EIOPA’s work on Big Data Analytics and Digital Ethics

19/05/2020
Institute des Actuaires

1

EIOPA's approach to InsurTech

EIOPA’s approach to InsurTech

Benefits

Risks

„Strike a balance between enhancing financial innovation and ensuring a well-functioning consumer protection framework and financial stability“

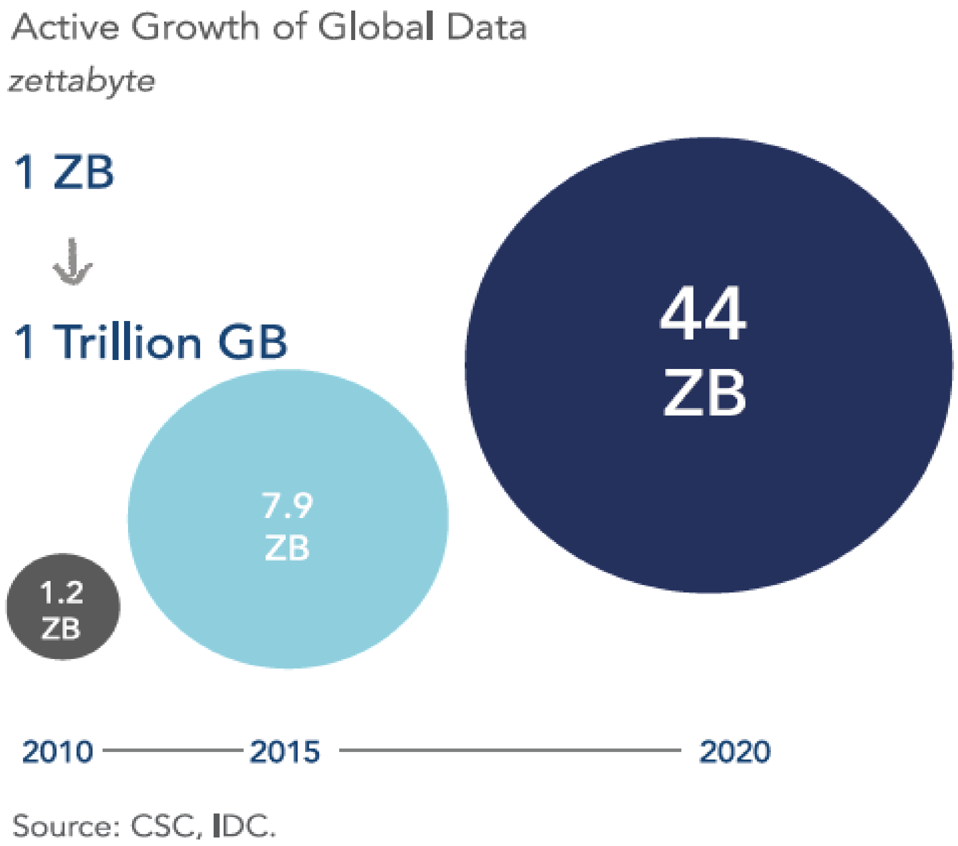
What do we want to achieve?

- **Protection of policyholders and beneficiaries and financial stability**
- **Promotion of sound financial innovation**
  - level playing field / technological-neutrality
  - understand shifting risks of new technologies and business models
  - strengthen supervision and cooperation between NCAs
- **Multistakeholder approach:** InsurTech Roundtables, Group on Digital Ethics
- **Multidisciplinary approach:** EIOPA InsurTech Task Force

2

Big Data Analytics in motor and health insurance: a thematic review

Increasing availability of data



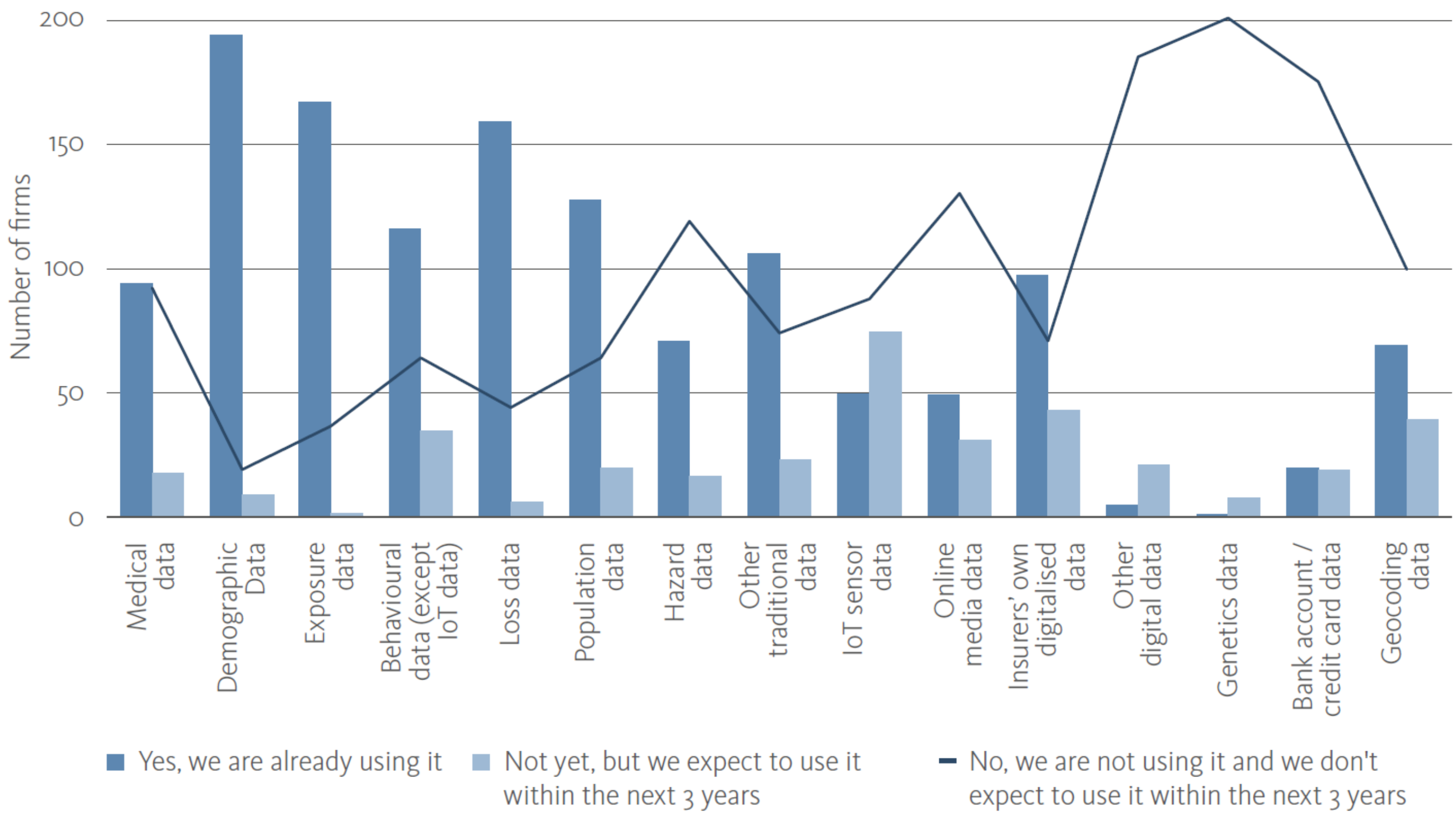
Source: Institute of International Finance

