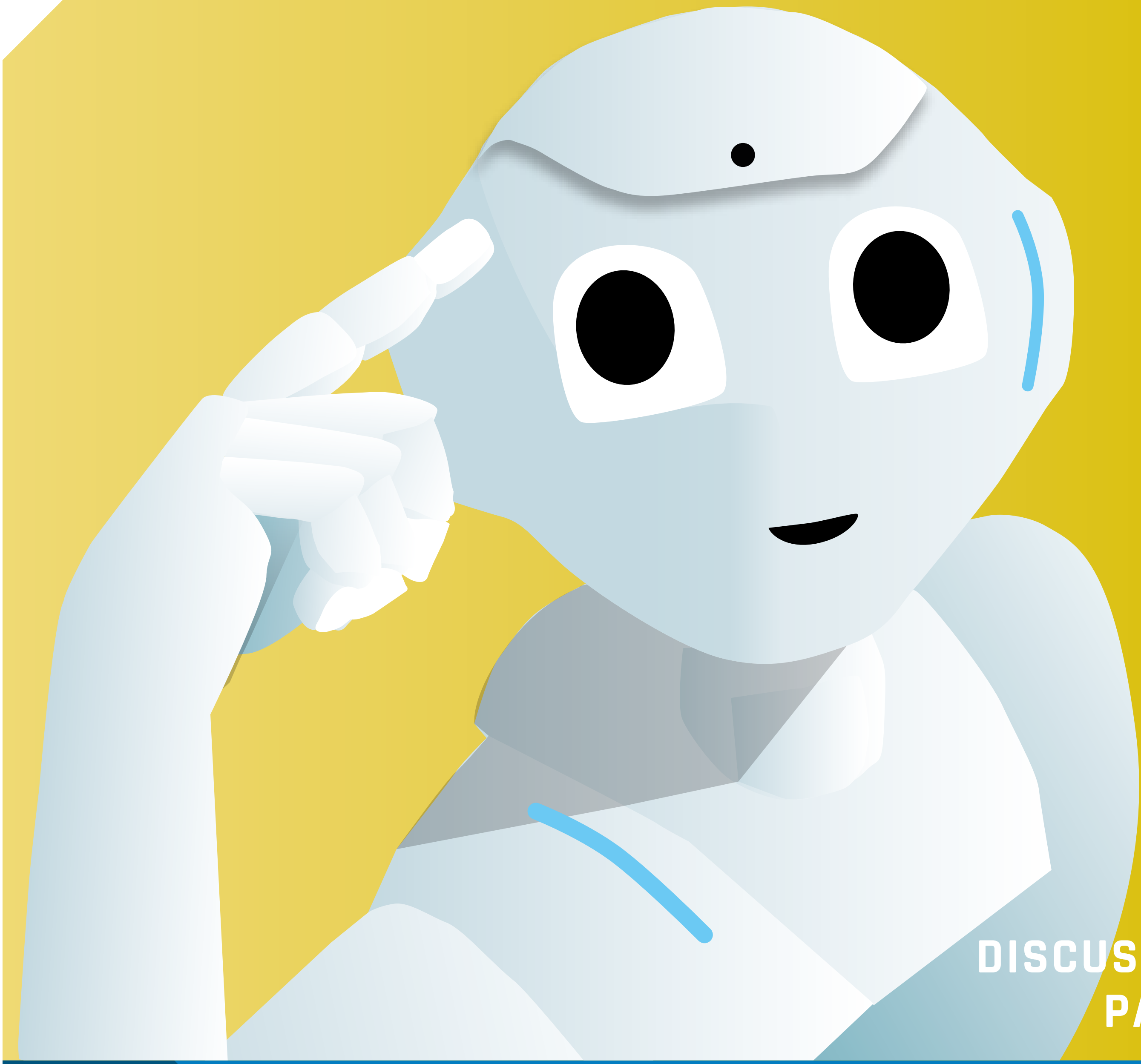


AAE
DISCUSSION
PAPER

WHAT SHOULD AN ACTUARY KNOW ABOUT ARTIFICIAL INTELLIGENCE?

JANUARY 2024

aae
actuarial association of europe

JANUARY 2024

This paper is a crucial guide for actuaries to navigate the evolving landscape of AI in their field. It highlights the role of AI in refining data analysis and introduces new methodologies for tackling actuarial challenges. Understanding these concepts helps actuaries maintain a competitive edge and adapt to changing industry norms.
The paper also sheds light on the ethical considerations and regulatory frameworks surrounding AI, crucial for responsible practice. For actuaries aiming to blend traditional skills with modern technology, this paper is an invaluable resource for staying relevant and innovative.
(*generated by ChatGPT*)

CONTRIBUTORS

JONAS HIRZ
ESKO KIVISAARI
PHILIPP MIEHE
CLAUDIO SENATORE
BOGDAN TAUTAN
FRANCESCO TORALDO

This paper is a discussion paper of the Actuarial Association of Europe. Any views expressed in this paper are views to stimulate and inform further discussion and should not be read as being representative either of the author's or contributors' individual opinions or of their employers or professional organisations or an agreed position of the Actuarial Association of Europe.

CONTENTS

1 PREFACE

Actuaries are living through the birth of large-scale adoption of Large Language Models (LLMs) like ChatGPT. Next to enormous promises, this technology carries risks and threats. It is yet to be seen where the journey with AI will exactly go and how fast societal adaption will push this topic forward.

By nature this paper is educational, i.e., it does not present any formal opinion that would have been approved by the AAE on the subject at hand. As the subject of the paper is under rapid development the AAE, while not seeking formal comments, will be interested in comments on this paper.

In the insurance industry, we see four key developments happening at the moment:

1. Existing risks and customers change due to AI (e.g., fraudulent claims through AI-generated pictures) potentially requiring adjustments of coverages, pricing, or risk sharing models.

2. New AI-related risks arise (e.g., claims caused by AI malfunction, or new operational risks) that require new products or approaches.

3. Competitive pressure increases as AI pushes efficiency across the full value chain, potentially forcing companies into applications of AI.

4. Regulators put more focus on AI which will require upskilling and organisational changes.

While the examples above concern insurance, we believe actuaries working not only in insurance but also more generally in business, pensions, social security and other areas need to familiarise themselves with novel tools to understand how they can revolutionise practices in the actuarial domain. Actuaries need to understand how they can, based on their long experience with complex models, bring their expertise to this novel area, and also warn of the possible dangers with models that might be used when assumptions they are based on are not valid or when the tools are used in areas where their validity has not been tested.

In addition to LLMs there are additional more general phenomena in the areas of Artificial Intelligence, and still more generally, in Data Science:

Data availability

- Data explosion or Big Data: Unprecedented amount of data in different areas in a variety of forms together with the data processing capacity needed to process this data.

- Pervasive digitalisation through continuously connected sensors everywhere: Real-time information on what is happening around (visibility on customers, business activities and market trends).

- Growing ecosystems across industries will lead to an increase in open-source data and data exchange, connecting e.g., insurers with banks, big techs, or other consumer brands.

Technology boost

- Enabling technologies with exponentially increasing efficiency: Processing power, storage, robotics ready for AI and automation, open-source AI software.

Market and consumer changes

- Changing consumer behaviour: Passion to engage with brands/companies anytime, anywhere, with the expectation of immediate service 24/7.

- New market forces in a ‘winner takes all’ economy: Tech giants challenging the markets. These AI-oriented players have access to top-notch talent, alternative data pools, immense funding from investors, and the ability to disrupt certain parts of the value chain.

- InsurTechs challenging incumbents.

- Sharing economy revolutionising the need to own.

- Virtual personal assistants or intelligent agents in customer service.

- Regulatory harmonisation trying to keep up-to-speed with different developments.

To make these concepts more concrete, let us fast-forward to the year 2030 and let us look at how journeys of different stakeholders could look like:

Isabel the actuary: She works in product development and is designing a new pay-as-you-drive product for self-driving cars. Having access to vehicle, driving, and claims data, she analyses the impact on claims depending on the chosen route, on choosing autopilot, and on weather conditions. Isabel designed this product together with the car manufacturer so that optimal routes are chosen automatically through the navigation system – associated premium savings are directly shown on the driver's screen. She embedded feedback loops to life and health products as well, so that clients with multiple covers can directly benefit from reduced claims expectations through discounts on their, e.g., accident cover. Additionally, Isabel can test the impact of this new product on forecasted internal and external Key Performance Indicators (KPIs) in real-time.

Nolan the sales agent: Nolan gets a trigger that one of his customers, Livia, got a new self-driving car. Hence, Nolan gives her a call and talks her through potential insurance solutions for her car to figure out what is best suited for her needs. Based on the ongoing call, the sales LLM provides real-time suggestions and quotes to Nolan to provide best-advice options. After learning, that Livia moved to a new apartment a few weeks ago, the sales LLM automatically updates Livia's policyholder profile with the latest data, only requiring Nolan to check the update and press ok.

