health insurance. This characterisation is somewhat contested with a counter-argument being that the distinction between stochastic and deterministic worlds remain pertinent. Stochastic risk will not disappear with AI, but the number of premium pools⁷ are increasing and becoming more granular⁸ and this could lead to financial exclusion of some high risk consumers. The situation could become more acute if Big Tech players were to enter the marketplace in view of the large amount of information they have about consumers or groups of consumers.

Moreover, more accurate consumer profiling techniques enabled by the use of the large amounts of data generated by individuals' behaviours online (social media in particular) or off-line (data harvested from smart wristbands, for example) would have a tendency to lift the "veil of ignorance"⁹ which underpins the pooling of insurance risks. The risk would be that algorithms using social media data, via the correlations created in datasets, end up laying down accepted set of norms for individual behaviours

from which we could only deviate by paying a higher insurance premium.

There are profound questions here on how risks are determined and the implications of a move from the collective and social determinants of behaviours to a model which implies the accountability of individuals. Accountability is an important concept here and the extent that individuals are made accountable for their lifestyle choices through the vector of insurance. Other risk factors, associated with the individual's environment or genetic makeup, would likely lead to inevitable discrimination and exclusion insofar as these are completely out of the hands of the individuals in question.¹⁰

Whilst well-structured regulatory approaches can mitigate risks, creating norms around the use of digital technologies by insurance firms is a complex task, the global nature of the insurance business means it traverses many jurisdictions and indeed, cultures. Moreover, ensuring non-discrimination and fairness for consumers of insurance companies in the context of digitalisation is a demanding task for regulators because of the diffuse nature of digital ethics. Of course, such goals should be reflected in the internal processes of insurance companies who have a moral duty in this regard.

The use of AI by insurance firms is an important part of the picture and this for Marchant 2019 "has many of the characteristics of other emerging technologies that make them refractory to comprehensive regulatory solutions."¹¹ The argument here is that the use of AI involves information flows across jurisdictions, from one set of professions to other groups of professionals and incorporates sets of practices that have not traditionally resided within the purview of the regulators.

The use of AI also threatens certain characteristics of what many ethicists and philosophers take to be intrinsic to human beings, including notions of dignity and autonomy. Such concepts do not fit comfortably into the current paradigm of financial regulation. This is to say, the ideas around controlling the normative impact of financial services as they pertain to the digital economy and surveillance are relatively new and not in current purview of regulators. This new environment might require new regulatory tools, which measure the capacity of the market to meet the needs of citizens (inclusion) with appropriate products (safe) for a fair price (competition).

5 AI HLEG (2019)

6 Bernardino, G., (2020)

7 The premium pool is defined as the subset of the risk pool (i.e. all insured of an insurance undertaking) that are paying the same premium

8 See page 33 of EIOPA (2019) and page 21 of FCA (2016)

9 Rawls, J. (1999)

10 CNIL (2017)

11 Marchant, G. (2019)

ETHICS AND FAIRNESS IN INSURANCE

The concept of fairness plays an important role in classic liberal thought as a manner of providing common ground for the resolution of conflict. In other words, fairness can furnish societal actors with a neutral space for the management of disputes. The idea of fairness is closely related to other key concepts that underpin any society such as justice, equal opportunity, freedom, trust, responsibility and accountability.

The idea of fairness spans multiple disciplines including sociology, law and politics and relates to the goal of the equal treatment of citizens and non-discrimination. That said, how fairness operates or *what is fair* remains highly contested and with much of the debate centring on the relationship between fairness and equality. Moreover, fairness has been linked to the concept of deservedness and the notion that people get what they deserve according to such attributes as hard work or particular skill-sets.

Within John Rawls' seminal work "A Theory of Justice", fairness plays a key role in the establishment of the "original position" in which citizens decide on the future shape of society without knowledge of their position in that society. Although not related to public policy-making in its widest sense, we can detect some resonance with the paradigm of risk-sharing and the practice of insurance in that on becoming part of the risk pool, participants are unaware of whether or not they will require compensation for an adverse event. Like insurance, fairness has an important temporal element as fair outcomes and risk profiles for citizens relate to life course and change over time. This is an important and often overlooked element as insurance provides security though different ages and is there to deal with changing fortunes.

In insurance, notions of fairness need to capture the interests of insurance firms, individual insured customers, the pool of insureds, and society as a whole. Their interests will impact how the concept of fairness is defined so, for example, insurance firms may stress their right to conduct the business of insurance freely within the legal bounds. Representatives of individual insureds may define fairness in this market as *inclusivity*. Representatives of the pool of insureds may stress actuarial fairness, according to which similar risks are treated similarly, so that the premium paid by individuals corresponds to their actual risk, taking into account that there are other factors that influence the premium (e.g. production costs). Society as a whole may put an efficient, well-functioning insurance market at the centre of its interests, as this fosters welfare and economic activity. The subject of fairness, responsibility and digital ethics in insurance markets has attracted a good deal of attention within the academy. A number of recently published papers attempt to deal with the issues around fairness and the use of AI by insurers.¹²

Many of the elements of the broad concept of fairness are reflected in existing professional practice and insurance and data protection regulation. The term "fairness" relates to requirements concerning the business conduct of insurance firms towards consumers. This includes policies on non-discrimination, access to insurance and the treatment of vulnerable consumers. Paradoxically, digitalisation represents both a challenge to establishing fairness in insurance and provides a means to implement more fairness in insurance in the future.

In the pursuit of fairness, some reflection is needed on whether it is "natural and inevitable" that the interests of stakeholders are inevitably opposed. If we consider the insurance firms and policyholders - the interests of both are intrinsically linked, the former cannot exist without the latter and the latter cannot have peace of mind without the former. The fact that the consumer seeks the best value for money and insurance firms seek to make sure the economic expectations of their owners are met (whether shareholders, investors, customers (e.g. in mutual insurers), or other stakeholders) means that their interests may not be aligned, but they also cannot

¹² Frezal, S. and Barry, L. (2019) and Holland, C.P., Mullins, M. and Cunneen, M., (2021) and Keller, B. (2020).

be described as completely opposing. The concept of fairness also pertains to competition between insurance companies.

Fairness in insurance is related to how the insurance market operates *for* society. Hence the fair operation of an insurance market is one where insurance firms offer reliable and effective insurance that is easy to compare and is marketed in a way that consumers can make informed choices. It would also be a market that provides insurance products that are essential for society. That could include a range of simple and affordable, default option premiums to afford unlucky and/or vulnerable people reasonable access. The notion of “essential insurance” is central here and speaks to challenges and demands from citizens/customers and public authorities. Here we see the important role that insurance plays in terms of societal responsibility and social equity in co-creating a market that offers opportunity and access to all citizens including vulnerable groups, to access essential products at an affordable price. On the other hand, unfortunately it is not always possible for insurance firms to offer insurance at an “affordable price” (e.g. flood insurance in heavily exposed areas, terrorist risk, pandemic risk etc.). In such cases it is a political and societal task to find solutions.

AI systems do not operate in an unregulated world. A number of legally binding rules at international, European and national level already apply or are relevant to the development, deployment and use of AI systems today. Legal sources include, but are not limited to: EU primary law (the Treaties of the European Union and its Charter of Fundamental Rights), EU secondary law (such as the Insurance Distribution Directive, Solvency II Framework, the General Data Protection Regulation, the Product Liability Directive, anti-discrimination Directives, or the Unfair Commercial Practices Directive), the UN Human Rights treaties and the Council of Europe conventions (such as the European Convention on Human Rights), and numerous EU Member State laws.

A precautionary approach also comes into play here. Many of the issues thrown up by the use of AI by insurance firms may have unwelcome long-term social consequences. For example, there is the potential that the adoption of more granular data sets by insurance firms may undermine ideas of risk-sharing and the principle of mutualisation and the belief in fairness that underpins it. AI may result in an unleashing of competitive forces in the insurance market that, rather than benefit citizens, result in more exclusion. The toolset provided by AI to insurance companies presents risks that will require regulatory and supervisory oversight. There are also some justified concerns over the advent of surveillance regimes whereby AI might be used to exclude certain cohorts from access to this important financial service which speaks to the need for non-discrimination.

The impact on human autonomy and subjectivities around freedom are important concerns and cannot be simply dismissed as dystopian fantasies. What the precautionary principle posits in the context of such uncertainty is an ethic of care and protection where we aim to safeguard post-war European values. With regard to digitalisation and insurance, we are in the early stages of adoption and hence this report is a timely intervention in the debates around fairness and non-discrimination. A new culture and understanding in the collection and use of data is to emerge. A fine-tuned perception of data categories (hard, structured, accurate, meta) should lead to a differentiated, appropriate treatment. In this regard, GDPR might require clarification of the concept of necessity and minimisation on specific insurance use cases, in a near future.

With BDA and AI in insurance, we find ourselves in a classic *pace problem* dilemma. We see this across a range of economic activities where the speed of technological developments is such that the regulatory and legal responses struggle to keep pace. It is a risk governance problem that is widely cited and implies the need for response from public policy makers, industry and civil society alike. To date, much of the focus has resided in material science and biotech, but the *pace problem* is now embedded in financial services.

Ideally, regulation should strive to keep pace with new technology and the *pace problem* is not a rationale for regulatory inaction or indeed self-regulation. That said, in the absence of clear sets of rules or a well-established body of law, be it hard or soft, there is a strong emphasis

on the development of a set of ethical guidelines to mitigate conduct risk.¹³ One approach to the challenge posed by the introduction of new technology is the so-called “precautionary principle”. In the context of financial services, this principle became more prominent after the financial crisis of 2008. However, balancing its implementation in the context of a dynamic market environment is a difficult task.¹⁴

In a fast-moving digital world, there is a strong imbalance between those who manage algorithms and data, and the data subjects, the latter struggling to exercise their rights. An example of this can be seen in attempts to exercise the right to obtain human intervention in the context of a decision made based on algorithmic processing, or the right to obtain information about the logic underpinning the operation of the algorithm.¹⁵ Another example is the current attempt by the European Commission¹⁶, among others, to conceptualise and build a digital identity for citizens that should provide them with the necessary sovereignty to fully manage their own data. Key here is the issue informed and freely given consent on the part of citizens/consumers.

The label “digital ethics” is relatively recent and dates from the first years of the 21st century. Its roots reside with the term “information ethics” which, like its successor, is a hybrid of disciplines, including philosophy, computer science and the social sciences. It concerns itself with the interface of the human with the digital realm and the need to preserve human dignity and manage changing power relations.¹⁷ Implicit here are the related concepts of fairness and non-discrimination. More recently, in terms of industry and the academy, we have seen the advent such of concepts as *ethically aligned design* and *corporate digital responsibility* (CDR) to mitigate risks posed by the use of digital technologies and data by business. In the case of CDR, this:

“Presupposes that ensuring the ethical design and uses of digital technologies and related data is not solely a technological challenge (e.g., developing algorithms for ethical reasoning). Rather, it requires organisations to develop a comprehensive, coherent set of norms, embedded in

their organisational culture, to govern the development and deployment of digital technology and data.”¹⁸

As Lobschat et al suggest, there is a need for industry to *internalise* the need for ethical reflection around the use and indeed misuse of digital technologies. One of the more imminent challenges relates to how to create sufficiently robust governance structures to achieve this aim and whether or not bespoke entities will be required to address ethical concepts such as fairness and non-discrimination. Such CDR-related values and norms share some principles and goals with Corporate Social Responsibility (CSR). CSR encompasses the economic, legal and ethical expectations that society has of organisations at a given point in time, and we propose that a similar perspective is inherent to any considerations of CDR as well.

However, particular attention will be needed to test the effectiveness of these new tools on competition and inclusion issues. Indeed, some participants doubt that this approach is sufficient to solve problems related to market structure and distorted competition. Arguably, the insurance sector will change more in the coming 10 years than it has in the past 50 years and all these changes represent both challenges and opportunities for insurance firms. They also present challenges and opportunities for insurance regulators and supervisors, who must adapt to an evolving landscape. In this process, it is important that insurance regulators and supervisors do not lose sight of policyholders’ interests. In seizing opportunities and overcoming challenges, positive consumer outcomes will always be the most fundamental measure of success.

3. APPROACH TAKEN, DEFINITIONS AND SCOPE

In the call for expression of interest of EIOPA’s Consultative Expert Group of Digital Ethics in insurance it was stated that the aim of the group was to develop a set of principles of digital responsibility in insurance seeking to promote the responsible use of new business models, technologies and data sources in the insurance sector. It was subsequently decided to narrow down the scope to the use of AI in the insurance sector, acknowledging the prevalent role that AI has in the digital transformation of the insurance sector and in order to align the work of the GDE with on-going cross-sectorial initiatives taking place

¹³ Carney, M. (2020)

¹⁴ O’Riordan, T. and Cameron, J. eds., (1994)

¹⁵ (CNIL) (2017)

¹⁶ European Commission (2020)

¹⁷ Capurro, R. (2017)

¹⁸ Lobschat, L., Mueller, B., Eggert, F., Brandimarte, L., Diefenbach, S., Kroschke, M. and Wirtz, J. (2019)

at European level. For this purpose, the report uses the definition of AI included in the recently published legislative proposal for a Regulation on the harmonisation of rules on artificial intelligence (Artificial Intelligence Act).¹⁹ However, following a technology-neutral approach and noting that to date there is no legal and specific definition of AI, the provisions included in this report are also relevant for other Big Data Analytics (BDA) processes used by insurance firms.

Moreover, the term “insurance firms” used in the report covers both insurance undertakings and insurance intermediaries, in so far as the large number of existing AI use cases in insurance can be used by both entities to support their day-to-day activities. The guidance provided in the present report therefore is addressed to both insurance undertakings and intermediaries when using AI in the respective areas of the insurance value chain where they are involved. Nevertheless, a distinction between insurance intermediaries and insurance undertakings is occasionally made when, due to the nature of a specific AI use case, insurance undertakings and intermediaries play a different role or are involved in their implementation in a different manner.

There is already a large array of AI ethical principles developed by public and private institutions at international, national, sectorial and individual insurance firm level. Each of these has its own idiosyncrasy and particularities reflecting the different stakeholders involved, its scope of application as well as the state-of-the-art of AI when it was published, noting that it is an evolving technology

which counts with an ever-increasing number of applications and where extensive research is on-going. However, as noted by the Berkman Klein Centre for Internet & Society at Harvard University,²⁰ these initiatives also have several commonalities, covering greater or less emphasis issues such as human rights, professional responsibility, human control of technology, accountability, fairness and non-discrimination and transparency and explainability.

The GDE has tried to leverage as much as possible on these initiatives, although given the nature and composition of this group, a particular focus was given to the cross-sectorial initiatives being developed at European level. More specifically, the GDE has tried to align its work as much as possible with the Ethics Guidelines for Trustworthy AI developed by the European Commission’s AI HLEG on April 2019, as well as the White Paper on AI subsequently published by the European Commission on March 2020.²¹ The AI HLEG’s Ethics Guidelines for Trustworthy AI identified seven requirements in order to ensure an ethical and trustworthy use of AI, which were used as a basis by the GDE to prepare the present report.

As it can be observed in Figure 2, the AI governance principles developed by the GDE closely follow the AI HLEG’s Ethics Guidelines for Trustworthy AI. Due to the already very large and ambitious scope of the report it was decided not to address environmental aspects, but on the other hand societal aspects are widely covered in the fairness and non-discrimination chapter. Similarly, there is not a stand-alone principle of accountability but it is explicitly mentioned in the principle of transparency

Figure 2 – Comparison between AI HLEG and GDE reports

Commission’s AI HLEG ethical guidelines for trustworthy AI	EIOPA’s GDE AI governance principles
Human Agency and Oversight	~ Human oversight
Technical robustness and safety	~ Robustness and performance
Privacy and Data Governance	~ Data governance and record keeping
Transparency	~ Transparency and explainability
Diversity, non-discrimination and fairness	~ Fairness and non-discrimination
Societal and environmental well-being	(Fairness and non-discrimination)
Accountability	(Transparency and Explainability / Data Governance and record keeping)
	Principle of proportionality

Source: EIOPA Consultative Expert Group on Digital Ethics in insurance

¹⁹ European Commission proposal 2021/0106 (COD) (2021)

²⁰ Fjed, J. Nagy A. (2020)

²¹ European Commission COM(2020) 65 (2020)

and explainability and it is also directly related with all the other principles, in particular the principle of data management and record keeping. Moreover, the principle of proportionality is also included in recognition of the wide number of AI use cases in insurance.

The GDE has further developed the high-level requirements developed by the AI HLEG, adapting them to the specificities of the insurance sector and focusing on those requirements that were deemed more relevant in insurance. As a result, the GDE has laid down six AI governance principles that the industry should consider when implementing or considering to implement AI within their organisations. The principles are followed by more detailed guidance seeking to clarify specific aspects or issues linked to those principles. More particularly, the report includes examples of how each of these principles could interplay with specific AI use cases in the insurance sector.

Each of the principles is covered in a separate chapter in this report. Each chapter begins with a short introduction, making reference to the most relevant legislative provisions in that specific area and also describing the topic and key challenges. The first principle addressed by the report is the principle of proportionality, which is a well-established principle in the insurance sector and also fully applicable in an AI context. The principle of pro-

portionality impacts the other governance measures covered in the remaining chapters, and in order to guide its application in an insurance and AI context, the GDE has developed an AI use case impact assessment framework.

The principle of fairness and non-discrimination covers a number of fairness and ethical dilemmas that may arise in the application of AI in insurance, and touches upon innovative approaches to fairness and non-discrimination like “fairness metrics”. Moreover, the principles of transparency and explainability explain the types of explanations that the different stakeholders in insurance need in specific AI uses cases, and also gives guidance to insurance firms on how to approach issues linked to the opacity of some AI systems (known as the “black-box” effect).

Furthermore, the principle of human oversight (often referred as “human in the loop”) represents a key governance measure for the responsible implementation of AI in insurance by ensuring a certain level of human oversight throughout the lifecycle of an AI system. Finally, the principle of data management and record-keeping and the one of robustness and performance address key issues arising from AI by providing extensions to the robust Model Risk Management (MRM) approach already in place in insurance to improve the governance and control mechanisms of critical models.

V. AI USE CASE IMPACT ASSESSMENT

Principle of proportionality: Insurance firms should conduct an AI use case impact assessment in order to determine the governance measures required for a specific AI use case. The AI use case impact assessment and the governance measures should be proportionate to the potential impact of a specific AI use case on consumers and/or insurance firms. Insurance firms should then assess the combination of measures put in place in order to ensure an ethical and trustworthy use of AI.

1. ASSESSING THE IMPACT OF AN AI USE CASE TO DETERMINE THE RELEVANT GOVERNANCE MEASURES

The principle of proportionality is a well-established principle in European insurance legislation. According to this principle, insurance undertakings should establish the necessary governance measures that are proportionate to the nature, scale and complexity of their operations (Article 41(2) Solvency II Directive). This also applies to the use of AI within the organisation. As shown in the introduction, there are several different types of AI use cases across the insurance value chain. Not all the AI use cases have the same impact on consumers and insurance firms. Therefore, the governance measures that firms need to implement to ensure an ethical and trustworthy AI differ from one use case to another; they should be proportionate to the characteristics (impact) of the specific AI use case at hand.

In order to help insurance firms assess the impact of a specific AI use case, the GDE proposes to follow the AI use case impact assessment below, which takes into account the impact of AI applications both on consumers as well as on insurance firms themselves, since indeed AI raises risks for both of them. Certainly, in case an AI systems produced erroneous predictions this could have both

prudential (e.g. model fails to price risks accurately) and conduct implications (e.g. potential biases could remain undetected leading to a higher risk of discrimination against certain groups). Insurance firms should, therefore, assess both implications and identify the responsible person or group of persons in their organisations (e.g. AI officer, Data Protection Office (DPO), end users, data committees etc.) to develop this impact assessment and document it and keep records of it.

The AI use case impact assessment itself should also be proportionate to the concrete AI use case; those cases that are likely to have a low impact on consumers and/or insurance firms would require a less thorough impact assessment than for those that can be reasonably expected to have a higher impact. Moreover, it is important to highlight that the AI use case impact assessment is just an initial stage of a broader AI governance framework. Once the insurance firm has assessed the impact of a specific AI use case, it will be able to determine the governance measures (i.e. transparency and explainability, human oversight, data management etc.) that need to be put in place across the lifecycle of the AI system in a proportionate manner. For example, if the AI use case impact assessment shows that a concrete AI use case has low impact (e.g. a simple Robotic Processes Automation process applied in back office operations), then the governance measures required for that AI use case would be very limited. Another example which is specific to the principle of transparency and explainability - all things being equal, the higher the impact of an AI use case, the greater the level of transparency and explainability measures the insurance firm should adopt, and vice-versa.

Notwithstanding the above, as further explained in the respective chapters in this report, there is not a straight-forward correlation between high impact and a high need of transparency and explainability. There are other governance measures such as human oversight ("human in the loop"), data management or robustness and performance that can mitigate the lack of explainability in certain situations. Therefore, once all the relevant governance measures have been implemented, insurance firms should assess once again the risks of AI uses and

Figure 3 – Process for assessing and implementing AI governance measures



Source: EIOPA Consultative Expert Group on Digital Ethics in insurance

determine whether the “mix” of governance measures is sufficient to ensure ethical and trustworthy AI systems.

Furthermore, there are also other dimensions that need to be taken into account other than the impact of an AI use case. Continuing with the example of transparency and explainability, in addition of the impact of an AI use case and combination of governance measures put in place, insurance firms should also take into account the context of the explanations that need to be provided, namely the recipient stakeholders of the explanations, since the explanations that need to be provided to consumers differ from those needed by auditors or supervisory authorities. The relevant dimensions for each governance measure are analysed in the relevant chapters of the GDE report.

2. AI USE CASE IMPACT ASSESSMENT FRAMEWORK

The GDE proposes that the impact of an AI use case is determined by the potential for harm²² (to an individual or to the insurance firm) on the basis of a two-pronged investigation into the severity of that harm and the likelihood that harm will occur. To keep the exercise proportionate, three levels of likelihood and severity have been

considered: High, Medium and Low. However, insurance firms can certainly define a greater number of levels shall they deem it appropriate. Combining the level of likelihood and the level of severity for a given AI use case, the level of risk is obtained as shown in the quadrant below.

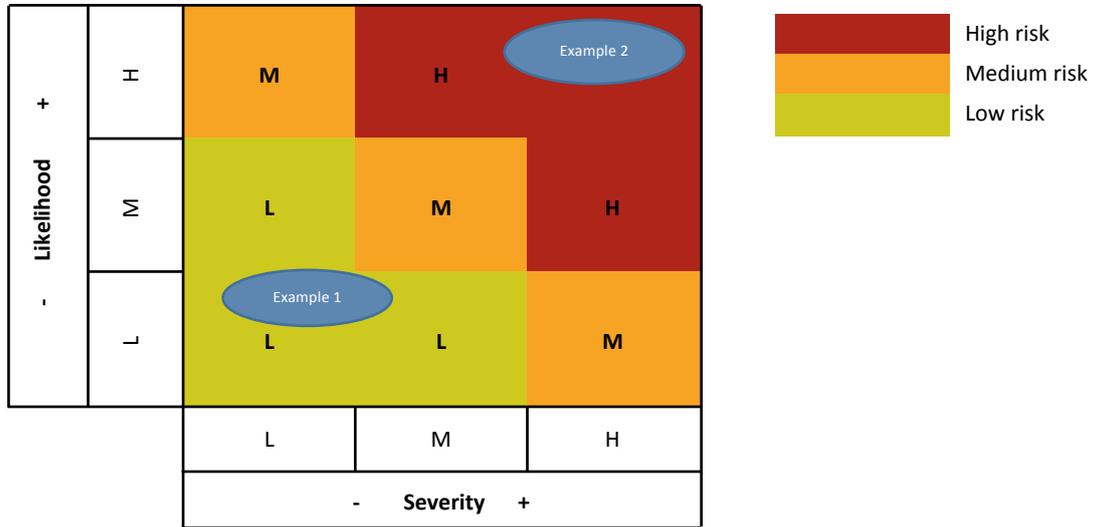
The likelihood and severity of harm of an AI use case is based on its potential impact on consumers and/or insurance firms. The proposed framework leverages as much as possible on already existing mechanisms; the assessment of impact on consumers closely follows Article 29 of the Data Protection Working Party Guidelines on the Data Protection Impact Assessment (DPIA),²³ in so far as the DPIA also follows a risk-based approach and aims to address similar risks than the ones arising from AI applications. The impact on insurance firms is fundamentally based on the risks that insurance undertakings regularly assess under their Own Risk and Solvency Assessment (ORSA) (Article 44 and 45 Solvency II Directive), and insurance intermediaries should also assess such risks when using AI. However, in addition to DPIA and ORSA considerations, the assessment also incorporates the recommendation from the AI HLEG to conduct an ex-ante fundamental rights impact assessment (FRIA),²⁴ being anti-discrimination and diversity considerations the ones that are deemed more relevant in an insurance and AI context. Other specificities of the insurance business and/or of AI applications are also incorporated in the analysis.

²² German Data Ethics Commission (2019)

²³ Article 29 Data Protection Working Party (WP 248 rev.01) (2017)

²⁴ AI HLEG (2020)

Figure 4 – Example of different impact of two AI use cases



Source: EIOPA Consultative Expert Group on Digital Ethics in insurance

The indicators proposed to be used as part of the AI use case impact assessment are shown in Figure 5 below (further developed in Annex II). By jointly assessing these indicators an overall assessment, judge based, can be obtained. However, not all the indicators necessarily have the same importance/weight. For example, a firm may consider that the risks of negatively impacting vulnerable consumers are more important than its reputational risk. It may even consider that only one indicator is enough to trigger an AI use case to be considered as high risk. It is up to the insurance firm using AI to responsibly assign to

each indicator the value/weight that they deem necessary in a responsible manner. Similarly, if case an AI use case is identified to have a high impact on the insurance firm but not on the consumer or vice versa, the governance measures should be adapted accordingly. For instance, if an AI use case has no impact on consumers less attention would need to be placed on removing potential bias in the training data of the AI system, but on the other hand the level of human oversight would arguably be similar if the AI use case had a high impact on consumers / individuals.

Figure 5 – AI use case impact assessment indicators

AI Use case Impact Assessment		
Impact on consumers	Impact on insurance firms	
Severity	Number of consumers affected	Business continuity
	Consumer interaction and interests	Financial Impact
	Types of consumers (e.g. vulnerable consumers)	Legal impact
	Human autonomy	Reputational impact
	Anti-discrimination and diversity	
	Insurance line of business relevance	
Likelihood	Evaluation or scoring, including profiling and predicting	
	Automated-decision making with legal or similar significant effect	
	Systematic monitoring	
	Model complexity/combining datasets	
	Innovative use or applying new technological or organisational solution	
	Type and amount of data used	
Outsourcing datasets and AI applications		

Source: EIOPA Consultative Expert Group on Digital Ethics in insurance

The severity of the harm for consumers that could potentially be sustained, for example as a result of a mistaken decision, depends on a number of factors, such as the number of consumers affected; the higher the numbers of consumers affected, the higher the impact of the AI use case. The significance of consumer's interests affected (e.g. monetary, non-material harm, health or legally protected rights) is also relevant, in particular in consumer-facing applications (as opposed to back-office operations where there is less interaction with consumers). The rights of underserved citizens (the ones who would like to access insurance but who have been rejected and/or received an unaffordable premium offer) should also be taken into account. Indeed it is relevant to differentiate between the different types of consumers affected, especially if those consumers are deemed to be in a vulnerable situation (e.g. old age, low level of studies, low income, etc.).

The impact on human autonomy is also relevant, since some AI systems used in some consumer-facing applications can have a direct impact on the behaviour and self-determination of consumers. It is also important to carefully assess the potential of unlawful discriminatory outcomes, for instance as a result of using imbalanced or biased datasets or complex algorithms capable of "reconstructing" protected characteristics even if they are not included in the training data. Finally, certain insurance lines of business are mandatory or are considered to be essential for consumers, such as motor, health or home insurance, and therefore these lines of business deserve especial care. Concerning the latter, different approaches may apply in different jurisdictions (e.g. in some jurisdictions home insurance is a requirement in order to be able to rent, or purchase a house, or some jurisdictions count with more robust Social Security systems than others).

