CONSULTATIVE EXPERT GROUP REPORT ON DATA USE IN THE INSURANCE SECTOR

17 November 2026

European Insurance and
Occupational Pensions Authority

CONTENTS

EXECUTIVE SUMMARY **4**

1. INTRODUCTION **7**

2. HOW INSURANCE WORKS **9**

2.1. Why insurance matters 11

2.2. The role of data in insurance 11

2.2.1. Why data matters in insurance 11

2.2.2. From traditional to data-driven insurance 13

3. DATA ETHICS **17**

3.1. From compliance to ethics in data use 17

3.2. Key ethical challenges in data use 18

3.2.1. Solidarity 18

3.2.2. Personalisation and mutualisation 20

3.2.3. Inclusion 22

3.2.4. Causality and correlation 24

3.2.5. Human autonomy, privacy and data security 25

3.2.6. Discrimination 26

4. SCENARIO ANALYSIS **28**

4.1. Purpose and approach 28

4.2. Data use scenarios 28

Limited data use scenario 29

Extensive data use scenario 29

Impact assessment dimensions 31

Scenario impact assesment 32

4.3. Use case A: Telematics in motor insurance 35

4.4. Use case B: Wearables in health insurance 37

5. CONCLUSIONS **39**

5.1. Recommendations for policymakers 39

5.2. Recommendations for supervisors 44

GLOSSARY **45**

ANNEX I – LEGAL FRAMEWORK **47**

ANNEX II – MEMBERS OF EIOPA'S CONSULTATIVE EXPERT GROUP ON DATA USE IN INSURANCE **52**

EXECUTIVE SUMMARY

Insurance relies on data to pool and differentiate risks, price uncertainty, design products, manage claims and capital, and provide services throughout the customer lifecycle. Technological progress has expanded the volume, variety, granularity and speed of information available to insurers. Traditional sources, including demographic information, claims histories and exposure data, are increasingly complemented by telematics, connected devices, digital interactions, and environmental and climate data. Together with greater computing capacity and more sophisticated analytical methods, these developments are changing how insurance products are designed, priced, distributed and serviced, and how data flows across the insurance ecosystem.

Against this background, EIOPA established the Consultative Expert Group on Data Use in Insurance, bringing together senior experts from insurance undertakings, pensions and reinsurance, industry and professional associations, academia, consumer organisations, civil society and consultancy firms. This report presents the outcome of the Group's work. It examines the opportunities, risks and trade-offs associated with increasing data use in private insurance and provides a principles-based contribution to the policy debate. The report focuses on data use rather than risks specific to artificial intelligence (AI), such as process automation, insufficient human oversight or model flaws. It nevertheless recognises that data, advanced analytics and AI are closely related. AI can process large volumes of complex information, while the quality, relevance, representativeness, governance and permitted use of data shape system performance and outcomes. AI is therefore considered where necessary to understand how it may amplify the benefits or risks of data use.

The report first examines data use in the sector from three complementary perspectives: the evolution from traditional to emerging data sources, data flows and interdependencies across the insurance ecosystem, and data use throughout the customer lifecycle. This provides a structured basis for identifying how data is transforming insurance activities, where opportunities and risks arise, which actors may be affected, and how implications may differ across the value chain and customer journey.

Building on this foundation, the report considers interconnected ethical questions concerning solidarity, personalisation, inclusion, fairness, autonomy, privacy, security and discrimination. The report recognises that these questions cannot be addressed solely through insurers' underwriting, pricing and data-use practices. The balance between risk differentiation and broader social objectives is also shaped by legislators, governments, regulators, supervisors and social security systems. Ethical outcomes therefore emerge from the interaction between market practices and public-policy choices.

A central ethical tension concerns solidarity. By pooling uncertain losses, insurance protects individuals and businesses against financial consequences they might otherwise be unable to bear. This reflects chance solidarity, since policyholders contribute before it is known whose risks will materialise. Insurance may also involve subsidising solidarity, where premiums do not fully reflect differences in expected risk and some groups support others. More granular data can improve risk differentiation, but may reduce subsidising solidarity and affect affordability and access. It does not eliminate chance solidarity or the need for pooling, since even accurate estimates cannot determine whether, when or how severely a loss will materialise for an individual.

This tension is particularly visible in personalisation. Better information may make previously difficult risks insurable and support more accurate pricing, prevention and risk mitigation. Greater granularity may, however, leave higher-risk, vulnerable or data-poor consumers facing higher premiums or exclusion. It is also important to distinguish risk-based pricing from differential pricing based on willingness to pay, customer inertia or characteristics unrelated to insured risk, which may raise concerns regarding transparency and fairness.

The relationship between data use and inclusion is similarly complex. Data-driven models may reproduce existing inequalities, facilitate the exclusion of higher-risk consumers or disadvantage those with limited digital footprints. Conversely, better data may support prevention and more tailored products, thereby extending access for some consumers. Inclusion should therefore be assessed by outcomes, including who benefits, who bears the costs, whether meaningful alternatives remain available, and whether improved access creates disproportionate disadvantages elsewhere.

Continuous and granular data collection also affects autonomy, privacy and data security. Telematics and wearables may support prevention and risk mitigation, but consumers may face pressure to share behavioural or health-related information where non-participation results in material disadvantages. Responsible data use therefore requires meaningful choice, purpose limitation, data minimisation, appropriate retention, robust security and clear accountability. Legal compliance is an essential baseline, but insurers should also consider whether their practices are proportionate, fair and consistent with consumers' reasonable expectations.

A scenario analysis brings these trade-offs together by comparing two stylised settings characterised by limited and extensive data use. The scenarios are not forecasts or preferred models, but a framework for assessing how differences in data availability, analytical capability, infrastructure, consumer acceptance and governance may affect consumers, insurers, supervisors and society. Two use cases (telematics and wearables) further illustrate these trade-offs.

Under limited data use, insurers rely mainly on traditional information and interpretable methods. This may support simpler products, broader pooling, easier explainability and lower privacy, cybersecurity and governance exposure. However, it may create blind spots, reduce the precision

of risk assessment, increase adverse selection, constrain innovation and limit the insurability of emerging or complex risks.

Under extensive data use, insurers combine granular, high-frequency information with advanced analytics and interconnected infrastructure. This may strengthen risk assessment, prevention and mitigation, efficiency, personalisation and capital allocation, while potentially extending coverage to some consumers. However, it may heighten privacy, cybersecurity, model, governance and third-party risks, increase pressure to share data and affect affordability or access for higher-risk and data-poor consumers. Competition effects are also mixed: wider data use may enable innovation, but advantages in data access, infrastructure and analytical capabilities may reinforce the position of larger firms, incumbents and platform operators.

Neither scenario produces uniformly positive outcomes. The policy objective should therefore be neither to maximise nor to minimise data use, but to ensure that data-driven innovation delivers better outcomes for consumers and society without compromising trust, fairness, insurability or financial resilience. Achieving this balance requires a calibrated approach that remains responsive to changes in markets, technologies and societal expectations.

To guide this approach, the report identifies five interrelated principles: access to insurance; fair and affordable outcomes; integrity, transparency and meaningful choice; prudential soundness; and safe progress. Together, these principles provide a proportionate and outcomes-focused framework for distinguishing legitimate and beneficial uses of data from those that may produce harmful or inappropriate outcomes. They also underpin the report's recommended policy approach. Policymakers should consider the social importance of the affected insurance product, the purpose and materiality of the data use, and its effects on consumers, insurers and society, with safeguards tailored to the relevant line of business and the nature and sensitivity of the data involved.