Types of data used by insurance firms (I)

Traditional data sources	New data sources enabled by digitalisation
Medical data (e.g. medical history, medical condition, condition of family members)	**IoT data** (e.g. driving behaviour (car telematics), physical activity and medical condition (wearables).
Demographic data (e.g. age, gender, civil and family status, profession, address)	**Online media data** (e.g. web searches, online purchases, social media activities, job career information)
Exposure data (e.g. type of car, value of contents inside the car)	**Insurance firms' own digital data** (e.g. interaction with insurance firms (call centre data, users' digital account information, digital claim reports, online behaviour while logging in to insurance firms' websites or using insurance firms' app)
Behavioural data (except IoT data) (e.g. Smoking, drinking behaviour, distance driven in a year)	**Geocoding data** (i.e. latitude and longitude coordinates of a physical address)
Loss data (e.g. claim reports from car accidents, liability cases)	**Genetics data** (e.g. results of predictive analysis of a person's genes and chromosomes)
Population data (e.g. mortality rates, morbidity rates, car accidents)	**Bank account / credit card data** (e.g. consumer's shopping habits, income and wealth data)
Hazard data (e.g. frequency and severity of natural hazards)	**Other digital data** (e.g. selfie to estimate biological age of the consumer)
Other traditional data (e.g. credit scoring, claim adjustment reports, information from the auto repair shops)	

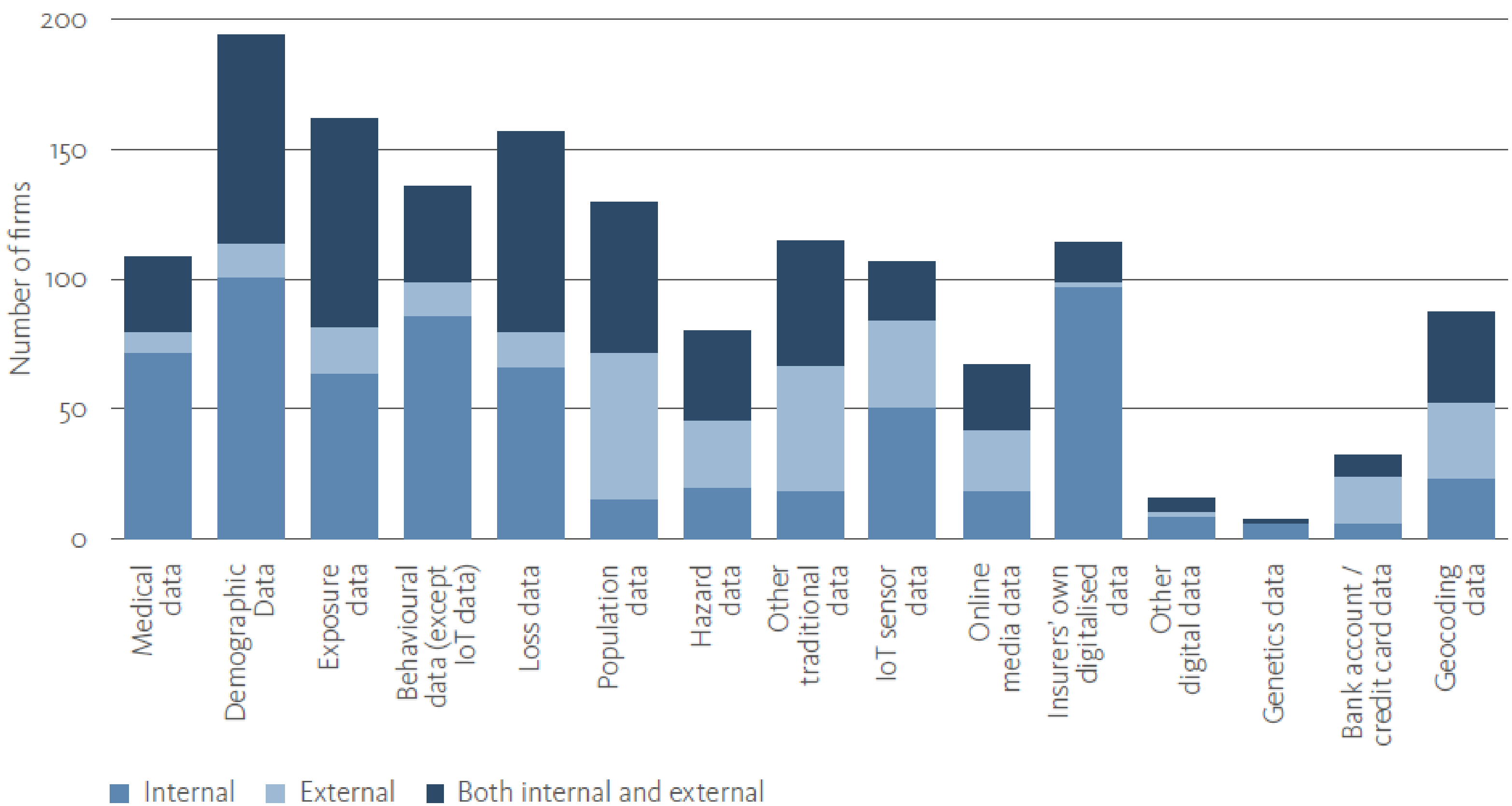
Source: The Geneva Association (the categorisation of types of data was slightly amended by EIOPA)[4]