The severity of the harm for insurance firms is determined by indicators like the impact on the firm's business continuity; if in the case the activity failed the insurance firm would incur a high risk of disruption to its core business, e.g. issuing policies, managing claims. The financial impact, including solvency risk, is also relevant and measured in terms of Gross Written Premiums or number of contracts. Finally, the legal and reputational consequences for the firm resulting from a mistaken decision should also be taken into consideration.

Finally, the likelihood that an AI use case could fail (if there are no adequate governance measures in place) depends on a number of factors such as the type of activity, with evaluation or scoring activities, including profiling and predicting, considered to be more risky from an eth-

ics perspective. The level of automation is also to be considered as a relevant factor to take into account, as well as the complexity of an AI system. The implementation of innovative business models and technologies are also considered to be more risky due to the lack of experience with dealing with them. The use of certain datasets (e.g. special categories of data as defined in Article 9.1 GDPR) are also a factor to take into consideration. Last but not least, outsourcing data sources or AI applications from third parties could potentially make more difficult the ability of insurance firms to ascertain the quality of the data or the outsourced tool, for instance due to intellectual property considerations.

VI. FAIRNESS AND NON-DISCRIMINATION

Principles of fairness and non-discrimination: insurance firms should adhere to principles of fairness and non-discrimination when using AI. They should take into account the outcomes of AI systems, while balancing the interests of all the stakeholders involved. As part of their corporate social responsibility insurance firms should take into account financial inclusion issues and consider ways to avoid reinforcing existing inequalities, especially for products that are socially beneficial. This includes assessing and developing measures to mitigate the impact of rating factors such as credit scores and avoiding the use of certain types of price and claims optimisation practices like those aiming to maximise consumers' "willingness to pay" or "willingness to accept". Fair use of data means ensuring that it is fit for purpose and respect the principle of human autonomy by developing AI systems that support consumers in their decision-making process. Insurance firms should make reasonable efforts to monitor and mitigate biases from data and AI systems. This may include using more explainable algorithms or developing fairness and non-discrimination metrics in high-impact AI applications. Insurance firms should develop their approach to fairness and keep records on the measures put in place to ensure fairness and non-discrimination.

According to Article 17(1), IDD, insurance distributors shall always act honestly, fairly and professionally in accordance with the best interests of their customers. This article is in line with the principle of fair processing of data of Article 5 GDPR. Furthermore, Article 20 (1), IDD specifies that any product proposed to the customer should be consistent with their demands and needs. Article 25 of the IDD and Delegated Regulation establish product oversight and governance (POG) requirements for insurance distributors, including the need to identify a target market and to regularly assess that the product remains consistent with the needs of the identified target market. The fairness principle is also recognised in Article 5 GDPR, and Article 5 of the

Unfair Commercial Practices Directive (UCPD). The latter considers unfair those practices that materially distort or are likely to materially distort the economic behaviour of an average consumer, highlighting the need to pay special attention to vulnerable consumers. Furthermore, the insurance industry has to comply with anti-discrimination legislation.

When considering fairness and non-discrimination in insurance, it is important to note that fairness is notoriously hard to define in general and abstract terms. It is a concept that varies according to context, interest group and over time. In Europe, for example, it was possible to use gender as a rating factor in insurance pricing and underwriting until prohibited by the *Test-Achats* ruling of the European Court of Justice,²⁵ even if there is a correlation between gender and risk in certain lines of business. The challenge of establishing clear principles to define fairness is also clear in the IDD and the GDPR. Both the IDD and the GDPR take into account procedural fairness and fair balancing of competing interests to mitigate harms on consumers / data subjects by providing specific safeguards and measures, but fall short of adopting a concrete definition of fairness.²⁶

The notion of fairness has different meanings for different stakeholders and for different AI applications. There is, for example, a clear difference between the meaning of fairness employed in an actuarial context and that used in the different approaches to healthcare insurance in member states across Europe and the United States.²⁷ As indicated by the AI HLEG, insurance firms need to find a balance between the interests of all the stakeholders involved.

The importance of certain insurance products for society is acknowledged by the existence of public insurance provisions where public transfers subsidise the cost of healthcare premiums, and other insurance products, including motor and household insurance, are mandatory in some jurisdictions. Insurance fulfils a wide variety of

²⁵ Judgement of the European Court of Justice, 1 March 2011, *Test-Achats*, Case C-236/09. See also Rebert, L., & Van Hoyweghen, I. (2015)

²⁶ Malgieri, G. (2020)

²⁷ Baker, T. (2010). Health Insurance, Risk, and Responsibility after the Patient Protection and Affordable Care Act. *U. Pa. L. Rev.* **159**, 1577.

functions, providing reassurance to consumers; security against significant financial consequences from death, unemployment or accident; a savings, pensions and investment-friendly environment by establishing a trust for the future. This enables insured customers to pursue risky, but beneficial activities, for them and for the community, that otherwise could not be exercised.

The important social role that insurance plays, highlights the need to promote financial inclusion. While the increasingly accurate and granular risks assessments enabled by AI systems have the potential to promote financial inclusion of some consumers, they could also make insurance more difficult to access for consumers classified as higher risk in competitive insurance markets or to consumers with more conservative approaches regarding the sharing of their personal data. Vulnerable consumers may be susceptible to harm and merit protection to avoid reinforcing existing inequalities. Insurance firms can contribute to financial inclusion but cannot offer solutions to entrenched social inequalities which are the preserve of public, governmental authorities.

Fairness in insurance also concerns the use of data and is key to ensure trust amongst stakeholders. Data collected should be appropriate for specific purposes and only the data necessary to meet those purposes should be collected. As new more sensitive personal data become available to insurers, such as DNA data, bank account or credit card data, IoT data collected from car telematics, health wearable devices, or social media data may reveal very sensitive aspects of individual's private lives and behaviours and therefore it is of utmost importance that it is used only in fair and appropriate ways.

This also means that insurance firms should be transparent about how they use the data and be able to appropriately explain these uses to consumers as well as to competent authorities. AI systems have the potential to support consumer decision-making process and insurance firms are in a position to develop and explain these benefits while respecting the principle of human autonomy. Potential benefits to different stakeholders have to be fairly balanced with negative effects, for example from the use of AI in some types of price and claims optimisation practices such as those seeking to maximise consumer's "willingness to pay" and "willingness to accept". Negative effects are especially critical when they cause harm to vulnerable consumers and protected classes.

This last point relates to the issue of unlawful discrimination. It is important to make reasonable efforts to monitor, and appropriately mitigate and/or remove biases in the

training and testing data to avoid these biases being reproduced in the outputs of AI systems. This also applies to the use of rating factors in insurance pricing and underwriting, beyond the specific cases in which the use of protected characteristics is permitted for risks assessment. The mitigation strategies should not hinder appropriate risk assessment. Monitoring training data for overt bias may not be sufficient when complex and opaque AI systems are trained on large datasets. In high-impact AI use cases, it may be preferable to consider other alternatives such as using more simple and explainable algorithms or using fairness and non-discrimination metrics to assess and monitor the outcomes of AI systems.

1. FAIRNESS AND NON-DISCRIMINATION IN INSURANCE

IT IS IMPORTANT TO ACKNOWLEDGE THE DIFFERENCE BETWEEN PRIVATE INSURANCE AND SOCIAL INSURANCE

An insurance collective, or pool of insured, is a risk community that generates particular forms of solidarity by redistributing risk without necessarily redistributing value within the portfolio or over time. In private insurance, the premium paid by consumers is paid for individual risk: high risk = high price and low risk = low price. Private insurance is not based on the principle of social compensation or subsidy whereby, for example, the young subsidise the costs of older people (risk) or wealthier customers cross-subsidise poorer customers (income). Solidaristic cross-subsidies of a kind do exist in private insurance between the members of the pool, where the premiums of customers who have not made claims is used to pay for those who have, but this happens only insofar as it is necessary to manage the risks in the pool in accordance with the law of large numbers. In fact, where premiums cover the cost of risk, the price for being protected is borne by each member of the pool individually. In addition, solidarity also takes place in private insurance when differences in risk between people are not fully represented in premiums. This form is instantiated by anti-discrimination law where law makers have agreed that some differences between people should not make a difference, for example in the cases of genetic information or gender.

Where societies have determined that certain forms of insurance are in the public interest, they may be made mandatory and/or organised as public or social insurance

schemes. In social insurance, individual cover may be decoupled from its production cost and redistributed as part of the total cost of insuring the whole pool on a *per capita* basis or in proportion to income. This is the case of statutory social security and welfare systems including some health insurance schemes. Indeed, country specificities are also relevant; for example, in Ireland, the use of community rating in health insurance also implies subsidising solidarity. Another example is the case of Denmark, where there are compulsory pension and life insurance schemes as part of terms of employment and based on collective agreements between social partners, which also imply subsidising solidarity.

This report focuses on private insurance, where the overall insurance portfolio has to be profitable and each insurance contract cannot be excessively unprofitable, such that the expected claim costs and other expenses should be covered or not too far in excess of the paid premium at the end of the contract.

SOUND AND TRANSPARENT GOVERNANCE PROCESSES ARE KEY TO ENSURE FAIRNESS AND NON-DISCRIMINATION

Sound and transparent data governance is key to ensure fair and non-discriminatory treatment of consumers. Chapter IX of this report explains in detail the data gov-

ernance measures that insurance firms should put in place throughout the AI system's lifecycle, including the need to ensure that the data is accurate, complete and appropriate for the purpose they are used. Chapter VII also highlights the importance of transparently communicating the use of data to consumers and obtaining their consent for the processing of his personal data (noting that Article 6 GDPR also foresees other legal grounds for the processing of personal data). Furthermore, the governance measures about robustness and performance as well human oversight described in Chapters X and Chapter VIII also aim to ensure a scientific and responsible way of working which would help address any shortcomings in the data. The present chapter focuses on the fair and non-discriminatory use of data and AI systems.

INSURANCE FIRMS SHOULD CONDUCT THEIR BUSINESS IN A FAIR MANNER WHEN USING AI AND MAKE REASONABLE EFFORTS TO TAKE INTO ACCOUNT THE OUTCOMES OF AI SYSTEMS

The difficulty of establishing a clear and stable definition of fairness is acknowledged in the approach taken by AI HLEG. This approach outlines instead two dimensions of fairness: a procedural dimension that focuses on governance and a substantive dimension that focuses on outcomes.

Figure 6 – Procedural and distributive fairness

Procedural Fairness (governance focus)	<ul style="list-style-type: none"> ■ Means the requirement of fair business conduct of the insurance firm vis-à-vis the consumer. ■ The IDD follows this approach, for instance by requiring; that distributors act honestly, fairly and professionally in accordance with the best interests of their customers (Article 17(i) IDD) that insurance products proposed to consumers are consistent with their demands and needs, that consumers are provided with objective information in a comprehensible form to allow them make informed decisions (Article 20(1) IDD). ■ Several of the governance measures put forward in this report aim to implement procedural fairness when using AI, for instance by promoting sufficient transparency and explainability to ensure accountability and effective redress mechanisms, or sound data management measures to make reasonable efforts to remove biases from the data used by AI systems.
Distributive Fairness (outcomes focus)	<ul style="list-style-type: none"> ■ Addresses the material outcomes of retail insurance distribution e.g. accessibility of cover at affordable prices and free of bias and discrimination. The AI HLEG also refers to the fair distribution of both benefits and costs. ■ Particularly relevant when using black-box AI systems to process large datasets where traditional scientific controls cannot easily be implemented (e.g. weight of variables, variance, asymmetry etc.). Despite its limitations, notably that correlation is not causation and model risk,²⁸ a growing area of research in the data science community seeks to develop fairness and non-discrimination metrics to assess the outcomes of AI systems. ■ The AI use case impact assessment in this report adopts this approach by proposing that insurance firms assess the impact of AI systems on groups of vulnerable consumers, on unlawful-discrimination, and the accessibility of certain lines of business that are important to financial inclusion.

Source: EIOPA Consultative Expert Group on Digital Ethics in insurance

²⁸ Model risk is defined as the risk arising from a lack of understanding of the circumstances under which the model may provide incorrect predictions

INSURANCE FIRMS SHOULD DEVELOP FAIR AI SYSTEMS BALANCING THE INTERESTS OF ALL STAKEHOLDERS INVOLVED

As suggested by the AI HLEG, insurance firms should, as part of their commitment to Corporate Social Responsibility (CSR) and bearing in mind the nature of free competition in the markets, find a balance between the various and changing interests of different stakeholders when considering the ethical challenges of AI and digitalisation. Firstly, there are the interests of consumers, who seek to obtain insurance coverage at an affordable price. Secondly, there are the interests of the insurance firm and its shareholders in sustaining a profitable business in competitive markets. Thirdly, and especially in mutual insurance business models, there are the interests of the pool of insured customers of an insurance firm which may differ from the ones of individuals customers (e.g. high-risk consumers will increase the premiums of the pool). Finally, societies have a vested interest in insurance as a means of providing security, health and welfare for their populations. Insurance provides ease of mind to consumers; security against the financial consequences of adverse outcomes of insured activities. This enables insured customers to pursue risky, but beneficial activities that otherwise could not be exercised. Societies not only consider the economic activity of insurance, but more importantly the economic and societal beneficial activities that insurance renders possible. As previously mentioned, this is reflected in the fact that some insurance products are mandatory, and some schemes are redistributive in spreading the costs across the pool regardless of risk profile/ production cost.

SPECIAL CONSIDERATION SHOULD BE GIVEN TO THOSE INSURANCE PRODUCTS THAT ARE PARTICULARLY IMPORTANT FOR FINANCIAL INCLUSION, ALTHOUGH THEIR RELEVANCE MAY DIFFER BETWEEN JURISDICTIONS

Certain insurance products are particularly important for financial and social inclusion. The GDE considers that the following insurance products are particularly relevant:

Figure 7 – Insurance products that are particularly relevant for financial inclusion

Type of product	Importance
Motor insurance	Lack of motor insurance can impact negatively the level of mobility required for employability as well as social minimum standard of living e.g. where public transport is inadequate
Health insurance	Inadequate health insurance can prevent access to adequate, basic health care, which will have a negative impact on individuals and societies
Household insurance	Has a very high protective effect against the loss of property, which can be particularly relevant for indebted families. In addition, in some Member States home insurance is a prerequisite to rent or purchase accommodation.
Third party liability insurance	Similar to home insurance, it has a very high protective effect against the loss of patrimony
Life insurance / pension provision	Provides security against poverty after retirement
Workers compensation insurance	Provides wage replacement and medical benefits to persons which are not able to work due to and injury suffered in the course of employment

Source: EIOPA Consultative Expert Group on Digital Ethics in insurance

Several EU Member States acknowledge the importance several of these products by making them mandatory and / or by organising them as social insurance. In social insurance, individual cover is decoupled from its production cost (namely risk) and the total cost of insuring the whole pool is redistributed on a *per capita* basis or in proportion to income. Social protection or welfare policies are the paradigmatic example of this. It is important to highlight that social insurance systems and mandatory insurance requirements vary across jurisdictions. For example, access to private health insurance or private retirement saving products are more relevant in jurisdictions that do not have robust Social Security systems, community rating (e.g. the Republic of Ireland) or collective agreements between social partners (Denmark), since the latter may provide an adequate financial “safety net”.

AI SYSTEMS HAVE THE POTENTIAL TO ENHANCE FINANCIAL INCLUSION FOR SOME CONSUMERS, BUT OTHER HIGHER-RISK CONSUMERS MAY FACE INCREASING CHALLENGES IN ACCESSING AFFORDABLE INSURANCE IN SOME LINES OF BUSINESS

A study sponsored by the European Commission²⁹ defined financial exclusion in insurance as “a process whereby people encounter difficulties accessing and/or using these products in the mainstream market that fit needs and enable them to lead a normal life in the society to which they belong”. In today’s digital society, the greater availability of data combined with increasingly powerful algorithms enable insurance firms to perform more granular risk assessments.

On the one hand, a better understanding of the risks in combination with risk-mitigation services can improve financial inclusion for some high-risk consumers who previously could not access affordable coverage. Examples include young drivers using telematics devices and patients with diabetes using health wearable devices. On the other hand, some high-risk consumers could encounter difficulties in accessing affordable insurance in markets where there is free competition. For example, DNA data could reveal previously unknown pre-existing medical conditions that could make it difficult for some consumers to access health or life insurance. People living in areas affected by climate change such as those more prone to suffer floods could face difficulties to access flood insurance as a result of increasingly granular risk assessments.

EIOPA’s 2019 thematic review on the use of Big Data Analytics in motor and health insurance found no evidence that an increasing granularity of risk assessments was at that time causing significant exclusion issues for high-risk consumers. However, insurance firms that participated in the thematic review expected the impact to increase in the years to come as a result of the increasing use of more accurate algorithms, new datasets, new rating factors, and a larger number of homogenous / premium risk pools.³⁰ In the household insurance line of business, a recent study from the Australian Competition and Consumer Commission found that more granular pricing approaches, in particular address-based risk assessments, had been a key contributor to a rise in premiums.³¹ Similarly, a study by the consumer organisation Consumentenbond found that

an increasing reliance by Dutch insurance firms on big data analytics when selling insurance contracts resulted in significant premium increases for many home insurance consumers.³²

Insurance firms can contribute to the financial inclusion of high-risk consumers in different ways. For example the “solidarity monitor” developed by the Dutch Insurance Association in 2017 aims to monitor the impact on premiums resulting from the increasing granularity of risk assessments in the Netherlands.³³ A voluntary restriction of insurance firms’ risk selection practices can also help up to a certain amount, given that it affects the insurance firm’s competitiveness *via-a-vis* its competitors and it could be against the interests of its shareholders.

In some countries where it is not specifically forbidden, some private insurance firms have committed to voluntarily restrict the use of DNA data in health and life insurance underwriting.³⁴ It is important that consumers which are not able or not willing to provide such data (or other types of sensitive data such as that derived from wearable devices), should still be able to obtain insurance cover. Insurance firms can also offer consumers access to more affordable insurance products if they use telematics devices, and provide them loss prevention / risk mitigation services to help consumers understand and mitigate their risk exposure (e.g. suggestions to drive safely or to adopt healthier lifestyles).

THERE ARE OTHER SOURCES OF CONSUMER’S VULNERABILITY THAT ALSO DESERVE SPECIAL ATTENTION

In addition to the high-risk consumers who could be impacted negatively by granular risk assessments, other groups of consumers may be especially susceptible to harm. These include groups of people who through, for example, poverty, physical and mental health disabilities, age etc., have reduced opportunities for societal participation. Such vulnerabilities can manifest themselves in different ways over time and may be exacerbated by digitalisation. A recent report from the OECD³⁵ concluded that groups of consumers with less access to, and experience with, the digital environment are increasingly vulnerable. In the insurance sector, a recent study on general insurance pricing practices conducted by the Financial

29 European Commission (2018)

30 The premium pool is the subset of the risk pool (i.e. all insured of an insurance undertaking) that are paying the same premium

31 Australian Competition and Consumer Commission (2018)

32 Consumentenbond (2018)

33 Dutch Association of insurers (2018)

34 UK Government and the Association of British Insurers (2018)

35 OECD (2019)

Conduct Authority in the UK found that 1 in 3 consumers who paid high premiums showed at least one characteristic of vulnerability, such as a low level of financial resilience or capability. The study also found that lower income consumers paid higher premiums for household insurance than those with higher incomes.³⁶

Figure 8 – Examples of types of vulnerable consumers in financial services

Type of vulnerability	Examples of groups with higher proportion of vulnerable consumers
Personal characteristics	<ul style="list-style-type: none"> ▪ Elderly people ▪ Low income or in poverty ▪ Low level of education ▪ Members of minorities ▪ Migrants ▪ Young people / students ▪ People living in areas affected by climate change (e.g. floods)
Life-time events	<ul style="list-style-type: none"> ▪ Unemployed people ▪ Homeless ▪ Divorced / single parents ▪ Over-indebted people ▪ People with records of payment default ▪ Prison inmates ▪ People injured in accidents (e.g. car accident) ▪ Victims of domestic violence
Health conditions	<ul style="list-style-type: none"> ▪ People with disabilities ▪ Hereditary medical conditions (e.g. based on genetics / DNA data) ▪ Pregnant women ▪ People experiencing mental health issues or undergoing therapy
Digital skills	<ul style="list-style-type: none"> ▪ Low level of digital skills ▪ Difficulties to access online / digital services

Source: EIOPA Consultative Expert Group on Digital Ethics in insurance

Many of the different factors and groups that can be considered as being vulnerable overlap, leading to what might be described as “intersectional vulnerability”. This

can mean that they will be faced with numerous, cumulative barriers to financial inclusion. In many cases, deeper investigation might reveal a spectrum of interconnected factors that contribute to a person being in a vulnerable situation. This makes it difficult to get out of this situation, as addressing one factor will not necessarily resolve them all. Indeed, while some of the types of vulnerability are related to behaviour (e.g. over-indebtedness or level of education) that could potentially be modified, others are intrinsic and cannot easily be modified (e.g. members of minorities groups and those with hereditary medical conditions).

Insurance firms can analyse and modify how their underwriting and pricing differentiation practices impact vulnerable consumers, although it is important to note that some characteristics of vulnerability mentioned in the table above are relevant risk factors for underwriting purposes (notably age is very relevant for life insurance and motor insurance). Examples of possible initiatives in this area may include not using credit scores for pricing and underwriting in motor insurance. With the use of telematics devices, insurance firms can also substitute attributes that cannot be influenced by customers (e.g. age) by attributes which can be more easily modified and have an effect on the risk (e.g. driving behaviour, lifestyle etc.). Insurance firms can also avoid using certain types of price and claims optimisation practices or put in place measures to mitigate their negative impact on vulnerable consumers. Moreover, insurance firms, could, with public subsidy where appropriate, design simple insurance products that cover basic protection needs at affordable prices for members of vulnerable / high risk groups.

CONSUMERS AND PUBLIC AUTHORITIES ALSO HAVE A ROLE TO PLAY

It is also important that consumers acknowledge their own responsibilities by behaving in a careful, sustainable, honest and responsible manner, taking preventative measures where necessary and respecting their premium commitments. Customer expectations should also be managed so that the links between price and risk are better understood.

Public authorities also have a role in fostering ethics and financial inclusion in the insurance market by developing adequate legislation (e.g. data protection, conduct of business legislation etc.) and supervising insurance firm’s compliance with existing regulations. In addition to the Social Security systems and welfare policies mentioned, a number of public-private pooling arrangements have

36 FCA (2019)

been used and secured by government back-stops. This is, for instance, the case of FloodRe in the UK, which provides affordable home insurance coverage to consumers living in areas prone to suffer floods.³⁷ In motor insurance, five EU Member States also provide coverage to high-risk drivers (e.g. taxi drivers, drivers with disabilities or drivers with a history of many accidents).³⁸

Moreover, some jurisdictions such as France have laws (loi Levain) restricting the type of data that can be used in health insurance³⁹. Other jurisdictions such as Denmark specifically do not allow the use of DNA data in life and health insurance. In Ireland, the system of community rating is applied in health insurance. Also some countries including France, Belgium, Luxembourg, the Netherlands and Spain have adopted or are in the process of adopting legal measures to limit the use of health data (and more particularly with regards to cancer survivors) when taking up credit insurance.

Public authorities could also potentially consider promoting financial inclusion in different ways such as by giving financial support to low income populations to allow them to get access to essential insurance products. They could also promote, for a set of essential insurance products, a “default option” to guarantee a large access to consumers at risk of exclusion. It has also been suggested that public authorities could define clear “objectives” to be reached by the industry as a whole as regards inclusion, for instance based on an assessment of the outcomes of AI systems on certain groups of vulnerable populations or protected classes.

FAIR USE OF DATA REQUIRES THAT IT IS APPROPRIATE FOR PURPOSE; SUBJECT TO INTELLECTUAL PROPERTY ISSUES, INSURANCE FIRMS SHOULD BE TRANSPARENT ABOUT HOW THEY USE THE DATA AND BE ABLE TO EXPLAIN THESE USES TO CUSTOMERS

Certain datasets can be especially sensitive and show very private behaviours and habits of consumers. This could be for example the case of health data used by health insurance firms, but also new datasets such as data collected by health wearables or car telematics devices,

³⁷ See Flood RE: <https://www.floodre.co.uk/>

³⁸ Spain (Consortio de compensacion de seguros), the Netherlands (De Vereende), Luxembourg (Pool des risques aggravées), Belgium (Bureau de tarification RC auto), Romania (Romanian Motor Insurance firms' Bureau (B.A.A.R.)) and Austria (Scheme for extraordinary risks)

³⁹ Jeanningros, H. & McFall, L. (2020)

or bank account and credit card information showing consumer's shopping habits. Provided that insurance firms have the adequate legal grounds to use such data (e.g. consumer consent), they can be used to provide valuable services to consumers, such as suggestions to improve driving skills or healthy lifestyles, or patient data may be essential to explore diseases and new healing concepts or to give hints for medical treatments to insurance customers. Bank account data or public posts in social media can be used to provide valuable offerings to consumers based on “life time events” e.g. when a consumer purchases a holiday package and he is subsequently offered travel insurance products.

However, they should not be used for purposes other than those for which they were collected (principles of purpose limitation and data minimisation in Article 5, GDPR) and against the interests of the consumer. For example, private habits such as “where do you go shopping” or “if you go to eat to fast food restaurants” or knowledge such as “who are the friends you meet”, “do you suffer from mental illness” may be derived from telematics or mobile phone data (e.g. geolocation data) but should not be used; such information should be actively excluded from data analysis models, for instance, for insurance pricing.

Similarly, social media data (for example blog posts, social media posts or photos etc.) should be avoided where such data are imprecise, manipulatable, and might lead to erroneous or unstable interpretations. Furthermore, life-time events derived from bank account data that could reveal consumer's vulnerabilities (e.g. purchases of medical products or legal expenses arising from a private or commercial dispute) should not be against the detriment of the consumer. The same would apply to other uses of data; leaving aside intellectual property issues, insurance firms should be transparent about how they use the data and be able to explain these uses to customers.

CONSUMERS NOT WILLING TO SHARE VERY PERSONAL AND SENSITIVE DATA ARE NOT STRICTLY NECESSARY FOR RISKS ASSESSMENTS SHOULD STILL HAVE ACCESS TO AFFORDABLE INSURANCE COVERAGE

Some consumers may not be willing to share some personal data which they deem very sensitive and private (e.g. health data collected from wearable devices). Insurance firms should explain to consumers that they may not be able to offer them certain features or products which require such data. However, consumers should never-

theless be offered alternative products providing them with adequate coverage,⁴⁰ as it is indeed commonly the case given that the penetration of usage-based insurance products is still very limited. Another example is the case of DNA data. On the one hand, if not already explicitly prohibited by national laws, consumers should not be requested to provide such data and/or should be able to access affordable cover in the market without providing such information. On the other hand, consumers should not take advantage of genetics analysis to purchase insurance products without transparently informing insurance firms about the fact that they have this information.

More generally, the availability of data from consumers with low digital skills or with difficulties / unwillingness to access digital services might be more scarce; insurance firms should be aware of these limitations when training AI systems and developing product offerings for this target market. Privacy rights should not come with a high price.

INSURANCE FIRMS SHOULD RESPECT THE PRINCIPLE OF HUMAN AUTONOMY BY DEVELOPING AI SYSTEMS THAT SUPPORT CONSUMERS IN THEIR DECISION-MAKING PROCESS AND AVOID UNFAIR MANIPULATION, INCLUDING UNFAIR NUDGING PRACTICES

Linked to the fair use of data described in the previous point and in line with the principle of human autonomy recognised by AI HLEG, AI systems used by insurance firms should support consumers in making better, more informed decisions when purchasing insurance products. They should not unjustifiably subordinate, deceive or manipulate consumers. This is particularly relevant in the area of behavioural economics, where AI systems can sometimes be deployed to shape and influence human behaviour through mechanisms that may be difficult to detect, since they may harness sub-conscious processes.