Drawing on the principles and trade-offs identified, the report supports a proportionate, outcomes-focused approach to incorporating data use into conduct supervision. This approach should build, where appropriate, on existing regulatory and supervisory frameworks while avoiding unnecessary administrative burdens. Supervisory activities may include monitoring exclusion and affordability and challenging insurers' data-use impact assessments across pricing, underwriting, claims handling and distribution. Attention should also be given to data-related aspects of explainability, bias and fairness testing, validation, meaningful human oversight, outcome monitoring and redress. As reliance on third-party data providers, AI systems and digital ecosystems increases, supervisors should monitor the resulting dependencies and ensure that insurers remain accountable for the products, services and decisions supported by these arrangements.

1. INTRODUCTION

Insurance has always relied on data to pool risks, price uncertainty and pay claims. In recent years, greater computational power, new and more granular data sources, rapid technological progress and increasingly sophisticated analytical tools have expanded insurers' ability to process data.

The continued expansion of data use could improve risk assessment, prevention and mitigation, product design, personalisation, financial inclusion, and risk management, thereby supporting better consumer outcomes, sector resilience and financial stability. At the same time, greater data use raises concerns about its potential impact on individuals, including issues that go beyond regulatory compliance and extend into the realm of ethics (such as discrimination, financial exclusion of vulnerable groups, reduced solidarity due to greater personalisation, and increased exposure to data protection risks)[1]. Additionally, adverse outcomes may also arise where insurers fail to make adequate use of available data, as they may lose competitiveness and miss opportunities to innovate, improve products and services, and better serve consumers. The critical role of insurance in society and its potential impact on individuals warrants careful consideration as data use continues to expand. Regulators and industry stakeholders therefore need to balance any benefits stemming from advanced data use with the protection of individuals at risk, while preserving fair, inclusive and competitive insurance markets.

To support its work on data use and digitalisation, EIOPA set up a Consultative Expert Group (CEG) to reflect on these developments and explore emerging dilemmas in the use of data in private insurance. The CEG brings together senior experts from insurance undertakings, pensions and reinsurance, industry associations, academia, professional associations, consumer organisations, civil society and consultancy firms. This report presents the outcome of its work, setting out key foundations, challenges, scenarios, findings and recommendations, while recognising the difficulty of drawing clear boundaries between beneficial and harmful uses of data. It also considers different forms of data processing and the transformative impact of artificial intelligence (AI), including the associated ethical considerations, building on questions explored in greater depth in EIOPA's report on AI governance principles.[2]

Given the evolving nature of both technology and the regulatory landscape, this report does not seek to assess the effectiveness of existing legislation, supervisory frameworks or regulatory guidance, nor does it provide a comprehensive review of applicable legal requirements. Rather, it

[1] See: Mullins, M., Holland, C.P. and Cunneen, M., 2021. "Creating ethics guidelines for artificial intelligence and big data analytics customers: The case of the consumer European insurance market". *Patterns*, *2*(10). DOI: 10.1016/j.patter.2021.100362

[2] Artificial intelligence governance principles: towards ethical and trustworthy AI in the European insurance sector – A report from EIOPA's Consultative Expert Group on Digital Ethics in insurance (EIOPA, 2021)

offers a principles-based analysis intended to support reflection on key issues, trade-offs and emerging risks associated with the increasing use of data, advanced analytics and AI in insurance. Recognising that technological developments often progress more quickly than formal regulatory frameworks, the report also highlights the role that governance practices, professional standards and industry good practices can play in promoting responsible and trustworthy data use in the sector. As such, the report should be viewed as a contribution to the ongoing discussion rather than a definitive assessment of a rapidly evolving field.

The analysis also recognises that insurance products, business models and consumer expectations are not static. Technological developments, new data sources and evolving risk landscapes continuously shape how insurance products are designed, priced, distributed and serviced. Therefore, assessments of data use should consider not only current market practices but also how innovation may change the nature of insurance products and consumer needs over time.

2. HOW INSURANCE WORKS

Insurance provides financial protection against uncertain events affecting people, property and savings. In exchange for premiums paid by policyholders, an insurer undertakes to cover certain costs or pay predefined benefits if specified events occur, such as illness, accidents, death or property damage. Insurance is commonly divided into life insurance, which covers risks linked to human life, such as death or disability, and non-life insurance, which covers risks such as property, motor and liability. Depending on the market, health insurance may fall under either category.

Insurance may be provided through social or private arrangements. Social insurance typically refers to insurance schemes established to achieve social protection objectives, namely providing coverage against fundamental risks, such as illness, unemployment and disability. Such schemes are typically mandatory, generally organised under a statutory framework, and may be financed through general taxation or through compulsory contributions earmarked for specific risks and benefits. While linked, the term public insurance is related but not synonymous. It refers primarily to the institutional nature of the provider or administrator of the scheme rather than its underlying purpose. Social insurance arrangements often arise where certain forms of protection are considered socially essential or where private markets alone cannot provide adequate coverage.

Private insurance, by contrast, is usually voluntary and provided by private undertakings operating in competitive markets, although some forms of cover may be mandatory. In practice, the distinction between social, public and private insurance is not always clear-cut, as many national systems combine elements of each.

At its core, insurance is a mechanism for pooling and sharing risk. Policyholders contribute premiums into a common pool before it is known who will suffer a loss. Those whose risks materialise are compensated from the pooled resources. This pooling of uncertain losses allows individuals and firms to transfer potentially severe financial consequences to a broader group and is one of the principal economic and social functions of insurance. Within this report, this mechanism is referred to as risk mutualisation, namely the process through which risks and their financial consequences are shared across a pool of policyholders. Rather than each individual bearing the full cost of losses they may incur, the costs of realised losses are distributed across the members of the pool through the payment of premiums.

While risk mutualisation describes the fundamental mechanism through which insurance pools and shares risk, it does not by itself determine how costs are distributed among policyholders or the extent to which different groups support one another within the pool. To better understand these different dimensions of risk sharing, it is useful to distinguish between chance solidarity, subsidising

solidarity and risk mutualisation. Risk mutualisation describes the pooling mechanism itself. Chance solidarity refers to the collective sharing of uncertain losses among members of the pool, since no policyholder knows in advance whether they will suffer a claim. Subsidising solidarity, by contrast, arises when premiums do not fully reflect differences in expected risk, resulting in transfers between lower-risk and higher-risk groups. While subsidising solidarity is often an explicit objective of social insurance systems, it may also arise in private insurance due to regulatory requirements, market conditions, data limitations, commercial considerations, or broader societal objectives.

For these mechanisms of risk sharing to remain sustainable over time, insurers must also be able to assess and manage the risks they assume. The viability of insurance therefore depends not only on how risk is pooled and shared across policyholders, but also on insurers' ability to estimate the likelihood and cost of future claims. This requires information. By using available information to distinguish between different risk profiles, insurers can set premiums, design products and manage their exposures. However, risk assessment is always conducted under uncertainty. Data may be incomplete, emerging risks may be difficult to model, and future events remain inherently unpredictable. As a result, premiums do not always perfectly reflect individual risk. Thus, insurance requires a balance between risk sharing and risk differentiation. Maintaining this balance is also important for limiting adverse selection and supporting the long-term sustainability of insurance pools, a topic discussed further in subsequent sections.

This balance is shaped not only by actuarial considerations, but also by commercial and regulatory factors. Given the inverted nature of its production cycle[3], the insurance sector relies on a rigorous regulatory and prudential framework, exemplified by Solvency II[4], to secure long-term commitments to policyholders. Moreover, while accurate risk assessment is important in private insurance, pricing and underwriting decisions are not determined by risk considerations alone. Market conditions, competition and customer needs also influence how insurers translate risk assessments into products and prices. Insurers must likewise meet regulatory requirements to ensure they can cover insured losses and withstand other adverse developments.