Livia the customer: After owning the car for few months, Livia drives to her parents' place, not using autopilot. Unfortunately, her windshield gets hit by a small rock. Her insurance app guides her to take photos. The car's diagnostics show that she can continue to drive but needs to go to a repair shop straight away – the nearest one, located only five minutes away, is shown on the navigation system. Once she arrives there, a replacement car is already available so that she is able to arrive to her parent's place just in time.

Peter from claims management: The company's GenAI assistant creates a full report within a few minutes including damage assessment, fraud potential, plan coverage, estimated payment, and the repair shop where Livia is taking her car. Peter reviews the report, conducts minor adjustments, and settles the claim shortly thereafter.

While these are made-up examples, many of these technologies already exist today and will become reality in the near future. Hence, rather than a burden, this is a unique opportunity for actuaries to strengthen their role, broaden their areas of influence, and safeguard responsible use of AI in critical areas. AI and adjacent trends will fundamentally transform how companies operate and will allow them to optimise sales, distribution, pricing, claims management, and many more areas. As many of these applications require actuarial expertise paired with data, business and communication skills, actuaries are optimally positioned to assume central roles and shape this change. Taking a stronger stance, one could argue that actuaries have the duty to step up in AI and protect what matters most – safeguarding welfare and protection of customers and societies.

Actuaries are traditionally well-trained to answer descriptive and predictive questions:

- WHAT are renewal rates?
- WHO are riskier clients?

AI generally means that predictions become cheaper (cheaper quick data collection, cheap storage, cheap computing power) which certainly helps actuaries in all of their work. According to economic theory, when the price of something drops, the demand for it will increase – and thus better and cheaper predictions are used more and also in novel areas.

While AI can improve answers to these actuarial core questions, its true added value lies within applications. Answering such 'HOW-questions' can create huge benefits for customers and societies:

- How can we target the right customers at the right time?
- How can we prevent policy lapse?
- How should our pensions and social security systems better react to changes in the societies and to changing demand?
- How can insurers and societies build a more inclusive environment without risk differentiation leading to direct or indirect discrimination, especially with intersectionally vulnerable minorities?

This paper, while not trying to tell everything crucial, attempts to give a condensed overview of the most important concepts connected to the area of Artificial Intelligence. We will then build on these concepts to highlight our ideas on how actuaries could make best use of AI. We hope it will be a good starter and pave the way for actuaries to learn more of the subject through references and recommended further readings.

2 OVERVIEW OF EUROPEAN REGULATION ON THE TOPIC

2.1 RISKS OF AI AND ROLE OF THE ACTUARY

AI applications pose several risks[1]. In this chapter we want to highlight a few pitfalls that actuaries should have in mind when handling data and AI:

- **Persuasive misinformation:** AI has the potential to generate highly persuasive but factually incorrect responses, leading to misinformation and misinterpretation. Requirement on explainability may help.

- **Discrimination:** The probabilistic nature of many AI applications can lead to unforeseen behaviours and capabilities during deployment, necessitating careful monitoring and management to ensure desired outcomes. Moreover, if models are naïvely trained on biased real-world data, these models can perpetuate and amplify existing biases if deployed without proper oversight, exacerbating social and cultural inequalities.

- **Data security and intellectual property:** Cloud-based training of AI models poses security risks as it involves the transmission of proprietary data, increasing the potential for data breaches and unauthorised access. In general, improper use of data and tools without adequate guidance and supervision can result in unintended consequences and ethical dilemmas.

- **Cybercrime:** The instant generation of convincing phishing emails and deepfakes facilitated by Generative AI enhances the ease of cybercrime, requiring heightened vigilance in online security.

[1] There are three main approaches to regulation of data and AI globally:
– European approach that starts from the rights of the consumer/individual,
– US approach relying on markets, and
– Chinese party controlled approach.

The US approach has its problems (e.g., issues around Meta, Google and especially X, formerly known as Twitter) and the Chinese approach probably is not good for innovations. Anu Bradford of Columbia University thinks (see The Brussels Effect: How the European Union Rules the World, and Digital Empires: The Global Battle to Regulate Technology) that the European approach at least currently has an upper hand in creating globally applicable regulation.

These risks require responsible ways on how to use and apply AI. While recent and upcoming regulation will embed principles to safeguard sustainable use of AI, actuaries will play a crucial role as well. Key success factors will include the following:

- A sound governance to navigate ethical, legal, and technological risks such as clear rules on where, when, and how to apply certain tools.
- A framework with clear roles and responsibilities to support decision making.
- Expertise, professionalism, and training around the topic of AI: The demands on the actuarial profession will increase because in the future it will be expected that actuaries understand AI models. This will require appropriate training and further education programs. In addition, as most employees will need have at least basic knowledge in AI, actuaries can and should contribute to these training initiatives.
- Suitable tools to monitor and manage AI risks.
- Algorithm robustness: actuaries should ensure the algorithms are running on secure and sound IT infrastructures, being less prone to operational risks such as algorithmic liability, hacks etc. The state and performance of algorithms need to be monitored regularly to exclude potential to cause harm.

2.2 GENERAL HORIZONTAL REGULATION IN THE EU

Direct discrimination occurs when a person is treated less favourably than another person simply because one of their protected characteristics is not the same. If the person's corresponding risk factor is not used by insurers, such discrimination can be completely avoided. See, for example, the Directive 2004/113/EC ('Gender Directive') (EC, 2004).

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

The disadvantage may be justified (e.g., using loss experience or size of the car in motor insurance) or not justified (e.g., using the colour of the car). More generally, it may be caused using proxy variables from the non-protected characteristics of policyholders (i.e., identifiable proxy), or opaque algorithms (i.e., unidentifiable proxy). See, for example, the Directive 2004/113/EC ('Gender Directive') (EC, 2004).

While the above are legally well defined terms, the following is not.

Algorithmic discrimination refers to the biased outcomes or decisions produced by algorithms and is usually considered a subset of indirect discrimination. European Insurance and Occupational Pensions Authority (EIOPA, 2019) conducted a thematic review on the use of Big Data Analytics (BDA) based on 222 participating motor or health insurers from 28 European jurisdictions. The thematic review revealed that 31% of insurance firms already actively used BDA tools and another 24% of firms planned to use them within the next three years. These new data analytics tools are generally used on pricing and underwriting, claims management and sales and distribution, whereas insurers have only taken limited approaches to ensure fair and ethical outcomes in the use of BDA in underwriting and pricing. See, for example, European Union Agency for Fundamental Rights report on 'Bias in Algorithms – Artificial Intelligence and Discrimination'.

2.3 HORIZONTAL REGULATION ON DATA AND DATA PRIVACY IN THE EU

The General Data Protection Regulation[2] is a Regulation in EU law on data protection and privacy in the EU and the other countries in the European Economic Area (EEA). The GDPR is an important component of EU privacy law and of human rights law, in particular Article 8(1) of the Charter of Fundamental Rights of the European Union. It also addresses the transfer of personal data outside the EEA. The GDPR's primary aim is to enhance individuals' control and rights over their personal data and to simplify the regulatory environment for international business. The regulation contains provisions and requirements related to the processing of personal data of individuals, formally called 'data subjects', who are located in the EEA, and applies to any enterprise—regardless of its location and the data subjects' citizenship or residence—that is processing the personal information of individuals inside the EEA.

The GDPR was adopted on 14 April 2016 and became enforceable beginning 25 May 2018. As the GDPR is a regulation, not a directive, it is directly binding and applicable, and provides flexibility for certain aspects of the regulation to be adjusted by individual member states.

[2] 2016/679, 'GDPR'

Personal data may not be processed unless there is at least one legal basis to do so. Article 6 states the lawful purposes are:

a. If the data subject has given informed consent for one or more specific purposes to the processing of his or her personal data;

b. To fulfil contractual obligations with a data subject, or for tasks at the request of a data subject who is in the process of entering into a contract;

c. To comply with a data controller's legal obligations;

d. To protect the vital interests of a data subject or another individual;

e. To perform a task in the public interest or in official authority;

f. For the legitimate interests of a data controller or a third party, unless these interests are overridden by interests of the data subject or her or his rights according to the Charter of Fundamental Rights (especially in the case of children).

If informed consent is used as the lawful basis for processing, consent must have been explicit for data collected and for each purpose data is used for (Article 7; defined in Article 4). Consent must be a specific, freely-given, plainly-worded, and unambiguous affirmation given by the data subject; an online form which has consent options structured as an opt-out selected by default is a violation of the GDPR, as the consent is not unambiguously affirmed by the user. In addition, multiple types of processing may not be 'bundled' together into a single affirmation prompt, as this is not specific to each use of data, and the individual permissions are not freely given.

The European Commission published on 21 April 2021 its draft regulation on artificial intelligence[3]. The Commission designed a horizontal regulatory framework that encompasses any AI system that touches the single market, whether the provider is based in Europe or not. The Artificial Intelligence Act uses a risk-based approach and sets up a series of escalating legal and technical obligations depending on whether the AI product or service is classed as low, medium or high-risk, while a number of AI uses are banned outright. At the time of writing there is agreement in the trilogue negotiations of the Parliament, the Council and European Commission on the draft regulation. The earliest possible adoption will be in 2024.