Consumer's decision making process is often not rational and is influenced by human biases such as favouring the known over the unknown, or processing information in a way that fits pre-conceived ideas, preferring the default option rather than making conscious choices, overestimating the probability of positive events, risk aversion, herding etc. AI systems based on behavioural economics assessments should not exploit consumer's biases but rather help consumers overcome them; influencing con-

sumer's private life / habits in the sense of manipulation should be avoided at all times.

For example, insurance firms can help consumers overcome their biases and make more informed decisions by transparently 'nudging' them towards healthier, safer or financially beneficial choices. Nudging can help at the product level: e.g. health insurance firms can nudge consumers towards healthier habits and improving of health. This may take place at various stages, e.g. through ongoing engagements or at the point of claims. This can be important to achieve higher customer satisfaction (e.g. choosing certain health centres, or repair shops). Nudging where insurers help influence the choices of consumers can be preventative and create value for insurance firms, consumers and society as a whole.

AI "recommender models," "robo-advisors" or "next best action models" can also be designed to nudge consumers towards buying more, or less, appropriate insurance coverage or types of insurance products they may not otherwise have considered (up-selling and cross-selling). In doing so, insurance should take into consideration the demands and needs of the target consumer and ensure that they are in line with the new offer. In such cases, insurance firms should try to follow the "likelihood to need" approach instead of "likelihood to buy".

Similarly, as explained in the use cases below, certain types of price and claims optimisation practices such as those aiming to maximise consumer's "willingness to pay" and consumer's "willingness to accept" should be regarded very critically from a fairness and non-discrimination point of view and therefore insurance firms should avoid such practices when they harm vulnerable consumers or protected classes.

INSURANCE FIRMS SHOULD MAKE REASONABLE EFFORTS TO REMOVE BIASES FROM DATA AND AI SYSTEMS

Plenty of different sources for bias exist. The clearest example is when a dataset contains protected attributes or apparently neutral proxy variables which closely correlate with those protected attributes. In addition, in supervised machine learning algorithms (reportedly 90% of all AI systems used to date) certain variables/datasets are labelled. This involves human judgements which could potentially reflect prejudices of the person that labelled the data (e.g. claims loss adjusters labelling a claim as fraudulent or not). Since those labels serve as ground truth, any contained bias gets reproduced and may be enforced in the

⁴⁰ EDPB Guidelines 01/2020 (2021)

final algorithm. Another, less obvious source of bias is the process of how the data was collected. If the data are not representative and do not reflect the real distribution, an AI system using these datasets for training will perform poorly.

AI systems excel in identifying patterns in data. Their major strength is the desired capability to find and discriminate classes in training data, and to use those findings to make predictions for new, unseen data.⁴¹ Any bias, errors, inaccuracies or mistakes in the data used to train the model, either accidental or intentional, will be reflected in the output of the AI system. The general assumption is that the more data are used, the more accurate becomes an algorithm and its predictions. When using a richer dataset, it clearly contains many correlations. However, not all correlations imply causality, and no matter how large the dataset is, it still only remains a snapshot of reality. It is crucial to notice and mitigate unwanted correlations, because otherwise the resulting trained algorithm may underperform in production when conditions change slightly, and worse, it may discriminate subgroups of the population when the predictions impact consumers.

It is not possible to completely remove biases from the data i.e. certain elements / groups of persons of a dataset will always be more heavily weighted and/or represented than others. For example, there will always be more men or more women driving certain types of vehicles. Nevertheless, insurance firms should make reasonable efforts to remove biases in the data used by AI systems. This includes removing from the datasets protected attributes (i.e. direct discrimination), although it is important to highlight that, as it is explained further below, for insurance underwriting purposes it is lawful to use some protected characteristics insofar as they are necessary

for adequate risk assessment (not for price optimisation practices). Proxies that could be correlated with protected characteristics should also be removed (i.e. indirect discrimination), unless their use is objectively justified by a legitimate aim and it is appropriate and necessary.

Figure 9 – Protected classes in EU Charter of Fundamental Rights and exemptions in national legislation for insurance risk assessments

Protected characteristic in Article 21 EU Charter of Fundamental Rights ⁴²	Allowed for insurance risk-based pricing and underwriting, with restrictions (depends on Member State's national law)
<ul style="list-style-type: none"> ■ Sex ■ Race ■ Colour ■ Ethnic or social origin ■ Genetic features ■ Language ■ Religion or believe ■ Political or any other opinion ■ Membership of a minority group ■ Property ■ Birth ■ Disability ■ Age ■ Sexual orientation ■ Nationality 	<ul style="list-style-type: none"> ■ age ■ disability ■ religion or belief⁴³ ■ sexual orientation⁴⁴

Source: EIOPA Consultative Expert Group on Digital Ethics in insurance

42 The terms of the charter are not directly binding to insurance firms, but they are binding to the institutions and bodies of the EU and national authorities only when they are implementing EU law. For example, the charter applies when EU countries adopt or apply a national law implementing an EU directive (e.g. the Insurance Distribution Directive) or when their authorities apply an EU regulation directly. Therefore the provisions of the charter may also be indirectly applicable to insurance firms, with the exceptions included in national law for risk-based underwriting. Moreover, in cases where the charter does not apply, the protection of fundamental rights is guaranteed under the constitutions of EU countries and international conventions they have ratified.

43 This is not allowed in all EU member states, but in some yes. See for example in Germany the national law on equal treatment (Allgemeines Gleichbehandlungsgesetz (AGG))

44 See previous footnote.

41 Ruf, B., Hirotsu, M., Detyniecki, M., Shire, N., Scharrer, R. (2019)

Moreover, in some traditional (multivariate) models insurance firms may also use protected characteristics (e.g. gender) as “control variables” to remove bias. They aim of including protected characteristics into the model would be to remove correlations amongst the predictive variables of the model in order to isolate each individual predictive variable’s unique contribution to explaining the outcome. This use of protected attribute information would be legal as it is necessary to detect and prevent unlawful discrimination. It may seem paradoxical at first sight: using protected attributes to remove bias in the model, but indeed appropriate information on protected attributes, where available (information about some attributes such as on ethnic origin or religion are not commonly available), can help ensure that outputs of algorithms are free of bias.

While the above-mentioned techniques may prove viable for traditional, deterministic algorithms used in insurance with a manageable quantity of data, it may be insufficient

for AI systems trained on “Big Data”. AI systems such as neural networks or deep learning can capture complex linear and non-linear correlations between the different data variables used, that are not obvious at first glance for the human eye, and which may provide unexpected links to protected attributes. This way, presumably non-protected attributes (or combinations of them) can serve as substitutes or proxies for protected attributes, and this issue is amplified due to the opacity of some AI systems.

Therefore, particularly for high-impact AI applications, insurance firms should assess the appropriateness of their predictor (algorithm and data required for the algorithm), i.e. if the algorithm can be made more parsimonious with regard to data and if the explainability with respect to the dependence on the protected characteristics can be improved. Another alternative is to review the outcomes of AI systems to identify potential biases. The box below shows an emerging research field in the data science community about how this could be done in practice.

OUTLOOK: ACTIVE FAIRNESS AND THE POTENTIAL USE OF FAIRNESS AND NON-DISCRIMINATION METRICS TO ASSESS THE OUTCOMES OF AI SYSTEMS

One of the challenges when trying to enforce fairness or non-discrimination in AI systems is that it is very difficult to understand and to observe the impact of any corrective measure implemented to improve the situation, without having a way to see what happens to the vulnerable population. Against this background, a new research field called “Fair Machine Learning” has emerged to find sustainable solutions for all sorts of fields of application. This area of scientific research, which is still at an early stage of development, questions the effectiveness of the current practice consisting of obtaining fairness and non-discrimination by simply trying to remove biases from the training data. Indeed, this approach may not be sufficient to guarantee a fair result since AI systems, such as neural networks or deep learning, are capable of finding non-linear correlations in the training data and therefore in some sense they are able to reconstruct the hidden (protected) information. To overcome these profound challenges, other approaches are required.

Researchers are therefore exploring solutions, which, as of today, suggest actively using the sensitive data or protected attributes as a control variable in order to make the outcome / impact on protected groups visible in a precise manner and subsequently mitigate any undue discrimination or unfair treatment. However, identifying vulnerable populations is a challenge in itself, since it comes with several associated risks, among them the misuse of this information, especially when the information is collected at individual level. Some sources of vulnerability may be identified with datasets commonly available at insurance firms (e.g. age, gender, income, location or level of studies), but other types of data are currently not collected for good reasons (e.g. data about ethnicity or religion) and its collection would probably require a broader societal debate and/or the involvement of public authorities to establish certain safeguards. One possible way to overcome these challenges is to assess the impact of a data input or rule for use (e.g. automated underwriting processes, price optimisation practices or rating factors with a limited causal link) in terms of non-discrimination by performing statistical analysis using

aggregated ethnic and income data at Zip code level which is publicly available in the Census.⁴⁵ Using such demographic descriptions at zip-code level is a way some insurance firms are already analysing their footprint. This approach, although interesting for observation purposes, should arguably not be used as a rule or objective. In fact, location-based compensation metrics (or using any other proxy) could lead to stronger discrimination, e.g. those vulnerable people living in statistically well-off areas.

As the research field matures, we expect to see solutions which overcome the risks associated with the direct collection of sensitive information, using approaches such as securely leveraging trusted third party for the sensitive data handling or approaches leveraging differential privacy. By exploiting sensitive data, it is possible to precisely quantify and steer fairness. To do so, it is necessary to define or use an existing fairness metric. It is worth mentioning that these metrics translate different nuances of what “fair” means. In the following table lists the most commonly used fairness definitions / metrics are outlined:

Figure 10 – Examples of fairness and non-discrimination metrics

Fairness metric	Description
Demographic Parity	The goal of “Demographic Parity” is to assign the positive outcome at proportionally equal rates to each subgroup of a protected class where the positive outcome refers to the favourable decision. ⁴⁶ For example, in the context of a recruitment scenario “Demographic Parity” could mean that male and female candidates are invited to job interviews at equal rates, proportionately to the number of applications.
Calibration	Another approach aims at equal positive and negative predictive values for all subgroups. ⁴⁷ Such calibration guarantees that the predictive values across subgroups correspond to the scores which represent the probability of predicting the positive or the negative outcome. For example, in a medical diagnosis scenario, a calibrated model could ensure equal levels of confidence in the predictions for patients of different gender or ethical backgrounds because the predictive values are comparable across all subgroups.
Equalized Odds	This fairness definition requires equal true positive and true negative rates for all subgroups. ⁴⁸ For example, where an insurance firms uses AI systems to scan through CVs and job applications in recruitment processes, “Equalized Odds” would ensure that the chances for men and women to be invited to the job interview are equal. ⁴⁹
Equalized Opportunities	This relaxed version of “Equalized Odds” is often used in practice because it reduces the computational complexity when working with large real-world datasets. “Equalized Opportunities” only requires the error rates for the favourable outcome to be the same but allows deviations for the unfavourable outcome. For example, in online marketing when the objective is to inform men and women at equal rates about an insurance offer, “Equalized Opportunities” could ensure that relevant segments of both groups are shown the information at equal rates. The rate of exposure to people for whom the offer is actually irrelevant may differ, however.
Individual fairness	All definitions mentioned above bind on a group level, based on one or several protected attributes. A completely different approach is “Individual Fairness” which abandons the idea of group memberships and suggests instead that any similar individuals should be treated similarly. For example, all the individuals with the same risk profile should pay the same premium for the same insurance product.

Source: EIOPA Consultative Expert Group on Digital Ethics in insurance

⁴⁵ See for example the study conducted by the Insurance Department of the State of Missouri on the impact of insurance-based credit scores on minority groups and low income populations at Zip code level: <https://insurance.mo.gov/reports/credscore.pdf>

⁴⁶ Dwork, C., Hardt, M., Pitassi, T., Reingold, O., Zemel, R. (2011)

⁴⁷ Crowson, C., Atkinson, E., Therneau, T., Lawson, A., Lee, D. and MacNab, Y. (2016)

⁴⁸ Hardt, M., Price, E. and Srebro, N. (2016)

⁴⁹ This fairness metric is already used by some companies such as LinkedIn: <https://engineering.linkedin.com/blog/2021/using-the-linked-in-fairness-toolkit-large-scale-ai>

Each of these metrics has its advantages and disadvantages. No silver bullet solution exists.⁵⁰ Historical attempts to achieve fairness based on purely algorithmic approaches have failed (see for example problems with Kantian ethics or utilitarianism). Therefore it is also true for the implementation of fair AI systems that sustainable solutions will strongly depend on the context of the application and the content of the desired notion of fairness. This one is a matter of context and needs to be chosen on a case-by-case basis.⁵¹ For high-impact AI applications, insurance firms could select the definition / metric of fairness that best suits their concrete AI use case in order to measure its outcomes. The promise of active fairness goes even further, researchers are exploring techniques that once the desired metric for a given application is settled, relevant constraints / guardrails could be introduced into the AI system to prevent the undesired effect on protected groups to happen at all, in some sense fair by design. However, although these investigations are quite interesting and promising, at this stage the solutions are not mature enough to be directly recommended. Moreover, using these results for regulation is premature since not only would it be in contradiction with other rights and regulations, but also because it reveals new questions, such as how to guarantee that the right metrics are chosen. Nevertheless, beyond the promise, we believe that active fairness is worth being considered as of today, since at minima, leads the debate to the heart of AI systems.

⁵⁰ Some European insurers are engaging with the Singapore Monetary Authority in phase 2 of the Veritas project to start systematically building fairness metrics for insurance.

⁵¹ Ruf, B. and Detyniecki, M. (2021)

DEFINING AN APPROACH TO FAIRNESS AND RECORD KEEPING

As explained above, insurance firms can address fairness and non-discrimination issues in insurance in many different ways. Each measure has its pros and cons. Starting by defining their approach to fairness in a responsible manner, which may vary from AI use case to another, insurance firms should assess their activities against this definition and implement in a responsible manner the measures they deem to be appropriate, in compliance with relevant legislation and taking into account the interests of all the stakeholders involved. Insurance firms should record/document all the measures they put in place. Transparently communicating such measures would reinforce consumer trust.

2. FAIRNESS AND NON-DISCRIMINATION IN SPECIFIC AI USE CASES IN INSURANCE

PRICING AND UNDERWRITING

Pricing and underwriting are at the core of insurance business. In the first place, pricing and underwriting take into

account the differentiation of insurance premiums with respect to claims risk (claims expectancy, frequency, and amount). Different consumers with different risk profiles pay different premiums. In addition, the insurance premium also needs to take into account the product design and corresponding benefits, terms and conditions, as well other acquisition costs (e.g. commissions paid to distribution channels and other overheads like taxes, salaries, etc.). Subsequently, both during the on boarding stage and at the renewal stage of the contract, the premium may be optimised by adjusting it to the price offered by market competitors, as well as to other non-risk based personal attributes of the consumer.

Correlation is not causation: actuarial / risk-based pricing in insurance should be based on rating factors with a risk correlation and a causal link in compliance with anti-discrimination legislation

As explained in more detail in Annex 3, insurance firms need to comply with anti-discrimination legislation at EU and national level. The current list of protected characteristics in the EU includes attributes such as nationality, gender, racial or ethnic origin, religion or belief, disability, age or sexual orientation. National anti-discrimination laws may differ from country to country in Europe, but they usually contain some provisions specific to insurance pricing that clarify which attributes are inadmissible for the purposes of risk assessment to determine price and

benefits (typically race, ethnic origin and sex) and which attributes are allowed if they are based on recognized principles of risk-adequate calculation (typically disability and age)⁵². Some Member States also allow the use of data on religion and sexual identity while other Member States have stricter national anti-discrimination laws. Pension risk or mortality risk cannot be assessed without taking into account the age of the consumer, and disability also materially influences morbidity risks. Religion or belief information may also be relevant for dowry insurance.

While direct discrimination can be avoided by not using protected attributes, indirect discrimination is more complex. Following the Test-Achats ruling barring the use of gender as a risk factor in calculating individual premiums or benefits, the European Commission issued guidance⁵³ clarifying that rating factors that correlate with gender, and thus can cause indirect discrimination, can be “objectively justified by a legitimate aim” when they are used for the estimation of risks / production cost. A similar approach should be used for other protected characteristics. The requirement that “the means are appropriate and necessary” is translated by the European Commission into the notion of “true risk factors in their own right”.⁵⁴

Figure 11 – Guidance on the necessary and appropriateness assessment of rating factors

Necessary and appropriateness assessment of rating factors and rating categories	
Necessary: Risk / claim correlation	<ul style="list-style-type: none"> Each rating factor used for risk differentiation should have a clear correlation with claims occurrence (i.e. risk).
Appropriateness: Causal Link	<ul style="list-style-type: none"> Each rating factor and subsequent rating categories (e.g. for the rating factor “job”, the rating categories could be “blue collar” or “white collar”, or more granular rating categories like “teachers, engineers, doctors, nurses etc.”) should have a causal link⁵⁵ between the rating factor or rating category and claims occurrence / risk. Each rating factor and rating category should have a valid explanation or rationale for different treatment of otherwise similarly situated consumers. Each rating factor and rating category should be in line with generally accepted actuarial principles^{56 57} AI systems used to predict risks based on a single or limited number of unconventional rating factors also raise significant concerns from a fairness and non-discrimination perspective.

Source: EIOPA Consultative Expert Group on Digital Ethics in insurance⁵⁸

52 Some Member States have more strengthened non-discrimination provisions, and at the same time other Member States allow price differentiation in insurance based religion and sexual identity

53 European Commission (2012/C 11/01) (2012)

54 The Commission explains this situation with the following examples: price differentiation based on the size of a car engine in the field of motor insurance should remain possible, even if statistically men drive cars with more powerful engines. On the contrary, it is not possible to price differentiation based on the size or weight of a person in relation to motor insurance (men are commonly taller and heavier than women).

55 There are different views about the definition of causal link, being addressed here by mentioning that each rating factor and rating category should have a valid explanation or rationale for different treatment of otherwise similarly situated consumers.

56 Actuarial Association of Europe (2021)

57 International Actuarial Association (2019)

58 Table is inspired on the New York Department of Financial Services (2019) Insurance Circular Letter No. 1

Insurance firms should assess and develop measures to mitigate the impact of rating factors such as credit scores, location, income, occupation or level of education on vulnerable populations and protected classes in those essential lines of business where they have a limited causal link

Certain rating factors, such as credit scores that may have a correlation with claims risk, (e.g. people with low credit scores on average make more claims than people with high credit scores) may not be appropriate in some lines due to their limited causal link (e.g. limited valid explanation or rationale between low credit score and suffering a flood). Furthermore, their use may have a disproportionate negative impact on vulnerable consumers, such as low-income populations or certain minority groups, and thus contribute to reinforcing existing inequality (this phenomenon is on occasion referred to as a “poverty premium”).⁵⁹ For example, when migrant populations arrive to a new country they reportedly have a low credit score because there is little credit-related information about them. Moreover, in times of economic crisis such as in the current Covid-19 pandemic, many people that have lost their jobs will see a worsening of their credit score. Indeed, statistically the variance in this group is very high, and the reason why they have a low credit score may differ widely and affect many honest and hardworking people..

It is important to note that there are different types of credit scores / credit reports; some use AI system to process personal information such as spending habits, shopping behaviour, mobile phone usage etc.; such behavioural information can be inaccurate and closely correlated with protected characteristics, and in addition consumers may not be aware that such information about them is used to calculate their credit score. Therefore, such information should not be used for pricing and underwriting purposes. Other credit scores rely on hard facts such as insolvency of customers, which could be related to a higher probability of non-payment, or attempting to commit fraud because of the need of money, or less careful life-style, etc. Some members of the GDE consider that this last type of third-parties credit scores may be used for pricing in certain lines of business, paying attention not to introduce disproportionately high prices to vulnerable consumers. Other GDE members consider that, if used at all, they should exclusively be used to prevent the risk of non-payment (i.e. not to increase the premium or to deny access), by developing adjusted payment processes (e.g. requesting the payment of the premium in advance).

As far as location data is concerned, for certain lines of business such as flood insurance there is a strong risk correlation and causal link; there is a valid explanation for a different treatment of otherwise similarly situated consumers (e.g. house next to a river or the sea as opposed to a house in a dry area). Some household insurance firms have started to make use of Big Data by substituting traditional Zip code data for more granular micro-zoning, such as geolocation and address-level data, in order to develop more accurate and granular risk assessments. The causal link of location data in other lines of business, such as motor insurance, may also be sound, although the causal link is more debatable when location data is used at a very granular level; claims frequency and severity in motor insurance are influenced by the traffic situation, road conditions, and repair costs at a certain location. Population density, criminality rates such as fraud, theft or vandalism (not relevant for basic motor third party liability insurance), presence of public transport, differences between big cities and rural villages, presence of mountains, rivers, bridges etc. are also relevant. However, traditional Zip code data used (or more granular address / area of residence / geolocation data used by some insurance firms) corresponds to where the consumers lives, which is not necessarily the same location as where the consumer drives, where there are more accidents or more traffic or roads in bad condition or where the repair costs are more expensive (i.e. it is a proxy).

For the sake of completeness, it is important to underline that some neighbourhoods (in cities like Copenhagen, Rome, Madrid or Frankfurt there are respectively 200, 78, 55 and 44 different Zip codes)⁶⁰ are predominantly inhabited by groups of protected classes or vulnerable populations, and therefore insurance firms should have measures in place to avoid significant price differences between adjacent micro-zones in order not to penalise inappropriately vulnerable groups and protected classes. Some insurance firms acknowledge these issues and do not use such rating factors (see the findings of EIOPA’s Big Data Analytics thematic review) or include some level of mutualisation in the premiums when they do use them. For instance, they do not charge the 20-40% premium increase that could correspond to the risk of claims for consumers with low credit scores or living in certain locations, also noting that the volume of the pool of consumers is also relevant from an underwriting perspective. It is also important to highlight that not using or limiting the use of rating factors that could negatively impact vulnerable consumers implies as a trade-off that the increased

59 Personal Finance Research Centre (2020)

60 Source: Google

risk would need to be mutualised by the pool of insured customers.

Finally, in some jurisdictions such as in the USA, several states have introduced legislation to forbid the use of rating factors in motor insurance that they consider not directly related to driving including credit scores, Zip code, occupation, marriage status or level of studies.⁶¹ Also in the USA the National Association of Insurance Commissioners (NAIC) has recently launched a special committee on “race and insurance”.⁶² In Belgium, insurance undertakings are required to publish on their website the criteria that they use to assess if they will offer the insurance to the consumer and the criteria they use to set the cost of the insurance policy for six different types of policies: motor, fire, life, health, legal assistance and civil liability insurance policies.⁶³

Rating factors that can be influenced by the consumer may be preferable in certain circumstances and AI systems can be used to find rating factors that better reflect risk

As mentioned in the previous point, rating factors such as age or zip-code are often closely related to claims and therefore have traditionally been used for pricing and underwriting purposes. However, they are a good example of “correlation” with limited “causality”: They do not accurately distinguish the risk of the individuals in that concrete risk segment. AI may identify segmentations which are more closely related to risk and that can be

influenced by the customer. For example, age is used in pricing for motor and health insurance. This proxy does not fully distinguish between cautious/ incautious drivers, or between those with healthy/unhealthy lifestyles. Should better risk factors be made available (e.g. speeding or calories consumption measured by telematics devices), this age categorisation could be reduced in the future.

Price optimisation practices such as those aiming to maximise consumer’s “willingness to pay” should be avoided when they unfairly harm consumers, in particular vulnerable consumers or protected classes and in lines of business that are essential for financial inclusion

On top of the “risk-based” actuarial tariff, and leaving aside premium adjustments to take into account re-insurance costs and other acquisition/production costs (e.g. overheads such as commissions paid to distribution channels, salaries of staff, technology costs etc.), some insurance firms adjust the premium to the market price and further optimise the final premium using a number of different techniques which are largely independent of the risk profile of the consumer. They are typically based on correlations (not causation) of risk relevant attributes, as well as other non-risk based factors such as income, level of studies, type of device used (brand of smartphone, tablet, desktop computer, ...), distribution channel, time of the day, location, apps downloaded etc. The main types of price optimisation practices / AI use cases are described in the table below.

61 https://www.treasury.gov/initiatives/fio/reports-and-notice/Document/2017_FIO_Annual_Report.pdf

62 https://content.naic.org/cmte_ex_race_and_insurance.htm

63 See Article 45 of the Belgian Law (Loi relative aux assurances, 4 April 2014) http://www.ejustice.just.fgov.be/cgi_loi/change_lg.pl?language=nl&la=F&table_name=loi&cn=2014040423

Figure 12 – Types of price optimisation practices

Type of price optimisation practice	Description
Market research & market competition analysis	Using market research for a line of business as well as for customer segments in order to determine the market position vis-a-vis competitors and then adjust the tariff to reflect market competition and to focus on the company's target groups (e.g. via targeted marketing campaigns and commercial discounts). Techniques used may for example be market research studies, also including web crawling techniques and mathematical methods to generate a market overview.
Churn models	Using mathematical methods including AI systems to identify customers under risk of churn. Consumers with a high propensity to shop around at the renewal stage can be given a discount from a commercial and actuarial point of view.
Customer life time value estimation models	Using mathematical models including AI systems to estimate claims expenses and premium income for customers over their whole customer relationship with the company, including possible up-selling and cross-selling of other products. This can then be used to decide on corresponding commercial discounts for certain customers or customer groups.
Price elasticity models	Using mathematical methods including AI systems to determine the customers' willingness to pay; the estimate of the customer's price elasticity then is used to adapt his price in the sense of the "highest possible premium", and this may be done both at the on-boarding stage as well as at the contract renewal stage
Individual real time price comparison	Using web crawling techniques in order to determine in real time the price the customer would have to pay at another insurance undertaking, or ask the customer his alternative price offers from other companies, in order to adapt the price in the sense of the "highest possible premium" or to select only those customers where a minimal discount is necessary.

Source: EIOPA Consultative Expert Group on Digital Ethics in insurance

Price optimisation practices are also used in other industries such as airlines, car rental, hospitality/lodging or e-commerce. They are also not necessarily new in the insurance sector, where for example agents and brokers have traditionally been allowed by insurers certain flexibility to offer commercial discounts to attract and retain consumers. Price optimisation practices may negatively impact vulnerable consumers, for instance when the consumers' vulnerability (e.g. elderly people or low financial capability) causes them to suffer from these practices. For example, loyal customers with a high price sensitivity or low propensity to shop around may not be aware that they may be paying much higher premiums than consumers with a similar risk profile but with higher price sensitivity or higher propensity to shop around (often referred to as "loyalty premium"). Moreover, behavioural data used in price optimisation practices can be correlated with protected characteristics and therefore increases the risks of indirect discrimination, particularly when processed with complex AI systems. Furthermore, some of these practices may be dangerous from a financial stability point of view, since the premium then is not risk adequate any more. Finally, it is debatable whether some of these practices can be in line with the ethical self-understanding and idea of insurance: people buy insurance to protect themselves against risks (often on a mandatory basis).

Therefore, certainly different standards apply compared to buying clothes, a plane ticket or a luxury car.

Price optimisation practices have drawn significant scrutiny from regulators, consumer organisations and industry, given the potential unfair treatment of certain groups of consumers. In the United States, the National Association of Insurance Commissioners (NAIC) published a White Paper⁶⁴ on price optimisation practices by insurance firms and a number of states subsequently prohibited or restricted the use of price optimisation techniques based on consumers' willingness to pay and/or their propensity to shop around at the point of renewal. In Ireland, the supervisor has opened an investigation analysing such practices.⁶⁵ In the United Kingdom, the UK's Financial Conduct Authority carried out a market investigation⁶⁶ into the pricing practices of insurance firms, after it found evidence that firms often deliberately targeted price increases on consumers who were perceived as less likely to switch. Following its investigation, the FCA proposed new rules to prevent firms from gradually increasing the renewal price to consumers over time, by requiring firms to offer a renewal price to consumers that

64 NAIC (2015)

65 Central Bank of Ireland (2020)

66 FCA (2020)

is no higher than the equivalent new business price for customers ordering through the same sales channel.