The way insurers assess and differentiate risk is also evolving. Advances in technology, analytics and data availability are enabling insurers to use increasingly granular information when assessing risks and making decisions. These developments create opportunities to improve risk assessment, risk prevention and mitigation as well as product design, but they also raise questions about the appropriate balance between risk differentiation, mutualisation, inclusion, data collection and use and affordability. Understanding this balance is essential for understanding the opportunities and

[3] The inverted production cycle of insurance means that the product is purchased and the premium is collected before the covered events occur and the ultimate cost of claims becomes known.

[4] Directive 2009/138/EC of the European Parliament and of the Council of 25 November 2009 on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II)

challenges associated with increasing data use in insurance, which are explored in the following sections.

2.1. WHY INSURANCE MATTERS

Private insurance plays an important economic and social role by helping individuals and businesses manage financial risks. It provides a safety net and prevents individual losses from turning into wider economic problems. By providing protection against uncertain events, it reduces the need for households and firms to maintain large precautionary savings and enables economic activity that might otherwise be considered too risky or unaffordable.

- **Life insurance** helps households manage the financial consequences of events such as death or disability, providing financial security and supporting long-term planning.
- **Health insurance** helps cover medical expenses that would otherwise be borne directly by the individual, therefore preventing medical bankruptcies (by covering gaps in public health systems that do not fully cover all expenses). Additionally, by providing access to private healthcare networks, it may help alleviate long waiting times in public health systems.
- **Non-life insurance** protects against a broad range of risks affecting property, including vehicles, and liability exposures, among others. For example, motor insurance covers the financial consequences of accidents involving vehicles. By doing so, it makes vehicle ownership viable, as a single accident could otherwise expose a household to liabilities far exceeding its financial capacity. Home insurance protects households against losses arising from events such as fire, theft or natural hazards and also plays a role in enabling access to a mortgage, as banks typically require it before granting a loan.

Through these functions, insurance supports financial resilience for individuals and businesses, contributes to economic stability and investment, and reduces the wider economic impact of unexpected losses. Given this societal role, preserving accessibility and affordability is important. The following sections examine how data supports the functioning of insurance and how emerging forms of data use may create both opportunities and challenges for insurers, consumers and society.

2.2. THE ROLE OF DATA IN INSURANCE

2.2.1. WHY DATA MATTERS IN INSURANCE

Insurance is fundamentally a mechanism for pooling and sharing risk. The sustainability of this mechanism depends not only on risk mutualisation, but also on insurers' ability to assess and differentiate risks. Data therefore lies at the core of insurance. It enables insurers to understand the

risks they assume and supports a wide range of activities, including product design, underwriting, pricing, claims handling, reserving, and capital management, among others.

Since insurance concerns uncertain future events, among others, data serves two interconnected purposes. First, it helps insurers to estimate the probability and potential cost of future claims. Second, it enables them to distinguish between different risk profiles and determine whether different products, premiums, terms or risk-management approaches may be appropriate.

Traditionally, insurers have relied on relatively stable and well-established sources of information, such as demographic characteristics, claims histories, medical information, property characteristics or driving records. However, technological progress has expanded the volume, variety and availability of data that can be used in insurance. As a result, traditional approaches are increasingly being complemented by new forms of data, advanced analytics and AI-driven techniques that can process information at a much greater scale and level of granularity.

These developments create opportunities: more granular information may allow insurers to distinguish between risks that would otherwise be treated as uniformly high-risk, potentially extending coverage to individuals or communities that might previously have been excluded. For example, in the context of natural catastrophes, satellite imagery, climate models, and real-time sensor data can enable more precise assessments of exposures. By moving from broad geographic exclusions to precise, property-level risk assessment, data can extend insurance coverage to communities and households that would otherwise remain uninsured. Similarly, continuous monitoring and real-time data may allow insurers and their policyholders to anticipate and prepare for extreme events via early warning systems (e.g. alerts or loss mitigation recommendations). Furthermore, data may also improve operational efficiency and claims management, including fraud detection and verification processes.

At the same time, the increasing use of data also raises new challenges: More detailed information can lead to more granular segmentations of risks, potentially affecting affordability and access to coverage. Greater reliance on personal and behavioural information may raise concerns regarding privacy, autonomy, transparency and discrimination[5]. Reliance on granular or incomplete data can misrepresent risk, while proactive risk programmes may limit autonomy or disadvantage lower-income consumers and high-risk customers. Data collection also increases privacy and cybersecurity risks, and firms that leverage data at scale may gain outsized market power, potentially reducing competition. The use of AI for processing information and computational tasks adds an additional layer of complexity, and while AI-specific risks—such as process automation, lack of human oversight, or potential model flaws—are not examined here, as this report focuses on data use, they

[5] See for example: Holland, Christopher and John, Daniel, "HUK-COBURG: The Implementation of an AI-Enabled Behavioural Insurance Business Model using Geo-Spatial Data" (2023). Hawaii International Conference on System Sciences. https://aisel.aisnet.org/hicss-56/li/research/4

should not be overlooked and may further amplify both the benefits and risks associated with increasing data use[6].

The way data is collected, analysed and used therefore influences who can access insurance, how risks are assessed and priced, how costs are distributed across policyholders, and the extent to which risks remain mutualised within an insurance pool. As data availability and analytical capabilities continue to expand, these questions become ever more important and form the basis for the ethical, supervisory and policy considerations explored in the following sections.

2.2.2. FROM TRADITIONAL TO DATA-DRIVEN INSURANCE

Technological progress has expanded both the availability of data and insurers' ability to analyse it. The growing use of data also means that information now flows through insurance markets in new ways and involves a wider range of actors, technologies and data sources than in the past. Understanding these data flows is essential for assessing both the opportunities and risks associated with increasing data use in insurance. The following subsections therefore examine insurance data from three complementary perspectives: data sources, the insurance ecosystem, and the customer lifecycle.

Perspective 1. Data sources: from traditional to emerging forms of data

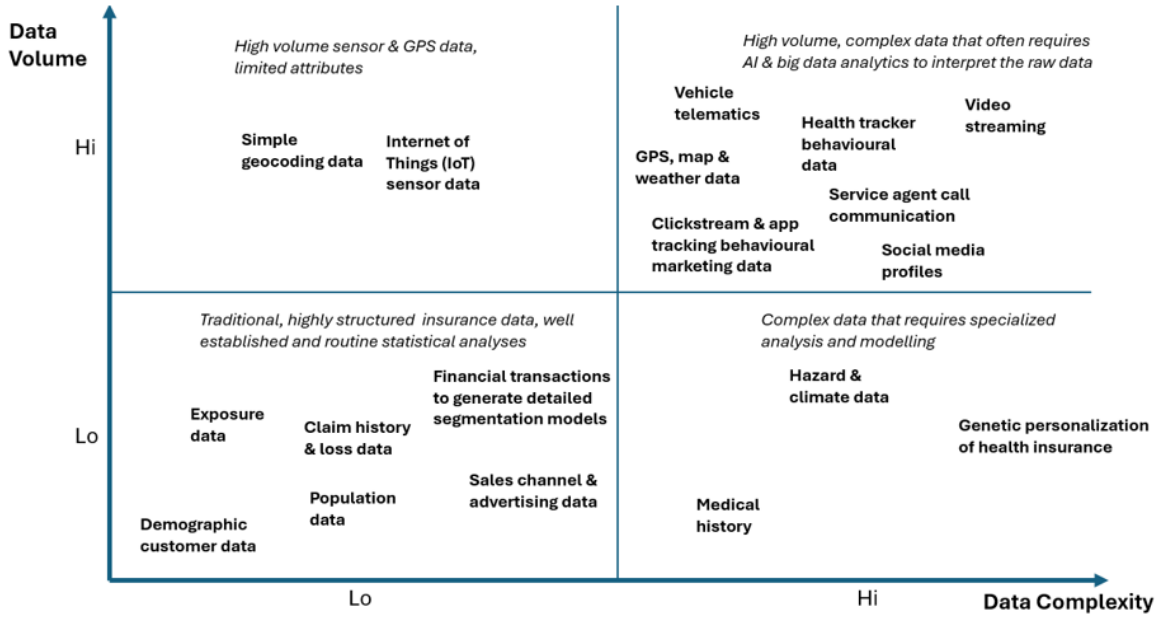
As mentioned, insurance relied primarily on relatively stable information collected when a policy was purchased and updated only occasionally. This data allowed insurers to assess risks, set premiums and manage their portfolios, but provided only a largely static view of policyholders and insured risks. These so-called "traditional data sources" are well-established forms of information that have been used in insurance for decades, including demographic characteristics, claims histories, medical information and property characteristics. They are typically structured, relatively limited in volume, and easy to interpret. These data are generally updated only when there is a change in the policyholder's circumstances, a change in the insured risk, the occurrence of a claim, or, in some cases, not at all. For example, a motor insurance claims history usually consists of a record of claims submitted over previous years, while demographic data may include characteristics such as age, address, gender, driving experience, occupation and education.