Types of data used by insurance firms (II)



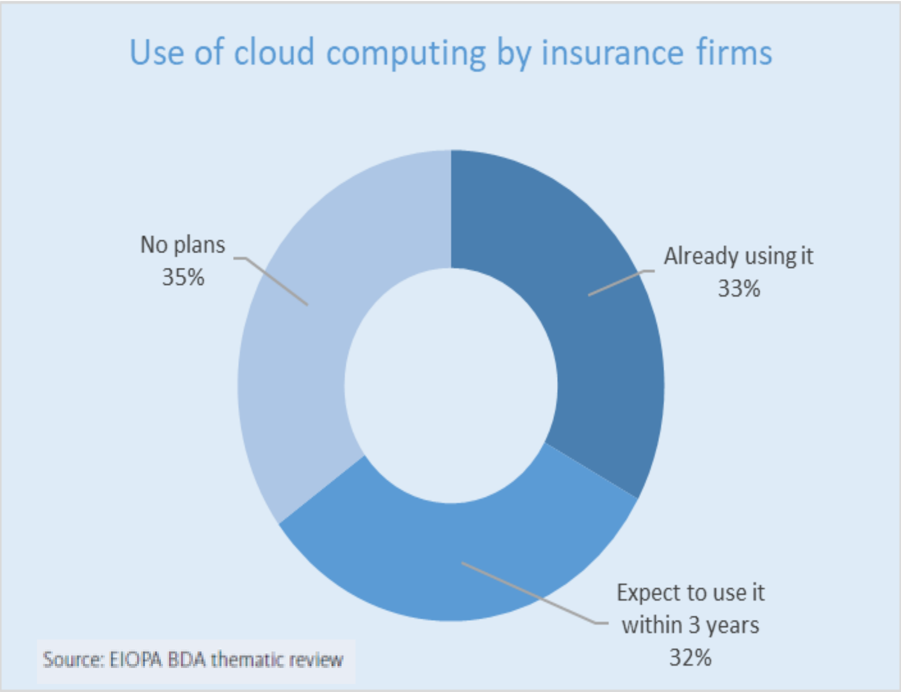
Source: EIOPA BDA thematic review

Data sources



Source: EIOPA BDA thematic review

BDA tools: cloud computing

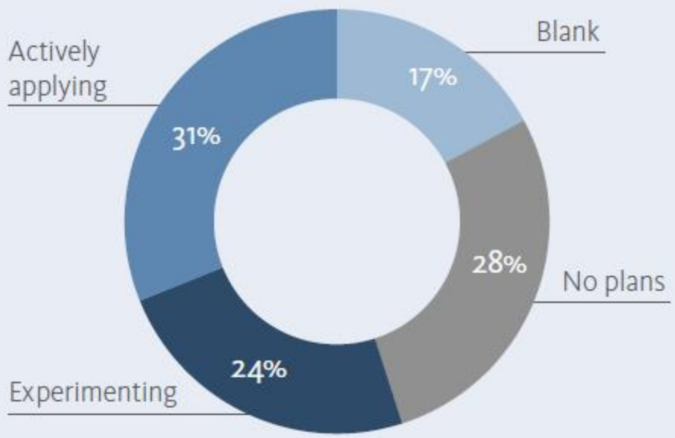


Service model	Services offered to clients
IaaS – Infrastructure as a Service	IT resources such as computing power, data **storage devices** or networks are offered as a service.
PaaS – Platform as a Service	Next to the infrastructure the provider offers on the platform **standardised interfaces**, which are used by services of the customer.
SaaS – Software as a Service	Wide range of services / applications offered (e.g**.** **machine learning**, contact data management, financial accounting, collaboration applications).

- Cloud computing is perceived as a **key enabler** of agility and data analytics