Some members of the GDE do not agree with the above-mentioned regulatory interventions and consider that based on the freedom of insurance firms to conduct a business they should be able to give commercial, marketing or underwriting discounts to consumers in order to try to acquire or retain them in the course of a commercial transaction. These types of discounts are influenced by many factors such as the insurance firm's strategy, willingness of agents/brokers to maximise the portfolio retention rate by applying extra-discounts to customers with a high risk of churn, willingness of the companies to ensure a long-term relationship with their customers and, in general to prevent that the "low risk" clients would move to a competitor. Indeed, the size of a portfolio is fundamental for insurers to work efficiently and the acquisition of new customers is expensive, therefore high customer retention is economically sensible. Similarly, the use of business intelligence techniques to set premiums in relation to one of its competitors is also a basic characteristic of competitive markets. Finally, some insurance firms also offer new customers a reduced premium at the on-boarding stage and then adjust it in subsequent years after they have a better understanding of their risk profile.

In view of the above, the GDE considers that a possible balanced solution in this area could consist on the following: for essential insurance lines of business, price elasticity models and individual real-time price comparison techniques used to maximise the price should be viewed critically from an ethical, fairness and also competition point of view and therefore avoided, both during the on-boarding of the consumer and at the renewal stage. On the other hand, the GDE considers that churn models, customer life time value estimation models, and targeted marketing campaigns and market competition analysis should still be possible, but insurance firms should make reasonable efforts to ensure that they do not disproportionately disadvantage vulnerable consumers and protected classes. Moreover, churn models should also not be used to increase the premiums of consumers less likely to shop around. The premium paid by consumers at the renewal stage should only be increased on the grounds of increased risks or increased costs (e.g. changes in the non-accident ratio, increasing healthcare costs, original premium include a commercial discount etc.), i.e. premium increases unrelated to increasing risks or increasing costs should be avoided, but premiums discounts for commercial and marketing purposes can take place with the safeguards mentioned above.

CLAIMS MANAGEMENT

Claims handling accounts for the insurance firm's highest cost whilst also presenting the greatest opportunity for satisfying customers and securing their loyalty. It is a moment of truth for customers which delivers on the promises made in underwriting and pricing. The management of the claims process has historically been an intensively manual activity. A growing awareness of how data led innovation can improve customer experience and the economics of claims has led to an increase in the use of Data Science in claims. Many organisations already have AI solutions to augment decision making in existing claims process flows; for instance in validation routines of a claim, NLP models can be used to interpret event descriptions to auto approve entitlement to cover. Valuation systems will also increasingly make use of AI Image Analytics to assess vehicle damage and connecting these two parts ordering and auto estimation systems to replace the need for an engineer's inspection. AI systems can also be used in the process of selecting (routing) the most appropriate specialist for each claim.

Insurance firms who are using AI systems continue to evolve their response to the practical, ethical and legislative issues around implementing and governing the use and deployment of increasingly sophisticated decision-making models. However not all decisions carry the same weight or impact and so a consideration of impact and context is critical. In the case of claims, many of the decisions will be executed by claims operators, who provide a backstop for each decision. Two use cases are analysed below from a fairness and non-discrimination perspective:

- **Compensation & Cash Settlement: claims optimisation practices aiming to estimate a consumer's "willingness to accept" a cash settlement offer should be avoided when they unfairly harm consumers, in particular vulnerable consumers or protected classes and in essential lines of business**

Contracts where a negotiated compensation is required to settle the claim, typically injury claims, are increasingly using AI systems to evaluate the compensation offer. AI systems can help speed up the compensation process for the benefit of the consumer. They will use a selection of variables to arrive at an agreed compensation value. AI systems would be expected to observe protected characteristics in the base calculation as well as seeking methods to remove bias in training data. Thorough testing of the model against a hold out, combined with ongoing

performance monitoring and human QA will help ensure fairness and transparency over the longer term. It is also worth noting that the backstop for compensation issues will typically be human intervention, and consumers will also typically have the opportunity to seek redress by following the insurance firm's complaints procedure, escalate to the regulator or seek legal help. The need for escalation being an important metric / KPI in considering the effectiveness of the models being deployed.

Figure 13 – Example of temporary injuries compensation values

INDEMNIZACIONES POR LESIONES TEMPORALES	
Tabla 3	
Tabla 3.A Perjuicio Personal Básico	
Indemnización por día	31,61 €
Tabla 3.B Perjuicio Personal Particular	
Por pérdida temporal de calidad de vida	
Indemnización por día (Incluye la indemnización por perjuicio básico)	
Muy Grave	105,35 €
Grave	79,02 €
Moderado	54,78 €
Por cada intervención quirúrgica	De 421,41 € hasta 1.685,67 €
Tabla 3.C Perjuicio Patrimonial	
Gastos de asistencia sanitaria	su importe
Gastos diversos resarcibles	su importe
Lucro cesante	su importe

Source: Dirección General de Seguros y Fondos de Pensiones⁶⁷

Case law or circulars from national supervisory authorities provide guidance of compensation values, but typically these will represent the edge case scenarios and there is often a range of settlement values for an injury rather than a single agreed and accepted value (see Figure 4 above). Fairness to non-claiming customers requires that costs are controlled and there is a duty for insurance firms to determine the level of the injury based on objective features of the incident and injury (e.g. via medical inspections) and consumer characteristics (e.g. based on salary and income to determine loss of profit). However, similar to some price optimisation practices, offering someone with the same characteristics, the same injury at the same time with the same product a materially different compensation is not appropriate; in order to determine the compensation offer, AI systems using behavioural data to estimate the consumer's price elasticity, probability to churn, live time value or more generally their "willingness to accept" a concrete offer should not be allowed, on the grounds of fair treatment and non-discrimination.

- **Fraud detection: Effective human oversight and redress systems are key to ensure fair and non-discriminatory outcomes of fraud detection practices, in particular when using complex unsupervised AI deep learning clustering techniques or network analytics**

Fraud detection in claims handling is dissimilar to fraud propensity detection in underwriting because in the latter the potentially fraudulent action happened in the past. In the claims management stage, is it about obtaining an indication of how much the insurance firm should spend to investigate if fraud was committed. Insurance firms typically use a wide array of tools for fraud detection purposes, including scenario based detection rules, a suite of simple business rules and decision trees created to identify specific events that are a high risk of fraud, e.g. a claim made within 2 weeks of policy inception. Insurance firms also may use more complex AI systems, including both supervised ML algorithms as well as unsupervised modelling such as deep learning clustering techniques or anomaly detection to identify potentially high-risk events.

From a fairness and non-discrimination perspective, being subject to an unjustified fraud investigation is cumbersome, time consuming and emotionally stressful, especially if the consumer is in a vulnerable situation (e.g. has recently suffered an accident or an important loss). On the basis of their legitimate interest to fight against fraud (Article 6 GDPR), insurance firms can use criminal records or similar databases without the consent of the consumer. Similar to other AI use cases, insurance firms should make reasonable efforts to remove potential biases in the training data. However this is not necessarily an easy task given the wide variety of datasets analysed for fraud detection purposes. For example, in some jurisdictions standard AI predictions based on criminal records have been found to be biased against certain minority groups.⁶⁸ In addition, if not handled in a responsible manner, certain fraud prevention techniques such as network analytics techniques could perpetuate existing patterns of vulnerability / discrimination through "network discrimination," whereby individuals are penalized (or rewarded) based on the characteristics of their family, friends or other members of their personal network.

Insurance firms therefore need to put in place relevant governance measures to mitigate such risks, such as establishing adequate levels of human oversight. A model of any kind should not be able to classify any customer or transaction as "fraudulent" without the involvement

⁶⁷ Dirección General de Seguros y Fondos de Pensiones (2020)

⁶⁸ ProPublica (2016)

of relevant staff, both at the pre-contractual stage (i.e. to reject access to a product) as well as during a claim. Fraud is considered a crime in most jurisdictions and therefore can only be applied following a formal investigation and legal process, i.e. fraud needs to be proved by insurance firms and always requires a certain level of human oversight and due legal redress mechanisms. Moreover, the accuracy of any anti-fraud modelling is of paramount importance both because a false-positive decision could impact a customer or claimant, and waste the time of the investigation team who have to prove or disprove the case, but also a false-negative result can cost the company a large amount of money and, if the fraud is egregious enough, could leave the company open to regulatory or legal censure due to failures in the control environment.

In order to measure the effectiveness of a fraud model the more effective metric is to use an acceptance that the

flagged policies or claims are worthy of further investigation by the appropriate teams. By optimising the flow of cases to the investigation teams, a good fraud model can both ensure that all customers and claimants are treated equally and also maximise the value of the investigation teams whose expertise is key to the prosecution of any illegal activity. It is also important to be aware that basing the accuracy of any model on the opinions of any individuals or teams within a business may introduce unintended biases in the results, for this reason it is also vitally important that all fraud detection models are tested for overall fairness independently of their accuracy at detecting malfeasance. Finally, to the accurate classification of a potentially fraudulent event a fraud detection model is also required to be explainable in that it should give sufficient information to an investigator to understand why a particular policy or claim has been identified for further investigation; model explainability would also help identify potential fairness and non-discriminatory issues.

VII. TRANSPARENCY AND EXPLAINABILITY

Principle of transparency and explainability:

Insurance firms should adapt the types of explanations to specific AI use cases and to the recipient stakeholders. Insurance firms should strive to use explainable AI models, in particular in high-impact AI use cases, although, in certain cases, they may combine model explainability with other governance measures insofar as they ensure the accountability of firms, including enabling access to adequate redress mechanisms. Explanations should be meaningful and easy to understand in order to help stakeholders make informed decisions. Insurance firms should transparently communicate the data used in AI models to consumers and ensure that they are aware that they are interacting with an AI system, and its limitations.

In accordance with Article 20 (i) IDD, insurance distributors shall provide the customer with objective information about the insurance product in a comprehensible form to allow the customer to make an informed decision. Furthermore, the principle of transparency is also recognised in Article 5 GDPR, which is further developed in Articles 13 and 14 GDPR, requiring firms to timely, appropriately and transparently inform consumers about how their personal data is processed. The GDPR also explicitly states that consumers should also be informed about the existence of automated decision making processes (often underpinned by AI technology), and, importantly, provide them with “meaningful” information about the logic involved, as well as the significance and the envisaged consequences of such processing.

Explainability and transparency issues are not new for the insurance industry or, more broadly, for financial services. The provision of transparent information to consumers is one of the cornerstones of financial services legislation, which can be illustrated with the requirements to provide to consumers simple and user-friendly key information documents⁶⁹ in order to enable them to

make informed decisions. However, transparency and explainability measures have limitations and need to be combined with complementary governance measures to ensure good outcomes for consumers. Conceptually, both explainability and transparency are related to the fairness and non-discrimination since the former are in many respects necessary prerequisites to reveal problems in the latter. As recognised by the AI HLEG, explainability and transparency are important aspects to ensure trust and accountability of AI systems, which goes to the heart of good governance in financial services. Explainability is therefore seen as a key building block in the construction of the ethical and trustworthy use of AI.

The issue of transparency and explainability in the context of AI and the digital economy is becoming increasingly important. In addition to the increasing use of new data sources (internal and external), types of data (e.g. IoT, social media, mobile phone data etc.) and data enrichment techniques, certain AI systems such as Neural Networks (NN) or Deep Learning (DL) can provide very accurate predictions but they can also be considered as a “black-box”⁷⁰ because the rationale of the outcome/prediction of the system is difficult to explain in a causal or deterministic manner. Indeed some/many AI systems autonomously learn from the training data which variables or combinations of variables are most useful in making predictions; the model is free to discover patterns between those variables that best model the data. As a result, it may be difficult to interpret and explain the causation of a decision and/or role/weight that a specific variable plays in the final decision if the outcome depends on complex linear and non-linear interrelations of all variables. This is currently an area of extensive research and there are already a number of supplementary explainability tools (e.g. SHAP, LIME, etc.) which help to understand how the AI systems function, albeit they still have some significant

⁶⁹ For instance the requirement to provide the Key Investor Information Document (KIID) under PRIIPS, or the Product Information Document in the IDD.

⁷⁰ Already some complex models using more traditional techniques like GLM (Generalised Linear Model) and GBM (Gradient Boosting Machine) can also be black boxes with their transparency and explainability challenges.

limitations which may lead to a false impression of understanding of the AI system.⁷¹

The use of AI systems presents both challenges and opportunities when we talk about transparency and explainability in insurance. For example, if “black-box” AI systems are used to model the pricing and underwriting of insurance contracts, it could be difficult for consumers to understand *why* she or he was denied cover or what they need to do to improve their premium, and therefore they would not be able to make informed decisions. They could also not be able to challenge wrong decisions and effectively exercise their right to redress. Moreover, the lack of explainability could also compromise the auditability of the system and increase the model risk, i.e. a lack of understanding of the circumstances under which the model may provide incorrect predictions. In insurance, this can have both prudential (e.g. model fails to price risks accurately) and conduct implications (e.g. potential biases could remain undetected leading to a higher risk of discrimination against certain groups).

From a different perspective, and possibly mitigating to a certain extent the model risk of the AI system, it can also be argued that the automation of certain decisions and processes powered by AI systems increase their transparency and explainability insofar that the automation of decisions previously made by humans is now more transparent because, provided that there are adequate records of data and methodologies used, they leave a digital trail and are therefore potentially replicable and easier to monitor. For example in the area of claims management, the decision to accept or reject an insurance claim has traditionally been done by claims handlers of the insurance undertaking and potential prejudices that these individuals could have could be left undetected. This would be more difficult if the decision is automated by means of an (explainable) AI system with predefined and duly traceable / documented parameters and inputs (see chapter IX). In these cases the AI HLEG highlights the importance of ensuring that consumers are informed of

the system’s capabilities and/or that they are interacting with a machine and not a human.

1. TRANSPARENCY AND EXPLAINABILITY IN THE INSURANCE SECTOR

The terms transparency and explainability are two inter-related terms addressing the type of information about an AI system that needs to be provided to the different stakeholders. In this report, transparency is broadly understood as providing information about the use, the nature or design of an AI system and the data variables and parameters used. Explainability is part of the concept of transparency and concerns the ability to explain the output of the AI system to a particular audience, in particular the weight / influence and causal relationship of a specific variable (or group of variables) in the final output. Both terms are closely related to the notion of traceability / record keeping, which is further developed in chapter IX.

INSURANCE FIRMS SHOULD ADAPT THEIR EXPLANATIONS TO SPECIFIC AI USE CASES

In line with the principle of proportionality and the AI use case impact assessment proposed in chapter V, the level and details of explainability should be adapted to the impact of a specific AI use case on consumers and/or the insurance firm. All things being equal, the higher the impact of an AI use case the greater transparency and explainability measures the insurance firm should adopt. However, in certain cases insurance firms may complement the limited explainability of some AI systems with other governance measures (see further below).

In addition to the impact of the specific AI use case, the nature / content / details of the explanations may vary from one use case to another. While in pricing underwriting explanations for consumers should be more detailed (see following point), in other AI use cases such as in fraud detection, this may not always be possible. In such cases, although some stakeholders such as supervisory authorities can have full access to all the necessary information, it may not be possible to provide totally detailed explanations to consumers to avoid compromising insurance firm’s legitimate interest to fight against fraud (including organised crime) and protect the pool. Indeed, insurance firms using AI for fraud detection purposes should be able

⁷¹ LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations) are two explainability techniques that aim to provide local explanations, i.e. an explanation about the behaviour of specific data points or regions in the input data (i.e. how they influence the output of the AI system). While they both have the advantage that they can capture multi-factor interactions in the dataset, they also have certain limitations. For example, SHAP’s method of sampling values assumes feature independence, which might not necessarily be the case and therefore affects the reliability of the explanation. Moreover, in LIME users can (subjectively) choose how to define the *proximity measure* for the “local” region of the model where the explanation applies; small changes in the scale of the *proximity measure* can lead to significantly different explanations.

to decide not to share certain information about the AI system it uses with certain stakeholders, in light of concerns over manipulation or exploitation.

INSURANCE FIRMS SHOULD ADAPT THEIR EXPLANATIONS TO THE DIFFERENT TYPES OF STAKEHOLDERS

Consumers have fairly different expectations compared to other stakeholders such as supervisors or auditors. Additionally, the Board needs a sufficient understanding of the AI system used as they can create significant public relations scandals and/or threaten the solvency of the undertaking. In addition, as previously mentioned it is clear that there are differences in what and how information can be disclosed to different stakeholders. For instance, supervisors should be entitled to receive comprehensive information from insurance firms, even when this consists of intellectual property rights, in order to be able, for instance, to audit the AI systems used by insurance firms. Nevertheless insurance firms' intellectual property should be respected and they cannot be put into a disadvantaged position vis-a-vis their competitors.

Certain stakeholders such as supervisors or auditors will require global/comprehensive explanations about the inner workings and logic of the AI system's behaviour and its component parts, i.e. its features, parameters and

interactions. This will notably allow them the possibility to better understand the circumstances under which a model can make incorrect predictions (i.e. model risk). For consumers (including insureds, claimants and prospective customers), notwithstanding the relevance of detailed information included in the terms and conditions of the insurance contracts, in certain contexts explanations may function best when they are counterfactual, i.e. how the AI system's outcome depends on a specific variable.

For example, in pricing and underwriting consumers should be clearly explained why a particular risk is not deemed acceptable, or what are the main rating factors that influence his premium. This transparency measure will contribute to reinforce consumer trust, and also allow consumers to adopt informed decisions and know what aspects of their behaviour they need to improve in order to obtain a better premium (noting that in some cases it will not be possible to change it, e.g. health pre-existing conditions).⁷² Under the same light, insurance intermediaries should be able to receive more detailed information from insurance firms in order to properly explain products to customers. Moreover, it can be argued that counterfactual explanations could potentially encourage some dishonest consumers to cheat in order to obtain better premiums, so insurance firms should carefully verify that the information provided by the consumer is accurate.

⁷² This is for instance the approach followed in the USA's Fair Credit Reporting Act; upon the request of a consumer for a credit score, the credit scoring agency should disclose to the consumer all of the key factors that adversely affected the credit score of the consumer in the model used, the total number of which shall not exceed 4. In Belgium, under Belgian law, consumers must have full transparency about which criteria are taken into account by insurance undertakings when taking out certain types of essential insurance policies. The law applies for six different types of insurance policies: motor, fire, life, health, legal assistance and civil liability insurance policies.

Figure 14 – Transparency and explainability information to be provided to different stakeholders when using AI in pricing and underwriting (The criteria with an asterisk are further developed in Chapter IX)

AI use case	Information to be provided	Types of stakeholders		
		Consumer	Auditor and supervisor	Board
Pricing and Underwriting	Is automated decision making or AI used?	x	x	
	What datasets are used	x	x	
	Why certain criteria are chosen for underwriting and pricing i.e. causal link	x	x	
	Counterfactual explanation - most influential rating factors	x	x	
	Reasons for using AI and consistency with corporate strategies / objectives*		x	x
	Description of how the model is integrated in the current IT system*		x	
	Staff involved in the design and implementation and core function groups*		x	x
	Data collection, preparation and post-processing methodologies*		x	
	Technical choices / arbitration and limitations / risks of the AI model chosen*		x	
	Code and data used to train and test the model*		x	
	Model performance, including KPIs*		x	x
	Model security measures*		x	
	Ethics and trustworthiness assessment*		x	x
	Documentation on compliance with regulation		x	x
	Certification by an independent body, disclosure of audit		x	x
System logic explained to a non-expert		x	x	
Implemented third-party technologies and risks		x	x	

Source: EIOPA consultative stakeholder group on digital ethics in insurance73

INSURANCE FIRMS SHOULD STRIVE TO USE EXPLAINABLE AI MODELS, IN PARTICULAR IN HIGH-IMPACT AI USE CASES

Insurance firms should endeavour to use as much as possible explainable AI models. This is particularly relevant when an AI application / use case has a significant impact on consumers and/or on the insurance firm; in such cases insurance firms should be able to sufficiently explain the AI model’s decision-making process, even if this could be at the expense of model performance.

Explainability is affected by the model choice and there is often a trade-off between model explainability and accuracy (e.g. decision trees are quite explainable but perhaps not as accurate as neural networks, which are much more opaque). Taking into account the work developed by the Information Commission’s Officer and the Alan Turing Institute,74 there are three main approaches that insurance firms may use for the creation of transparent / explainable AI models as of today:

- Use of explainable AI systems: Some AI systems such as decision trees or linear regression are generally highly explainable and therefore classical scientific

controls could be used to explain their functioning and rationale. Such AI systems are not always ideal for the learning task (e.g. image or speech recognition), yet if explainability is more important than the accuracy of the result, these methods should be prioritised.

- Use of black-box systems, combined with supplementary explainability tools: In case insurance firms consider that for a certain specific task the use of black-box systems is a suitable option, it is a good practice to support their use with supplementary explainability tools such as LIME or SHAP75 in order to meet an adequate level of understanding about the underlying rationale of the system. The limitations of the supplementary explainability tools used should be duly documented to the extent possible and accompanied with comprehensive model testing, noting that on occasions these tools can provide a wrong sense of certainty.
- Hybrid methods – use of “challenger models”: Explainable AI systems can be combined with challenger black-box systems which are trained on the same data for the purpose of future selection / engineering, comparison and insight. Challenger models can increase the explainability of the production model, but if their insights (e.g. engineered features)

73 This table is inspired on the work of Van den Berg, M. and Kuiper, O. (2020)

74 Information Commission’s Officer and the Alan Turing Institute (2020)

75 See footnote 71

are incorporated into the production models, their rationale should be appropriately documented and explained.

Regardless which of the explainability approaches is followed, insurance firms should bear in mind that when unjust adverse impact occur, accessible accountability mechanisms should be foreseen that ensure adequate redress.

INSURANCE FIRMS MAY COMBINE MODEL EXPLAINABILITY WITH OTHER GOVERNANCE MEASURES

Besides model explainability there are also other governance measures that insurance firms can put in place in order to ensure ethical and trustworthy AI systems. Therefore, in case the economic benefits of the use of a black-box AI systems undoubtedly outperform the risks, and/or in the absence of alternatives (e.g. for image or speech recognition), insurance firms may compensate the lack of explainability with other governance measures such as an enhanced level of human oversight (human-in-the-loop) and data management throughout the AI model lifecycle. For processes where insurance firms move towards automated model building and deployment with limited human oversight, insurance firms should adopt greater model explainability.

As discussed in the fairness and non-discrimination chapter, compensating for the lack of explainability by including a “human in the loop” may not be sufficient if the AI system provides unjustified biased outputs because it is trained with biased datasets. For example, in fraud detection an AI system could inadvertently guide loss adjusters into targeting with more scrutiny, say, some ethnic groups if the AI system were trained with biased data. Therefore, in addition to having adequate levels of human oversight, insurance firms need to establish robust data management processes in such AI use cases. Insurance firms need additionally statistical tests to reveal where possible discrimination could arise.

INSURANCE FIRMS SHOULD DEVELOP MEANINGFUL AND EASY TO UNDERSTAND EXPLANATIONS

Notwithstanding the importance of detailed and accurate information included in the terms and conditions of the insurance contract, the recipients of the explanations of the output of an AI system need to be able to understand

the information they are given. Therefore, insurance firms should be able to adapt complex mathematical logic and outputs of AI systems into simple and easy to understand explanations.

Explanations should be meaningful; they need to help stakeholders in making informed decisions (e.g. by providing consumers information about the main rating factors that influence their premium). Simplistic explanations, often useful for non-technical stakeholders such as consumers, may not be useful for stakeholders with a more technical background, who would need more comprehensive information (e.g. auditors) in order to be able to perform their duties.

INSURANCE FIRMS NEED TO TRANSPARENTLY COMMUNICATE THE DATA USED IN AI MODELS

The products offered by insurance companies are based upon data disclosed by persons, processes or external sources/vendors. The data is afterwards manipulated by a person, process or system. Finally, the data is consumed by the customer by buying a product or service from an insurance firm. In compliance with Articles 13 and 14 GDPR, insurance firms have to inform consumers about the types, sources and purposes of the use and processing of personal data in the AI models. The GDPR also recognizes the right of consumers to ask insurance firms to delete their personal data (“right to be forgotten”).

Article 6 GDPR lays down 6 legal grounds for the processing of personal data, the need to obtain consent from the customer for certain types of processing being particularly important from a transparency perspective. As highlighted by the Article 29 Working Party Guidelines on consent,⁷⁶ the consent provided by the consumer should be free, specific, informed and unambiguous. Indeed, consumers should be provided with the necessary information and, via an appropriate process, be enabled with the capacity to understand the key information the consumer needs in order to choose the right option. This information should ideally not be provided in long and legalistic terms and conditions of the insurance contract, but rather on dedicated privacy notices with clear and simple language and/or via an adequate on-line environment in the insurance firms’ website or app, building a coherent journey for the consumer who should have a good understanding of the consequence of his/her choices.

⁷⁶ Article 29 Data Protection Working Party (wp259rev.01) (2016).

More particularly, in the case of pricing and underwriting, consumers should be informed about which the main rating factors that would affect the premium in order to reinforce trust, enable them to adopt informed decisions and adapt their behaviour. Consumers need a clear understanding of the types of data that play a role in the insurance firms' decisions concerning them, to be able to verify whether the data considered by the insurance firm is relevant and accurate. Rating factor transparency would also contribute to reinforce consumer trusts and prevent insurance firms from relying on data points that consumers could potentially consider sensitive, intrusive or potentially discriminatory. Furthermore, and while it may not be directly related to the use of AI, cyber security incidents affecting their personal data should also be communicated to the affected individuals in a timely fashion.

INCREASING PERSONALIZATION OF PRODUCTS AND SERVICES COULD POTENTIALLY AFFECT THE COMPARABILITY OF INSURANCE PRODUCTS IN THE FUTURE, ALTHOUGH EXISTING TOOLS MAY PROVE TO BE SUFFICIENT

The increasingly availability of new types of data allows insurance firms to better profile consumers and develop increasingly tailored products and services adapted to the needs and characteristics of consumers. The greater variety of products and services, which certainly entails many benefits for consumers, could on the other hand reduce the comparability between products and services, which is an important transparency tool for consumers to make informed decisions. However, existing tools such as 'comparison websites' are widely used by consumers in some lines of business and therefore, despite their limitation (see next paragraphs), can help consumers compare different products. In addition, reforms introduced in European insurance legislation (IDD, PRIIPS) in recent years have placed significant focus on promoting the comparability of insurance products, namely with the requirement to provide to consumers before the conclusion of the contract key information documents using a standardised format and consumer friendly language.

Finally, it is important to acknowledge that transparency / comparability tools also have their limitations. For example, experience shows that consumers do not always carefully read the terms and conditions of their insurance policy. In addition, private comparison websites tend to focus excessively on price and typically have a limited

market coverage⁷⁷. Indeed existing tools should be combined with alternative solutions to ensure good consumer outcomes. In this regard it has been suggested that public comparison websites should be promoted to address the deficiencies of existing tools. Moreover, specifically with regards to the processing of personal data, to complement the "consent" requirement (i.e. transparency requirement) of the GDPR, relevant developments are taking place in the area of Personal Information Management Systems (PIMS), which are new products and services aiming to help individuals to have more control over their personal data and manage and control their online identity.⁷⁸

INSURANCE FIRMS SHOULD INFORM CONSUMERS WHETHER OR NOT THEY ARE INTERACTING WITH AN AI SYSTEM AND ITS LIMITATIONS⁷⁹

The use of chat-bots to interact with consumers in non-sensitive processes (e.g. Q&As or to help consumer navigate through their website) as well as the use of robot-advisors to provide advice to consumers (e.g. about investments options in life insurance) are increasingly common in the insurance sector. Whenever such tools are used by insurance firms, it is important that consumers are aware that they are interacting with an AI system and not a human. Consumers should also be provided with meaningful and timely information about the system's capabilities and limitations, and to the extent possible, consumers should be allowed to request the intervention of an employee at some point of the process. In the specific case of robot-advisors, consumers should also receive a basic understanding of the algorithms behind the recommendations (e.g. level of risk appetite, preference for "green" products or certain industries / types of securities etc.) as well as an indication whether human advice is also provided by the firm and how it can be accessed.