The emergence of connected devices, digital platforms and new analytical capabilities has expanded both the volume and variety of data available to insurers. Compared with traditional insurance data, these newer forms of information can provide a more detailed and dynamic view of policyholders, insured assets and risks, as data can now be generated continuously. Today, insurers can complement traditional sources of information with data generated through connected devices

[6] For information on AI-specific risks, see for example: Artificial intelligence governance principles: towards ethical and trustworthy AI in the European insurance sector – A report from EIOPA's Consultative Expert Group on Digital Ethics in insurance (EIOPA, 2021)

(such as vehicle telematics, sensors or wearable devices), digital services and online interactions. As a result, data is becoming more granular, more frequent and, in some cases, available in near real time. However, it is also more complex to process and interpret, making it increasingly important to understand how information is collected, stored and analysed[7]. As shown in **Figure 1,** one way to characterise this evolution is to distinguish insurance data according to its volume and complexity.

Figure 1. Typology of insurance data by volume and complexity



Source: C. P. Holland (2026), Typology of Insurance Data, author's original research and analysis.

The bottom-left quadrant of **Figure 1** represents traditional insurance data, which is typically low in volume and complexity, structured and relatively easy to analyse using well-established statistical techniques. Examples include demographic data, claims histories and exposure information. At the opposite end of the spectrum, the top-right quadrant contains data that is both high in volume and high in complexity. Examples include vehicle telematics, wearable-device data, video streams and social media data, which often require advanced analytical techniques, including AI and big data analytics, to extract meaningful insights. The remaining quadrants illustrate that volume and complexity do not always evolve together. Some data sources generate large volumes of relatively simple information, such as sensor or GPS data, while others—such as detailed medical histories,

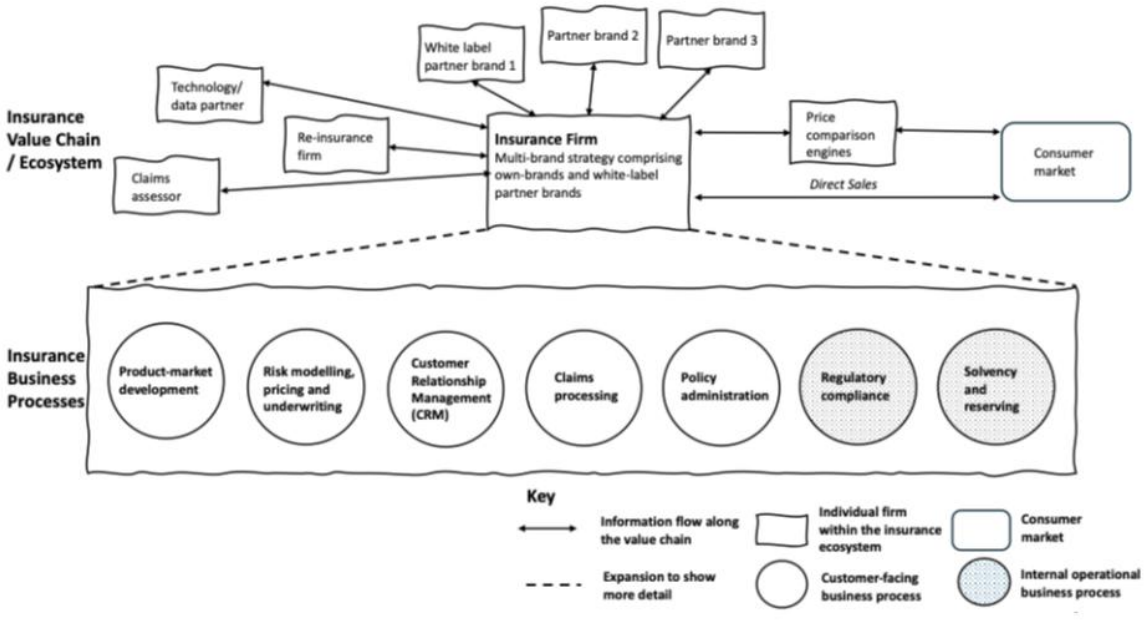
[7] EIOPA's thematic review on Big Data Analytics in motor and health insurance, highlights insurers' increasing use of new forms of data, advanced analytical techniques and digital technologies across the insurance value chain. The review also illustrates how these developments are creating new opportunities for product design, risk assessment, prevention and customer engagement.

genetic information, or hazard and climate data—may be more limited in volume but require sophisticated modelling because of their complexity.

Perspective 2. The insurance ecosystem: how data flows across insurance markets

The growing use of data is changing not only the type of information used by insurers, but also how information is generated, shared and used across the insurance ecosystem. **Figure 2** shows an illustrative diagram based on an analysis of key insurance business processes by IBM[8] and adapted to include economic partners[9]. The main functional areas within an insurance firm are shown in the large rectangle, grouped into customer-facing and internal business processes. The external value chain is captured by the broader context of relationships with economic agents such as partner brand organisations, reinsurance firms, customers, and distribution channels. The term "insurance ecosystem" is used to emphasise that it is not a linear value chain, but a network of separate companies that collaborate to form an ecosystem offering insurance services to customers. Overall, insurance increasingly relies on a network of interconnected actors that may include policyholders, insurers, reinsurers, distributors, technology providers, data providers and other specialised partners. As a result, data now flows through insurance markets in more complex ways than in the past.

Figure 2. Data flows in the insurance ecosystem



Source: C. P. Holland (2022)

[8] IBM (2016) *IBM Insurance Process and Service Models, General Information Manual*.