[3] Artificial Intelligence Act, the AI Act.

The original draft regulation includes into the high risk category essential private and public services, including access to financial services such as credit scoring systems. I.e., it did not put insurance or pensions into the high risk bucket. The updated draft of the AI Act, released in November 2021, classifies 'AI systems intended to be used for life insurance purposes' under the high-risk category. Specifically, it refers to 'AI systems intended to be used for insurance premium setting, underwritings and claims assessments.' The details of the trilogue agreement are not available yet, but at least savings instruments in life insurance will be considered high risk.

For AI applications in the high risk category there will be the following requirements:

- Design in line with requirements: Ensure AI systems perform consistently for their intended purpose and are in compliance with the requirements put forward in the Regulation.

- Conformity assessment: Ex ante assessment of conforming with the respective requirements.

- Post-market monitoring: Providers to actively and systematically collect, document and analyse relevant data on the reliability, performance and safety of AI systems throughout their lifetime, and to evaluate continuous compliance of AI systems with the Regulation.

- Incident report system: Report serious incidents as well as malfunctioning leading to breaches to fundamental rights (as a basis for investigations conducted by competent authorities).

- New conformity assessment: New conformity assessment in case of substantial modification (modification to the intended purpose or change affecting compliance of the AI system with the Regulation) by providers or any third party, including when changes are outside the 'predefined range' indicated by the provider for continuously learning AI systems.

2.4 VERTICAL REGULATION IN THE EU

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), and the upcoming e-privacy Directive (ePD). Such legislation is intended to include a governance framework around AI as well. However, given the nature of AI and its implications, especially on the ethics of using data and digital technologies, the frameworks do not suffice or reflect all concerns. For example, Article 41 (1)[4]

[4] 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. 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.

In essence, EIOPA relies on the Ethics Guidelines for Trustworthy AI developed by the European Commission’s High Level Expert Group on AI (hereinafter AI HLEG) see (EIOPA, 2021). These principles are highlighted in the following table:

Proportionality Principle	Conducting AI use cases in order to determine the governance measures required for insurer’s specific AI applicability. Based on its impact on policyholders and firms, insurers can assess, which measure to put in place.
Fairness and non-discrimination	Avoid practices, which influence consumer’s willingness to pay or willingness to accept, using data fairly. Developing fair algorithms, socially responsible, and metrics that prove transparency of AI applications among societies is one of the concerns of this principle.
Transparency and explainability	If high-impact use cases are used, insurers should be able to explain the applicability of AI, and use algorithms that combine model transparency and governance measures.
Human oversight	Assigning and documenting the roles and responsibilities for the staff involved with AI algorithms and processes.
Data governance and record keeping	Use accurate and complete data in AI systems, and comply with data protection laws such as GDPR. Data storage and management is also of concern here.
Robustness and performance	Using robust AI algorithms, minimising their potential to cause harm. The performance of algorithms has to be monitored and deployed on secured IT infrastructures, being less prone to operational risks (hacks, algorithmic liability etc.)

3 ARTIFICIAL INTELLIGENCE AND DATA SCIENCE IN A NUTSHELL – MAIN CONCEPTS CLARIFIED

3.1 DATA SCIENCE (DS)[5]

Data Science (DS) can be considered to give operational tools to solve business problems through machine learning or other data-driven models, extracting information and knowledge from structured or unstructured data. Seen from a business perspective, DS requires the translation of a business problem into a research and analysis project and then transform it, again with the help of data into a practical solution. A DS project can be represented through a process of data analysis and interpretation that must be seen as iterative rather than linear, subject to continuous verification.

Data Science is useful in answering five types of fundamental questions:

- How much or how many? (regression)
- Which category? (classification)
- Which group? (grouping)
- Is it strange? (anomaly detection)
- Which option should be taken? (recommendation/prescription)

From a business perspective, these questions become, for example:

- Who are the best customers?
- Why are they buying ‘that’ product (see, e.g., the PayPal example in this document)?
- How to predict whether a customer will buy another type of product?
- Why have certain customers not been buying for a long time?

Data science does not necessarily require sophisticated algorithms and multi-core cloud computing but (depending on the problem) solid understanding of the business problem and data, good data handling skills and ability to bring it into action in the organisation.

[5] https://en.wikipedia.org/wiki/Data_science consulted November 2023

3.2 ARTIFICIAL INTELLIGENCE, MACHINE LEARNING, DEEP LEARNING, AND GENERATIVE AI

The terms Artificial Intelligence, Machine Learning and Deep Learning are terms that are often used synonymously. The simplest and most effective way to explain the difference is to refer to the Chinese box system where one domain is a component of the one before it. Machine Learning (ML) is a domain of Artificial Intelligence (AI) and Deep Learning (DL) in turn is a domain of Machine Learning. ML is simply a way of achieving AI and DL is one of the many approaches related to ML. In other words, one can consider AI as the basic discipline and ML and DL the techniques, or rather, the models that enable its application. More formally, we can have the following definitions:

- AI systems involve all those operations that are characteristic of the human intellect and performed by computers. They also perceive their environment while collecting and making use of data and they reason on the knowledge gained from this data. Such operations include planning, language understanding, object and sound recognition, learning and problem solving.

- ML is an area of AI that focuses on the ability of machines to receive a set of data and learn on their own, modifying algorithms as they receive more information about what they are processing. ML is thus a way of 'educating' an algorithm so that it can learn from various situations. Education, or even better training, involves the use of huge amounts of data and an efficient algorithm to adapt (and improve) according to the situations that occur.

- DL is one of the approaches to ML that originates from brain morphology and functioning, i.e., the interconnection of various neurons. DL uses huge models of neural networks with various processing units; it exploits computational advances and training techniques to learn complex patterns through huge amounts of data. Common applications include image and speech recognition.

An additional issue to mention is that Generative AI utilises Deep Learning algorithms to create new data, such as realistic images and coherent texts, for different applications. Combined models like ChatGPT enable more natural interactions. Language models like LLMs, trained on vast text data, generate meaningful content by predicting the next word based on context, playing a vital role in Generative AI's ability to produce coherent and relevant text.

3.3 MACHINE LEARNING DEEP DIVE – SUPERVISED, UNSUPERVISED AND REINFORCED LEARNING

ML has at its base a series of different algorithms that, starting from primitive notions, will know how to make a specific decision rather than another, or perform learned actions over time. Depending on the type of algorithm used to enable machine learning, i.e., on how the machine learns and accumulates data and information, one can subdivide ML into three different learning systems: supervised, unsupervised and reinforcement learning. See (Wüthrich & Merz, 2022) for an extensive overview of modern machine learning methods.

3.3.1 Supervised Learning

Supervised learning consists of providing the machine's computer system with a series of specific and codified notions, i.e., models and examples that allow it to build a real database of information and experiences. In this way, when a machine is faced with a problem, all it has to do is to draw on the experiences stored in its system, analyse them, and decide what answer to give on the basis of already codified experiences. This type of learning is, in a way, provided already packaged, and the machine only has to be able to choose which is the best response to a stimulus given to it. In short, the operation of this algorithm consists of a goal/outcome variable (or dependent variable) that must be predicted by a given set of predictors (independent variables). Using this set of variables, we generate a function that maps the inputs to the desired outputs. The training process continues until the model reaches the desired level of accuracy on the testing data. Examples of supervised learning include regression, decision tree, random forest, KNN (The k-nearest neighbours algorithm is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point), logistic regression, etc. Algorithms that make use of supervised learning are used in many fields, from medicine to speech identification: they have the ability to make inductive hypotheses, i.e., hypotheses that can be obtained by scanning a series of specific problems to obtain a suitable solution to a general problem.

3.3.2 Unsupervised Learning

In unsupervised learning the information entered into the machine is not tagged or paired with results. The machine is able to draw on certain information without having any examples of its use and, therefore, without having knowledge of the expected results depending on the choice made. The machine itself must catalogue all the information in its possession, organise it and learn its meaning, its use and, above all, the result to which it leads. In unsupervised learning the machine, will have to organise the information intelligently and learn which results are best for different situations that arise. The operation of this algorithm is characterised by the fact that we have no target or outcome variable to predict/estimate. Examples of unsupervised learning include a-priori algorithm (an algorithm for frequent item set mining and association rule learning over relational databases), K-means (k-means clustering is a method of vector quantisation, originally from signal processing, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, i.e., cluster centres or cluster centroid, serving as a prototype of the cluster), etc.

3.3.3 Reinforcement Learning

Reinforcement learning is the most complex learning system, which requires the machine to be equipped with systems and tools to improve its learning and, above all, to access and process the characteristics of its environment. In this case, therefore, the machine is provided with a series of support elements, such as sensors, cameras, GPS, etc., which enable it to detect what is happening in its surroundings and make choices to better adapt to its environment. The machine is trained to make specific decisions, exposing it to an environment in which it continuously trains itself through trial and error. It learns from past experience and tries to acquire the best possible knowledge to make accurate business decisions. An example of reinforcement learning is a Markov decision-making process.