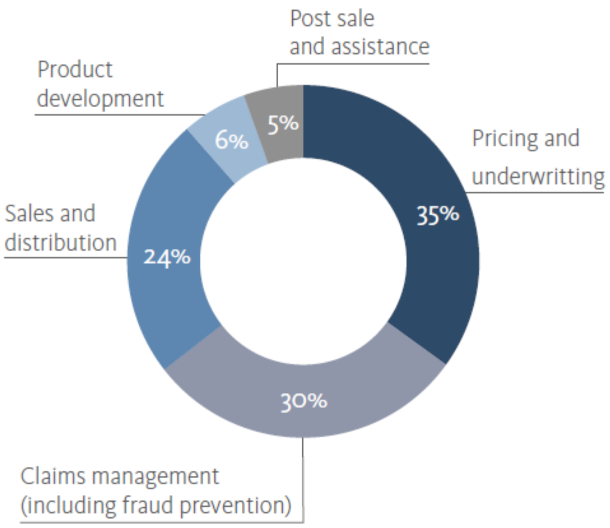
BDA tools: AI and ML

Figure 7 – Usage of BDA tools such as AI and ML



Source: EIOPA BDA thematic review

Figure 8 – Usage of BDA tools such as ML and AI across the value chain



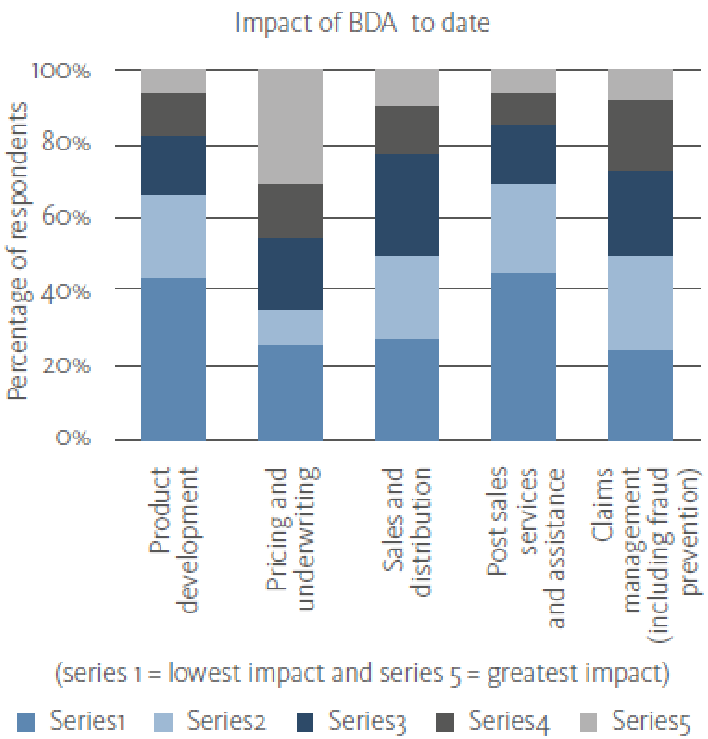
Source: EIOPA BDA thematic review

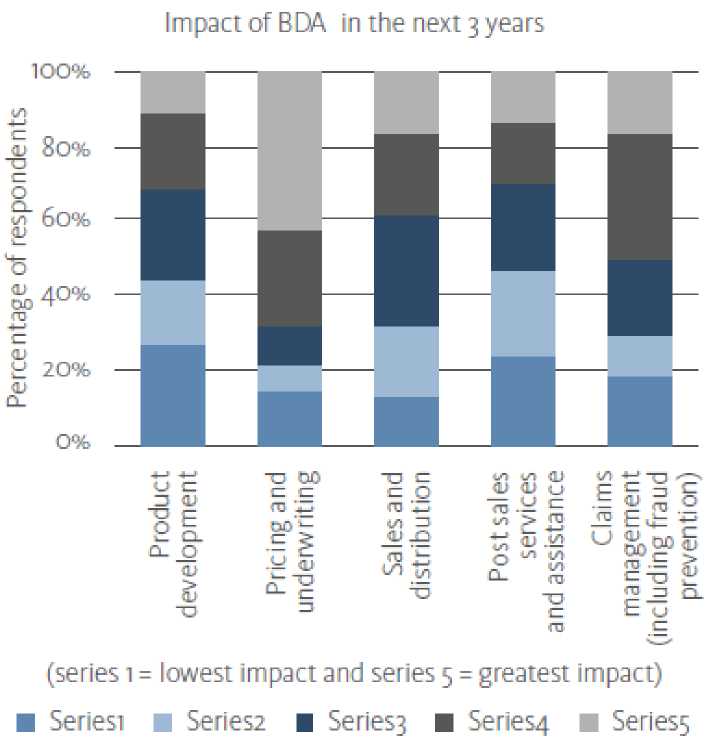
AI / ML use cases in insurance

Use Case	Output
Churn models	Use of ML churn models for the prediction of consumer's propensity to shop around at the renewal stage, which can be useful for pricing and underwriting (e.g. for price optimisation in combinaiton with a demand price-elasticity analysis) or for servicing the customer (e.g. "Next Best Action" approach)
Chatbot	Enable "human like" conversations with consumers by analysing customer unstructured data via text or voice with the use of natural language processing and other ML algorithms
Sentiment Analysis	Evaluate the sentiment in feedback provided by consumers to transform it into usable information to help improve customer satisfaction and engagement
Electronic document management	Robotic process automation (RPA) – Deep learning networks used for automatic classification of incoming documents of unstructured data (e.g. emails, claims statements), routing them to the correct department
Claims management	Optical character recognition (OCR) - Deep learning networks used to extract information from scanned documents such as images from damaged cars to estimate repair costs
Fraud prevention	Analysis of fraudulent claim patterns based on FNOL data provided by the consumer
Product development	Use of ML and graph database in predictive modeling for the identificaiton of disease development patterns
Pricing and underwriting	BDA tools used in motor and health insurance for processing large quantities of data from different sources, often on a real-time basis (e.g. quote manipulation), using a wide array of statistical techniques