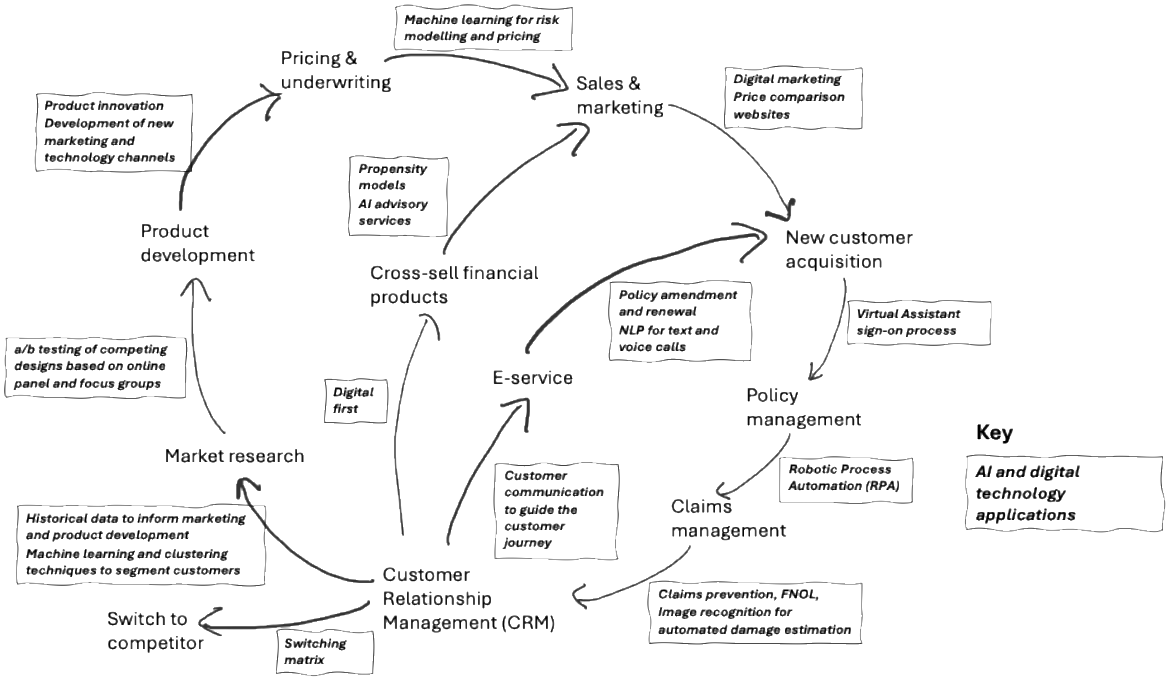
[9] Holland, C.P. (2022) 'Artificial Intelligence (AI) and Digital Transformation in the Insurance Market: A Case Study Analysis of BGL Group', *Proceedings of the 55th Hawaii International Conference on System Sciences*. Available at: https://doi.org/10.24251/HICSS.2022.553.

Perspective 3. The customer perspective: data throughout the insurance lifecycle

A third perspective focuses on the customer journey. Insurance involves a series of interactions between the customer and the insurer, ranging from product design and distribution to underwriting, policy administration, claims handling, renewal and, ultimately, switching providers. Data plays an important role throughout this lifecycle, supporting decision-making and service delivery at each stage.

Traditionally, many of these interactions relied on information collected at specific points in time, such as when a policy was purchased or a claim was submitted. New forms of data allow information to be updated more frequently and, in some cases, continuously. As a result, data is progressively supporting a more continuous relationship between insurers and policyholders, influencing how insurance products are designed, delivered and serviced over time. **Figure 3**, based on case study research, illustrates this customer lifecycle perspective by showing how information can support interactions throughout the insurance customer journey[10].

Figure 3. **Customer lifecycle model of information**



Source: C. P. Holland (2022)

[10] Ibid.

3. DATA ETHICS

3.1. FROM COMPLIANCE TO ETHICS IN DATA USE

Given the societal role of insurance, there is an expectation that the sector operates with high standards of conduct, professionalism, and integrity. These expectations apply particularly to the use of data, which, as discussed above, plays a central role in insurance and is shaping the sector's evolution.

These expectations are reflected in a comprehensive regulatory framework. Binding rules, such as the General Data Protection Regulation (GDPR)[11], the AI Act[12], the Solvency II Directive[13] and the Insurance Distribution Directive (IDD)[14], establish requirements relating to data processing and protection, governance, transparency, discrimination and consumer protection. Alongside these legal obligations, soft law instruments, such as supervisory guidelines, recommendations, codes of conduct and industry standards, also influence how data is used in practice by providing additional interpretation, guidance and expectations beyond legally binding requirements[15] (see Annex I).

Compliance with legal and regulatory requirements provides an essential baseline, but it does not always resolve every question arising from the use of data. Data ethics therefore constitutes a distinct, complementary dimension that focuses on the broader impact of data collection, processing and use on individuals and society. In particular, data ethics is relevant where regulation sets objectives or principles but leaves room for interpretation in implementation. In such cases, organisations may choose to do the minimum necessary to remain compliant, or they may seek to uphold both the letter and the spirit of the rules. Ethical considerations help inform these choices.

The boundaries between hard law, soft law, and data ethics are not fixed; rather, they are shaped by interpretation and application and are subject to ongoing reflection and refinement. This nuanced approach recognises that the use of data in insurance is influenced by a combination of legal, ethical, and societal considerations and therefore requires a balanced framework—one that protects policyholders' rights and interests while allowing insurers to responsibly leverage data-driven insights.

[11] Regulation (EU) 2016/679 (GDPR)
[12] Regulation (EU) 2024/1689 (Artificial Intelligence Act)
[13] Directive 2009/138/EC (Solvency II)
[14] Directive (EU) 2016/97 (Insurance Distribution Directive, IDD)
[15] Artificial intelligence governance principles: towards ethical and trustworthy AI in the European insurance sector – A report from EIOPA's Consultative Expert Group on Digital Ethics in insurance (EIOPA, 2021)

In insurance, ethical questions arise because underwriting and pricing require differentiation between risks, which frequently means differentiating between individuals. While some degree of risk differentiation is necessary for insurance to function, important questions remain regarding how such differentiation should be carried out and what limits should apply.

At first sight, determining what constitutes an ethical approach on the use of data might seem straightforward (i.e., it might seem easy to distinguish between ethical and unethical practices). In practice, however, ethical assessments are more complex. While it is important to balance the interests of different stakeholders, it is equally important to recognise that ethical valuations, and the resulting policy choices, may differ depending on the choice among different ethical frameworks or schools of thought in ethics. For example, deontological approaches (a rules-based approach like Kantian ethics) may lead to different, and at times seemingly conflicting, conclusions than consequential approaches (which focus primarily on outcomes and impacts). Both approaches are reflected in EU regulation. For instance, the ban on the use of gender in insurance pricing is typically a result of applying a deontological approach, whereas the Product Oversight and Governance process of the IDD results from a consequential assessment of consumer outcomes. Therefore, discussions on ethics benefit from clarity on the approach chosen, as otherwise different stakeholders will find it difficult to find any common ground.

3.2. KEY ETHICAL CHALLENGES IN DATA USE

Many of the ethical tensions discussed in this section should not be understood solely as questions for insurers to resolve through their underwriting, pricing or data-use practices. The balance between risk differentiation, solidarity, inclusion, affordability and individual autonomy is ultimately influenced by broader public policy choices. Legislators, governments, regulators, supervisors and social security systems all shape the environment within which insurers operate. Ethical outcomes therefore arise from a continuous interaction between public policy and market practices. Where society wishes to preserve certain forms of solidarity, ensure access to essential insurance, or limit the use of particular data sources, these objectives may require deliberate policy interventions in addition to responsible insurer conduct.

3.2.1. SOLIDARITY

Insurance is a mechanism for sharing uncertain losses across a pool of policyholders. By spreading the financial consequences of losses that materialise for some across the wider pool, it allows individuals and businesses to face risks that would otherwise be difficult or impossible to bear alone. This risk-sharing function reflects the social role of insurance. Yet views differ on whether, and to what extent, private insurance should also preserve forms of solidarity that go beyond the simple

pooling of uncertainty, a question that turns on the distinction presented in section 2 between chance solidarity and subsidising solidarity.

As mentioned, chance solidarity refers to the collective bearing of uncertain losses within an insured pool: policyholders contribute premiums before it is known who will suffer a loss, and those whose risks materialise are compensated from the pooled resources. This form of solidarity is inherent in insurance and reflects its social function, as it allows individuals to obtain financial protection against adverse outcomes that would be difficult to face individually.

Subsidising solidarity arises when premiums do not fully reflect actuarially expected losses, resulting in transfers between risk groups. Where chance solidarity is a structural feature of pooling, subsidising solidarity reflects decisions about how premiums are structured and the extent to which risk differences are recognised in pricing. By limiting the extent to which premiums mirror individual expected risk, subsidising solidarity can help ensure that higher-risk or potentially vulnerable consumers remain able to afford cover. In such cases, part of the cost associated with higher-risk individuals is borne collectively by the wider pool of policyholders.