3.4 DEEP LEARNING DEEP DIVE

Those wishing more information about Deep Learning are advised to check, e.g., https://www.linkedin.com/pulse/what-generative-ai-llm-luis-escalante. Some concepts there are:

- Convolutional Neural Networks (CNNs): often used for image recognition and classification tasks by applying filters to extract features that will be passed through a series of layers to produce a classification output.

- Recurrent Neural Networks (RNNs): these are commonly used for sequential data and work by processing one element of the input sequence at a time and using the previous state of the network to inform the processing of the current element. RNNs are useful for language modelling, speech recognition, and sentiment analysis tasks.

- Generative Adversarial Networks (GANs): used for generative tasks, such as image synthesis and text generation, by training two neural networks, one to generate fake data (Generator) and the other to discriminate between real and fake data (Discriminator). The two networks are trained in opposition to each other with the goal of improving the generator’s ability to generate realistic data.

- Transformers: designed specifically for natural language processing tasks by using self-attention to selectively focus on different parts of the input sequence, allowing them to process long sequences of text efficiently so it can be used for language modelling, machine translation, and question answering.

3.5 GENERATIVE AI AND LARGE LANGUAGE MODELS (LLMS)

Generative AI, with its ability to create text, images, audio, and more in seconds, is poised to transform various industries. Its power surpasses existing enterprise technologies, and its capabilities have rapidly advanced in the past few years. Generative AI can already engage in human-like conversations, generate diverse styles of images from text, and even pass exams. It can create high-quality videos and compose music across instruments and genres.

As there are risks such as intellectual property infringement and biased outputs, responsible AI approaches need to be taken to mitigate them. It needs to be remembered that an LLM does not ‘understand’ the problems it tackles – instead it basically does a statistical analysis and predicts based on the training material (Internet content) the probable answer to the question posed to it. Organisations need to define rules, integrate Generative AI into workflows, and train teams to use these tools responsibly. Early experimentation and ongoing learning will be crucial for long-term success with Generative AI.

In the insurance industry, Generative AI can revolutionise the full value chain, e.g., claims management, underwriting, and customer service. By leveraging Generative AI, insurance companies have the potential to reduce time, save costs, and improve accuracy in various processes. However, to realise the full value of Generative AI, organisations need to address roadblocks such as limited understanding of value, data governance challenges, digital platform limitations, skill fragmentation, and change management.

As Generative AI becomes more prevalent, it will lead to the emergence of new roles, changes in operating models, increased productivity, and revamped talent acquisition and performance management practices. Personalised training will be essential to equip employees with the skills to effectively use Generative AI tools.

3.6 DATA SCIENCE IN PRACTICE

We can identify seven key steps for the practical application of data science:

- **Ask a precise question:**
  - Describe the problem identified.
  - Answer needs to be measurable.
  - Start small with a simple question.
  - Have a hypothesis in mind.

- **Select and collect data:**
  - Generate a single point of truth.
  - Take GDPR into account.
  - Normalise.
  - Distinguish between training data and test data.

- **Clean data:**
  - Check data quality (missing values, sums are correct, outliers, duplicates, blanks, errors...).
  - Harmonise scale if necessary.

- **Explore and transform data:**
  - Remove redundant or duplicate features.
  - Remove noisy features (e.g., with few data points).
  - Keep your hypothesis in mind.

- **Prepare infrastructure and tools:**
  - Choose appropriate infrastructure and tools for the problem ahead.
  - Try to keep things as simple as possible.
  - Make sure that every step will be repeatable.

- **Train, test and assess model:**
  - No hard coding.
  - Comment your code.
  - Understand your model.
  - If possible, provide ranges .
  - Always test your trained model.
  - No in-sample testing (keep training data and test data separate).

- **Infer business actions:**
  - Check that your result gives an answer to your question.
  - Visualise your results.
  - Iterate the process and refine/extend your results.
  - Make your message clear, simple and precise.

3.6.1 Ask a precise question

This first step is crucial to have clarity on which problem you actually want to solve and on how to measure impact and success. For this step, business and actuarial knowledge is often inevitable as otherwise measurement of impact may lead to wrong conclusions or may disregard risks.

3.6.2 Select and collect data: retrieve the 'raw data' needed for the identified problem

This stage of the process requires some attention because it involves both thinking a priori about what data will be needed, and the actual 'retrieval' of data from a plurality of sources (both internal but also external datasets). Many elements herein resemble to anything actuaries traditionally do with data analysis. However, while the danger of duplication is recognised, it seems reasonable to restate here also the obvious parts. The data may be structured data (e.g., from databases and internal applications of the company, such as a CRM or an industrial application, e.g., for managing production or warehouse management) or unstructured data (text, images, videos from e-mails, documents, collaboration platforms, but also from external sources such as social networks, open document repositories, web pages, etc.). A crucial element of this phase is the verification of regulatory compliance.

3.6.3 Clean data: processing data for analysis

The Data Cleaning (or Data Preparation) phase consists of the act of manipulating and pre-processing raw data from a variety of sources and in different formats, cleaning it, harmonising it and transforming it into data that can be used by analysis tools.

3.6.4 Explore and transform data

Data Exploration phase means an initial 'exploratory analysis'; in essence, statistical tests are carried out, initial analyses are made and the first Data Visualisation techniques are tested. From here we begin to see the concept of an iterative and non-linear process. In the Data Exploration phase data errors may emerge in the data or an intervention can be needed that 'leads back' to the previous phase of data cleaning and preparation. Closely related parts of the Data Exploration phase are experimentation and modelling, i.e., the process of identifying and building the analysis model for solving the specific problem identified in the very first phase of the entire Data Science process. These phases involve the 'fine-tuning' and validation (including the choice of algorithms and their possible 'tuning') of the analytical model. The model is then tested by exploiting the transformed data and, based on the generated output (i.e., the insights obtained), and its performance and effectiveness is tested in terms of accuracy of information and actual value with respect to the decision-making process.

3.6.5 Prepare infrastructure and tools

Optimal infrastructure and tools depend on the problem at hand. In many cases, it is advisable to start simple and use tools that you and/or your organization commonly use. If precision or a new feature is the decisive factor, then scale to new tools. Always make sure that the used infrastructure and tools comply with regulation and internal rules for data handling.

3.6.6 Train, test, and assess model: Advanced Analytics and Machine Learning

At this point in the process algorithms are used to analyse data, discover hidden patterns or extract interesting knowledge from this data. 'Typical' operations of this phase are parameter identification, processing, modelling and pattern evaluation. One defines here how to extract actual value from large volumes of information, choosing algorithms and 'training' methods to search for patterns in the data (e.g., with machine learning), as well as the form of representation or the set of representations with which the information is to be extracted (classification rules, decision trees, regression classification, regression, clustering, etc.). An important part of this phase of the Data Science process is also to provide business users with all the necessary elements (both quantitative and qualitative) to be able to access information and knowledge that is truly relevant to the problem identified, the possible applicable solution and the effectiveness with respect to the business decision.

3.6.7 Infer business actions: communicating and visualising results

Data visualisation actually comes into play several times during various phases of a typical Data Science process. The 'final' phase of the process concerns the communication of the results derived from the analyses, understood as the visualisation of such results through analysis systems that must be made available and usable easily by business users. Here Data Visualisation and Data Storytelling, i.e., advanced data analysis systems make it possible to Infer in a visual/graphical sense, among hundreds and thousands of pieces of data (of different formats and structures) information, correlations and patterns. The aim is to unearth 'a story' hidden in this data that can only come 'to the surface' through advanced analysis and can become usable for business users, without specific technical skills, precisely thanks to Data Storytelling and information visualisation.

In addition to the stages of the Data Science process just described, it is essential to avoid the mistake of thinking that once the analysis system has been put into production the process is concluded. Models and algorithms do not perform indefinitely. It is therefore necessary to continue to monitor the performance of the models with respect to the business objectives.

The above steps are perfectly consistent with what is proposed by the methodological framework CRISP-DM (Cross Industry Standard Process for Data Mining) which represents an industry standard and has become popular for its flexibility and possibility of customisation.

3.7 AI-DS IN BUSINESS USE IN INSURANCE – EXAMPLES

In the following, we provide an non-comprehensive list of applications of AI along the insurance value chain to show the breadth of its applicability already today. See, for example, (Seehafer, et al., 2021) for an overview on selected actuarial case studies. A lot of these examples require actuarial input or interpretation:

- **Product development and pricing:**
  - Non-linear pricing and other data-driven pricing approaches, with potential benefits for insurability.
  - Personalised product development based on other than classical sources of data, better reflecting individual needs for the benefit of the customers.
  - Coverage of new risks and leveraging new data sources (e.g., cyber, robotics), in particular to provide solutions for new consumer trends .
  - AI-supported product design leveraging LLMs based on market and consumer data, reflecting strategic appetite and market conditions.
  - Drafting of Terms and Conditions through LLMs.
  - Dynamic quote modification in real-time.
  - Wording and price and/or renewal fine-tuning based on client feedback.