Source: EIOPA BDA thematic review

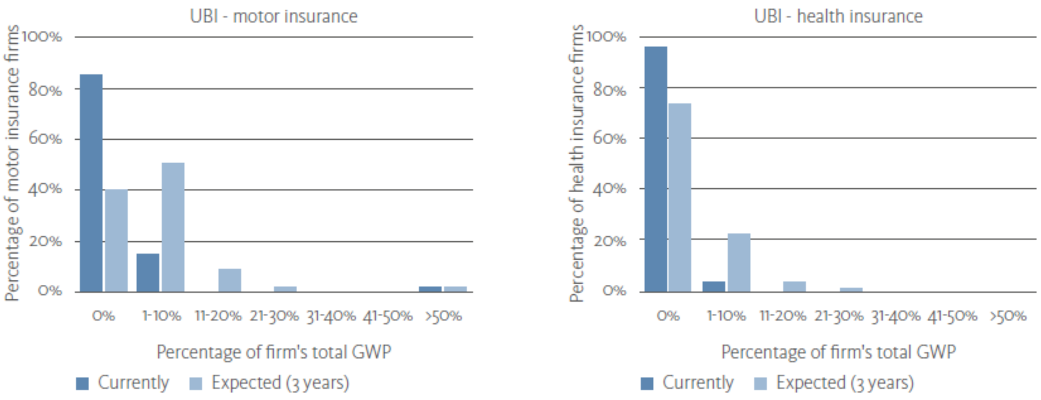
Impact on the value chain





Usage-based insurance policies

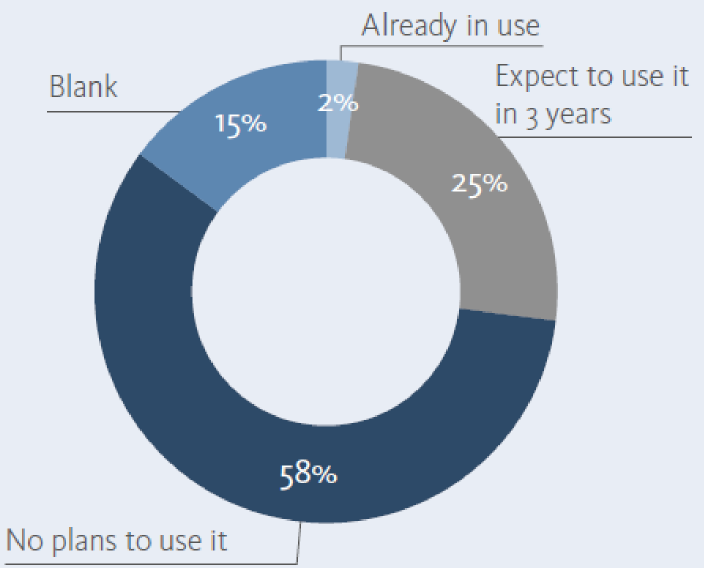
PENETRATION OF UBI PRODUCTS IN MOTOR AND HEALTH INSURANCE



Line of business	Type of telematics device	Type of data collected (depends on the telematics device)	Types of services offered (depends on the telematics device)
Motor insurance	On board device (OBD) dongle or "black box", mobile phone app, GPS, emergency message plug, forward facing cameras ("dash cams")	Average speed, maximum speed, acceleration and braking habits (G-forces), geolocation, distance travelled, time of travel (e.g. day or night), number of journeys, crash reports, battery and engine condition, cornering, lane changes	**Risk mitigation and prevention:** premium discounts based on driving habits, preventive push-notifications or alerts (e.g. black-spot roads or bad weather conditions or battery and engine breakdown problems), travel statistics reports, driving coach recommendations, treats and vouchers for good driving behaviour **Assistance:** road assistance in case of accident or car theft, emergency call in case of accident (ecall)
Health insurance	Wearable bracelets and other fitness trackers, mobile phone app, smart watch	Heart beat rate, blood pressure, blood oxygen level, activity data (e.g. sports or step counter), hours of sleep, geolocation, food and water consumption, calorie consumption, glucose level.	**Risk mitigation and prevention:** rewards for healthy habits, health activity reports, diabetes management **Assistance:** medical assistance services in case of accident, safety alarm for elderly (e.g. BDA tools can predict falls from anomalies in usage/activity patterns)

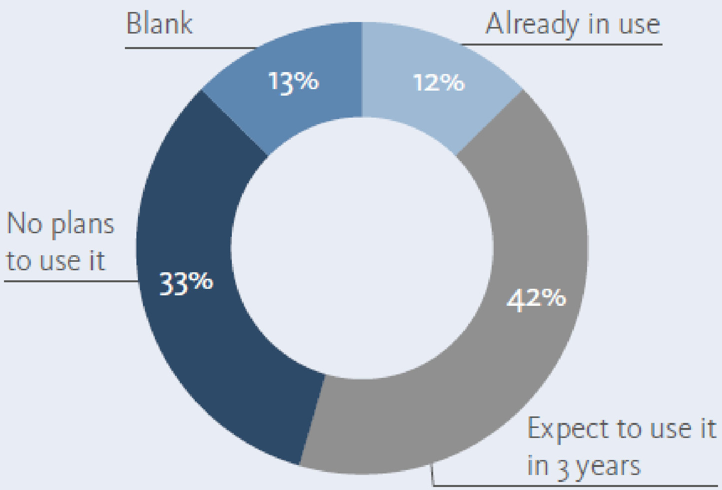
Robo-advisors and Chatbots

Figure 14 – Robo-advisors penetration in insurance



Source: EIOPA BDA thematic review

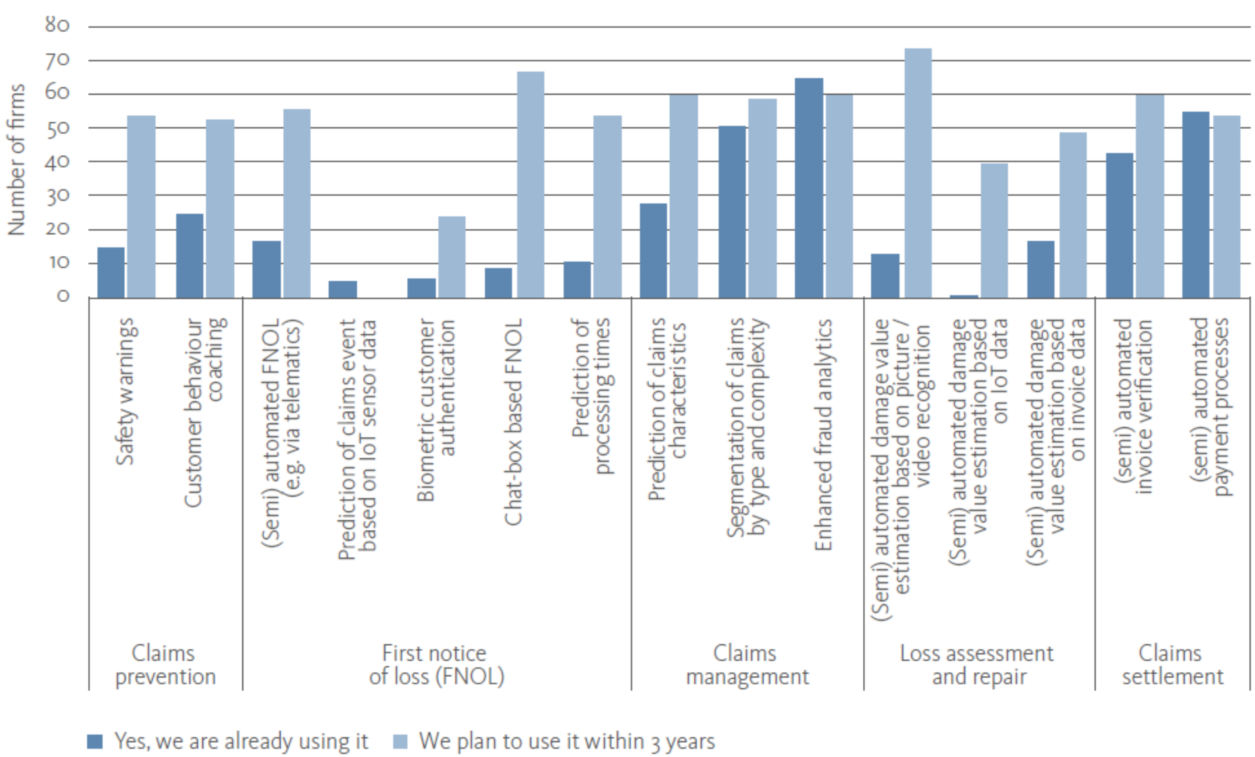
Figure 15 – Chatbots penetration in insurance