However, in private insurance, this form of solidarity is constrained by the voluntary nature of participation. If lower-risk consumers consider premiums too high relative to their own risk, they may leave the pool or switch provider, increasing the average risk in the pool and potentially raising premiums further. This dynamic is closely linked to adverse selection. If insurers cannot distinguish risks sufficiently, higher-risk individuals may be more likely to seek or retain cover, while lower-risk individuals may opt out.

This does not mean that subsidising solidarity is absent from private insurance and it would be misleading to suggest that private insurance can, or necessarily will, eliminate subsidising solidarity altogether. Risk differentiation is never perfect: data limitations, modelling uncertainty and emerging risks mean that premiums do not always fully capture individual expected losses. Moreover, private insurers are not always required to set actuarially exact premiums, and they may preserve elements of subsidising solidarity because of competitive pressures, long-term customer strategies, or their institutional identity. Such choices might, however, be limited by regulation, market conditions and prudential requirements. Under the Own Risk and Solvency Assessment (ORSA) of Solvency II, for instance, firms should identify loss-making or subsidised risks, assess their impact on the undertaking's solvency position, and consider how continued losses could affect future capital needs.

Additionally, subsidising solidarity in private insurance may arise in cases of market interventions, such as tax benefits, mandatory insurance, or price controls, all of which can blur the line between private and public insurance. A new concept in this area is also that of 'advantageous solidarity'. This means that less risky individuals are often more risk-averse compared to others, leading them

into paying higher premiums than their risks would require. They will thus subsidise more risky insureds.

This distinction between chance solidarity and subsidising solidarity clarifies why risk differentiation is such an ethical challenge in data-driven insurance. The question is not simply whether risks should be differentiated, but how far that differentiation should go, and which forms of solidarity should be preserved, reduced or eliminated. The answer is not determined by actuarial considerations alone. It also reflects broader societal and public policy choices regarding the desired balance between risk-based pricing, affordability and inclusion.

Where subsidising solidarity is the objective, risk mutualisation may resemble social insurance, with limited or no differentiation between individual risks. Where chance solidarity is the main objective, greater data availability can support the risk-based mutualisation typical of private insurance, by allowing insurers to assess risks more accurately.

The way data is accessed and used can also affect solidarity across generations and social groups. More granular data may help insurers identify shared systemic risks, such as climate change, pandemics or natural catastrophes, and support prevention, resilience and insurability. At the same time, if data use leads to excessive segmentation, it may reinforce existing inequalities or shift costs between groups in ways that undermine affordability and access. Insurers therefore face ethical choices about whether data use strengthens or weakens the collective function of insurance.

3.2.2. PERSONALISATION AND MUTUALISATION

As discussed above, insurance requires some degree of risk differentiation. The increasing availability of data, combined with AI and other advanced analytical tools, makes this differentiation more granular. This can support more accurate risk assessment and pricing, but it also raises ethical questions about how far personalisation should go, which data should be used, and whether the resulting outcomes remain fair, proportionate and consistent with the social role of insurance.

In principle, more information can improve the assessment of risk. If insurers are better able to identify relevant risk factors and distinguish causal drivers from weaker correlations, they may price risks more accurately and extend coverage to consumers or risks previously considered difficult to insure. For example, telematics can allow safer young drivers to access lower premiums than would be possible under broad age-based pricing approaches. At the same time, more granular risk assessment may also contribute to risk prevention by providing consumers with incentives and feedback that encourage safer or healthier behaviours.

However, greater granularity also has downsides. It may enable insurers to focus on the most profitable risks while leaving higher-risk, vulnerable or data-poor consumers[16] with limited access, higher prices or no cover. In this sense, personalisation can contribute to exclusion where the benefits of better risk assessment are not matched by safeguards on access and affordability. It may also create incentives to collect increasingly detailed personal or behavioural data, raising concerns about privacy and human autonomy. For example, health insurers may offer wellness programmes or wearable-based products that reward consumers who share behavioural data such as physical activity levels, sleep patterns or other health-related indicators. Consumers whose data suggests healthier behaviour may benefit from discounts, while those unwilling or unable to share equivalent data may face higher premiums. In such situations, greater personalisation may reduce cross-subsidisation between risk groups, but it may also raise questions regarding fairness, affordability and the treatment of consumers who cannot benefit from data sharing or do not want to share their data.

A distinction should be drawn between risk-based pricing and differential pricing. Risk-based pricing aims to improve the assessment of insured risk by using information that is relevant to the likelihood or severity of losses, while differential pricing usually refers to setting different premiums according to characteristics such as willingness to pay, likelihood of switching provider or consumer inertia. For example, a customer returning to a quote after several days may be identified as less likely to switch and offered a higher premium than a new customer with the same risk profile. Similarly, third-party data such as online purchasing patterns, social media activity or geolocation metadata may be used to infer lifestyle, income or willingness to pay rather than insured risk. Another typical form of differential pricing is the so-called *loyalty premium*, where long-standing customers are charged more because they are considered less likely to switch provider. Regulators have already made critical statements on certain differential pricing practices that may result in unfair outcomes for consumers[17]. The ethical question for insurers is whether such a pricing practice serves consumers in a sustainable manner.

Risk-based pricing, when based on objective and actuarially relevant risk factors, is generally regarded as consistent with the principle of actuarial fairness. By contrast, differential pricing, which relies on behavioural or individual-specific data, raises broader concerns regarding solidarity, transparency and the fair treatment of policyholders, as it may go beyond risk-based differentiation. Of course, supervisory authorities and consumer protection bodies can help mitigate harmful practices where their mandates and the applicable regulatory framework allow them to do so, and competitive pressure may also discipline certain market behaviours. However, neither regulation nor competition can resolve all ethical questions. Insurers must still consider how their use of data

[16] Data-poor consumers are individuals for whom insurers possess limited relevant data, whether due to a short insurance history, low digital footprint, lack of access to data-generating technologies, or a decision not to share personal data.

[17] Supervisory Statement on Differential Pricing Practices in Non-Life Insurance Lines of Business, EIOPA-BoS-23/076, 16 March 2023

affects the societal role of insurance, particularly where personalisation may influence access, affordability, autonomy or trust.

Increasing granularity of data use is often linked to concerns about "demutualisation". This concept is sometimes used to suggest that personalisation could undermine the pooling function of insurance by making each policyholder's premium so individualised that mutualisation becomes obsolete. This concern should be treated carefully. More granular data may reduce some forms of subsidising solidarity, because premiums can more closely reflect expected claims. However, it does not eliminate chance solidarity or the need for risk pooling[18].

Even if insurers had very accurate knowledge of the probability distribution of a given risk, they would still not know whether, when or how severely the risk would materialise for a particular individual. Insurance covers actual losses, not expected losses. The purpose of the pool is therefore not only to estimate expected claims, but to ensure that the financial consequences of realised losses are borne collectively. All policyholders benefit from protection against uncertainty, even though only some will ultimately suffer a loss and receive compensation. In this sense, personalisation and individualisation do not, by themselves, collapse the pooling mechanism or make mutualisation obsolete.

Decisions regarding the acceptable limits of personalisation are not solely technical or commercial matters. They also reflect broader societal preferences concerning solidarity, privacy, fairness and consumer protection. Public authorities may therefore determine that certain uses of data should be restricted or subject to specific safeguards, even where they might improve predictive accuracy.

3.2.3. INCLUSION

More effective methods of risk assessment can increase both inclusion and exclusion. On the one hand, AI systems and more granular data can make it easier to identify high-risk consumers, potentially leaving some individuals without cover or facing prohibitively high premiums. On the other hand, better data and analytical tools can improve the understanding of risks and make risks previously considered uninsurable insurable, without undermining the prudential soundness required from insurers. They can also support prevention and risk mitigation, enabling consumers to reduce their risk exposure and, in some cases, access more affordable coverage.