- **Marketing, distribution, and CRM:**
  - AI-based campaign design and management, e.g., optimising which customers to best approach with which products, via which channels, with which messages.
  - Improved client experience through better-informed sales staff and best offers, enabled through LLMs trained on all previous client interactions.
  - Improved lead generation for sales staff, partners, and direct channels, leveraging multiple data sources, e.g., daily smart lead lists for agents or real-time leads (next best offer) during calls via LLMs.
  - Renewal pricing optimisation considering customer profile, price elasticity, past interactions, and other factors, see (EIOPA, 2023).
  - Real-time and fully integrated quoting for brokers and agents.
  - Churn detection (along different types of churn) and suggesting prevention measures based on previous experience, sentiment analysis, and expert experience
  - Improving customer/agent/broker segmentation through multiple data sources to better steer business decisions.
  - AI-based derivation of customer lifetime value to define room for CRM measures such as discounts.

- **Underwriting***:*
  - Leveraging of external (big) data sources to improve underwriting decisions.
  - Pattern recognition among claims, generating underwriting insights.
  - AI-based risk assessment, e.g., pre-damage check in Own Damage using image recognition, documents analysis in risk assessment.
  - AI-based underwriting fraud detection.
  - Digital brokerage technology such as 'UnderwriteGPT'.

- **Claims and benefits management:**
  - Automated claim assessment, determining degree of damage and instant generating of appraisal forms.
  - Automated claims categorisation, e.g., determining automated vs. manual handling, granularity of information needed from customer.
  - Self-learning fraud model leveraging AI and analysis of structured and unstructured data, as well as prevention.
  - Automated claims management and prevention using Internet of Things (IoT) and LLMs to extract relevant data from e.g., images and emails.
  - Claims management through AI, considering customer experience, adjacent extra costs (legal), and potential renewal options.
  - AI-based single loss reserving considering various data sources, e.g., claims documents, call protocols, submitted material, pictures.
  - AI-driven expert network management and control leveraging also integrated scorings.
  - Direct re-imbursement powered by AI.
  - Improving customer satisfaction through simplifying complex claims processes, e.g., assisted by virtual/augmented reality tools, 3D modelling, AI to scan documents and retrieve relevant info to register claims in the systems.

- **Operations and IT:**
  - Automation of administration and services, e.g., through LLM-powered chat-bots and apps.
  - Acceleration of day-to-day tasks through speech-to-text, machine translation, summarising of text, LLMs for quick text generation, personalisation of emails, classification of text and documents etc.
  - Conversational interface, allowing customers to review policy, submit claims, and track status.
  - Code review and debugging through LLMs.
  - Automatic coding and software development through LLMs.
  - Conversational IT ticketing.

- **Finance and actuarial:**
  - Data-driven risk analyses to improve the use of reinsurance.
  - Legacy product simplification within migrations.
  - Automated reports and document generation.
  - Real-time financial forecasting.
  - Automated invoice processing.
  - Automated tailoring of investor reports and communication.
  - Automated regulatory changes monitoring and alerting.
  - Automated model documentation, validation, development (coding).
  - Real-time monitoring of KPIs and KRIs, and scenario testing.

- **ALM and Investment:**
  - Research engine based on unstructured data.
  - Portfolio adjustments/hedging based on news surveillance.
  - Automated monitoring of market trends.
  - Summarisation of market reports.
  - Optimisation of Strategic Asset Allocation (SAA) and Tactical Asset Allocation (TAA), manager selection.
  - Better proxy modelling of assets, liabilities, and interactions.

4 WHEN TO APPLY AI – HOW TO AVOID THE OVERUSE AND THE ISSUES OF AI

Next to regulatory constraints there needs to be a business case showing value for the insurer (financial, risk minimising, increasing customer satisfaction, data quality) and a dedicated owner, responsible for implementation, maintenance and monitoring. Agile implementation can help to quickly show a proof of concept before rolling out on a bigger scale, increasing buy-in from management and reducing sunk costs.

The value of AI lies within the combination of business sense, data handling and modelling. Actuaries are perfectly positioned and enabled to unlock this value. Hence, the following anecdote from Matt Lerner from PayPal about an intern who solved a problem should serve as a prime example. PayPal faced the challenge of losing around one million merchants annually – previous analyses had not been able to give an answer as to where this is coming from. To narrow down the reasons behind this churn, they first excluded account closures and focused on 'going dark' accounts, which indicates disengagement from the product. They also ruled out 'one-and-dones' and addressed onboarding issues for new entrants. They further excluded false positives and non-regretted churn, narrowing down the analysis to well-behaved, non-seasonal merchants. By shifting their focus to larger merchants, they were able to refine their approach and detect various smaller issues that led to churn. By creating a predictive model, they now flag merchants at risk along these smaller issues and proactively manage through their customer service.

This example should highlight that understanding data, reducing noise, and linking it back to business reality are key success factors in business-related problems. Here specifically, understanding the true reason of churn has been key to create the appropriate action plan for operations with significant value generated.

So, what does this mean for actuaries specifically? Actuaries can and should actively shape and drive certain AI and data science initiatives. Looking at successful projects in that area, there are a few, while somewhat generic, best practices that increase the chance of success and are straight-forward to apply:

1. Work interdisciplinary, engaging the right people – impact improves with diversity of skills and experience. An actuary can act as a linking pin, responsible for all applications and exchange of skills to work well together.

2. Start with a well-defined business question you try to solve and write down hypotheses on its impact on strategy, operations, and financials. In most applications, the more the question targets prescriptive aspects, the higher the value of the answer will be. Align with management.

3. Once this is clear, collect, modify and interpret the data you need to answer this question, of course considering data protection requirements. As shown in the example of PayPal, it is of utmost importance the data matches the business question, because even when applying the 'best' model, conclusions can be wrong if data is not properly understood.

4. Then, design the model considering factors of responsibility, explainability, simplicity, accuracy, predictability and potentially more. Leverage data scientists if needed.

5. Consider the risks that are embedded in answering this business question (e.g. insurability, wrong incentivisation, customer inclusion and satisfaction and cannibalisation), in the data (e.g. data protection, data security), and in the model (e.g. explainability, discrimination) and be concrete on limitation of applicability and results.

6. Implement and run the model with repetitive iterations. Extract early output in an agile way, even if first results will not be perfect. This will improve and shorten the systems development cycle.

7. Interpret, challenge, and validate results, putting particular focus on previously highlighted risks. Iteratively refine results and document.

8. Answer the question you asked at the beginning and communicate results in an understandable way to management:
   - Be clear on the impact on strategy, operations, and financials.
   - Explain deviations from initial hypotheses.
   - Highlight areas of concern and limitations.
   - If applicable, provide next steps on how to roll out, generalise, or transfer the learnings.

5 AI EXPLAINABILITY (XAI) – FROM BLACK BOX TO A WHITE, MODEL AGNOSTIC APPROACH

AI explainability, also known as interpretability or transparency, refers to the extent to which the decisions and predictions made by an AI system can be understood and explained by humans. This is an important consideration for a variety of reasons, including accountability, trust, and fairness. Recall that explainability is indispensable in the assessment of indirect discrimination.

AI explainability is an important topic also specifically in the context of ML models. It refers to the ability to understand and interpret how an ML model reaches its decisions. While explainability is crucial for the insurer or other undertaking (e.g., in making sure the model fulfils ethical requirements), the lack of it will make it difficult for other users to trust the decisions it makes and for regulators to ensure that it is operating ethically.

Specifically in the actuarial context in the last years there have been some proposals aimed at establishing a framework for explaining ML models to go beyond the confines of the linear models.[6]

There are several different categories of indicators that have been proposed for measuring the explainability of AI systems. These include:

- **Intelligibility:** This refers to the degree to which the internal workings of an AI system are understandable to users. This can be assessed by looking at the complexity of algorithms and models used by the AI system, as well as the clarity and transparency of any explanations provided by the system.

- **Auditability:** This refers to the ability of an AI system to provide a clear and traceable record of its decision-making process, allowing users to understand how and why the system reached a particular conclusion. This is important for ensuring accountability and trust in the system.

- **Transparency:** This refers to the extent to which an AI system reveals its inner workings and decision-making processes to users. This can be achieved through techniques such as feature visualisation and sensitivity analysis, which allow users to see how the system is using different inputs to make its predictions.

- **Explainability** (and an indicator to measure explainability): This refers to the ability of an AI system to provide clear and understandable explanations for its decisions and predictions to users.

[6] Towards Explainability of Machine Learning Models in Insurance Pricing – Kevin Kuo, Daniel Lupton - 03/2020.

- **Trustworthiness:** This refers to the degree to which an AI system is perceived as reliable and trustworthy by users. This can be influenced by a variety of factors, including the system's accuracy, reliability, and explainability.

Those categories of indicators are applied/developed, as we will see in the next paragraphs, in five different XAI macro-areas[7]: 1) Feature Interaction and Importance 2) Attention Mechanism 3) Dimensionality Reduction 4) Knowledge Distillation and Rule Extraction 5) Intrinsically Interpretable Models.

AI interpretability and explainability play a pivotal role in the actuarial field, offering valuable insights and enabling effective decision-making. Model understanding provides actuaries with essential tools to:

- debug ML models effectively;

- identify potential issues in data preprocessing, feature engineering, or model training.