Source: EIOPA BDA thematic review

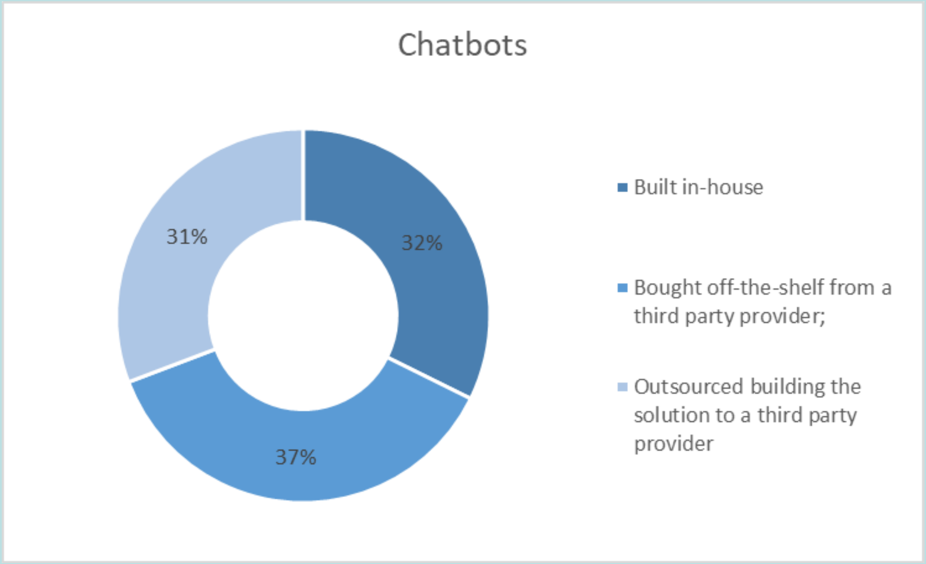
Claims management

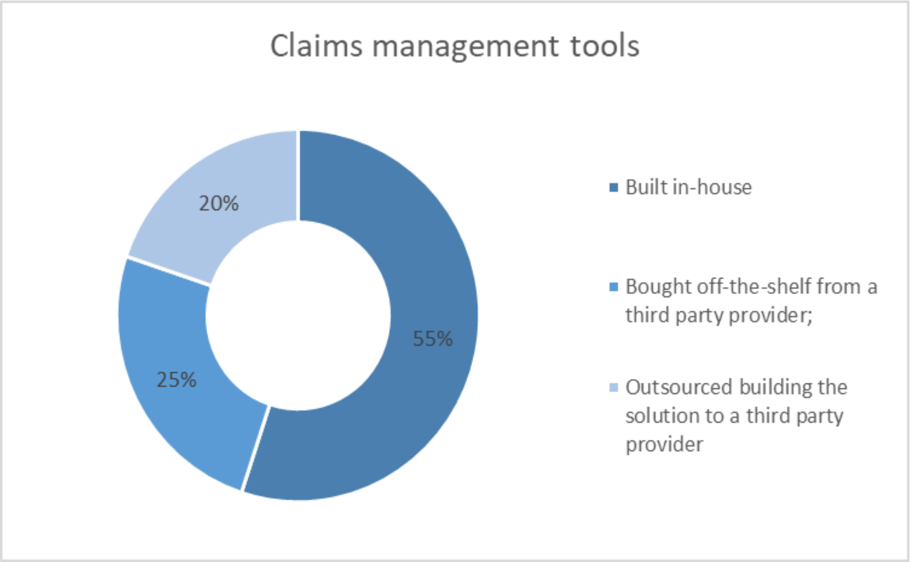
Figure 17 – Use of BDA in claims management



Source: EIOPA BDA thematic review, based on the classification of claims management processes from McKinsey&Company [14]

Third-party service providers



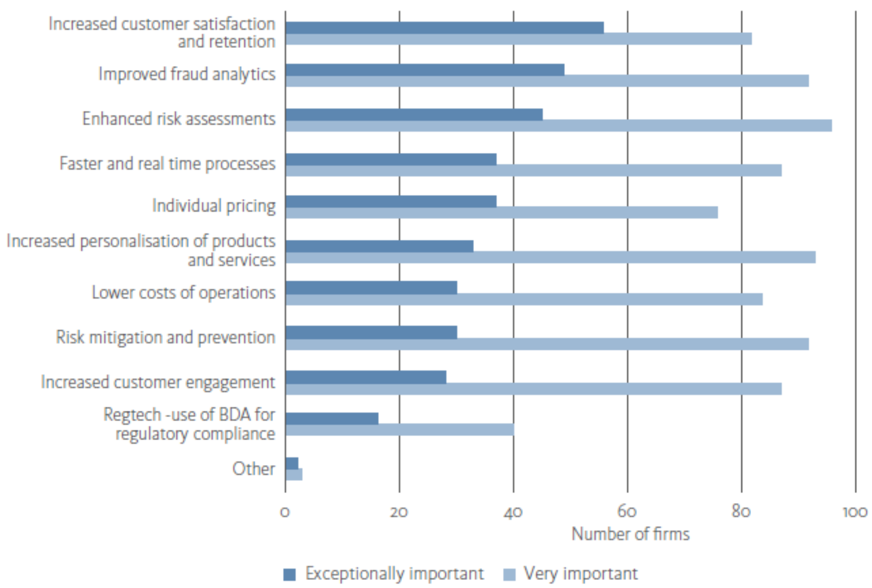


Motor insurance rating factors

Motor insurance rating factors categories	Examples of rating factor included in this category	Influence on final premium (approximation)*	Type of information provided
Driver details	Age of driver, mileage, car usage	High	Perceived as having a direct causal link
Vehicle details	Horsepower, car model, car value	High	
Claims and experience	Bonus malus, year of obtaining the driving license	High	
Cover	Type of cover, deductibles	High	
Driver behaviour	Driving score, acceleration, telematics data	High	
Loyalty	Multi-subscription, renewal, tenure with company	Low	Perceived as having an indirect link to risk behaviours – more likely elasticity
Location	Postal code, region, area of residence	High	
Affluence	Credit score, kind of home ownership, occupation	Low	
Distribution	Sales or distribution channel	Low	
Non-risk (not captured)	E-mail address,** customer marketing opt-out preference, quote manipulation	Low	
Other	Miscellaneous	Low	Not clear, excluded