In this context, exclusion is both an ethical and societal concern. In insurance, exclusion may arise when consumers are unable to obtain cover or can only access it on terms that are unaffordable or unsuitable. This may contribute to broader social exclusion, particularly where the insurance product is closely linked to participation in economic or social life. At the same time, it is important

[18] The concept of 'demutualisation' has found a novel use in connection to individualisation of insurance risks. Traditionally the concept has meant converting a mutual insurer into a limited liability company. This document does not address at all this traditional mainstream meaning of the concept.

to recognise that private insurance may not always be the most appropriate or effective mechanism for addressing all forms of exclusion. A pragmatic approach would be to focus on measures that align with the strengths and capabilities of the private insurance market, while recognising the complementary role of public policy (for instance, addressing certain protection gaps through social security arrangements, public support mechanisms or mandatory coverage schemes).

The relationship between data use and inclusion is therefore complex. Data-driven models can contribute to exclusion where algorithmic decisions rely on historical data that reflects existing social or economic inequalities. In such cases, insurance may reproduce or amplify pre-existing disadvantages. Similar concerns may arise for consumers with smaller or less reliable digital footprints, including older people, persons with lower digital literacy or individuals without access to connected devices, who may also be penalised by models that rely heavily on digital behaviour or online data.

At the same time, better data can support inclusion by enabling insurers to develop more inclusive products. For example, usage-based motor insurance may allow safer young drivers to pay lower premiums based on how they actually drive, rather than being priced mainly by reference to their age group. Similarly, more granular risk data may help insurers distinguish between risks within a broad category that would otherwise be treated as uniformly high risk, thereby extending coverage to consumers or areas that might otherwise be excluded. Claims and demographic data, for example, can inform microinsurance solutions for low-income or elderly consumers. Data may also help detect changing vulnerability over time, allowing insurers to offer adjusted payment schedules, policy modifications or other flexible solutions where consumers face temporary hardship due to illness, unemployment or other life events. Inclusion in insurance can therefore take different forms. It may mean granting access to insurance on fair and affordable terms despite differences in risk level. It may also mean applying differentiated treatment where this supports access, for instance through tailored products or adjusted underwriting criteria.

Inclusion may also result from preventive and proactive approaches. Wellness programmes, proactive care plans and risk-mitigation strategies can encourage healthier behaviours, reduce losses and support long-term financial resilience, provided that human autonomy and meaningful choice are respected.

However, both data use and measures intended to promote inclusion can create ethical trade-offs. Restricting or prohibiting the use of relevant data may protect one group, but may also shift costs to others (including other vulnerable consumers). Similarly, subsidised premiums for one group may improve affordability for that group while increasing premiums for another. In some cases, private insurance markets can expand access through better data, improved risk assessment and more tailored products. In others, certain risks may be difficult to insure sustainably, for example where

losses are highly concentrated or severe. Where society nevertheless considers protection against such risks desirable, public authorities may choose to intervene via policy measures.

Inclusion should therefore be understood as a shared responsibility, influenced by the interaction between insurer decisions and public policy choices. The ethical challenge is therefore to assess inclusion not only by intention, but by outcome. Insurers and policymakers should consider who benefits, who bears the cost, whether alternatives remain available, and whether a measure improves access and affordability without creating disproportionate disadvantages elsewhere.

3.2.4. CAUSALITY AND CORRELATION

Insurance relies on identifying relationships between data points and insured risks. Some of these relationships are causal: a factor contributes to the likelihood or severity of a loss. Others are merely correlated: two variables move together without one necessarily causing the other. Causality does not mean certainty. Rather, it means that a factor increases or decreases the probability of a particular outcome. For example, smoking does not make lung cancer inevitable, but it increases the probability of certain health risks. Correlation, by contrast, describes a relationship between variables without establishing cause and effect. For instance, ice cream sales and drowning incidents may both increase during summer, but buying ice cream does not cause drowning (both are influenced by a third factor, such as warmer weather). For this reason, causal links are often considered a fairer basis for insurability and pricing decisions than purely statistical associations.

This distinction is particularly important as insurers progressively use AI and advanced analytics to identify patterns in large and complex datasets. Such tools may detect correlations that improve prediction but do not necessarily explain the risk. If these correlations are treated as causal, they may lead to unfair or inaccurate conclusions, unjustified exclusions, or indirect discrimination against certain groups. In addition, reliance on spurious correlations may result in mispricing and undermine insurers' risk management and financial soundness.

The causality/correlation distinction also matters for questions of responsibility. Where a risk factor reflects behaviour that individuals can reasonably influence, such as certain driving behaviours or smoking, its use in pricing may be easier to justify. Where a factor reflects circumstances outside the individual’s control, such as socio-economic hardship or exposure to climate risk, stronger caution is needed. In such cases, granular data may improve prediction, but it may also shift costs onto individuals for risks they cannot realistically avoid or control.

At the same time, correlations should not be dismissed altogether. Insurance has always relied on statistical relationships, and some predictive variables may be useful even where the underlying causal mechanism is not fully understood. Limiting certain data uses too broadly may reduce the ability of insurers to assess risks accurately, support prevention, identify emerging risks or expand insurability. However, the more indirect, opaque or sensitive the variable, the stronger the need for

proportionality, explainability and outcome monitoring. Thus, the ethical challenge is not to require perfect causal proof for every data point, but to ensure that the link between the data used and the insured risk is relevant, proportionate and explainable. Insurers should pay particular attention where variables are opaque, are difficult for consumers to understand or challenge, are outside consumers' reasonable control, or may act as proxies for protected or vulnerable characteristics.

3.2.5. HUMAN AUTONOMY, PRIVACY AND DATA SECURITY

Ethical concerns regarding privacy become more significant as data collection grows more continuous and granular. For example, wearable devices may provide information on exercise, sleep patterns or heart rate, while telematics can capture detailed behavioural information about driving practices. Although participation in such programmes is voluntary, consumers may perceive strong financial disincentives if they choose not to participate. While telematics can have a positive impact in terms of risk prevention and mitigation, it also raises questions about whether consent remains genuinely free and whether individuals can realistically refuse to share more and more detailed personal information.

Thus, the increasing use of data in insurance raises fundamental questions about individual autonomy, privacy, and the security of personal information. These concerns are addressed by a layered regulatory framework spanning the GDPR, the EU AI Act[19] and sector-specific instruments such as Solvency II and the IDD — providing a robust, though not exhaustive, baseline for responsible data use. However, legal compliance is not always sufficient and ethical data use requires insurers to consider whether their practices are fair, proportionate and respectful of the policyholder as an individual.

This is especially relevant where insurers have discretion over how much data to collect, how long to retain it and how broadly to use it. In such cases, the responsible approach is not necessarily to use all data that is available or legally accessible, but to collect and retain only what is necessary and justified for the relevant purpose. A practical test is whether a well-informed policyholder would reasonably expect, understand and consider fair the proposed use of their data. This reflects both the spirit of data protection law and the fair treatment obligations under the IDD.

At the same time, technology evolves, and data may reveal new insights into causal relations, risk drivers and consumer outcomes that are difficult to anticipate at the point of collection. This does not justify open-ended data use, but it does mean that governance should allow purposes and safeguards to be reassessed over time. In all, responsible data use in insurance requires not only legal compliance but a genuine commitment to the interests of policyholders as individuals whose autonomy, privacy, and security deserve active protection.

[19] The requirements of the GDPR and the AI Act are summarised in Annex I

In practice, this means ensuring that consent to share behavioural data is truly voluntary, that human oversight of AI-assisted decisions is substantive rather than formal, that data breaches are communicated promptly, and that personal data is retained only if necessary for its stated purpose. These are not merely regulatory obligations; they are expressions of the trust that policyholders place in insurers when they share their most sensitive personal information, and of the responsibility that comes with it.