This empowers actuaries to rectify these issues and enhance model performance.

Additionally, model understanding is instrumental in detecting biases that may lead to disparate impacts on certain groups. By uncovering such inappropriate biases, actuaries can take corrective measures, mitigating discrimination and upholding fairness in decision-making processes.

Furthermore, in actuarial applications where ML models' predictions significantly impact individuals' financial well-being and livelihoods, model understanding becomes a vital tool to provide recourse. By revealing the reasoning behind predictions, individuals can comprehend the factors influencing their outcomes and challenge decisions if necessary. This transparency empowers individuals to seek redressal for potentially unjust outcomes.

In high-stakes actuarial scenarios, trust in ML model predictions is paramount. Model understanding enables actuaries to evaluate the reliability of model outputs and make informed decisions based on the model's reasoning. By verifying the soundness of the model's features with respect to auxiliary criteria, actuaries can gain greater confidence in using the model's predictions for critical decision-making.

[7] Explainable Artificial Intelligence (XAI) in Insurance – Systematic Review

5.1 LOCAL AND GLOBAL EXPLANATIONS

Local explanations are specific to a single prediction made by an ML model. For example, a local explanation might show which input features had the greatest influence on a model's decision to classify an image as a dog (e.g., LIME, i.e., Local Interpretable Model-agnostic Explanations).

Feature importance is a measure of how important each input feature is to an ML model's predictions. For example, an ML model might assign a higher importance to an image's colour than its shape when classifying objects (e.g., Attribution Maps, SHAP).

Global explanations provide an overview of an ML model's behaviour across multiple predictions. For example, a global explanation might show which input features are generally important for an ML model's decision-making process (e.g., Permutation Importance).

Counterfactual explanations: These are explanations that show how a model's prediction would change if one or more input features were altered. For example, a counterfactual explanation might show how an ML model's prediction of a customer's creditworthiness would change if his/her income were increased (e.g., Partial Dependency Plot).

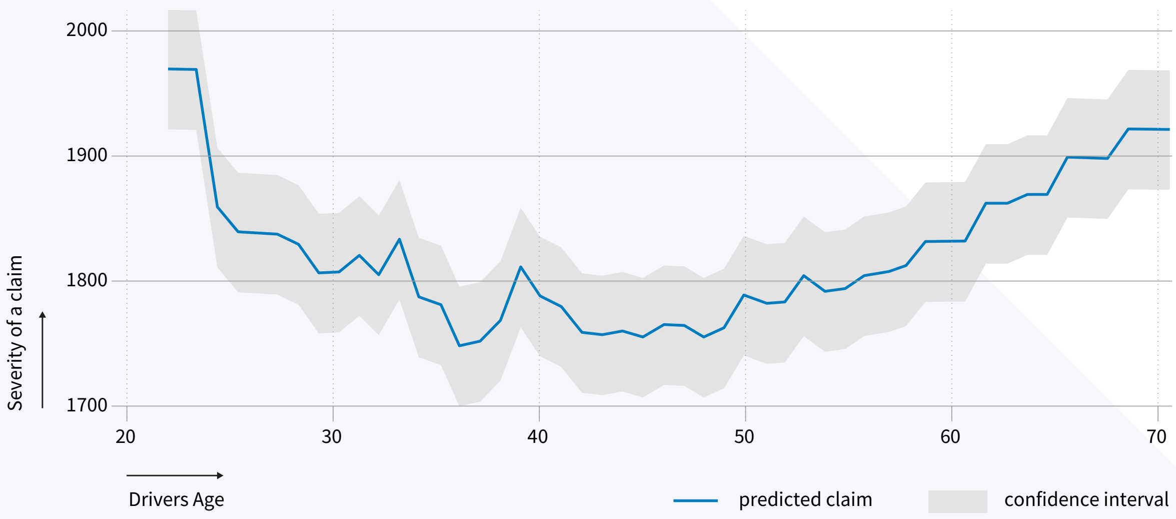
5.2 PARTIAL DEPENDENCY

The partial dependency plot (PDP) is a way to visualise the marginal effect of a feature on the predicted outcome of a ML model. While every statistical indicator (also the relativity plot for the classic GLM) has some assumptions, for the PDP the main one is the independence: the PDP assumes that the feature of interest is independent of other relevant features, except when explicitly varied in the plot.

An illustrative application of a PDP within the insurance context involves an exploration of how the probability or severity of a claim evolves concerning variations in a specific feature, such as the age of the policyholder. In this instance, we will employ a severity model, specifically a Random Forest model, to forecast the mean claim amount. This PDP serves as a visual tool to elucidate the relationship between the age of policyholders and the predicted average claim amount.

To create a PDP it is necessary first to fit a model. This could be a decision tree, random forest, or any other type of model. Once the model is trained, it can be used to make predictions on a grid of values for the feature of interest. For example, if we are looking at the effect of age on the predicted probability of a claim, we could make predictions for a range of ages from 20 to 70. Next, we plot the predicted severity of a claim on the y-axis and the age of the policy holder on the x-axis. This will show us how the predicted probability changes with age.

FIGURE 1



Source: Example performed using public available car dataset – French Use Case

The Figure 1 shows the relationship between the policyholder age and the predicted target, i.e., average claim severity, averaging out the effects of the other inputs. For interval inputs, the 95% confidence interval for the average target prediction is indicated by the shaded band around the line.

One limitation of the PDP is that it only shows the marginal effect of a single feature on the predicted outcome. It does not take into account the potential interactions between features. To understand the interactions between features, we would need to use more advanced techniques such as partial dependence plots with interaction terms.

From a mathematical point of view the structure of the PDP is a function that maps a single feature to the target variable. This function is typically constructed by fixing the values of all other features in the model, and then averaging the predicted values of the target variable for a range of values of the feature of interest.

5.3 INDIVIDUAL CONDITIONAL EXPECTATION (ICE)

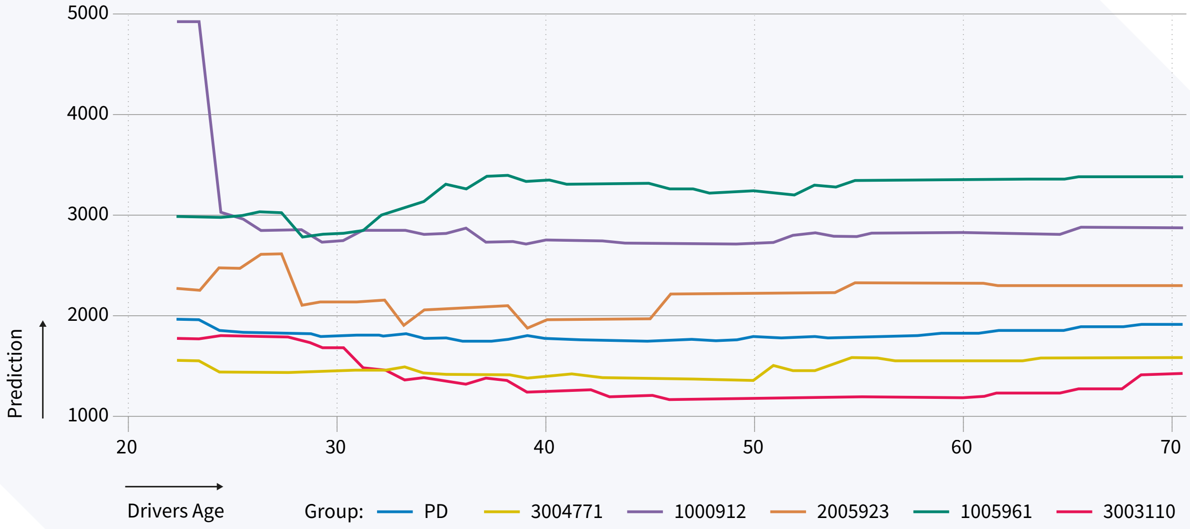
The individual conditional expectation (ICE) explainability indicator is based on the idea that a model is more interpretable if its predictions can be accurately explained by a small number of input features. An ICE plot visualises the dependence of the prediction on a feature for each instance separately, resulting in one line per instance, compared to one line overall in partial dependence plots. A PDP is the average of the lines of an ICE plot.

In the context of insurance the ICE can be used to evaluate the interpretability of a ML model that predicts the likelihood that a policyholder will file a claim based on his/her individual characteristics. To use the ICE, we first calculate the ICE curve for each input feature, which shows the relationship between the feature and the model’s predictions.

If the ICE curves are relatively simple and can be accurately explained by a small number of features, this indicates that the model is highly interpretable, as its predictions can be easily understood in terms of the input features. On the other hand, if the ICE curves are complex and cannot be accurately explained by a small number of features, this indicates that the model is less interpretable, as its predictions are less transparent and more difficult to understand.

Following the same example done above:

FIGURE 2: ICE PLOT



Source: Example performed using public available car dataset – French Use Case

The Figure 2 shows the partial dependency and the relationship between policyholder age and the predicted target for each individual observation.