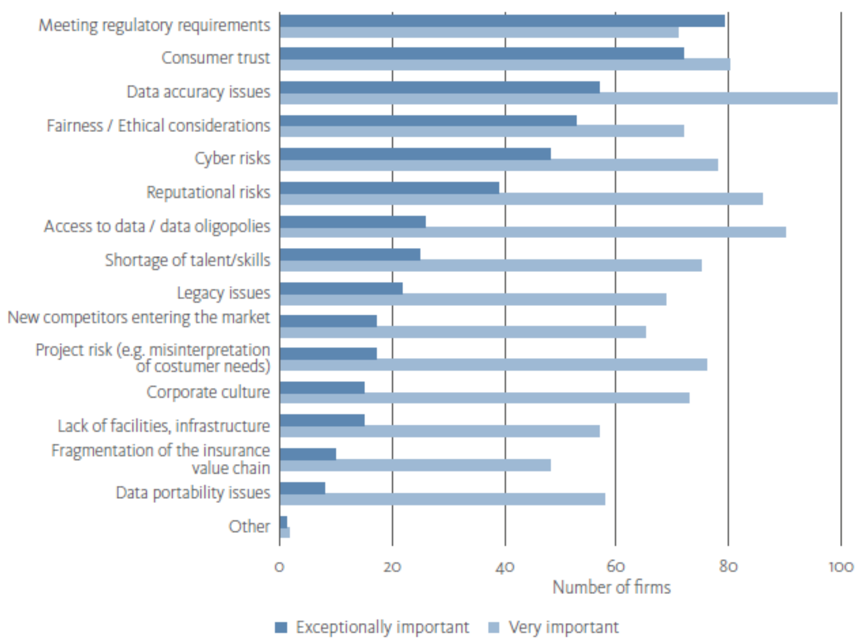
BDA opportunities and risks

Figure 32 – BDA opportunities according to insurance firms

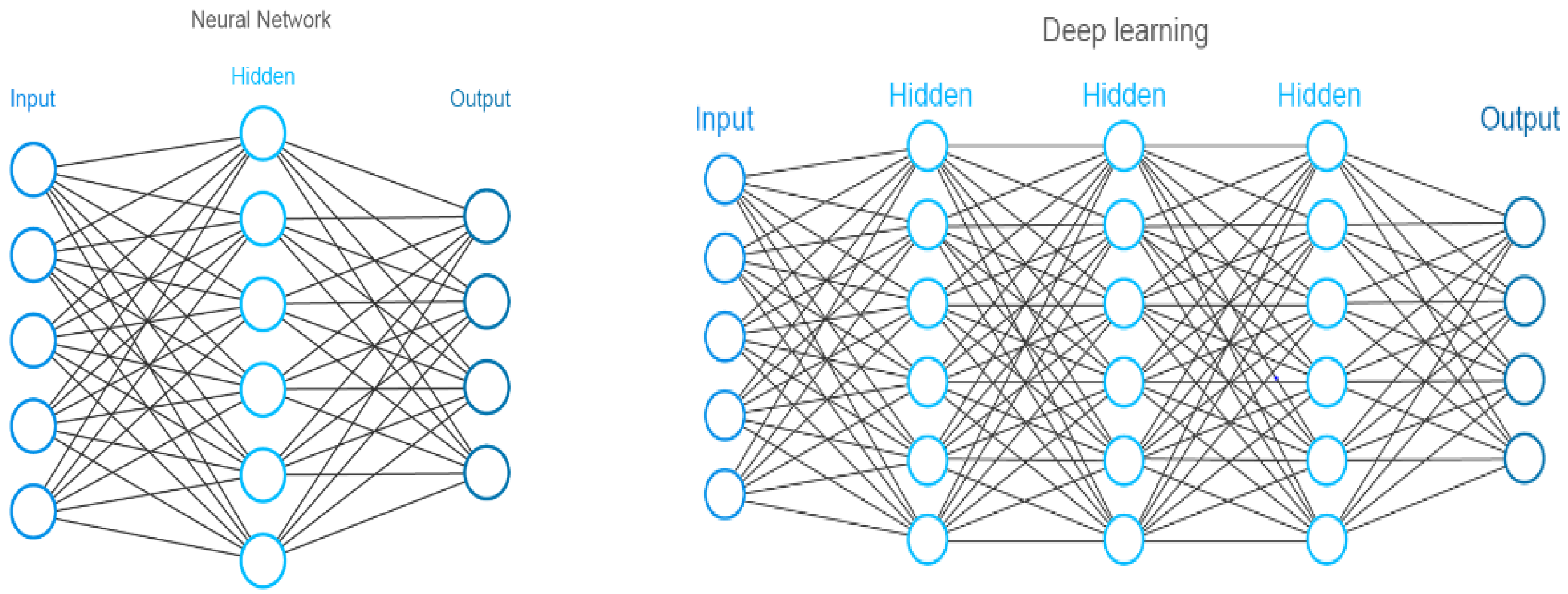


Source: EIOPA BDA thematic review

Figure 33 – BDA challenges according to insurance firms



Machine learning: Artificial Neural Networks



- Issues around accuracy, transparency, auditability and explainability of **(black-box) AI / ML algorithms**

Fair / ethical use of data

	Lesser desire to act	Greater desire to act
Who is harmed by price discrimination?	Wealthier consumers - eg time poor, cash rich	Consumers with characteristics which might be deemed vulnerable (eg low income, old ag, etc.)
How much are these individuals harmed?	Profitability difference between consumer segments is minimal and is immaterial to the harmed segment	Significant profitability differences and the harm has a significant adverse effect on the segment affected
How significant is the pool of people harmed?	Very small minority	Significant group of consumers
How are firms price discriminating?	Transparent and based on behaviour which consumers can easily change (eg switching)	Hidden and based on intrinsic characteristics which consumers cannot easily change (eg personal characteristics)
Is the product/service essential?	Product/service is considered non-essential but desired by some consumers	Essential product/service (eg current account or motor insurance)
Does society view the price discrimination as egregious/socially unfair?	Little concern expressed about practices and firm behaviour widely accepted	Persistent and broad-based concern expressed and firm behaviour seen as poor conduct