77 EIOPA (2014). Report on good practices on comparison websites: https://register.eiopa.europa.eu/Publications/Reports/Report_on_Good_Practices_on_Comparison_Websites.pdf

78 EDPB TechDispatch (2020)

79 The GDPR already foresees this requirement to a certain extent; pursuant to Art. 13(2)(f) GDPR, controllers must, at the time when the personal data are obtained, provide the data subjects with further information necessary to ensure fair and transparent processing about the existence of automated decision-making and certain additional information

2. TRANSPARENCY AND EXPLAINABILITY IN SPECIFIC AI USE CASES IN INSURANCE

PRICING AND UNDERWRITING

Insurance firms should be able to explain to regulators and auditors that the principles behind the tariff model are sound, and consumers should be informed of the main rating factors that influence the premium in order to reinforce trust, enable them to adapt their behaviour, and adopt informed decisions.

This area requires a high level of transparency and explainability of systems, models and data used. The reasons around the need of explainability in tariffs ratebooks vary; one of the most relevant is to assure transparency towards the regulator (and within the insurance firm itself): in accordance with existing regulation, it is important that insurance firms can prove and explain that the principles behind the calculation of the tariff are solid and sound (e.g. in order to guarantee the solvency requirements of the insurance firm, although continuous monitoring and control on an aggregated level can suffice in certain cases based on the principle of proportionality). The need for transparency/explainability is one of the reasons why GLM models are normally used to come up with tariffs ratebooks; they are simple to explain and allow performance sensitivities / “what if” analyses in a rather simple way.

The need for explainability also has primary importance from the consumer fairness perspective. Without it, consumer would not know the exact systems behind their tariff premium, therefore they should be informed as to the main data/factors that affect the tariff premium; this is also valid in case of “non-traditional” factors like scores of driving behaviour used in telematics tariffs. In the future, insurance tariffs could increasingly use features such as “emergency braking system activity” collected via telematics devices installed in cars or similar lifestyle features collected via health wearable devices. Such features will likely need to be processed by AI “black-box” systems, meaning that features can only be explained roughly but not in detail with the current state-of-the-art of explainability techniques. Insurance firms will therefore need to establish alternative governance measures around the use of such AI systems after conducting an AI use case impact assessment (e.g. taking into consideration issues such as whether the features processed by AI black-box systems have a major or minor impact on the premium paid by the customer) and taking into account the information needs of the recipient stakeholders.

NEXT BEST ACTION MODELLING

These types of AI systems would affect customer’s decisions in a more indirect manner and therefore may require a lower level of transparency and explainability, although insurance firms should count on other AI governance measures and ensure compliance with IDD provisions, including the need to adequately identify the target market.

We refer to “next best action” modelling to sum up in a single use case the streams of statistical models (e.g. customer life-time value estimation models) connected to up-selling and cross-selling solutions. The aim of these types of systems is to identify groups of customers, which have a higher/lower probability to need and/or buy a certain insurance product or service.

Compared to other AI use cases in insurance, these types of statistical models would affect customer’s decisions in a more indirect manner, they should rather enable the insurance firm to identify and propose the “most appropriate insurance solution to the right customer”, which is compliant with the IDD directive and could, in certain cases, imply synergies with the concept of “target market”. Indeed, these AI use cases should aim to follow a “likelihood to need” approach instead of “likelihood to buy”. Because of the aim of these systems, it is not always necessary to ensure the use of easily explainable statistical methods; the purpose of these models, in fact, is just to propose a certain solution/product to the consumer who may or may not decide to buy.

It is important to specify that “no-need of high level of transparency in systems” does not mean “low level of control over the data used”; the insurance firm must count on other AI governance measures to ensure ethical and trustworthy outcomes of AI systems, including assuring that all data used to calibrate models/systems are allowed from a regulatory, privacy and compliance point of view

FRAUD DETECTION

Transparency and explainability of fraud detection solutions using AI black-box systems to give indications of fraud to the claim handlers may be low, but should be complemented with alternative governance measures

We refer here only to fraud detection solutions using AI black-box systems to give indications/suggestions to the claim handlers on which claims to prioritise for further investigations. In such cases, the fraud detection score

should not have a direct impact on the consumer prior to a human intermediation/investigation i.e. fraud will always need to be proved.

Similarly with the previous use case (next best action models), the aim of a fraud detection system is to provide suggestions to the insurance firm on how to prioritise a certain action and, consequently, increase efficiency in internal processes. For this reason, it is not always necessary to ensure a high level of explainability for the statistical methods chosen to perform the solution. However, as explained in the fairness and non-discrimination chapter of this report, in addition to including a “human in the loop” insurance firms should also ensure that the AI systems are trained to the extent possible with unbiased datasets to avoid consistently providing claims handlers with biased recommendations. Adequate accountability and effective redress systems should also be in place.

Moreover, introducing an explainable fraud detection score has the advantage of facilitating the levels of acceptance, comprehension and usability of claims handlers/adjusters: they could be much more effective in their investigations if the score is easily explainable. Once again, it has to be specified that “no-need of high level of transparency in systems” does not mean “low level of control over the data used”; all regulatory, privacy and compliance principles must be respected when selecting the data to be used.

CLAIMS MANAGEMENT: “OPTICAL CHARACTER RECOGNITION” (OCR) AND “IMAGE PROCESSING TECHNIQUES”

It should be possible to use to these “black-box” AI systems to process images and text given when there are no practical alternatives and the benefits seem to outpace the risks, but similar to other AI use cases complementary governance measures should be in place to ensure ethical and trustworthy AI systems

Optical Character Recognition (OCR) software enables insurance firms to process claims by automating scanning and data extraction of an insurance claim submitted by the customer; AI systems go through hand written or printed data from scanned claims forms and detect errors or flag critical fields of the claims form. Similarly, low explainable AI systems such as deep neural networks / convolutional neural networks can be used to process datasets of images and identify patterns within them in an efficient manner. These techniques can improve the

efficiency, timeliness and accuracy of the claims management process.

Given that there are no practical alternatives to these “black-box” AI systems to process images and text, other governance measures such as enhanced human oversight and data management should be in place in order to ensure robust and trustworthy processes. A high level of human oversight may not be necessary for small claims or internal administrative processes using OCR or image processing techniques. However, human oversight would be necessary when the claim is above certain thresholds or during the first stages of implementation of these novel techniques in an organisation in order to ensure that the AI system is functioning as expected. The potential use of image processing techniques to process certain sensitive datasets (e.g. faces of individuals or images of human organs) and/or for certain tasks (e.g. pricing and underwriting in life and health insurance) would also immediately trigger the need of additional governance measures.

In claims management, additional transparency measures could include providing consumers with ex-ante practical information about the systems’ capabilities and limitations, the level of automation (e.g. whether there is human still involved/in control or if the claims settlement offer is fully automated) or practical specifications of image requirements needed (size, colouring, angle, side etc.). Consumers should also be informed whether submitting the image of a damaged car is mandatory or whether it is possible to request a human expert to examine the damaged car. Furthermore, it should be disclosed whether the data from the image is enhanced with consumer’s personal (behavioural) data to determine the final loss refund offered and if he/she agrees with this.

VIII. HUMAN OVERSIGHT

Principle of Human Oversight: Insurance firms should establish adequate levels of human oversight throughout the AI system's life cycle. The organisational structure of insurance firms should assign and document clear roles and responsibilities for the staff involved in AI processes, fully embedded in their governance system. The roles and responsibilities of staff members may vary from one AI use case to another. It is also important that insurance firms assess the impact of AI on the work of employees and provide staff with adequate training.

Article 41 (1) of the Solvency II Directive requires insurance undertakings to establish an adequate transparent organisational structure with a clear allocation and appropriate segregation of responsibilities and an effective system for ensuring the transmission of information. More specifically Solvency II's governance requirements also foresee the creation of governance functions (i.e. the audit, actuarial, compliance and risk management functions), which, in addition to the designation of a Data Protection Officer (DPO) in compliance with Article 37 GDPR, already provide several "lines of defence" to address potential issues arising from AI. However, there are other staff members (e.g. management board, data scientists, end users etc.) that play important roles in the design and implementation of AI systems within an organisation.

As recognised by the HLEG, establishing adequate levels of human oversight of such AI systems and processes helps ensuring that an AI system works as intended and does not cause adverse effects. Human oversight (often referred as "human in the loop" or "human on the loop" or "human in command")⁸⁰ is defined as some form of direct human involvement in the design, operation, maintenance, adaptation or application of AI systems. It is important to note that although AI increasingly enables the automation of tasks and processes, there will always be a certain level of human involvement in the deploy-

ment of AI systems along the different stages of the AI model lifecycle; humans can be involved in the selection and cleaning of the data used to train the AI system, in the selection of the most suitable AI algorithms to perform a specific task, in the calibration and operation of the model or in the monitoring of the outcomes and updating of the model.

The selection of the relevant staff with the adequate background, experience and training is particularly relevant in an AI context, where processes and tasks traditionally performed by humans become increasingly automated with the use of AI systems. Some fully automated AI systems may act as "black-boxes" and are therefore harder to audit/explain/understand, which raises a number of challenges, including a lack of understanding under which circumstances an AI system can provide inaccurate predictions. Including some form of human oversight contributes to more robust governance frameworks, as humans can intervene to verify, ratify or correct the prediction made by an AI system. For example, a highly complex rating model that automatically selects the best rating factors from a pool of lawful and approved variables might remain too complex to be easily understood. Developers might know how that type of AI model theoretically works, but having to explain how it arrives at specific outputs might be challenging. In these cases, human intervention might be advisable to challenge the model and ensure that it is not producing unethical or unintended outcomes due to unnoticed biases or correlations in the data.

Both non-AI and AI systems have their own advantages and disadvantages. On the one hand, AI systems are capable of performing more complex tasks in a more accurate, efficient and faster manner than systems without AI. On the other hand, specific human judgement can bring nuances to a decision-making process that automated processes are not able to capture, such as breadth of contextual knowledge and understanding, emotions, feelings, values or common sense. For example, insurance firms may decide to establish caps or guardrails to the outputs of AI tools in order to ensure ethical outcomes (e.g. not charging vulnerable consumers excessively). The decisions of both humans and AI systems can both be biased, the latter due to potential bias in the training data, and

⁸⁰ Depending on the level of human control, some publications differentiate between "human in the loop" or "human on the loop" or "human in command". In this document they are all broadly referred as human oversight. References included in the report to "human in the loop" should also be understood as human oversight in the broad sense.

the former due potential prejudices that they may have. An adequate balance of human-AI collaboration seems therefore desirable for certain tasks, noting that there are different levels of human oversight; from those where the human operator is in complete control, to those where they play a more passive or monitoring roles.

From a different angle, the increasing automation of processes has an impact for the work and employability of the staff of insurance undertakings. On the one hand, AI systems can assist them in performing tasks that previously were not possible or to develop their work more efficiently, such as in the case of the use of AI for fraud prevention. On the other hand, their tasks can fundamentally change and some workers can be displaced by automated processes. This raises the issue of the need to take into consideration its impact for the work of employees and provide staff with adequate training to enhance their digital skills.

1. HUMAN OVERSIGHT IN THE INSURANCE SECTOR

INSURANCE FIRMS SHOULD ESTABLISH ADEQUATE LEVELS OF HUMAN OVERSIGHT TAKING INTO ACCOUNT THE IMPACT OF SPECIFIC AI USE CASES AND OTHER GOVERNANCE AND CONTROL MEASURES IN PLACE

The selection of a higher or lower level of human oversight should be proportionate to the nature, scale and complexity of the risks inherent to the specific AI use case (intended use and potential impact) and taking into account the combination of governance measures in place around that specific AI use case. For example, for those processes where insurance firms move towards automated model build and deployment with limited human oversight, insurance firms should adopt greater model explainability or other governance measures such as enhanced data management or measures aiming to ensure the robustness and performance of AI systems, especially for high-impact AI uses cases. On the contrary, insurance firms may compensate for the lack of explainability of certain AI systems with enhanced levels of human oversight and data management throughout the AI model lifecycle.

ESTABLISHING ADEQUATE HUMAN OVERSIGHT THROUGHOUT THE AI SYSTEM'S LIFECYCLE

In the design phase, it is important that human developers take reasonable steps to remove any bias from the training datasets. Also in the design phase of certain AI use cases such as in pricing and underwriting, it is advisable to embed into the AI system operational constraints (e.g. guardrails), limiting the extent to which it can automatically change or make a decision without being first checked by a member of staff.

Once the AI system is in production, oversight shall focus on the day-to-day operations/processes of the AI system, monitoring and controls of the AI-system, adjustments and incident handling according to the previously established procedures, such as in the case of the guardrails. Processes can also include having relevant staff reviewing previously designed metrics on the outputs of the AI systems, for instance to monitor the impact of an AI system on certain groups of vulnerable consumers.

In the production phase, other human oversight measures can include prescribing that the output of the AI system does not become effective unless it has been previously reviewed and validated by a human. For example this could be the case of the use of AI in claims management for claims that are above certain thresholds, or in the case of fraud detection, since consumer fraud always needs to be proved by the insurance firm and should count with adequate redress mechanisms.

For other AI applications such as in pricing and underwriting, after the design and testing of the AI system, the output of the AI system can become immediately effective without the intervention of staff from the insurance firm. However, in addition to establishing relevant controls (e.g. analysing the AI system's outputs to identify unintended non-discrimination or unexpected results), insurance firms should keep adequate records of the training data and modelling techniques, to ensure transparency, auditability and that consumers have adequate redress mechanisms.

INSURANCE FIRMS SHOULD CLEARLY DEFINE IN POLICY DOCUMENTS THE DIFFERENT ROLES AND RESPONSIBILITIES FOR THE STAFF INVOLVED IN AI PROCESSES

Insurance firms should document their AI activities into policy documents, either in a standalone AI policy strategy, or by updating other relevant policy frameworks (e.g. risk management strategy, ICT-strategy etc.). These policy documents should set out processes, roles and responsibilities of the different staff members involved in the implementation of AI within the organisation. The relevant policy documents should be reviewed and updated periodically, in particular if there are material changes in the use of AI within the organisation. Processes should be in place to ensure a corporate culture and training which guarantees ethical values, including awareness of AI activities and their challenges and potential pitfalls. There is also a need to establish adequate processes to ensure sufficient communication, reporting and documentation towards administrative, management or supervisory bodies and vice versa, especially for those high-impact / material AI use cases. Accountability frameworks also need mechanisms to allocate responsibility for AI systems during their development, implementation and use, underpinning any outcomes.

Bearing in mind that governance and risk management processes should be proportionate to the potential impact and intended use of AI systems, and noting that insurance firms are free to choose the organisational structures that better suits their corporate structures and/or specific national requirements, the following roles and responsibilities are considered particularly relevant for the responsible deployment of AI in insurance.

› Management board

Management board level executives should have sufficient understanding about how AI is used in their respective organisations and their potential risks. They

should bear the ultimate responsibility for the use of AI in an insurance firm. The management board should be updated on a regular basis to enable understanding of AI deployment and use, in particular for those AI use cases that are significant given their intended use and potential impact. The risk management framework approved by the management board should encompass AI activities which should be tested by the relevant audit function.

The management board should provide clear internal communication about the vision and policy of using AI within the organisation, for example by means of developing an AI strategy and/or updating existing ones (e.g. IT strategy, risk management strategy etc.). The relevant policy documents should provide a framework for the application of AI in the organisation (e.g. goals, principles, values, processes, requirements, responsibilities, etc.).

› Compliance function / legal department

The laws and regulations underpinning AI solutions are constantly evolving so it is important to have legal or compliance monitoring of the regulatory guidance and legislation in this area. There are a number of legal frameworks that insurance firms must already obey when deploying an AI tool, for example Solvency II and GDPR. There may be national laws that need to be taken into consideration or European guidelines.

The responsible legal department / compliance officer should monitor the catalogue of deployed tools and ensure that they meet the standard of regulatory changes. Policies should be updated and communicated to comply with the regulatory framework.

› Risk management function

Insurance firms should review the extent to which their control, testing and feedback loop criteria of their risk management frameworks remain relevant and valid with AI, reflecting the expected reduction of manual interventions in AI systems. For example, some insurance firms

may consider automating some elements of model governance and sign-off to more fully realise the benefits (e.g. to stop manual oversight becoming a constraint on pace in the market). Finally, controls around internal and external data are expected to need enhancements given the increasing use of novel datasets and to ensure the underlying data is fit for purpose and free from prohibited biases.

› **Audit function**

A key part of the governance framework is risk management and internal audit. Auditability enables organisations to assess both the quality and efficacy of the algorithms and the effectiveness of the governance process. Trustworthiness is a key foundation of AI activities and positive audit findings can contribute to the trustworthiness of the technology and how insurers use it. Any negative outputs can be reported on and assessed by the risk team to enable appropriate controls to be embedded.

› **Actuarial function**

The actuarial function holder⁸¹ should consider the impact that AI activities have in the actuarial function or in areas the actuarial function significantly relies on. While the actuarial function holder does not need to be an expert in AI, the role holder must be able to demonstrate sufficient understanding of the AI techniques in use in the actuarial function or affecting actuarial activities to effectively challenge, assess potential risks, including ethical issues, and be able to make informed decisions in relation to them. It is important that sufficient and comprehensive communication is ensured.

The actuarial function holder should consider how principles and recommendations / requirements for AI apply to other actuarial work where complex modelling is used and similar risks and challenges are likely to appear, specifically in relation to unfairness, discrimination, transparency and explainability.

› **Data Protection Officer**

Organisations will be aware of their data protection obligations which apply to many AI applications. In particular, GDPR considerations must be built into the design of any tool, building and testing, the deployment of the application and the outcomes produced. The Data Protection

Officer or person responsible for data protection should be engaged at the outset and a Privacy Impact Assessment undertaken.

› **AI / Data / Ethics Committees**

Taking into account that each insurance firm should decide the adequate governance structure that fits their respective organisations, a good practice is to establish multidisciplinary and diverse committees comprised of cross-functional experts spanning actuarial, risk, compliance, data protection, legal and technology responsible for overseeing the use of AI in the organisation. Some large organisations are exploring (or have already set up) ethics committees in recent years. Given the overlap between ethics and AI deployment, these committees could align or merge with AI committees or Data committees. These committees should provide added value to the deployment of AI within an organisation and avoid becoming too formalistic, meeting in regular intervals to review AI systems' development, deployment or procurement and serve as an escalation point for evaluating risks and dependencies.

› **AI Officer**

Proportionate to the complexity, scale, potential impact and intended use of the organisation's AI developments, there may be value in appointing an AI Officer or similar (e.g. Data Officer) to ensure the consistency and coherence of the work undertaken by the business and technology teams. The AI officer may be a new function / position created within the organisational structure of the organisation (system of governance), or may be incorporated into the responsibilities of already existing functions (e.g. actuarial function, data protection officer, head of the IT department etc.). The AI Officer would be required to have sufficient expertise to provide oversight and advice to all functions and could coordinate the certification of activities.

As shown in the previous point about AI committees, effective deployment of AI within an organisation is dependent on collaboration between multiple functions. The advantages of appointing a person responsible for AI activities is that their primary focus can be ensuring that each team is aware of the policies and governance framework that apply. They can also monitor and coordinate all use cases of AI activities throughout the organisation and act as a central point of expertise for the organisation. A potential disadvantage could be that parallel responsibilities could be established by appointing a person

⁸¹ Key function holders are the persons responsible for a key function as opposed to persons having, carrying out or performing a key function

without being fully embedded in the existing system of governance.

› **Developers of AI systems**

Data scientists, actuaries and those responsible for the deployment of AI systems should have specific domain knowledge and receive regular training on how to implement ethical and trustworthy AI systems. These training / guidelines should promote the notion that the use of any AI solution should adhere to the fundamental ethical principles of fairness, transparency, explainability, accountability, robustness and accuracy and the respect for human autonomy.

Teams developing AI activities should ensure that the ethical principles are embedded into those project proposals and delivery plans where they play a role. In addition, the teams should maintain a catalogue of AI tools to enable other functions to review compliance on an ongoing basis. This catalogue and corresponding controls documentation should be available to senior management and control functions.

› **End users**

End users of AI systems in insurance such as claims loss adjusters or marketing professionals may not have a programming or mathematical background but they also play an important role in the development and implementation of ethical and trustworthy AI systems. On the one hand, they need to share their domain knowledge with AI developers so that they can develop AI systems that are fit for purpose. During the design phase they can also be involved in other tasks, such as in the labelling of datasets used in supervised learning AI systems (e.g. labelling past claims as fraudulent or not). During the implementation phase, they should also provide feedback to AI developers about any shortcomings of the AI systems or suggestions to improve it.

› **IT Security Officer**

Appropriate technical measures will need to be deployed to protect the security of the data and the model. The person (or team) responsible for security should ensure that data model security is embedded within the information security management framework. Like data protection, this should be regularly assessed throughout the lifecycle of the development and deployment of the application.

› **HR Officer**

As many of the governance requirements hinge on training, HR could be engaged to use appropriate training modules and ensure that both new joiners and existing staff are trained as appropriate. In collaboration with the relevant experts within the organisation, HR should also devote some attention to ensuring that the appointed people detailed above have the appropriate skillset and expertise to perform their role. But of course, necessary training can also be set up by data scientist departments or others on their own.

› **Other staff members and stakeholders**

All relevant staff within the organisation should have knowledge on how AI activities are applied across the organisation, what the output of such deployment is and what its strengths and limitations are. Teams responsible for utilising AI should have received relevant training to reduce the likelihood of deployment errors.

In some organisations the shareholders of the firm and/or the Supervisory Board may also play a role in the implementation of AI within the organisation. While their role will vary from one organisation to another, they will typically have a more passive role in the form of being provided proper information about the uptake of AI in the organisation, for instance in the annual report prepared by a Management Board.

THE ROLES AND RESPONSIBILITIES OF STAFF MEMBERS MAY CHANGE FROM ONE AI USE CASE TO ANOTHER

Depending on the specific AI use case and its materiality, the role of the different staff members may also change. For example, the Management Board may need to approve those AI applications that are more material before they are deployed into production, but not for other AI applications with limited risk and impact. Noting once again that insurance firms should establish the organisational structure that better fits its business model, the following table illustrates an example of the different involvement of staff members during the design phase an AI application: Approval (A), Consultation (C) and Information (I):

Figure 15 – Example of involvement of different staff members during the development phase of different AI application depending on their materiality

	Management / executive Board	Head of IT department	Developers of AI systems	Data protection officer (DPO)	AI / data officer	Compliance function	Risk management function	Audit function	Actuarial function
<i>High Impact</i>	A	A	A	C	A	A	C	I	C
<i>Medium Impact</i>	I	A	A	C	A	C	C	I	C
<i>Low Impact</i>	I	I	A	C	C	C	I	I	I

Source: EIOPA Consultative Expert Group on Digital Ethics in insurance

ASSESSING THE IMPACT OF AI ON THE WORK OF EMPLOYEES AND PROVIDE STAFF WITH ADEQUATE TRAINING

When insurance firms decide to use AI within their organisations, they need to take into account the impact of AI for employment, worker's right, digital skills and competence. The transition to work with AI systems should be fair from the employee perspective and due consideration should be given to human rights and the human factor of employees (e.g. when AI systems are used to evaluate, predict and guide the performance of insurance intermediaries or sales representatives).

Insurance firms should also transparently communicate to their employees the implications of the use of AI for their jobs and the skills they will need to acquire in the digital age. When employees are working in combination with AI, they should be provided with adequate training to ensure they have the necessary competence level and skills needed to perform the relevant tasks. Where relevant, union representatives should also receive appropriate training so that they can understand the implications of AI for the organisation.

2. HUMAN OVERSIGHT IN SPECIFIC AI USE CASES IN INSURANCE

PRICING AND UNDERWRITING

Compared to other AI use cases, pricing and underwriting should be subject to a proportionally high level of challenge and oversight throughout the AI system's lifecycle. Human oversight should apply to areas where it is more effective by making use of domain expert knowledge or incorporate information that did not play a role in the development of AI pricing and underwriting systems.

Human oversight may be used in the design, development, calibration, and testing of the pricing AI systems. Furthermore, the monitoring and recalibration processes are important tasks for staff experts. This becomes very relevant in case an insurance firm is transitioning towards "black-box" AI systems, affecting transparency and explainability, bearing in mind that insurance firms usually have in place consumer complaints processes or more thorough screenings of an individual, which allow for interaction with staff from a service team. Experts might be tasked with analysing the model's outputs to identify unintended non-discrimination or unexpected results. They may also scrutinize the model's factors to identify proxies for unlawful or discriminating factors.

In using AI for pricing, the challenges are similar to those for traditional models: ensuring robustness and performance, non-discrimination and lawfulness, ethical aspects such as fairness or an adequate level of explaina-

bility. New AI techniques may need new approaches and methods for calibration, testing, validation and monitoring – adapted to the respective techniques and new business processes. However, the general setting of human oversight from traditional pricing will also add value in the context of AI. Insurance firms may, for example, require experts to analyse the outcomes from the AI system and consider whether it is performing as expected. Consumer feedback could also be considered to gain an understanding of how the new AI pricing system is impacting individual customers rather than on aggregate.

While the principles remain the same, the complexity and frequency of consumer interaction can change significantly – for example in telematics motor insurance – which will lead to the necessity of developing processes to adequately address these challenges. This could be, for instance, a more frequent analysis of consumer feedback and model performance, to reduce the potential impact of a pricing model not working as intended or incurring in unintended discrimination.

Automation is an important aspect of pricing and it also plays an important role for fairness in the sense that similar risks, should be priced similarly. Therefore, human decisions with respect to the price in case of an individual consumer may not be sensible. That is, such a level of human oversight may not be more desirable from an ethical point of view.

CLAIMS MANAGEMENT

Human oversight for AI applied to the management of claims should be proportionate to the impact the claim has on the consumer. Small claims or those with a non-material impact on consumers may benefit from low level of oversight. When assessing if a claim has a significant impact on consumers, insurers should consider non-financial factors too and take that into account when deciding the extent of human oversight required.

Automated AI claims management systems have the potential to generate significant benefits for insurance firms and consumers: reduced costs, faster payment of claims, more accurate loss estimates, higher retention rates and better consumer experiences. While the potential benefits are clear, there are also risks and these could have a significant impact on consumers depending on the nature of the claim. Areas where AI systems could cause consumer harm are in the refusal of a legitimate claim, inaccurate repair estimates, negotiation of unfair settlement prices for damages as well as to unreasona-

bly restrict consumer choices in the options for repairing a vehicle among others. The first step to prevent harm is to design a fair process and to involve human experts in the development, implementation and testing of the AI system. An appeals / redress process that involves human oversight therefore seems a reasonable precaution to mitigate these potential consumer harms.

Insurers can follow a proportionate approach where there are specific triggers based on the potential impact on consumers (not just monetary terms) that define where human oversight is required. For example, insurers may decide that human oversight is required in all motor insurance claims exceeding a certain monetary threshold, involving vulnerable consumers, claims suspected of fraud or require the immobilisation of the consumer's vehicle for some time that would cause significant disruption to the consumer.

Human oversight is also likely to be important to deal with fraudulent claims considering the significant impact on consumers if wrongly suspected of fraud. Automation yields economic and time benefits, and human oversight ensures sensible dispute resolution and effective monitoring and evaluation of AI performance.

LOSS PREVENTION

Insurance firms can play a significant role in preventing losses by providing notifications and recommendations to consumers. Despite its benefits, there are still risks to be considered and any AI driven or automated system for such actions should be subject to appropriate human oversight based on the influence it could have on individual's behaviour and lifestyle.

Automated loss prevention through the use of pro-active intervention to influence consumer behaviour has the potential to reduce the number of accidents and claims, and therefore improves consumer safety. Reduced claims will ultimately lead to lower premiums in a well-functioning competitive market.