5.4 SMALL PERTURBATION

The Small Perturbation explainability indicator is a method for evaluating the interpretability of a ML model. It is based on the idea that a model is more interpretable if small changes to the input data result in small predictable changes to the model's predictions.

To use the Small Perturbation explainability indicator, we first choose a small, fixed perturbation value. Next, we make small changes to the input data by adding or subtracting the perturbation value from each feature. For each perturbed input, we calculate the model's prediction and compare it to the original prediction.

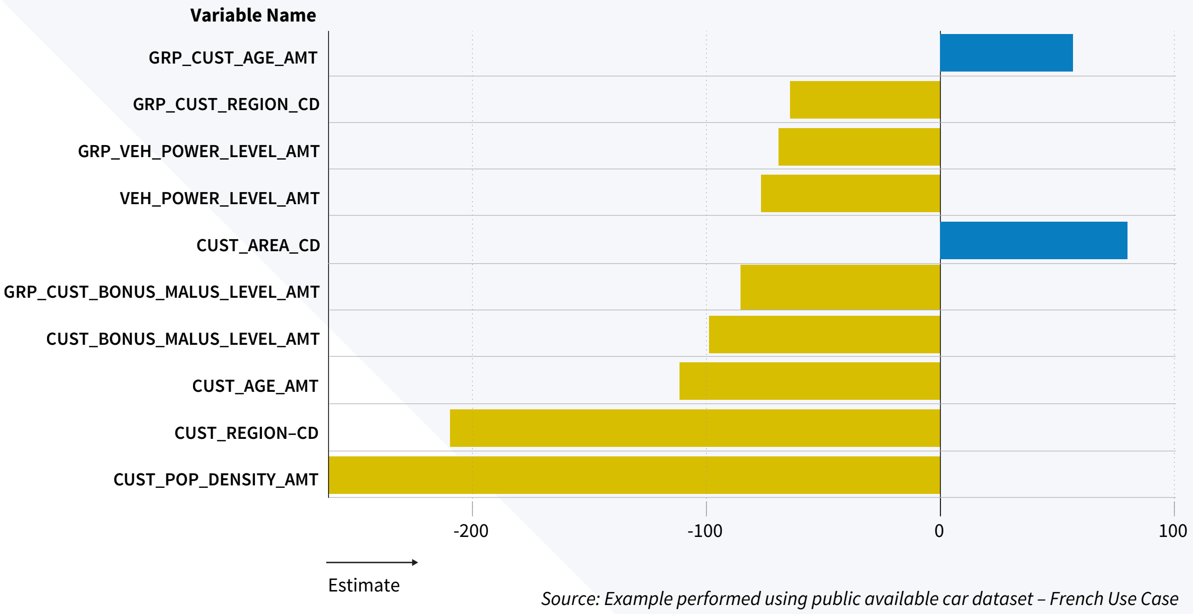
If the model's predictions are largely unchanged by small perturbations to the input data, this indicates that the model is highly interpretable, as small changes to the input data result in small, predictable changes to the model's predictions. On the other hand, if the model's predictions change significantly with small perturbations to the input data, this indicates that the model is less interpretable, as small changes to the input data can result in large, sometimes unpredictable changes to the model's predictions.

5.5 LOCAL INTERPRETABLE MODEL-AGNOSTIC EXPLANATIONS (LIME)

The local interpretable model-agnostic explanations (LIME) is a technique used to explain the predictions of machine learning models in a way that is interpretable to users. The basic idea behind LIME is to fit a simple, interpretable model to the neighbourhood around the prediction. Once the interpretable model has been fit it is used to explain the prediction. Following the same example used above for the Partial Dependence, we suggest to use a Random Forest model to predict the average claim amount.

We can use LIME to explain the model's prediction for a specific data instance, specifically in the graph reported below we see the LIME result for an instance whose predicted claim amount is 1.732€, the LIME results are:

FIGURE 3: LIME CHART



Source: Example performed using public available car dataset – French Use Case

The Figure 3 displays the regression coefficient (estimate) for the inputs selected in the local surrogate linear regression model for fitting the predicted value of the target Average Claim Amount for each individual observation. The inputs are ordered by significance in the chart (GRP stands for grouped), with the most significant input for the local regression model appearing at the bottom of the chart.

In the instance considered as example in the Figure 3 the most important variable, contributing to the prediction, is the CUST_POP_DENSITY_AMT (the population density where the policy holder lives), the second most important is CUST_REGION_CD (the region where the policy holder lives[8]) etc.

In this way, LIME allows us to explain the predictions of complex machine learning models in a simple, interpretable way. This can be useful for understanding the model's behaviour and for detecting any potential biases or errors in the model's predictions.

5.6 SHAPLEY ADDITIVE EXPLANATION SHAP

The Shapley additive explanation (SHAP) value for a given input feature in a ML model is derived using a method called the Shapley value from game theory. The Shapley value is a measure of the contribution of each player in a cooperative game and it has been adapted for use in machine learning to measure the contribution of each input feature to a model's predictions.

To calculate the SHAP value for a given input feature, the following steps are performed:

- The model's prediction is calculated for the input data.
- The average prediction of the model is calculated across all possible inputs.
- The difference between the model's prediction and the average prediction is attributed to the contribution of the input feature.

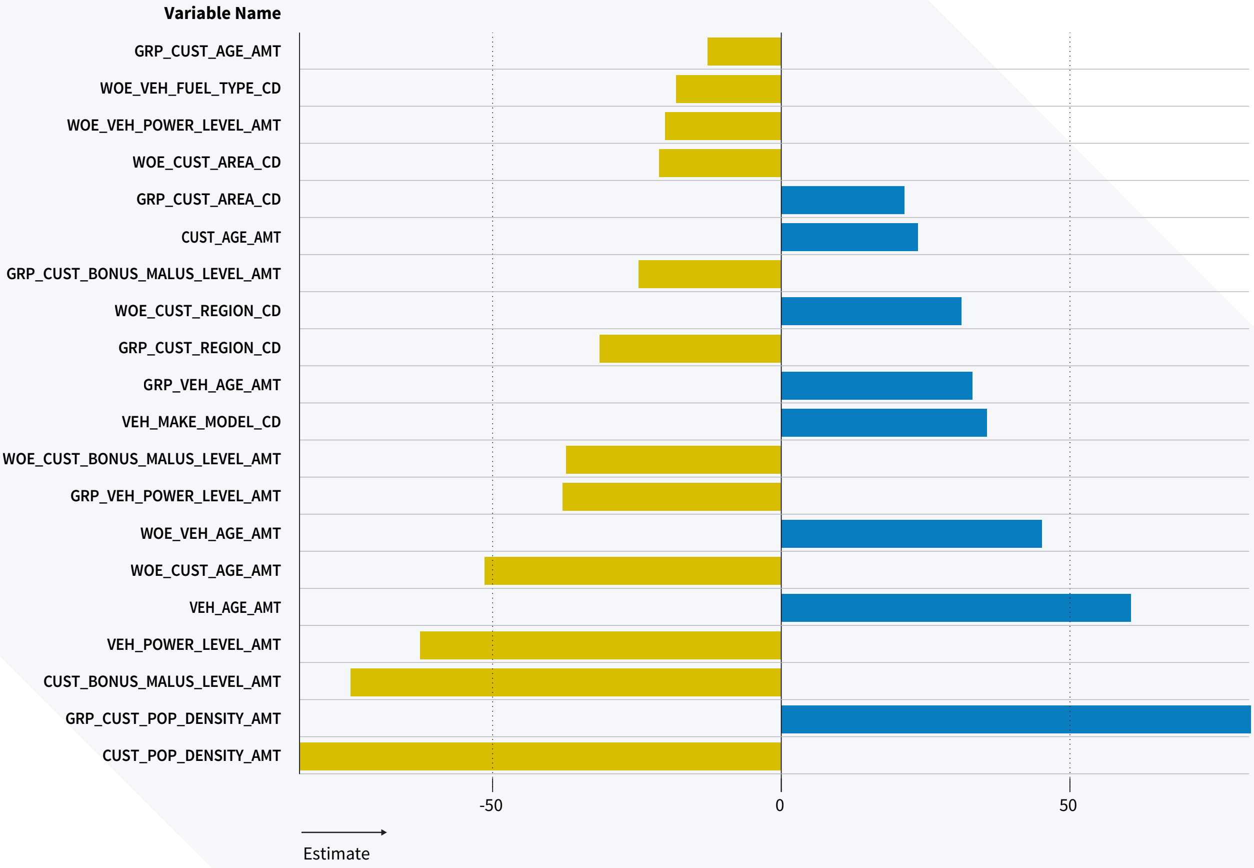
This process is repeated for each input feature to calculate the SHAP value for each feature. The SHAP values for the input features can then be used to evaluate the interpretability of the model, as described below.

In general, the SHAP value for a given input feature measures the degree to which the feature contributes to the model's predictions and can be used to evaluate the interpretability of the model. If the SHAP values for the input features are relatively small and consistent in direction, this indicates that the model is highly interpretable, as its predictions can be accurately explained by the contributions of individual input features. On the other hand, if the SHAP values are large and inconsistent in direction, this indicates that the model is less interpretable, as its predictions are less transparent and more difficult to understand.