Source: UK's Financial Conduct Authority††

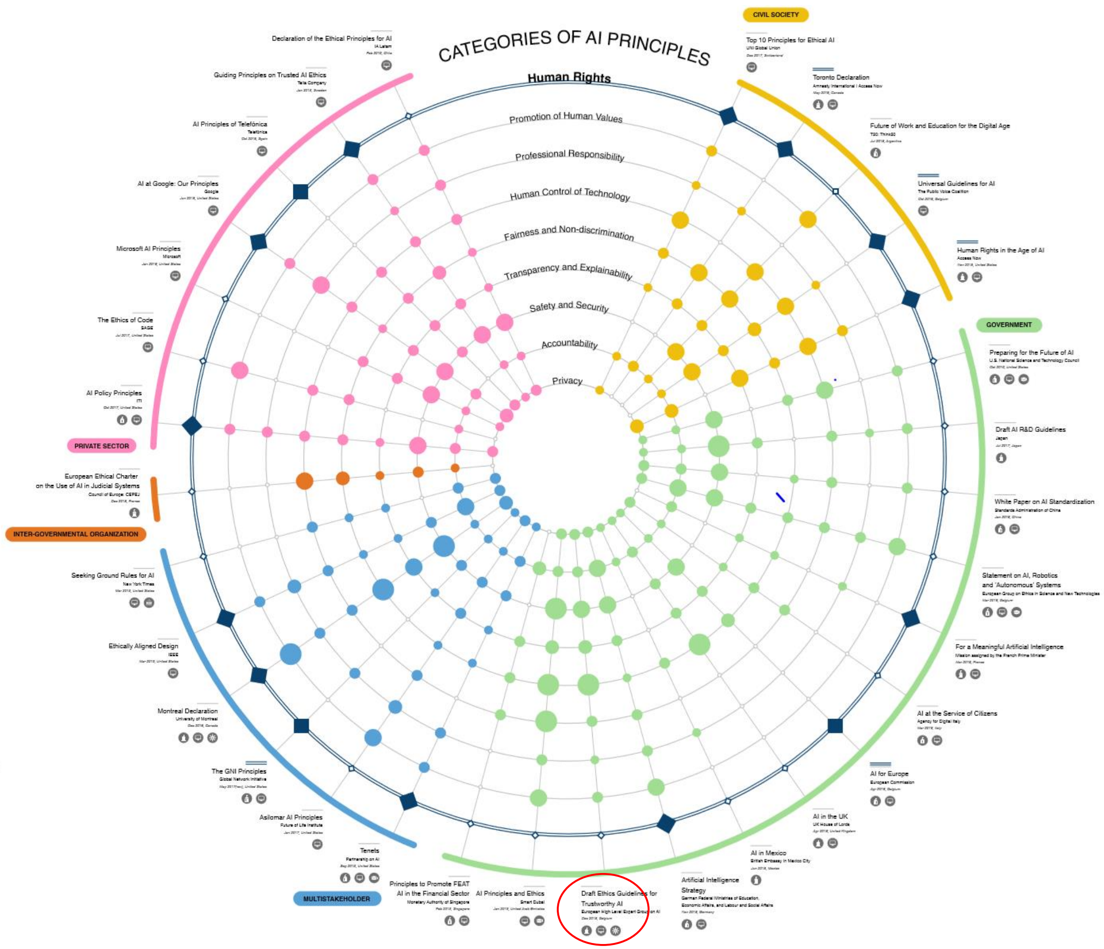
3

Digital ethics in insurance

Consultative Expert Group on Digital Ethics in insurance

- About the Group
  - Created in October 2019
  - 40 stakeholders from the insurance industry, consumers, and academics
- Objective:
  - Enhance trust in the use of new business model, data sources and technologies in insurance by defining a set of (non-binding) principles of digital responsibility in insurance
- Scope:
  - Specific to the insurance sector
  - Address new technologies, data sources and business models
  - Focus on pricing and underwriting, but also other areas of the value chain
  - Retail consumers prioritised
- Timeline
  - Work is expected to be finalised in Q3/Q4 2020

Existing public and private initiatives on AI principles



The European Commission's HLEG AI Ethical Guidelines (June 2019) defined 7 requirements for ethical and trustworthy AI:

I. Human Agency and Oversight
II. Technical robustness and safety
III. Privacy and Data Governance
IV. Transparency
V. Diversity, non-discrimination and fairness
VI. Societal and environmental well-being
VII. Accountability

Source: Harvard Law School: https://clinic.cyber.harvard.edu/2019/06/07/introducing-the-principled-artificial-intelligence-project/

Principles of digital responsibility in insurance

- WS1: Fairness and non-discrimination
  i. Diversity, non-discrimination and fairness
  ii. Societal and environmental well-being

- WS2: Transparency and explainability
  i. Transparency
  ii. Accountability

- W3: Governance
  i. Human agency and oversight
  ii. Technical robustness and safety
  iii. Privacy and data governance

Assess how HLEG requirements can be **applied in the insurance sector** in a proportionate manner and in concrete use cases across the insurance value chain

Adapted to concrete AI use cases

Take into account opportunities and challenges

Type	Category – task	Explainability (work in progress)	Accuracy (work in progress)
Naïve Bayes Algorithm	Supervised Learning - Classification	High	Low
Linear Regression	Supervised Learning - Regression	High	Low
Logistic regression	Supervised learning – Classification	High	Low
Support Vector Machine	Supervised Learning - Classification	Low	High
Decision Trees	Supervised Learning – Classification and Regression	High	Low
Random Forest	Supervised Learning – Classification and Regression	Low	High
Gradient Boosting trees	Supervised Learning – Classification and Regression	Low	High
Artificial Neural Networks (NN) and Deep Learning Networks (DL)	Supervised Learning - Reinforcement Learning – classification and regression	Low	High
K Means Clustering	Unsupervised Learning - Clustering	Low	High

Some issues being discussed in the GDE

- What does **fairness in pricing** mean in an insurance and AI context?
- What type of **explanations need to be provided to the different stakeholders**?
- Should insurance undertakings using AI develop and internal **AI strategy / guidelines**?
- What should be the **role of actuaries**?

Thank you

Julian Arevalo Carreno
Senior expert on financial innovation, EIOPA
Julian.arevalo@eiopa.europa.eu