In a big data environment based on the dynamic exchange of telematics data, the insurance firm has access to a rich set of driver information that can be analysed and interpreted using AI technologies to evaluate risk on a continuous basis – this creates a set of analytical outputs that can be used to push notifications and advice to the driver, e.g., to change characteristics of their driving behaviour (speed, acceleration, route choice), offer incentives for improvements in driving score and general coaching

advice to improve driving performance. There is also the possibility of giving suggestions and advice for vehicle maintenance and safety checks based on telematics information. It is important that consumers receive such advice on a voluntary basis, and in those cases where consumers choose not to heed the advice, it should not have negative consequences in case of a claim.

Insurance firms should apply human oversight in a proportionate manner to the potential impact these activities could have on the customer and whether they are practicable. For example, suggesting drivers to start their journey to work two hours in advance because it would be safer might not be reasonable regardless of how true it is. Insurance firms cannot anticipate all possible scenarios in order to provide precise guidance to their staff in relation to when human oversight should apply. Instead, there should be guiding principles that are well understood by all involved, which indicate what should be escalated for additional oversight. There should also be regular monitoring of which notifications / recommendations were subject to human oversight and which were not so that errors can be identified as well as potential misalignment in the interpretation of agreed principles.

IX. DATA GOVERNANCE AND RECORD KEEPING

Principle of data governance of record

keeping: The provisions included in national and European data protection laws (e.g. GDPR) should be the basis for the implementation of sound data governance throughout the AI system lifecycle adapted to specific AI use cases. Insurance firms should ensure that data used in AI systems is accurate, complete and appropriate and they should apply the same data governance standards regardless of whether data is obtained from internal or external sources. Data should be stored in a safe and secured environment and, in particular for high-impact use cases, insurance firms should keep appropriate records of the data management processes and modelling methodologies in order to enable their traceability and auditability.

In addition to the references to fair and transparent processing of personal data already included in chapters VI and VII, Article 5 GDPR includes other relevant principles such as the principles of data accuracy, as well as the principle of purpose limitation, requiring firms that any further processing of personal data is compatible with the original purpose. Also noteworthy is the principle of data minimisation, according to which firms must only collect and process the personal data that is necessary. In accordance with Article 30 GDPR, insurance firms should maintain records of all the processing activities of personal data under their responsibility. Moreover, insurance legislation also contains relevant data quality requirements, such as Article 82 Solvency II Directive in relation to the calculation of technical provisions, which is further developed in Delegated Regulation 2015/35.⁸² The Delegated Regulation also includes data quality requirements for the calculation of Solvency Capital Requirement using internal models (Articles 121(3) and 231), and for the calculation of technical provisions (Article 19).

The use of data and data analytics is not new for the insurance sector; they have historically been used by insurance firms to assess and underwrite risks, price insurance policies or pay insurance claims. However, in today's digital

society and economy, there is an increasing availability of new sources and types of data (e.g. IoT data, image data or social media data), which can be processed by increasing powerful and complex AI systems, bringing several opportunities, but also some challenges.

The AI HLEG notes that it is of fundamental importance that firms ensure the quality and integrity of the data to prevent potential harm arising from AI models. Certainly, AI algorithms rely heavily on the training data. The calibration and structure of the model is determined by the input data, and therefore any bias, errors, inaccuracies or mistakes in the data used to train the model, either accidental or intentional, will be reproduced by the AI algorithm. This can be illustrated with the metaphor “garbage in, garbage out”, which holds both for traditional models as well as AI models.

Data quality is not only important for ethical reasons, but also for business performance. For example if the training data is not sufficiently complete and diverse, the risk of overfitting the model increases i.e. when the trained model predicts very good predictions for the training data, but does not do the same when confronted with new data points in the testing and production phase. This can cause that when the model is put into production it will deliver erroneous results for insured persons with characteristics that differ from those in the training set applied.

In addition to data quality and integrity, the AI HLEG recognises the importance of respecting privacy and data protection throughout the entire AI model lifecycle in order to allow consumers trust in the data gathering process. More particularly, emphasis is made in the need to have protocols governing the access to datasets within organisations, which is of particular importance for the insurance sector in view of the sensitive data handled in some lines of business.

Finally, considering the difficulty of understanding the functioning of an AI system (the so-called black-box effect), and thus the associated difficulties in auditing them, it is important to keep relevant records of the data used in the AI models as well as the modelling method-

⁸² European Commission Delegated Regulation (EU) 2015/35

ologies. As recognised by the AI HLEG, this will facilitate the auditability and explainability of AI systems.

1. DATA GOVERNANCE AND RECORD KEEPING IN INSURANCE

INSURANCE FIRMS SHOULD ADAPT THE DATA GOVERNANCE AND RECORD KEEPING MEASURES TO THE IMPACT OF SPECIFIC AI USE CASES

Similar to other governance measures (e.g. explainability or human oversight), in application of the principle of proportionality insurance firms should establish the relevant data governance and record keeping measures that are proportionate to the potential impact of the specific AI use case at hand. Those AI applications that are expected to have a higher impact should count with stricter data governance measures, and lower impact applications should have less onerous requirements on a proportionate basis.

INSURANCE FIRMS SHOULD USE COMPLIANCE WITH GDPR AS THE BASIS FOR SOUND DATA GOVERNANCE

The General Data Protection Regulation (GDPR) already includes comprehensive requirements on data governance concerning the processing of personal data by insurance firms (also applicable to other sectors). The use of non-personal data (i.e. not regulated by GDPR), in particular if it could impact a consumer because of its usage, should also have high data quality standards. Insurance firms have to comply with all the data governance requirements included in the GDPR, including having a lawful reason for processing personal data, such as having obtained the informed consent from the consumer. Where applicable, compliance with data management provisions in insurance legislation (e.g. for the calculation of internal models or technical provisions) should also be complied with.

Article 5(1) GDPR includes other important principles from a data management perspective such as the principle of purpose limitation, requiring firms to inform individuals about the specific purposes for processing the data and ensuring that any further processing is compatible with

the original purpose. Firms must also only collect and process the personal data that is necessary to fulfil that purpose (principle of data minimisation).

INSURANCE FIRMS SHOULD ENSURE THAT DATA USED IN AI MODELS IS ACCURATE, COMPLETE AND APPROPRIATE

AI is heavily dependent on data. Thus, to ensure good, meaning reliable and accurate ML systems good quality data is necessary. More particularly, accuracy, completeness, and appropriateness of data is key for building AI models, which deliver reliable and stable results. These are basic requirements of scientific work well established in the GDPR and Solvency II. In particular, Article 231 of Delegated Regulation 2015/35 specifies these data requirements, which can be summarised as follows:

Figure 16 - Data quality requirements for high-impact AI applications

Accurate	<ul style="list-style-type: none"> ▪ no material errors ▪ consistent utilisation ▪ recorded in a <i>timely</i> manner and consistently over time
Complete	<ul style="list-style-type: none"> ▪ sufficient historical information ▪ all relevant parameters used ▪ No relevant data excluded without justification
Appropriate	<ul style="list-style-type: none"> ▪ consistent with the purposes for which it is to be used ▪ estimations made on the basis of the data do not include material estimation errors ▪ data is consistent with the assumptions underlying the modelling techniques

Source: EIOPA Consultative Expert Group on Digital Ethics in insurance⁸³

SOUND DATA GOVERNANCE SHOULD BE APPLIED THROUGHOUT THE AI MODEL LIFECYCLE

As explained in the introduction, errors and bias can be found in the collection, processing and application of data, which if not corrected there would later be reflected in the output of AI models. It is therefore important that insurance firms should have in place sound data govern-

⁸³ Based on the data quality requirements for data used in certified internal models under Solvency II (Art. 231 of Delegated Regulation 2015/35 based on Article 121 (3) Solvency II Directive)

ance at every stage of the design and implementation of an AI algorithm:

- › **Data collection:** During the data collection phase, insurance firms should carefully select the types of data and data sources that are suitable for the task performed by the AI algorithm. In particular, collection of data should consider a proper coverage in terms of diversity (including paying attention to the absence of data from non-digital populations), but also all relevant scenarios in particular for adverse situations where possible.
- › **Data preparation:** After having collected the data, insurance firms need to process it so as to ensure that the data is accurate, complete and appropriate before it is used in the AI model. This phase involves understanding the nature of the data, its characteristics, format and quality (data exploration), and subsequently “cleaning” the dataset to address possible data quality issues, for instance by removing duplicated data, invalid data or completing missing values. In particular, insurance firms should pro-actively remove potential bias from the training data of AI models, not only by not using protected attributes such as gender or ethnicity (“fairness by blindness”), but also other variables that could be proxies of such protect attributes when their use is not justified (for further details see chapter VI). Data transformation (“feature engineering”) should be traceable where transformations have a significant impact on data and models (see below). Moreover, for business performance purposes during this pre-processing phase it is also important to ensure that the AI model does not over fit or under fit the training data.
- › **Post processing:** Finally, it is also important to assess the outcome of the AI algorithm from a data quality perspective, including seeking to reduce potential discriminatory biases from the trained models. Particular attention should be given to new types of data, for instance, coming from wearables, telematics, social media, imaging or other external data, such as credit scoring. All of the above applies and should be enforced with close attention, because of their novelty and the fact that behavioural data can be highly correlated with protected characteristics. For this type of data a “correction/verification” loop is of outmost importance. Insurance intermediaries can play an important role in the prevention for the use of poor quality data since, in their advising activ-

ities, they would be able to detect such potential poor data usage, in all areas of the value chain where they are involved.

THE SAME DATA QUALITY STANDARDS SHOULD BE APPLIED TO DATA PURCHASED FROM THIRD PARTY DATA VENDORS

AI systems used by insurance firms typically combine the use of data from internal sources (i.e. provided directly by the consumers or generated by insurers) and/or external sources (e.g. provided by credit rating agencies, public repositories or research centres). When making use of data from thirds parties, insurance firms should apply the same data quality standards that they apply to their own datasets. In particular for high impact applications, the use of data from third parties (including programming libraries, credit scoring agencies and similar sources) should be avoided if they do not meet proportionate requirements of transparency and explainability about the assumptions or methodologies used to process it. If insurance firms would still decide to make use of such datasets, they should not only rely on legal clauses included in contractual agreements with third-party data providers, but rather carry out their own data quality checks through the AI system model lifecycle to ensure the absence of bias or low quality of the data that outsources from third parties. Solvency II requirements for outsourcing are also relevant when relying on third party models and data.

DATA USED IN AI MODELS SHOULD BE HANDLED AND STORED IN A SECURE MANNER

Taking into consideration that insurance firms often handle sensitive types of personal data, in particular in certain lines of business such as health or life insurance, in line with principle of integrity and confidentiality of Article 5 GDPR, it is important that they have protocols in place to store the data in a secured and safe environment. Among other things, the protocols should require the data to be stored in restricted areas so that only the appropriate business users are able to access it. In doing so, due consideration should be given to the different types of data used by insurance firms; more sensitive data such as health data should count with stricter security and access requirements than less sensitive datasets such as anonymised data collected by insurance firms through their websites.

INSURANCE FIRMS SHOULD KEEP APPROPRIATE RECORDS OF THE DATA AND THE MODELLING METHODOLOGIES TO ENSURE THEIR REPRODUCIBILITY / TRACEABILITY

Considering the difficulty of understanding the functioning of an AI system (the so-called black-box effect), and thus the associated difficulties in auditing them, insurance firms should keep relevant records of the data used in the AI models as well as the modelling methodologies. This would

allow, on the one hand, to trace back decisions and verify decisions in case they could eventually be disputed, and on the other hand, avoid misuse of models by inattention.

In this context, and duly taking into account the principle of proportionality, for those high-impact AI applications (for which it is recommended to have repositories with all deployed models in the organisation), the main attributes of the model, regardless of whether it is developed internally or if it is outsourced from third parties, should be recorded as described in the table below:

Figure 17 – Record keeping requirements for high impact AI applications

Record	Description
Reasons for using AI	Explanation of the business objective/ task pursued by using AI and its consistency with corporate strategies / objectives. Explanation how this was implemented into the AI system. This would help avoid misuse of the AI system and enable its audit and independent review.
Integration into IT infrastructure	Description of how the model is integrated in the current IT system of the organisation and document any significant changes that could eventually take place
Staff involved in the design and implementation of the AI model	Identify all the roles and responsibilities of the staff involved in the design and implementation of the AI model as well as their training needs. This would allow to ensure accountability of the responsible persons.
Data collection	Document how the ground truth was built including how consideration was given to identifying and removing potential bias in the data. This would include explaining how input data was selected, collected and labelled.
Data preparation	Records of the data used for training the AI model, i.e. the variables with their respective domain range. This would include defining the construction of training, test and prediction dataset. For built (engineered) features, records should exist on how the feature was build and the associated intention.
Data post processing	Description of processes in place to operationalize the use of data and to achieve continuous improvement (including addressing potential bias). Records should specify the timing and frequency of data improvement actions.
Technical choices / arbitration	Document why a specific type of AI algorithm was chosen and not others, as well as the associated libraries with exact references. The limitation / constraints of the AI model should be documented and how they are being optimised alongside their supporting rationale. Ethical, transparency and explainability trade-offs that may apply together with their rationale should also be recorded.
Code and data	Record the code used to build any AI model which goes to production/exploitation. Exclusively for high impact applications, insurance firms should record the training data used to build the AI model and all the associated hyper parameters, including pseudo-random seeds. If this requirement proved to be too burdensome, insurance firms may put in place alternative measures that ensure the auditability of the AI model and the accountability of the firm using them.
Model performance	Explanations should include, inter alia, how performance is measured (KPIs) and what level of performance is deemed satisfactory, including scenario analysis and timing and frequency of reviews and / or retraining of the model. Ethical, transparency and explainability trade-offs that may apply together with their rationale should also be recorded.
Model security	Describe mechanisms in place (or make reference to) to ensure the model is protected from outside attacks and more subtle attempts to manipulate data or algorithms themselves: how robust is the model to manipulation attacks (especially important in auto ML models)
Ethics and trustworthy assessment	Description of the AI use case impact assessment i.e. the potential impact on consumers and/ or insurance firms of the concrete AI use case. Explain how the governance measures put in place throughout the AI systems lifecycle address the risks included in the AI use case impact assessment and ensure ethical and trustworthy AI systems.

Source: EIOPA Consultative Expert Group on Digital Ethics in insurance

2. DATA GOVERNANCE AND RECORD KEEPING IN SPECIFIC AI USE CASES IN INSURANCE

PRICING AND UNDERWRITING

Thorough checks need to be performed on data quality/unbiasedness, retraining/re-estimation AI algorithm and data reconciliation as well as records of data source/preparation/post-processing and model performance updated. Specifically for price optimisation practices typically requiring more complex AI systems, behavioural data used in these models demands retraining and re-estimation more often and therefore data validation and reconciliation practices as well as record keeping practices need to be regularly updated.

The acceptance criteria an insurance firm uses, to determine the risk exposure and premium that needs to be charged, is presented in the initial moment of contact with the applicant. From this moment on (personal) data provided by the applicant in order to enable the specific insurance coverage arrives at the insurance firm in the data collection phase. Subsequently, the data is prepared by means of cleaning, aggregation, transformations and possible enhancements such that it can serve as an input for the specific pricing model (GLM, ML or Robo-pricing). Within the AI system the model is then trained and tested (fine-tuning) until the model is considered suitable for deployment and use. The arrival of new data could lead to a new AI model life-cycle following the steps outlined above.

In the data preparation phase, data quality needs to be high (accurate, complete and appropriate) and the data needs to be unbiased. In the subsequent training and testing phase of the AI algorithm potential biases can also be found and removed to guarantee data quality. In the training and testing phase reverse engineering needs to take place to achieve data reconciliation to ensure that the data is not contaminated as a result of errors/mistakes. Furthermore, data reconciliation on a broader level within the AI model life-cycle can assist in understanding and explaining the 'black-box' model both internally as well as externally. Within this entire process privacy and data proportion in line with GDPR needs to be assured as well as guidelines in terms of access and rights to use/edit highly sensitive data. Pricing and underwriting use case are high impact cases as a result of which records need to be kept

of the reasons for using AI, integration into IT infrastructure, staff involved in the design and implementation of the AI model, data collection, data preparation, data post processing, technical choices / arbitration, code, model performance, model security and impact assessment.

Enhancing risk assessments by combining traditional and new data sources (IoT data) leads to changes in the data preparation phase. Data quality and unbiasedness needs to be assured within the new combined dataset. The AI algorithm needs to be retrained and re-estimated after which data reconciliation needs to take place with regards to the new combined dataset. Records should be updated about data collection of existing and new data sources both separately and aggregated. Also records of the data preparation and data post-processing phase should be updated. The impact of the new data source and model performance should also be updated.

Price optimisation practices based upon individual behavioural data creates the need for a more dynamic and adaptable (complex) AI model life-cycle. Data quality and unbiasedness are again crucial, however individual behavioural data demands retraining and re-estimation more often as well as data reconciliation. The latter becomes more important in this case since the dynamic feature introduces additional complexity in the AI system. Since the AI algorithm is optimised every iteration of new behavioural data is processed and records about data collection, preparation and post-processing need to be updated. Records of technical choices/arbitration, model performance and impact assessment need to be updated too.

CLAIMS MANAGEMENT (IMAGE RECOGNITION)

The data collection phase is updated with image provided by the claimant, therefore, the claimant needs to abide to these requirements in order to ensure high data quality and unbiasedness. Records about the reasons for using AI and integration into IT infrastructure are particularly important.

The claims management process initially demands data and/or images from the claimant after which internal and external (third-party) data is leveraged to process the claim. This is part of the data preparation phase where high data quality and unbiasedness is demanded. Subsequently, the data is used to train/test the AI algorithm that processes the claim. Thereafter, the claimant receives information on the claim settlement possibilities. This

basically concerns the AI model life-cycle with the general requirements mentioned above.

Image recognition, optical character recognition (OCR) and automated repair estimation and settlement affects the data collection phase. In this phase the requirements need to be specified with regards to the image. As opposed to data an image can be of (too) poor quality, framed or nudged and subsequently be interpreted incorrectly by the AI system. Therefore, the claimant needs to abide to these requirements in order to ensure high data quality and unbiasedness. Thereafter the data is prepared and additional checks can be performed such that the data, and especially images, conform to requirements. In case the process is fully automated, records need to be kept of the reasons for using AI, integration into IT infrastructure, staff involved in the design and implementation of the AI model, data collection, data preparation, data post processing, technical choices / arbitration, code, model performance, model security and impact assessment.

Preventive push-notifications, travel statistics reports, driving coach recommendations, driving score evaluation and rewards for good driving behaviour affect the entire AI model life-cycle as the interaction between the insurance firm and consumer becomes more frequent. Each exchange of telematics data leads to new analysis and therefore new push-notifications, statistics, scores and recommendations. This requires high data quality and unbiasedness of the telematics data. This is a high impact case therefore, records need to be kept of the reasons for using AI, integration into IT infrastructure, staff involved in the design and implementation of the AI model, data collection, data preparation, data post processing, technical choices / arbitration, code, model performance, model security and impact assessment.

LOSS PREVENTION

**Checks on data quality/ unbiasedness, retraining/ re-estimation AI algorithm and data reconciliation.
Records of data source/preparation/post-processing and model performance (automatically) updated.**

Once an AI model has been trained and tested and it is deployed for use, the AI algorithm can dynamically update its estimates based on frequently updated data. The newly generated estimates can provide for a more accurate assessment of the risks as well as notifications, statistics and recommendations as guidance to prevent losses from occurring. The data that is newly collected needs to be of high quality and unbiased.

X. ROBUSTNESS AND PERFORMANCE

Principle of Robustness and Performance:

Insurance firms should use robust AI systems, both when developed in-house or outsourced to third parties, taking into account their intended use and the potential to cause harm. AI systems should be fit for purpose and their performance should be assessed and monitored on an on-going basis, including the development of relevant performance metrics. It is important that the calibration, validation and reproducibility of AI systems is done on a sound manner that ensure that the AI systems outcomes are stable overtime and/or of a steady nature. AI systems should be deployed in resilient and secured IT infrastructures, including against cyber-attacks.

Article 25 GDPR includes the requirement for firms to ensure data protection by design and by default, which means that insurance firms need to integrate data protection into all data processing activities and business practices, from the design stage right through the life-cycle of the AI systems. With regards to insurance legislation, Solvency II also includes relevant provisions related to robustness and performance in the context of internal models and the calculation of technical provisions in insurance (Delegated Regulation 2015/35). The International Association of Actuaries has also developed International Standards for Actuarial Practice to ensure the robustness and accuracy of models.⁸⁴ Finally, once the proposal for a Regulation on digital operational resilience for the financial sector and amending Regulations (DORA)⁸⁵ enters into force it will introduce new operational resilience and IT security requirements.

The calibration, validation and documentation of mathematical models is a crucial and well-established step in the insurance sector, which is strongly rooted in data-led statistical analysis. While the level of automation, opacity, non-linearity or dimensionality of AI systems bring a number of new challenges from a model robustness and per-

formance perspective,⁸⁶ insurance firms can leverage on their mathematical expertise (e.g. by including extensions to the robust Model Risk Management (MRM) approach already in place in some insurance organisations for critical models) in order to ensure the fitness of methods and AI models used.

As highlighted by the AI HLEG, one of the key requirements for AI to be trustworthy is that it is robust. This applies in the technical and ethical sense, ensuring that AI operates in a reliable manner that does not cause harm. To do so, one must consider the context and environment in which AI works from an operational and societal viewpoint.

Performance (including prediction accuracy) plays a significant role in achieving robust AI. A high performance AI system will generally provide greater confidence in the reliability of its results, which facilitates the prevention of unintended harm and its deployment in a controlled environment. However, technical performance is not enough and it is also necessary to ensure that AI is used in the manner in which it was intended.

In addition, data is a key contributor to robustness and performance. Chapter X of this report already addresses the importance of data being unbiased, complete and accurate but for the purposes of robustness and performance, training data must also be representative of the populations targeted by AI models and consistent with the data that is available when in use.

⁸⁴ International Actuarial Association (2016)

⁸⁵ European Commission (COM/2020/595) (2020)

⁸⁶ Bussman, Giudici, Marinelli and Papenbrock, (2021) and Springer, Giudici and Raffinetti (2021)

1. ROBUSTNESS AND PERFORMANCE IN INSURANCE

THE POTENTIAL IMPACT OF A SPECIFIC AI USE CASE WILL DETERMINE THE LEVEL OF PERFORMANCE REQUIRED

Similar to the other governance measures described in this report, the required minimum robustness and performance measures of an AI system are closely linked with its intended use and the potential to cause harm (see chapter V for further details). The greater the risk of causing unintended harm of a specific AI use case, the stricter performance requirements that it must be observed, and vice versa. However other governance measures such as the level of human oversight of an AI application should also be taken into account.

THE AI SYSTEM CHOSEN SHOULD BE APPROPRIATE FOR ITS INTENDED USE

As a first step, insurance firms should clearly define what is the intended task / objective that it aims to achieve with a specific AI system. On the one hand this will allow defining the adequate performance metrics for that specific AI system (see further below), and on the other hand it will ensure that the model is not used for a purpose different for what it was created. If the scope or nature of the purpose changes then so must the AI system. For certain high-impact AI applications insurance firms may need to be able to sufficiently explain the AI model's decision-making process, even if this could be at the expense of model performance.

INSURANCE FIRMS SHOULD ASSESS AND MONITOR THE PERFORMANCE OF AI SYSTEMS ON AN ON-GOING BASIS AND TAKING DUE CONSIDERATION OF THEIR LIMITATIONS AND POTENTIAL SHORTCOMINGS

AI generally performs well when its predictions are close to the actual observations. It relates to its ability to make correct or accurate predictions. While errors are inevitable, insight on the nature, scale and likelihood of these errors is an important part of AI development and performance assessment. It helps making an informed use of AI and supports adequate risk management, enabling

modellers to assess whether the risk of the AI systems not performing adequately are within a defined risk appetite.

There must be awareness of the model's shortcomings and under which circumstances these materialise. This might require restrictions in model use to ensure that AI is deployed only under circumstances that lead to desired levels of performance or that minimise underperformance.

PERFORMANCE METRICS SHOULD BE ADAPTED TO THE OBJECTIVE PURSUED AND THE NATURE OF THE DATA USED

Performance metrics (accuracy, recall, precision etc.) depend on the nature of the data used and the intended application of the AI. For example, in classification AI systems used in fraud detection, insurance firms should decide if the objective is to maximise the prediction accuracy (number of fraudulent claims detected), reduce the number of false positives (legitimate claims wrongly labelled as fraudulent) or false negatives (claims labelled as legitimate which in the end are fraudulent). Depending on the objective, the metric used will be different. It is also important to monitor the performance of an AI system with regards to vulnerable consumers and develop appropriate metrics for this purpose (see chapter VI). Finally, also as explained in the previous chapter, the selection of performance metrics must be documented alongside its rationale.

SOUND DATA MANAGEMENT IS KEY TO ENSURE THE PERFORMANCE OF AI SYSTEMS

While the topic of data management was already covered in the previous chapter, it is important to highlight the critical role of data in model performance and therefore due consideration must be given to its integrity and fitness for purpose, considering whether it is accurate, complete, unbiased and representative. It is important to address any significant gaps in these areas and assess their potential implications in the model's usability. Data that is suitable and adequate for one task might not be for another and this assessment should be carried out every time a new AI model is developed. Data quality and integrity controls therefore contribute strongly to model performance.

When developing AI models, rationale should be provided on why the data used is deemed adequate, how

any gaps have been addressed and the potential implications of such measures. Data quality and integrity procedures should be adapted based on the results of previous controls so that they continuously incorporate lessons learned during the development of AI models.

The manner in which AI models are applied can have a significant impact on their performance if it is not consistent with how they were intended to be used. The use of AI must be consistent with the use that was considered when the model was developed. This must take into account any model limitations and the context and environment in which the AI model was intended to be used.

SOUND MODEL CALIBRATION, VALIDATION AND REPRODUCIBILITY MEASURES SHOULD BE IN PLACE

AI systems put into production should be, where appropriate, retrained, recalibrated and revalidated periodically. This would be particularly the case in the event of significant changes in the input data, relevant external factors, and/or in the legal or economic environment. The criteria for significant model change should be well documented for each AI application, and should become more stringent as materiality of the AI system increases.

In order to ensure that the AI system is fit for purpose, insurance firms should have structured validation processes in place even when AI systems are frequently updated or retrained. Insufficient validation procedures may enhance the model risk (i.e. that the model does not do what it was designed for), given that retraining can cause the model to change considerably.

The validation process of high-impact AI systems should include the use of scenario analyses and stress tests. Insurance firms should also determine the minimum frequency for revalidations for each AI system. However, in the context of fully automated AI systems which are constantly and automatically updated (replacing the traditional approach of periodic manual review and model update), the usefulness of periodic validations is limited and therefore validation could instead focus on the justification of model outcomes and the process through which a model is continuously adapted. Adequate controls should be in place (e.g. guardrails) to avoid that the AI system places excessive emphasis to short-term patterns that could affect the model's performance over time.

AI SYSTEMS SHOULD PRODUCE STABLE OUTCOMES OVER TIME

AI is robust when it can be used with confidence and behaves as is intended. This applies to a wide range of items from system IT security (including cybersecurity) to model stability, including data protection. Indeed an AI model that reliably and consistently provides similar predictions for similar inputs can be considered stable and therefore robust. This means that its parameters and estimations do not vary significantly when the data used for its training changes. The presence of outliers in the training data can affect a model's robustness, extreme data that could belong to a different statistical process would potentially affect robustness too, clustered and auto-correlated data, etc. Due consideration should be given to the data used for AI training, to ensure that different iterations or subsets of such training data would not result in significantly different AI models or key model parameters. Developers should document any actions taken in this regard during model training and testing, together with their rationale.