[8] It is good to note here that, as neighbourhoods can have a strong link to ethnic background, we might have here an example of indirect discrimination based on ethnicity, or at least an ethnicity premium.

Still following the example considered in the previous paragraphs on the prediction of the average claim intensity, we see in Figure 4 an example of displaying the SHAP Value:

FIGURE 4: CHAP VALUE



Source: Example performed using public available car dataset – French Use Case

For each individual observation, an input's Shapley value is the contribution of the observed value of the input to the predicted value of the target AVERAGE_CLAIMS_AMT. The inputs are displayed in the chart ordered by importance according to the absolute values.

6 IMPLICATIONS OF AI AND DATA SCIENCE ON EDUCATION

Since the beginning of the actuarial science, actuaries have pioneered sophisticated mathematical techniques and employed innovative tools to enhance the precision when estimating uncertainty. The realm of AI encompasses various domains, and by enhancing the educational syllabus, actuaries can develop new skills, deepen their knowledge, and expand their proficiency. With the advent of ‘big data’ concepts and the application of data science techniques, as elucidated in earlier sections, there is a widespread effort among financial institutions to extract greater insights from data, and apply these findings to new domains.

Remarkably, AI spans across a diverse array of applications, delving into every facet of insurance value chain. Consequently, possessing a greater depth of knowledge than ever before and the ability to engage in interdisciplinary work are highly important.

Actuarial education in Europe, as given by the AAE’s core syllabus, aims at covering all elements that are considered by the IAA – the International Actuarial Associations’ syllabus. In both cases the revised Bloom’s Taxonomy classification (Krathwohl & Anderson, 2001), see Figure 5, is being used as a guidance to create and allocate correctly new courses.

FIGURE 5: BOOM’S TAXONOMY

Cognitive Knowledge	1. Remember	2. Understand	3. Apply	4. Analyze	5. Evaluate	6. Create
A. Factual	A1	A2	A3	A4	A5	A6
B. Conceptual	B1	B2	B3	B4	B5	B6
C. Procedural	C1	C2	C3	C4	C5	C6
D. Metacognitive	D1	D2	D3	D4	D5	D6

This taxonomy facilitates the definition of learning objectives by employing a verb, denoting an action (e.g., apply), and an object, typically a noun like Factual. The action signifies the cognitive process, while the object describes the intended knowledge acquisition for students. These levels provide ample flexibility to encompass various AI topics, tailor courses accordingly, and develop new ones that align with the specific requirements of a given audience. The existing syllabus of AAE already acknowledges Data and Systems as a predefined category. Within this category, subjects like data analysis, data visualisation, neural networks, and decision trees are already addressed.

Standard programs and continuing professional development (CPD) courses are challenged by rapid developments in AI and Data Science. This is not only caused by the emergence of new actuarial methods and by being able to extract more from data, but also from advances in other professional domains, such as IT. Actuarial institutes, universities, and other education providers to be aware of the developments and the possible integration of new courses that align with the needs and progress of AI and Data Science landscape.

To define and provide some examples, the applicability of AI can be seen through three main areas when considering the possible need to form a new syllabus and adapting the CPD programme.

Firstly, the widely discussed topic of **Data**, which has evolved from the general use of the term ‘Big Data’ to a more specific focus on areas within Data Science, such as ‘Data management’ and the ‘Use of alternative data’. Some examples of knowledge related to ‘Data’ that need to be considered include:

- Designing and structuring databases.
- Setting up data infrastructures, including administration of data servers.
- Data quality management.
- Methods of creating data models.
- Implementation of different data storage systems.
- Making use of external and alternative data, through external sources, or devices (IoT) to enhance calculations.

The second topic worth consideration falls under the category of **Technical modelling** which, of course, is something that all actuaries have in their repertoire. However, this time, it will involve acquiring new technical skills and a thorough understanding of the interaction between hardware and software components. Areas of interest might include:

- Imperative programming / Object Orientated Programming – through programming languages such as C#, Python, R etc.
- Implementing Machine Learning algorithms.
- Understanding and implementing Large Language Models.

- Understanding the hardware needed when deploying algorithms.
- Setting up digital twin environments.
- Use of advanced IT testing standards to ensure models are implemented correctly.
- Properly document actuarial models and algorithms – ensuring cross platform transferability.

Thirdly, the topic of **Visualisation and Reporting**, will help in understanding the data that is being used, and monitoring it, highlighting the results from algorithms and ensuring the business needs are met. Here are some examples:

- Building knowledge graphs – connecting data to business needs.
- Analysing data using visual methods, as an example see session 5.
- Automation and standardisation of reports, especially regulatory ones.
- Building dynamic dashboards, example: product pricing, financial reporting.
- Ensuring explainability and transparency of models.

Current syllabus, as implemented by the actuarial institutes at a local level in Europe, already considers some of the aspects mentioned above, covering all these three categories. However, many of the current courses involve the use of pre-existent tool packages, with pre-built functionality, and cover only an array of methods and algorithms. Referring to some of the topics mentioned above, and making use of Bloom’s taxonomy guidelines, new educational programmes could be considered. For instance, one might consider a course titled, ‘Insurance data modelling’ as part of the Data category, where two types of courses can be formed – one providing a more general overview and understanding – a B2 level, and one that focusses more on implementation and creation of data models – C6 for example.

As AI world is moving rapidly, we encourage actuaries to actively pursue learning opportunities to keep up with all the developments. We anticipate that the syllabus could undergo significant changes in coming years, helping the growth of knowledge actuaries need to absorb in order to be able to apply AI and Data Science methods correctly. Moreover, applying regulatory requirements rigorously, conducting activities ethically and creating transparency will help in keeping an open dialogue between all stakeholders, bringing new opportunities in our professional domain.

REFERENCES

EC. (2004). Council Directive 2004/113/EC. Publications Office of the European Union.

EIOPA. (2021). Artificial Intelligence Governance Principles Towards Ethical and Trustworthy Artificial Intelligence in the European Insurance Sector.

EIOPA. (2023). *Supervisory statement on differential pricing practices in non-life insurance lines of business.* https://www.eiopa.europa.eu/system/files/2023-03/EIOPA-BoS-23-076-Supervisory-Statement-on-differential-pricing-practices_0.pdf.

Krathwohl, D., & Anderson, L. W. (2001). *A revision of Bloom's Taxonomy of Educational Objectives Bloom's Taxonomy.*

Seehafer, M., Nörtemann, S., Offtermatt, S., Transchel, F., Kiermaier, A., Külheim, R. & Weidner, W. (2021). *Actuarial Data Science.* De Gruyter STEM.

Wikipedia. (s.d.). Data science as per November 2023. https://en.wikipedia.org/wiki/Data_science.

Wüthrich, M. V., & Merz, M. (2022). *Statistical Foundations of Actuarial Learning and its Applications.* Springer Actuarial.

FOR FURTHER READINGS: BOOKS AND PAPERS

Ajay Agrawal, Joshua Gans, Avi Goldfarb (2022), Prediction Machines - The Simple Economics of Artificial Intelligence, Harward Business Review Press.

Bradford Anu (2019), The Brussels Effect: How the European Union Rules the World, Oxford.

Bradford Anu (2023), Digital Empires: The Global Battle to Regulate Technology, Columbia.

Douglas Adams (1979 etc.), The Hitchhiker's Guide to the Galaxy, Pan Books.

EIOPA (2023). Supervisory statement on differential pricing practices in non-life insurance lines of business. supervisory statement (europa.eu).

Harvard University (2023). Explainable Artificial Intelligence - From Simple Predictors to Complex Generative Models, https://interpretable-ml-class.github.io/.

Henry A. Kissinger, Eric Schmidt, Daniel Huttenlocher (2022), The Age of AI, John Murray.

THE ACTUARIAL ASSOCIATION OF EUROPE

The Actuarial Association of Europe (AAE), founded in 1978 under the name of Groupe Consultatif Actuariel Européen, is the Brussels-based umbrella organisation, which brings together the 37 professional associations of actuaries in 36 countries of the EU, together with the countries of the European Economic Area and Switzerland and some EU candidate countries.

The AAE has established and keeps up-to-date a core syllabus of education requirements, a code of conduct and discipline scheme requirements, for all its full member associations. It is also developing model actuarial standards of practice for its members to use and it oversees a mutual recognition agreement, which facilitates actuaries being able to exercise their profession in any of the countries concerned.

The AAE also serves the public interest by providing advice and opinions, independent of industry interests, to the various institutions of the European Union - the Commission, The Council of Ministers, the European Parliament, ECB, EIOPA and their various committees - on actuarial issues in European legislation and regulation.

aae
actuarial association of europe

ACTUARIAL ASSOCIATION OF EUROPE

SILVERSQUARE NORTH
BOULEVARD ROI ALBERT II 4
B-1000 BRUSSELS
BELGIUM
☎ +32 2 274 06 61
✉ INFO@ACTUARY.EU
WWW.ACTUARY.EU