Model robustness can be evaluated using statistical metrics that measure the distance between model outcomes. These metrics should be selected while considering the model's intended use and desired outcomes. Attention should be given to how the model performs through time and whether there are significant differences in its output for recent and older data. Another approach to assess robustness is through model redundancy. A model can be considered robust when its output is similar to that of other, independent models. Parallel model running and back testing are useful tools to evaluate robustness from these points of view.

INSURANCE FIRMS SHOULD DEVELOP RESILIENT IT SYSTEMS AND INFRASTRUCTURES

Traditional models used by insurance firms are often coded as rules in production systems. However, AI systems are algorithmic, and therefore require more computation power (e.g. using cloud infrastructure). Insurance firms should therefore consider upgrading their IT infrastructure in order to address potential IT software and hardware constraints when implementing AI solutions.

Moreover, AI systems deployed in production for automatic use or with limited human oversight can cause significant harm before this is noticed and addressed. Cyber-attacks or adversarial attacks (data inputs an

attacker has intentionally designed to cause the model to make a mistake) can affect the trust placed on AI systems and therefore resilient security measures should be in place to be able to address such situations.

FALL-BACK PLANS SHOULD BE IN PLACE IN CASE THE AI SYSTEM DOES NOT PERFORM AS INTENDED

Especially for high-impact AI applications, insurance firms should establish fall-back plans in case AI systems do not perform as intended or are victims of a cyberattack or adversarial attack. Indeed fall-back plans (e.g. switching from a statistical to a rule based procedure) are an important contribution to model robustness, as they ensure that users can still operate in some capacity should the use of an AI model be interrupted.

AI SYSTEMS OUTSOURCED FROM THIRD PARTIES SHOULD COUNT WITH SIMILAR ROBUSTNESS AND PERFORMANCE REQUIREMENTS

There are a wide range of service providers (often the same ones that offer cloud computing storage services) that offer off-the-shelf AI applications to insurance firms. While insurance firms are ultimately responsible for the AI applications that they outsource (see Article 49 Solvency II), the service providers shall ensure that the AI applications that they commercialise count with the highest quality standards and provide insurance firms with sufficient information to enable them having an adequate understanding of their functioning and the limitations of the AI systems.⁸⁷ Moreover, insurance firms should also bear in mind potential concentration of risks arising from high dependence on a specific service provider, being in these cases important to maintain and test exit strategies for outsourced solutions, in particular when they are material.

⁸⁷ The information to be provided by third party service providers to insurance firms should be similar to the information requirements included Article 13 European Commission 2021/0106 (COD) (2021).

2. ROBUSTNESS AND PERFORMANCE IN SPECIFIC AI USE CASES IN INSURANCE

PRICING AND UNDERWRITING

Guidance for what constitutes adequate pricing model performance depends largely on the risk being priced, the data available and the impact of a potential lack of model performance.

Risks subject to large uncertainty will generally result in less accurate models and there are cases where insurance firms will accept model performance that would not be sufficient when modelling risks subject to less uncertainty. Rationale as to why the attained performance is deemed acceptable should be provided to inform stakeholders of significant modelling challenges and the risk's underlying uncertainty. The identification of the main sources of uncertainty is recommended where possible in order to assess what actions (if any) can be taken to reduce such uncertainty.

Lack of data or not having access to relevant data can also add to lack of model performance. This may be the case when pricing a new line of business for which there is no prior experience or when pricing a well-known risk but in a different jurisdiction, which could result in significant changes to how the risk materialises. Consideration should be given to these factors and how they affect model performance. Expert judgement might be required to provide an indication on how data shortcomings affect model performance (e.g. increasing volatility, likely under or overestimating the cost of the risk for a specific cohort, etc.). Any actions taken to reduce the impact of data shortcomings should be clearly identified and disclosed in addition to whether data quality is expected to improve with time or if it is an inevitable feature of the risk.

Special focus should be given to the consequences of inadequate model performance and information should be provided to stakeholders that enables them to understand the consequences for the business and its consumers. The threshold for what is considered adequate model performance should be based on the potential consequences of the model not performing as intended. For example, a new business line under controlled growth and which represents a small portion of an insurance firm's portfolio is more likely to tolerate lower model performance than one representing the majority of an insurance firm's business where significant mispricing would

have significant consequences in solvency and financial stability.

Pricing models in particular should require significant robustness. Different prices for seemingly similar risks should be understood and clearly explained. Similarly, significant differences in prices from one period (e.g. one year) to the next should have a well-documented rationale rather than be generically attributed to changes in the data. Where data has changed significantly, this should be disclosed to stakeholders and, where possible, accompanied by an estimation of the impact each significant data change had in the model.

Insurance firms should consider, within the business continuity management, what fall-back options should apply in the case of their pricing models being suspected of significant under performance, discrimination, etc. and requiring their present model to be withdrawn from production at a short notice.

CLAIMS MANAGEMENT: OPTICAL CHARACTER RECOGNITION (OCR) AND IMAGE PROCESSING TECHNIQUES

Required model performance should be commensurate with the role OCR or image processing techniques play in insurance firms' claims management processes and the impact they ultimately have on claimants, considering, if applicable, any human oversight that may be part of the process.

While OCR techniques usually provide extremely high rates of accuracy in many instances, they might still result in a significant number of errors when applied to documents including many words and there is the risk that key information such as names or addresses could be affected by such errors. Nevertheless, this could be acceptable if an insurance firm uses this information to supplement existing client data or where there is adequate human oversight. Processes that rely on extensive automation might require additional measures to ensure adequate performance such as multiple models that help overcome each other's weaknesses, an automated process to ask claimants for confirmation or clarification where the

model has been proven to show increased rates of error, etc.

Image processing techniques might be used for triage rather than to quantify actual compensation and since this process would already incorporate human oversight, the requirements for robustness could be lower.

Image or document quality could significantly affect model performance and insurance firms should be aware of this shortcoming and allow for that in their processes. This could result in a combination of requesting better quality data from loss assessors or claimants and passing the claim and its documentation to be analysed by an individual.

LOSS PREVENTION

The importance of these initiatives lies, in most cases, on how the models are applied and the potential impact on consumer behaviour, therefore performance and robustness requirements should be considered with that in mind.

When considering the potential impact on consumer behaviour, it is important to see how the recommendations or incentives provided by the insurance firm could influence potential claims arising from failure to adhere to them. Consumer perception of an insurance firm's likely position in this regard will play an important role on how they act.

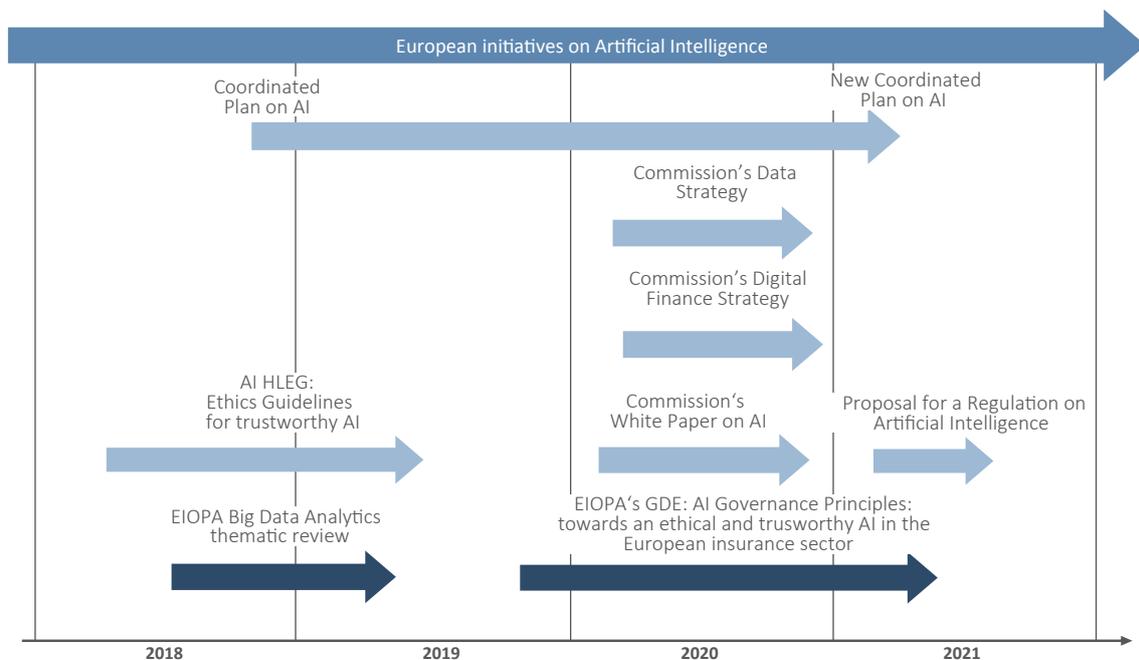
For example, if an insurance firm suggests reducing any driving to only essential trips due to weather concerns, insurance firms must consider how this could potentially be in conflict with government guidance or whether consumers might see it as a prohibition from travelling that could result in claims being declined if not followed. Consumers might also wonder how they can justify that travel was essential in the case of a claim. Where insurance firms take these loss prevention measures into consideration when accepting or declining a claim, then there must be very high requirements for performance and robustness. If it is simply used to provide advice and it is clearly disclosed as such, then performance and robustness requirements can be more relaxed.

XI. CONCLUSION

In recognition of the important role that AI is expected to play in shaping the digital future of European societies and economies, the European Union has adopted several initiatives in recent years aiming to address the opportunities and challenges of AI. The European Commission's Coordinated plan on AI⁸⁸ provides a high-level overview of these initiatives, which include several initiatives to strengthen the EU's AI capabilities through different funding and research tools. Among other initiatives, it also foresees the creation of AI testing and experimentation facilities, as well as a network of European Digital Innovation Hubs that will help SMEs and public administrations to take up AI. The Coordinated plan on AI also makes reference to the Commission's Data Strategy⁸⁹ and related initiatives, which ultimately aim to facilitate the free-flow of data across the EU, which is fundamental for the development of AI.

Strengthening the up-take of AI is also one of the key priorities identified by the Commission's Digital Finance Strategy,⁹⁰ which also acknowledges the risks arising from the use of AI and foresees the possibility of developing regulatory and supervisory guidance on the use of AI applications in finance. Building on the AI HLEG's Ethical Guidelines for Trustworthy AI and the subsequent Commission's White Paper on AI, the proposal for a Regulation on Artificial Intelligence⁹¹ published by the European Commission in April 2021 seeks to address these risks by developing a legislative framework to ensure ethical and trustworthy outcomes for certain high-risk AI applications. Insurance-related activities are currently not included among the high-risk AI applications that will need to comply with the requirements of the legislative proposal, which now needs to be deliberated by the European Parliament and the Council.

Figure 18 – Overview of some of the key EU initiatives on AI



Source: EIOPA Consultative Expert Group on Digital Ethics in insurance

88 European Commission COM(2021) 205 (2021)

89 European Commission COM/2020/66 (2020)

90 European Commission COM(2020) 591 (2020)

91 European Commission 2021/0106 (COD) (2021)

In addition to the cross-sectorial initiatives mentioned in the previous paragraphs, specifically for the insurance sector EIOPA published the key findings of its thematic review on the use of Big Data Analytics in motor and health insurance in May 2019. This fact-finding exercise showed that around one third of the participating insurance firms were already adopting a wide variety of AI use cases across the different areas of the insurance value chain, and another third of them were experimenting with this technology. The thematic review concluded that BDA / AI represents many opportunities for the insurance sector, both for consumers as well as for insurance firms; from more efficient and automated processes to more personalised products and services or to faster claims processing techniques. However, the thematic review also identified a number of challenges which are not necessarily new for the insurance sector, such as the ones related to transparency as well as to fairness and non-discrimination.

As a follow-up to the thematic review, the GDE multidisciplinary stakeholder group was created, among other things because the challenges arising from AI often imply relevant trade-offs that go beyond purely regulatory or supervisory considerations, such that EIOPA considered a wider examination of these challenges would be useful. Taking into account the opportunities and challenges of AI, the GDE has developed an ethical and trustworthy AI governance framework which seeks to enable stakeholders from the insurance sector to harness the benefits arising from AI, while at the same time addressing the challenges in a proportionate manner. The proposed framework recognises the freedom of insurance firms to select the combination of governance measures that better adapts to their respective business models, and more particularly to the concrete AI use cases that they aim to implement, but at the same time highlights those areas requiring special consideration so as to promote trust in the use of AI by insurance firms. It is aimed at being a useful toolbox that can be used to rise to the opportunities and challenges arising from the use of AI.

Adequate governance measures need be implemented throughout the complete AI system's lifecycle in a proportionate manner. This should be done by assessing the potential impact that a concrete AI use case may have on insurance firms and/or consumers. The higher the impact of a concrete AI use case, the more robust governance framework that is required, noting that an ethical and trustworthy governance framework is achieved by a combination of measures and not by a single / stand-alone one. The principle of fairness and non-discrimination highlights that as part of their corporate social responsibility

insurance firms should bear in mind that there are some insurance lines of business which are particularly important for societal and financial inclusion. Insurance firms should also be aware that, if not handled in a responsible manner, AI has the potential to reinforce existing inequalities by disproportionately impacting consumers that due to their personal circumstances are already in a vulnerable situation. Making relevant efforts to assess the outcomes of AI systems is therefore crucial.

Insurance firms should also strive to use explainable AI systems, in particular for high-impact AI applications, although in certain cases the lack of explainability may be compensated with alternative governance measures. Insurance firms should also establish adequate levels of human oversight across the AI system's lifecycle and assign clear roles and responsibilities amongst their staff and provide them with adequate training. Data management and record keeping is key to ensure the accountability of insurance firms, and therefore they should make reasonable efforts to removing bias in the training data and keep relevant records of the modelling methodologies used and how datasets were processed. Furthermore, the principle of robustness and performance highlights the need to monitor the prediction accuracy of AI systems using relevant metrics, and they should be deployed in resilient and secured IT infrastructures.

To conclude, it is important to bear in mind that insurance firms need to comply with all applicable legislations at all times, which may influence the selection of AI governance measures for a concrete AI use case.

Moreover, it is expected that the AI governance principles for ethical and trustworthy AI proposed in this report will need to be reviewed in the coming years in the light of the on-going extensive developments in the field of AI, including the development of tools for tackling potential risks arising from the development of high risk AI use cases. The issues touched on by the development of AI are very varied and can have profound effects on the way we understand insurance and our society today, and therefore require continuous dialogue between all the different stakeholders from the insurance sector and beyond.

ANNEX 1 – BENEFITS AND ETHICAL CHALLENGES OF AI USE CASES ACROSS THE INSURANCE VALUE CHAIN

Figure 19 – Benefits and ethical challenges of some AI use cases in insurance

Insurance Value Chain	Business Process	AI Systems	Insurance Firm Benefits	Consumer Benefits	Main ethical issues
PRODUCT DESIGN & DEVELOPMENT					
	<i>Market research</i>	Historical consumer and survey data to inform new product development	Innovative products to grow sales and retain consumers, enable new entrants into insurance	Increased innovation and competition	Use of data from various sources where the scope of data processing has been insufficiently clearly explained to consumers (e.g. social media data or usage-based apps (waze, health trackers etc) where the purpose of selling data for insurance purposes has not been clearly explained.
	<i>Product development</i>	Novel products, e.g. parametric, usage-based insurance and behavioural insurance	Increased likelihood of successful launch and growth of new insurance product-markets	Wider choice of suitable insurance services, improved user interfaces and better value insurance	Consumer bias towards the existing consumers of incumbent firms; non-digital consumers information not taken into account Certain types of behavioural data can be highly correlated with protected characteristics
	<i>Digital (video) identification</i>	Video facial recognition Passport document authentication	Higher customer satisfaction; reduction of password-reset requests and generally reduced call centre activity	Faster, easier self-authentication; Usability at any time, very little time required, low personal requirements	Need to have access to digital tools (internet, computer, tablet, mobile phone etc.)

PRICING AND UNDERWRITING

Insurance Value Chain	Business Process	AI Systems	Insurance Firm Benefits	Consumer Benefits	Main ethical issues
	<i>Risk assessment</i>	<p>(a) Traditional data sources, e.g. consumer information, exposure data and loss data to estimate risk profile for groups of consumers</p> <p>(b) New forms of big data, e.g. behavioural data, Internet of Things, telematics, personal tracker data and GPS to inform behavioural analytics</p>	<p>(a) Traditional risk assessment to inform underwriting decisions</p> <p>(b) New forms of big data, e.g. behavioural data, Internet of Things, telematics, personal tracker data and GPS to inform behavioural analytics</p>	<p>(a) The risk is an informed assessment based on the characteristics of similar consumers within a broadly homogeneous group</p> <p>(b) Micro-segment risk profiles may benefit consumers with lower individual risks than their 'average' segment but may identify higher-risk sub-groups; there is the potential for risk mitigation and interactive communication between the insurance firm and the consumer</p>	<p>Algorithmic bias from limited training data</p> <p>Use of non-risk data in pricing may be difficult to justify and explain</p> <p>Risk of exclusion of small groups of high-risk consumers</p> <p>Privacy and surveillance issues alongside concerns over citizen autonomy</p> <p>While not directly related to pricing, ethical issues could arise around data management and re-selling and commoditisation of personal data</p>
	<i>Price optimisation</i>	<p>(b) Micro-segment / personalised pricing based on individual non-risk behavioural data (e.g. to estimate price elasticity, lifetime value and propensity to churn) and market competition analysis</p>	<p>(a) Increasing profitability and consumer retention</p> <p>(b) Accuracy of pricing that can be linked much more precisely with individual risk profiles and other personal characteristics</p>	<p>(a) pricing can be seen to be broadly 'fair' based on belonging to a market segment</p> <p>(b) personalised pricing can be supported by detailed and transparent explanations that may also mitigate risk through dynamic pricing based on behaviour</p>	<p>Low explainability of AI black box models may not allow explaining to consumers which are the rating factors affecting their premium</p> <p>Use of non-risk behavioural data in pricing may be difficult to justify and explain</p> <p>Negative impact on vulnerable consumers (e.g. old age, low level of studies, low income populations) with lower price elasticity and/or propensity to churn, often due to the personal characteristics that make them part of vulnerable groups</p> <p>Information asymmetry concerns</p> <p>Non-risk factors used in pricing may not be justified, e.g. social media data or clickstream data</p>

Insurance Value Chain	Business Process	AI Systems	Insurance Firm Benefits	Consumer Benefits	Main ethical issues
	<i>Sales and marketing</i>	Digital marketing techniques based on the dynamic analysis of online search behaviour	Adaptive digital marketing to capture consumer attention and increase online sales through insurance websites and digital search intermediaries	Ease of use and less time to search and evaluate competing insurance services, potentially achieving better consumer value and more suitable insurance products	Potential to exploit consumers to coerce them into buying unnecessary or expensive insurance Exclusion of vulnerable groups and how to manage protected characteristics at this stage in value chain
	<i>New consumer acquisition</i>	Virtual Assistant and Chatbots that utilise Natural Language Processing (NLP) and insurance ontologies to support communication	Facilitate consumer journey, particularly in the online channel, lower cost to serve consumers, and increased automation of Know Your Consumer (KYC) information requirements	Immediate online help that is context specific to improve the relevance of the information and reduce the time for the consumer journey on a 24/7 availability	Potential for automation of incorrect or harmful advice Atomisation of offerings means that product offerings risk becoming <i>financial advice</i> <i>Risk of biased consumer selection in new consumer acquisition strategies</i>
	<i>Consumer retention:</i>	Consumer churn and switching models	Smart competitor intelligence that can be used to defend market share and measure the likelihood of switching behaviour based on price and product attributes	Better prices for consumers that are willing to negotiate and navigate price noise	Exploitation of consumer characteristics such as inertia to reduce competition and artificially high margins
	<i>Consumer Relationship Management (CRM)</i>	AI-managed proactive consumer communication and cross-selling of related services	Better consumer service across multiple channels at lower cost, differentiated offerings, improved consumer retention and increased revenue per consumer	Relevant and timely communication, ease of access to related insurance services and value benefits from bundled products and services	Lock-in effects on consumers may reduce their propensity to search and find better products and services, and risk of biased and /or incorrect advice
	<i>Behavioural analysis of consumers</i>	Behavioural risk analysis in areas such as health and automotive	Better understanding of risk and pricing, and mitigation of risk through pro-active communication with consumers to nudge behavioural improvements	Competitive pricing and rewards for lower risk behaviour and reduction of adverse outcomes	Unintended consequences such as consumers gaming the system may lead to unsafe behaviour, e.g. exercising when feeling ill, and being distracted by insurance 'safety' advice when driving Monitoring of behaviour could develop into surveillance, which leads to data privacy issues, e.g. confidentiality about visit locations on a GPS route Alters power relations between consumer/citizen and insurance provider Insurance firms become intrinsically linked to non-financial actors – i.e. more lightly regulated

SALES, DISTRIBUTION & MARKETING

Insurance Value Chain	Business Process	AI Systems	Insurance Firm Benefits	Consumer Benefits	Main ethical issues
E-SERVICE	<i>Insurance policy management</i>	Robotic Process Automation (RPA)	Improve the performance of legacy systems and reduce administrative costs for routine activities, e.g. email correspondence, broker quotes, policy amendments and management of legal documents	Lower prices and better e-service, e.g. policy amendment and speed of response to routine enquiries	Small, but could result in unfair bias in the rules-based processing of policy documents, claims and e-service
	<i>Consumer self-service through multiple channels</i>	NLP, Voice recognition, insurance ontology maps and Virtual Robots to facilitate consumer self-service and offer advice	Chatbots use a variety of AI technology to support routine communication regarding all aspects of e-service and reduce the cost to serve consumers	E-service support and advice are available through Internet, call-centre and mobile app 24/7 and reduces the amount of time and effort required to manage insurance policies and keep them up to date and meet current requirements	Excludes or discriminates against consumers with poor technology knowledge and skills, and potential for incorrect advice and guidance that could result in unsuitable policy amendments and purchases
	<i>Consumer complaints</i>	Automated resolution of a high proportion of consumer complaints through all marketing channels	Reduced administration costs	Timely resolution of complaints	Risk of algorithmic bias in the resolution of a complaint to the detriment of specific consumer segments
LOSS PREVENTION	<i>Coaching</i>	Provide diagnostic advice and coaching based on AI analytics from health, home and automotive telematics, e.g. suggest exercise and driving behaviour changes	Provide higher value services to a broad consumer base and maintain an economically viable cost/income ratio	Receive valuable, advanced analytics advice that is based on relevant big data from a large insurance community that allows improved behaviour leads to lower risk of accidents and reduced insurance premiums and better understanding of personal risk profile	Behavioural nudging could lead to loss of individual freedoms and independent decision-making Risk of incorrect advice that could lead to an adverse outcome for the consumer

Insurance Value Chain	Business Process	AI Systems	Insurance Firm Benefits	Consumer Benefits	Main ethical issues
	<i>First notice of loss (FNOL)</i>	Automated SOS calls via telematics, prediction of accidents from IoT sensors and GPS	Automated collection of accident data to inform the handling of the claims process	Speed of response from emergency services	Negligible other than economic cost of false alarms
	<i>Loss reserving</i>	Use machine learning to estimate losses, in particular of high-frequency claims	More accurate financial forecasts of significant loss events improve financial planning and reporting	Gain surety of future payment through better estimates and allocation of claims money to specific large loss events	Negligible to the consumer
	<i>Claim processing</i>	AI image recognition to estimate repair costs repairs in household property insurance, business premises and automotive	Reduced cost of expert assessors and faster turnaround of claims	Reduced time to settle the financial element of the claim and arrange and to improve the quality of the claims settlement offer (e.g. repair cars in garages with long time guarantee on the quality of the repair)	There is the potential to withhold the true value of the claim, e.g. through the short-term inducement of faster payment. Automated systems may be more opaque to stakeholders
	<i>Fraud detection</i>	AI fraud detection to identify anomalies in individual claims and search for fraud patterns across all claims	Reduced payments to fraudulent and illegal claims to reduce overall costs	Easier and faster claims for legitimate insurance claims, and lower premiums from lower industry costs in the insurance market	A false accusation of fraud could result from algorithmic bias based on previous fraudulent claims and therefore disadvantage specific groups
	<i>Automated payments</i>	Integrated payment systems that link automatically to the workflow	Reduced administration costs from closed-loop systems and reduced fraud	Faster payments for claims and invoiced work	Negligible to the consumer where bank account penetration is high

Source: EIOPA Consultative Expert Group on Digital Ethics in Insurance²³

CLAIMS MANAGEMENT

ANNEX 2 – AI USE CASE IMPACT ASSESSMENT FRAMEWORK

SEVERITY

• IMPACT TO CONSUMERS

Number of consumers affected: The rationale for this indicator is that the higher the number of consumers affected the higher the severity of potential harm of an AI use case, and vice versa.

- › High: the use case has a direct impact on a high number of consumers.
- › Medium: the use case has a direct impact on a reduced number of consumers
- › Low: use case has no direct impact on consumers or only a very limited number of them

Consumer interaction and interests: The rationale is that certain use cases that might affect consumer's rights and interests – either by the level/type of interaction and/or consumers vulnerability - have more potential harm.

- › High: use case is implemented in a process that interacts directly with consumers (from an insurance undertaking or throughout an insurance intermediary, etc.) and/or on essential consumer interests (e.g. monetary, non-material harm, health or legally protected rights)
- › Medium: use case is implemented in a process that has a moderate impact on consumer rights and obligations under the contract and/or entering into a contract and/or on essential consumer interests (e.g. monetary, non-material harm, health or legally protected rights).
- › Low: use case is implemented in back office operations with no material consumer impact.

Types of consumers affected (including vulnerable consumers): the rationale is the power imbalance between vulnerable consumers and insurance firms, meaning the individuals may be unable to easily exercise

their rights and/or protect their interest because of their own situation of vulnerability.

- › High: Potential negative impact on consumers with characteristics which might be considered vulnerable (e.g. old age, low level of studies, low income etc.)
- › Medium: Not likely to have a significant impact on consumers with characteristics which might be considered vulnerable (e.g. old age, low level of studies, low income etc.)
- › Low: Potential positive impact on consumers with characteristics which might be considered vulnerable (e.g. old age, low level of studies, low income etc.)

Human autonomy: the rationale is that some AI systems, especially those used in consumer-facing applications, can potentially have a significant impact the behaviour and self-determination of consumers.

- › High: AI systems can shape and influence the behaviour and/or self-determination of consumers through mechanisms that may be difficult to detect for consumers, for instance because they harness sub-conscious processes, and could potentially lead consumers to decisions that are not on their best interest.
- › Medium: AI systems can shape and influence the behaviour and/or self-determination of consumers, but in a manner that is transparent and easy to understand for consumers and will likely lead to positive outcomes for consumers (e.g. risk prevention).
- › Low: AI system has no significant impact on the behaviour and/or self-determination of consumers.

Anti-discrimination, Diversity and Fairness: AI systems run the risk of creating or perpetuating discrimination and bias when processing personal information. AI systems can also make predictions which frequently turn out to be incorrect, lead to disparities in outcomes between groups or use personal data in ways which individuals would not reasonably expect.

- › **High:** Use of large new datasets, which are imbalanced (e.g. the training data has a greater proportion of datasets in favour of a particular ethnicity or gender), or uses datasets that reflect past discrimination (e.g. criminal records in some jurisdictions) or complex AI algorithms that can capture non-linear correlations in the training data and therefore in some sense they are able to reconstruct the hidden (protected) information.
- › **Medium:** Use of not excessively large datasets which to a great extent have already been used in insurance in the past and some efforts have been done to remove biases from the training and ensure that the dataset is not imbalanced. The processing of the datasets is done by complex AI algorithms supported by relevant governance measures (e.g. human oversight, use of supplementary explainability tools, metrics to monitor the outcomes of AI systems etc.).
- › **Low:** Use of simple and explainable algorithms trained on small balanced datasets with which the insurance firm already has extensive experience using them and where reasonable efforts have been made to ensure that the dataset is sufficiently representative and free of bias.

Line of business relevance for consumers: The rationale is that certain lines of business are considered to be more essential for consumers than others, which is often reflected on the fact that they are mandatory

- › **High:** use case is applied in lines of business that are considered to be essential for consumers and/or mandatory (e.g. motor insurance, health insurance, household insurance)
- › **Medium:** use case is applied in lines of business that are not essential for consumers / mandatory but important for certain groups of consumers (e.g. travel insurance)
- › **Low:** use case is applied in small lines of business that are not essential for consumers / mandatory (e.g. mobile phone insurance, home appliances extended guarantees)

• IMPACT ON INSURANCE FIRMS

The rationale is that use cases that are implemented in operations that are essential for the insurance firm and/or can have a material impact of the insurance firm financials have more potential harm

Business continuity:

- › **High:** the use case is implemented in a critical activity (i.e. if the activity fails the insurance firm will incur in high risk of disrupting its core business, e.g. issuing policies, managing claims, etc.)
- › **Medium:** the use case is implemented in a moderately sensitive activity (i.e. if the operation fails, there is a workaround to continuing the core business that might be in place for a short period of time)
- › **Low:** the use case is implemented in a low sensitive operation (i.e. if the operation fails no core business activities are affected)

Financial Impact:

- › **High:** Failure of the use case results in a material impact to the financial commitments of an insurance undertaking (e.g. large number of contracts / Gross Written Premiums, solvency ratios will be affected)
- › **Medium:** Failure of the use case has a short-term financial impact but an improvement plan can be put in place to mitigate impact to stakeholders.
- › **Low:** Failure of the use case has no material financial impact

Legal impact:

- › **High:** Failure of the use case results in a violation of legal commitments with the potential of critical impact of an insurance firm (i.e. high-end sanctions from supervisors and/or criminal liability and/or major civil liability)
- › **Medium:** Failure of the use case results in a violation of legal commitments with the potential of a certain impact of an insurance firm (i.e. low-end sanctions from supervisors and/or civil liability)
- › **Low:** Failure of the use case results in no violation of legal commitments or the violation will have low or no impact of an insurance firm

Reputational impact:

- › **High:** Failure of the use case would likely have e.g. a mass media impact affecting insurance firm's reputation
- › **Medium:** Failure of the use case would likely have a specialized media impact (insurance, business, etc.)
- › **Low:** Failure of the use case would likely not affect insurance firm's reputation

Likelihood

Evaluation or scoring, including profiling and predicting, especially from “aspects concerning the data subject’s performance at work, economic situation, health, personal preferences or interests, reliability or behaviour, location or movements” (recitals 71 and 91 GDPR). Also when the processing in itself “prevents data subjects from exercising a right or using a service or a contract” (Article 22 and recital 91 GDPR)

- › High: AI use case performs risk scoring, fraud management, usage of external databases or consumer behaviour (web navigation, social networks, etc.) for consumer profiling
- › Medium: AI use case provides recommendations based on consumer preferences and with consumer consent
- › Low: No evaluation or scoring

Automated-decision making with legal or similar significant effect: processing that aims at taking decisions on data subjects producing “legal effects concerning the natural person” or which “similarly significantly affects the natural person” (Article 35(3)(a) GDPR)

- › High: Machine is autonomous in the making decision process taking decisions on behalf of the insurance firm
- › Medium: Machine is providing recommendations to humans for decision-making
- › Low: No involvement in decision-making

Systematic monitoring: processing used to observe, monitor or control data subjects, including data collected through networks or “a systematic monitoring of a publicly accessible area” (Article 35(3) (c)) 15 GDPR).

- › High: Data is collected without consumer awareness
- › Medium: Data is collected under consumer consent (i.e. connected car)
- › Low: No systematic monitoring

Model complexity/combining datasets- The rationale is that high complex models are more difficult/impossible to be exhaustively tested and hence the likelihood to find unexpected situations in production is higher (causing the use case to fail and cause the undesired harm)

- › High: Complex models such as those involving multi-factor interactions or non-determinate prediction

of a future reality or using a high number of variables (i.e. high dimensionality)

- › Medium:
- › Low: Low complex models such as those involving a simple deterministic depiction of reality or a simple probabilistic appraisal using small amounts of variable (i.e. low dimensionality)

Innovative use or applying new technological or organisational solution. The rationale is that new logic cannot be compared to precedent performance and the likelihood to find unexpected situations is higher until the use case reach maturity (and same as above higher potential of failure and harm)

- › High: Machine is implementing a new process/actions (e.g. new products, new advisory, etc.)
- › Medium: Machine is implementing current processes with new business rules (e.g. a new way of assessing risk score)
- › Low: no new processes/new rules are involved (e.g. automating existing processes)

Type and amount of data used. The rationale is twofold, high sensitive data has more likelihood to cause more harm and new data has higher potential of failure with the same rationale of new processes (please note that data concerning vulnerable data subjects is already considered in the Severity section)

- › High: Special categories of data as defined in Article 9.1 GDPR, as well as personal data relating to criminal convictions or offences as defined in Article 10 GDPR, or data that has a high risk of being a proxy of these types of special categories of data (e.g. life style data, bank account and credit card data) or use of data that was never managed before in the organisation and is going to be processed massively or internal sensitive data (tax data, financial data, HR data, etc.)
- › Medium: Personal data other than the special categories defined in Articles 9.1 or 10 GDPR or use of data that though existing but was not processed or not processed massively
- › Low: Non-personal data or not in the above categories

Outsourcing of datasets and AI applications. The rationale is that quality of data from external data sources is more difficult to control and hence the likelihood of the use case failure to perform as expected is higher. Similar

considerations would apply concerning AI applications outsourced from third party services providers. Intellectual property considerations may represent a barrier.

- › High: Mainly external data sources where it is not possible to effectively ascertain how the data has been collected / manipulated / processed
- › Medium: A mix of internal and external data sources
- › Low: Only internal data sources allowing the insurance firm to ensure that data has been processed in a responsible manner across the data value chain.

ANNEX 3 – NON DISCRIMINATION REGULATORY FRAMEWORK IN INSURANCE

Non-discrimination means that consumers are not unduly disadvantaged as a consequence of carrying protected characteristics. More specifically the EU law distinguishes between direct and indirect discrimination.

Direct discrimination means different treatment in comparable situations on grounds of a protected characteristic.

Indirect discrimination means where an apparently neutral provision, criterion or practice would put persons with one value of a protected characteristic at a particular disadvantage compared with persons with a different value of the same characteristic, unless that provision, criterion or practice is objectively justified by a legitimate aim and the means of achieving that aim are appropriate and necessary.

The current list of protected characteristics in the EU is included on Article 21 EU Charter of Fundamental Rights and includes characteristics such as nationality, sex, racial or ethnic origin, religion or belief, disability, age and sexual orientation. For a state or public authority as actor, almost all differentiation based on protected characteristics is considered discriminatory. For private actors, discrimination is dealt with in a fragmented way. A more comprehensive equal treatment between persons irrespective of religion or belief, disability, age or sexual orientation has been drafted by the EU commission in 2008 but has not been resolved.

In insurance on the EU level, currently only disadvantaging due to nationality ([Art. 18 TFEU](#)) and due to sex is considered discrimination, compare [COUNCIL DIRECTIVE 2004/113/EC \(Gender Directive\)](#). Some member states have more comprehensive non-discrimination provisions. Moreover, equal treatment regarding racial or ethnic origin is required for the access to insurance in [Directive 2000/43/EC](#). This would be widened to religion or belief, disability, age and sexual orientation in the above-mentioned Commission proposal. Regulating the access to insurance could be viewed to limit the right of the insurance firm to not offer a contract if a product exists, unless other eligible provisions, criteria or practices, e.g.

detected fraud propensity, would justify denying the contract. Moreover, the price of that contract could be too high for the consumer to afford it. Apart from these non-discrimination requirements, the insurance firm is free to decline to offer the cover.

In insurance, some of the protected characteristics are obviously necessary for an adequate risk assessment and for estimating the expected production cost of the cover at inception. Pension risk or mortality risk cannot be assessed without taking age into account. Disability materially influences morbidity risk. Therefore, a comprehensive implementation of material non-discrimination provisions beyond access to insurance is not foreseen or discussed.

The Gender Directive contains a member state option to allow “proportionate differences in individual’s premium ... where the use of sex is a determining factor in the assessment of risk based on relevant and accurate actuarial and statistical data.”, Art. 2 Par.2.

However, the European Court of Justice found this provision invalid in the “Test-Achats” [case](#). Thereupon, EU Commission has issued its “[Guidelines on the application of Council Directive 2004/113/EC to insurance, in the light of the judgment of the Court of Justice of the European Union in Case C-236/09 \(Test-Achats\)](#)”. These contain detailed explanations what the Commission considers admissible. Particularly, marketing and advertisement to influence the gender mix of the pool of insured remains possible. Factors that correlate with gender and thus can cause indirect discrimination can be “objectively justified by a legitimate aim”. Appropriate estimation of insurance production cost qualifies as a legitimate aim. The requirement that “the means are appropriate and necessary” is translated into the notion of “true risk factors in their own right”.

ANNEX 4 – LIST OF ACRONYMS

AI	Artificial Intelligence
AI HLEG	European Commission's High Level Expert Group on AI
ANNs	Artificial Neural Networks
BDA	Big Data Analytics
CRM	Customer Relationship Management systems
DPIA	Data Protection Impact Assessment
DPO	Data Protection Officer
EIOPA	European Insurance and Occupational Pensions Authority
ESAs	European Supervisory Authorities (EBA, ESMA and EIOPA)
FNOL	First Notice of Loss
GDE	EIOPA Consultative Expert Group on Digital Ethics in insurance
GDPR	General Data Protection Regulation
GLM	Generalised Linear Models
GWP	Gross Written Premiums
IDD	Insurance Distribution Directive
IoT	Internet of Things
ML	Machine Learning
NLP	Natural Language Processing
OCR	Optical Character Recognition
ORSA	Own Risk and Solvency Assessment
RPA	Robotic Process Automation
SCR	Solvency Capital Requirement
UBI	Usage-based insurance, e.g. telematics tariffs in motor insurance

ANNEX 5 – MEMBERS OF EIOPA’S CONSULTATIVE EXPERT GROUP ON DIGITAL ETHICS IN INSURANCE

Name	
Andreas Hufenstuhl	PricewaterhouseCoopers
Antti Talonen	University of Helsinki
Chris Holland	Loughborough University
Chris K. Madsen	Aegon
Christian Hugo Hoffmann	Syntherion
Cristina Bellido Andújar	VidaCaixa
Daniel John	HUK-COBURG
David Wassong	Bleu Piment Consulting
Desislav Danov	FinTech Guardian
Edoardo Carlucci	Better Finance
Esko Kivisaari	Actuarial Association of Europe
Fernando Acevedo Frías	Independent consultant
Florian Pons	Institute des Actuaire
Gemma Garriga	Euler Hermes
Gert Meyers	Tilburg University
Jasper De Meyer	BEUC
Jaya Handa	Liberty Mutual
Jens-Daniel Florian	Marsh
Jimmi Prah	PFA Pension
João Torres Barreiro	Independent consultant
Lars Gatschke	Verbraucherzentrale Bundesverband e.V.
Liisa Halme	Trade Union Pro
Liz McFall	University of Edinburgh
Lutz Wilhelmy	Actuarial Association of Europe
Malika Larbi	Europ Assistance Holding
Marcello Zacchetti	Cattolica Assicurazioni

Name

Marcin Detyniecki	AXA
Martin Mullins	University of Limerick
Mirko Kraft	Coburg University of Applied Sciences and Arts
Olivier Jérusalmy	Financial Inclusion Europe
Owen Morris	Aviva
Paolo Stefano Giudici	University of Pavia
Pedro Écija Serrano	Grant Thornton
Petra Žárská	Spoločnosť Ochrany Spotrebiteľov S.O.S
Philippe Cotellet	Airbus
Piotr Czublun	CZUBLUN TRĘBICKI Law Firm
Raymon Badloe	Achmea
Rui Ferreira	Zurich
Thomas Brenøe	Insurance & Pension Denmark
Virginia Antonini	Reale Mutua

GLOSSARY

Artificial intelligence	<p>Artificial intelligence means software that is developed with one or more of the techniques and approaches listed in the following paragraph and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with.</p> <p>The approaches mentioned in the previous paragraph include the following: (a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning; (b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference/deductive engines, (symbolic) reasoning and expert systems; (c) Statistical approaches, Bayesian estimation, search and optimization methods.⁹³</p>
Artificial Neural Networks (NN) and Deep Learning Networks (DL)	<p>Algorithms that operate with an input layer, one or more unknown hidden layers, and an output layer. In what is known as a “feedforward network”, the information flows from the input layer, through the hidden layer into the output layer. NNs also learn by example and through experience (“backpropagation”).</p> <p>An important difference between NN (including DL) and other types ML algorithms is that the latter often use manual feature extraction, that is, human programmers determine which features the machine learning software should use in making its predictions. On the other hand, in NN the algorithm itself learns from the data which features are most useful in making predictions.</p>
Big Data Analytics (BDA)	<p>Large volumes of data that can be generated, processed and increasingly used by digital tools and information systems for making predictive, descriptive and prescriptive analysis. This capability is driven by the increased availability of structured data, the ability to process unstructured data, increased data storage capabilities and advances in computing power.</p>
Corporate social responsibility	<p>Corporate social responsibility is the responsibility of enterprises for their impact on society and, therefore, it should be company led. Companies can become socially responsible by integrating social, environmental, ethical, consumer, and human rights concerns into their business strategy and operations as well as by following the law⁹⁴</p>
Human oversight	<p>Human oversight is defined as some form of direct human involvement in the design, operation, maintenance, adaptation or application of AI systems.</p>

⁹³ Definition based on the Commission’s proposal for a Regulation 2021/0106 (COD) laying down harmonised rules for Artificial Intelligence

⁹⁴ Definition used by the European Commission: https://ec.europa.eu/growth/industry/sustainability/corporate-social-responsibility_en

Insurance firms	In this report the term insurance firms refers to both insurance undertakings and insurance intermediaries. The guidance provided in this report is addressed to both insurance undertakings and intermediaries when using AI in the respective areas of the insurance value chain where they are involved. Nevertheless, a distinction between insurance intermediaries and insurance undertakings is occasionally made when, due to the nature of a specific AI use case, insurance undertakings and intermediaries play a different role or are involved in their implementation in a different manner.
Internet of Things (IoT)	Is the networking of telematics devices, vehicles, buildings, and other items embedded with electronics, software, sensors, wearables actuators, and network connectivity that enable these objects to (a) collect and exchange data and (b) send, receive, and execute commands
IoT-based insurance products	Insurance products based on IoT sensor devices to measure consumer's behaviour and environment to perform risk assessments and price discount rewards. For instance, this would be the case of Pay-As-You-Drive (PAYD) and Pay-How-You-Drive (PHYD) products in motor insurance, or Pay-As-You-Live (PAYL) products in health insurance. Sometimes also referred in this report as "telematics tariffs".
Machine learning	Machine learning (ML) is the ability of computers to learn from data through appropriate algorithms. This allows them to build a model of their world and better solve their intended tasks. Approaches of ML can be characterized by the dimensions of the task (differentiating fundamentally between classifications, regression and clustering), the data types (special approaches exist for example for text, language and image data) and the algorithms (how is the problem solved technically).
Personal data	Personal data means any information relating to an identified or identifiable natural person; an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person ⁹⁵
Price optimisation	Adjustments to the risk-based / actuarial price to create the final price offered to consumers taking into account the market price as well as a number of different techniques such as price elasticity models, churn models or life time value estimation models which are largely independent of the risk profile of the consumer. Adjustments to the risk-based / actuarial price to take into account re-insurance costs and other acquisition/production costs (e.g. commissions paid to distribution channels, salary of staff etc.) are not considered as price optimisation practices in this report.
Rating factor	Any factor that is involved in the process of pricing of an insurance policy, and influences the premium paid by the consumer.
Transparency and explainability	In this report transparency is broadly understood as providing information about the use, the nature and/or design of an AI system and the data variables and parameters used. Explainability is part of the concept of transparency and concerns the ability to explain the output of the AI system to a particular audience, in particular the weight / influence and causal relationship of a specific variable (or group of variables) in the final output

⁹⁵ Article 4(i) GDPR

REFERENCES

- Actuarial Association of Europe (2021). Code of professional conduct, Common Principles of Ethical and Professional Behaviour accepted by the Actuarial Associations affiliated to the Actuarial Association of Europe. https://actuary.eu/wp-content/uploads/2017/10/Code_of_Conduct_22092017_FINAL.pdf
- AI HLEG (2019). Ethics Guidelines for Trustworthy AI. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- AI HLEG (2020). Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment. <https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment>
- Australian Competition and Consumer Commission (2018). Northern Australia Insurance Inquiry. <https://www.accc.gov.au/system/files/Northern%20Australia%20Insurance%20Inquiry%20-%20First%20interim%20report%202018.PDF>
- Article 29 Data Protection Working Party (WP 248 rev.01) (2017). Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is “likely to result in a high risk” for the purposes of Regulation 2016/679. https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236
- Article 29 Data Protection Working Party (wp259rev.01) (2016). Guidelines on Consent under Regulation 2016/679 https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=623051
- Baker, T. (2010). Health Insurance, Risk, and Responsibility After the Patient Protection and Affordable Care Act. *U. Pa. L. Rev.* **159**, 1577
- Bernardino, G. (2020). Challenges and opportunities for the insurance sector in Europe. In *Annales des Mines-Realites industrielles* (No. 1, pp. 99-102). FFE.
- Bussman, Giudici, Marinelli and Papenbrock, (2021). “Explainable AI in credit risk management”, *Computational Economics*, 57, (203-216)
- Carney, M. (2020), Reith Lecture series. <https://www.bbc.co.uk/programmes/mooopy&t>
- Capurro, R. (2017). Ethics and Robotics. <http://www.capurro.de/ethicsandrobotics.html>
- Central Bank of Ireland (2020). Differential Pricing: The Economics and International Evidence. Developed by Shane Byrne and Yvonne McCarthy. [https://www.centralbank.ie/docs/default-source/publications/financial-stability-notes/no-10-differential-pricing---the-economics-and-international-evidence-\(byrne-and-mccarthy\).pdf](https://www.centralbank.ie/docs/default-source/publications/financial-stability-notes/no-10-differential-pricing---the-economics-and-international-evidence-(byrne-and-mccarthy).pdf)
- CNIL (2017). How can humans keep the upper hand? https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf

Consumentenbond (2018). Home insurance premiums are rising due to the use of big data. <https://www.consumentenbond.nl/nieuws/2018/premies-woonhuisverzekeringen-stijgen-door-gebruik-big-data>

Crowson, C., Atkinson, E., Therneau, T., Lawson, A., Lee, D. and MacNab, Y. (2016). Assessing calibration of prognostic risk scores. *Statistical Methods in Medical Research*, 25(4):1692–1706

Dirección General de Seguros y Fondos de Pensiones (2020). Tablas indemnizatorias Baremo 2020. <http://www.dgsfp.mineco.es/es/Regulacion/DocumentosRegulacion/Tablas%202021.pdf>

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., Zemel, R. (2011). Fairness Through Awareness. <https://arxiv.org/pdf/1104.3913.pdf>

Dutch Association of insurers (2018). Solidarity Monitor, The 2nd measurement. https://www.verzekeraars.nl/media/5732/solidariteitsmonitor-2018-eeenmeting-versie-22-oktober-2018_eng.pdf

EDPB Guidelines 01/2020 (2021). Guidelines on processing of personal data in the context of connected vehicles and mobility related applications. https://edpb.europa.eu/system/files/2021-03/edpb_guidelines_202001_connected_vehicles_v2.o_adopted_en.pdf

EIOPA (2014). Report on good practices on comparison websites: https://register.eiopa.europa.eu/Publications/Reports/Report_on_Good_Practices_on_Comparison_Websites.pdf

EIOPA (2019). Big Data Analytics in motor and health insurance: a Thematic Review. https://register.eiopa.europa.eu/Publications/EIOPA_BigDataAnalytics_ThematicReview_April2019.pdf

EDPB TechDispatch (2020). Personal Information Management Systems. https://edps.europa.eu/sites/default/files/publication/21-01-06_techdispatch-pims_en_o.pdf

European Commission (2012/C 11/01) (2012). Guidelines on the application of Council Directive 2004/113/EC to insurance, in the light of the judgment of the Court of Justice of the European Union in Case C-236/09 (Test-Achats). <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2012:011:0001:0011:en:PDF>

European Commission Delegated Regulation (EU) 2015/35 of 10 October 2014 supplementing Directive 2009/138/EC of the European Parliament and of the Council on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II). https://eur-lex.europa.eu/eli/reg_del/2015/35/oj

European Commission (2018), Financial services provision and prevention of financial exclusion, <https://www.bristol.ac.uk/media-library/sites/geography/migrated/documents/pfrco807.pdf>

European Commission COM(2020) 591 (2020). Digital Finance Strategy for the EU. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0591&from=EN>

European Commission (COM/2020/595) (2020). Proposal for a Regulation on digital operational resilience for the financial sector. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0595>

European Commission COM/2020/66 (2020). A European strategy for data. <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1593073685620&uri=CELEX-%3A52020DC0066>

European Commission COM(2020) 65 (2020). White Paper on Artificial Intelligence – A European Approach to Excellence and Trust. https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

European Commission 2021/0106 (COD) (2021). Proposal for a Regulation on the harmonisation of rules on artificial intelligence (Artificial Intelligence Act). <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence-artificial-intelligence>

European Commission COM (2021) 205 (2021). Coordinated Plan on Artificial Intelligence 2021 Review, <https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review>

FCA (2016). Feedback statement on Call for Inputs on Big Data in retail general insurance. <https://www.fca.org.uk/publication/feedback/fs16-05.pdf>

FCA (2019). General insurance pricing practices. Interim Report. <https://www.fca.org.uk/publication/market-studies/ms18-1-2-interim-report.pdf>

FCA (2020). General insurance pricing Practices. Final Report. <https://www.fca.org.uk/publication/market-studies/ms18-1-3.pdf>

Fjed, J. Nagy A. (2020). Principled Artificial Intelligence. The Berkman Klein Centre for Internet & Society at Harvard University. <https://cyber.harvard.edu/publication/2020/principled-ai>

Frezal, S. and Barry, L. (2019). Fairness in uncertainty: Some limits and misinterpretations of actuarial fairness. *Journal of Business Ethics*, pp. 1-10

German Data Ethics Commission (2019). Opinion of the Data Ethics Commission. https://www.bmjv.de/DE/Themen/FokusThemen/Datenethikkommission/Datenethikkommission_EN_node.html

Hardt, M., Price, E. and Srebro, N. (2016). Equality of Opportunity in Supervised Learning. pages 1–22

Holland, C.P., Mullins, M. and Cunneen, M., (2021). Creating Ethics Guidelines for Artificial Intelligence (AI) and Big Data Analytics: The Case of the European Consumer Insurance Market. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3808207

Information Commission's Officer and the Alan Turing Institute (2020). Explaining decisions made with AI. <https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-artificial-intelligence-1-0.pdf>

International Actuarial Association (2019). International Standards of Actuarial Practice (ISAPs). <https://www.actuaries.org/iaa/IAA/Publications/ISAPs/IAA/Publications/05IS-APs.aspx?hkey=334b21a7-a3ac-4e0e-8294-3cbc755ab14a>

International Actuarial Association (2016). International Standard of Actuarial Practice 1A Governance of Models (ISAP 1A). https://www.actuaries.org/IAA/Documents/CTTEES_ASC/Final_ISAPs_Posted/ISAP_1A_Final_December2018_Web.pdf

Jeanningros, H. and McFall, L. (2020). The value of sharing: Branding and behaviour in a life and health insurance company. *Big Data @ Society* **7**, 205395172095035–15

Judgement of the European Court of Justice, 1 March 2011, Test-Achats, Case C-236/09

Keller, B. (2020). Promoting Responsible Artificial Intelligence in Insurance, Geneva Association. <https://www.genevaassociation.org/research-topics/new-technologies-and-data/promoting-responsible-artificial-intelligence-insurance>

KPMG (2021). Thriving in an AI World. <https://info.kpmg.us/news-perspectives/technology-innovation/thriving-in-an-ai-world/ai-adoption-accelerated-during-pandemic.html>

Lobschat, L., Mueller, B., Eggers, F., Brandimarte, L., Diefenbach, S., Kroschke, M. and Wirtz, J. (2019). Corporate digital responsibility. *Journal of Business Research*

Marchant, G. (2019) “Soft Law” Governance of Artificial Intelligence. <https://aipulse.org/soft-law-governance-of-artificial-intelligence/>

Malgieri, G. (2020). The concept of fairness in the GDPR: a linguistic and contextual interpretation. <https://dl.acm.org/doi/abs/10.1145/3351095.3372868>

Mckinsey (2020). The state of AI in 2020. <https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/global-survey-the-state-of-ai-in-2020>

Missouri Department of Insurance (2004). Insurance-Based Credit Scores: Impact on Minority and Low Income Populations in Missouri. <https://insurance.mo.gov/reports/credscore.pdf>

NAIC (2015). Price Optimization White Paper. Casualty Actuarial and Statistical (C) Task Force https://www.naic.org/documents/committees_c_catf_related_price_optimisation_white_paper.pdf

New York Department of Financial Services (2019). Insurance Circular Letter No. 1. https://www.dfs.ny.gov/industry_guidance/circular_letters/cl2019_01

OECD (2019). Challengers to Consumer Policy in the Digital Age. <https://www.oecd.org/sti/consumer/challenges-to-consumer-policy-in-the-digital-age.pdf>

O’Riordan, T. and Cameron, J. eds., (1994). Interpreting the precautionary principle (Vol. 2). Earthscan.

Personal Finance Research Centre (2020). The Poverty Premium: A Customer Perspective. <https://fairbydesign.com/wp-content/uploads/2020/11/The-poverty-premium-A-Customer-Perspective-Report.pdf>

ProPublica (2016). How We Analyzed the COMPAS Recidivism Algorithm. <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>

Rawls, J. (1999). A Theory of Justice. Harvard University Press. p. 118.

Rebert, L., & Van Hoyweghen, I. (2015). The right to underwrite gender. The Goods & Services Directive and the politics of insurance pricing. Tijdschrift voor Genderstudies, 18(4), 413-431.

Ruf, B. and Detyniecki, M. (2021). Towards the Right Kind of Fairness in AI. <https://arxiv.org/abs/2102.08453>

Ruf, B., Hirot, M., Detyniecki, M., Shire, N., Scharrer, R. (2019). "Regulating Machine Learning: where do we stand? AXA Rev. https://axa-rev-research.github.io/static/AXA_WhitePaper_RegulatingML.pdf

Springer, Giudici and Raffinetti (2021). "Shapley-Lorenz explainable artificial intelligence", Expert systems with applications, 167, Elsevier.

TECHNGI research group, Loughborough University. www.techngi.uk

UK Government and the Association of British Insurers (2018). The Code on Genetic Testing and Insurance. <https://www.abi.org.uk/globalassets/files/publications/public/genetics/code-on-genetic-testing-and-insurance-final.pdf>

Van den Berg, M. and Kuiper, O. (2020). XAI in the Financial Sector A Conceptual Framework for Explainable AI (XAI), Hogeschool Utrecht, Lectoraat Artificial Intelligence <https://www.internationalhu.com/research/projects/explainable-ai-in-the-financial-sector>

Getting in touch with the EU

In person

All over the European Union there are hundreds of Europe Direct information centres.

You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union.

You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696 or
- by email via: https://europa.eu/european-union/contact_en

Finding information about the EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications at:

<https://publications.europa.eu/en/publications>. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1952 in all the official language versions, go to EUR-Lex at: <http://eur-lex.europa.eu>

Open data from the EU

The EU Open Data Portal (<http://data.europa.eu/euodp/en>) provides access to datasets from the EU. Data can be downloaded and reused for free, for both commercial and non-commercial purposes.

**EUROPEAN INSURANCE AND
OCCUPATIONAL PENSIONS AUTHORITY**

Westhafenplatz 1,
60327 Frankfurt am Main, Germany



Publications Office
of the European Union