Furthermore, the growing volume and sensitivity of data held by insurers also increase cybersecurity and operational resilience risks. Data breaches involving health, financial or behavioural information may expose individuals to fraud, identity theft or other harms. Ethical data use therefore requires not only lawful collection and processing but also robust safeguards to ensure that sensitive information is adequately protected throughout the data lifecycle.

3.2.6. DISCRIMINATION

As insurers increasingly rely on large and complex datasets to assess risk, concerns also arise as to whether they produce fair outcomes. While insurance inevitably involves differentiating between risks, not all forms of differentiation are ethically or legally acceptable.

Direct discrimination arises when consumers are treated less favourably in access, pricing, claims handling or service quality on the basis of protected or otherwise sensitive characteristics[20]. Certain forms of differentiation are clearly prohibited, such as the use of gender in insurance pricing. Beyond protected attributes, insurers are also subject to broader expectations on the fair treatment of consumers. Immutable characteristics play a special role when considering access to insurance cover that may be regarded as essential. Certain types of insurance can be considered essential insofar as a lack of coverage may lead to exclusion or significant disadvantage for affected individuals[21]. However, where insurance is understood to fulfil such an essential function, insurers should, to the extent possible, seek ways to maintain access to affordable coverage.

In practice, the most significant ethical challenge associated with the use of data often relates to indirect or statistical discrimination (where there is no intention to discriminate, but apparently neutral variables may disproportionately disadvantage certain groups). Variables such as postcode, occupation, educational background, device ownership, purchasing behaviour or online activity may improve predictive accuracy while also acting as proxies for protected characteristics or socio-economic vulnerability.

[20] Protected attributes refer to personal characteristics that are safeguarded by law or policy against unfair or prejudicial treatment. Common protected attributes include age, gender, race or ethnicity, disability status, health condition, religion, sexual orientation, and sometimes socioeconomic status. These attributes are recognized because discrimination based on them can perpetuate inequality and exclusion, especially for vulnerable groups.

[21] Which particular form of insurance is considered essential will depend on the jurisdiction concerned, including the design of its social security system, as well as individual circumstances.

Differential treatment may raise concerns of indirect discrimination where it places certain groups at a disadvantage without justification. The ethical challenge is therefore not limited to whether a variable predicts risk, but whether the resulting differentiation can be objectively justified in light of the broader objectives of insurance. For example, the use of postal codes may provide useful information about risk exposure and expected claims costs, while also acting as proxies for protected characteristics. Their use should therefore be assessed not only in terms of predictive performance, but also in terms of necessity, proportionality and fairness. Where such variables are used, insurers should be able to demonstrate that they serve a legitimate objective and that the same objective cannot reasonably be achieved through less discriminatory means.

Ethical concerns around discrimination may arise even where regulatory requirements have been formally met. For example, under the IDD and the POG Delegated Regulation, insurance product manufacturers should ensure that the product is aligned with the characteristics of the target market, and that taking into account these characteristics, no consumer is adversely impacted. Such assessment could include the risk of unfair discrimination which may emerge through what is sometimes referred to as an ethnic penalty, where factors such as location, occupation, education, income, or other socio-economic indicators act as proxies for ethnicity or other protected characteristics, resulting in systematically less favourable outcomes for certain groups. Where indirect or statistical discrimination is identified, insurers should assess whether the practice pursues a legitimate objective and whether the means used are necessary to achieve that objective. Consideration should also be given to the nature of the affected characteristic, including whether it is immutable or whether policyholders can reasonably influence the relevant attribute over time.

Data use and advanced analytical techniques may play an important role in identifying and mitigating model biases and unintended discriminatory outcomes, which can emerge from the interaction of complex processes and systems within insurers. Ex post data analysis can become a valuable tool for assessing outcome fairness[22].

[22] In fact, the AI Act foresees the collection of sensitive data for this purpose.

4. SCENARIO ANALYSIS

4.1. PURPOSE AND APPROACH

Building on the previous sections, which set out the role of data in insurance and the main ethical questions it raises, this section examines how different levels of data use may affect consumers, insurers, supervisors and society. The purpose of the scenario analysis is to provide a structured way to explore the trade-offs that may arise as data availability, analytical capabilities, customer acceptance and governance arrangements evolve.

As seen, the current insurance landscape is characterised by a gradual expansion of data-driven practices. Insurers increasingly complement traditional data sources with selected new forms of data and advanced analytics, while consumer trust, privacy considerations, technological constraints and regulatory requirements continue to shape adoption. The analysis proposed in the following sections compares two stylised scenarios: one characterised by limited data use and another by extensive data use. These scenarios are deliberately contrasted in order to make the potential benefits, risks and policy tensions more visible and they should be read as analytical tools to identify the most significant positive and negative impacts across five assessment dimensions (consumer, insurer, regulator, data ethics and macroeconomic effects). Selected use cases are included at the end of the section to make the scenarios more tangible. They show how differences in data availability, sharing arrangements and governance maturity may affect practical insurance applications, while also linking the scenario analysis to the report's broader conclusions on access, affordability, trust, and safe innovation.

Due to the absence of comparable quantitative evidence, the analysis is qualitative and draws on desk research and expert input from the Consultative Expert Group. The findings should therefore be understood as illustrative rather than exhaustive and other industry experts may identify different or additional implications. Furthermore, this chapter does not assess whether existing regulations and guidance sufficiently mitigate risks or facilitate benefits associated with varying levels of data use, as that evaluation is beyond its scope and would also need to reflect aspects specific to national markets.

4.2. DATA USE SCENARIOS

This section compares two stylised scenarios for data use in insurance: a limited data use scenario and an extensive data use scenario. The two scenarios differ across several dimensions: data

availability, analytical methods, technological infrastructure, skills, regulation and consumer acceptance.

Limited data use scenario

The limited data use scenario describes an insurance market operating under constrained data-sharing conditions and a narrow operational perimeter for new forms of data use. Although scientific and technological progress continues, insurers can rely only on a limited range of data, shaped by privacy-focused regulation, cautious customer consent, incomplete standardisation of data exchange and fragmented interoperability.

In this scenario, pricing and underwriting depend primarily on traditional data sources such as demographic information, policy and claims history, exposure data and internal records, supplemented by publicly available or aggregated contextual information, such as environmental, climate and population-level indicators. Most emerging data categories, including vehicle telematics, IoT sensor data, behavioural signals from mobile or wearable devices, online and social media activity, genetic information and continuous health monitoring, are either inaccessible, legally restricted or available only in highly aggregated forms.

Data acquisition and exchange remain fragmented, often relying on intermediated arrangements with limited interoperability. This increases integration costs and reduces transparency. Insurers therefore favour interpretable and easily auditable analytical methods, including generalised linear models, regression techniques and credibility approaches. Advanced machine-learning models are used only selectively, either because the available data does not justify more complex modelling or because governance requirements would be disproportionate. Technology infrastructure remains largely legacy-based and batch-oriented, making real-time processing and high-velocity data integration impractical.

Extensive data use scenario

The extensive data use scenario describes a highly standardised and interoperable insurance data ecosystem, with clear governance frameworks for lawful access, accountability and oversight. In this scenario, customers are more willing to share a wide range of personal and contextual data where the value proposition is clear, safeguards are credible and mechanisms for control are meaningful.

Insurers can integrate a broader range of data beyond traditional sources, including high-frequency telematics, IoT sensor streams, health-monitoring data, behavioural and digital interaction signals, financial indicators, and advanced environmental and climate datasets. Sensitive data categories, potentially including genetic information, may be used only where legally admissible and subject to strict conditions such as purpose limitation, data minimisation and accountability.

The richness and velocity of data allow insurers to use advanced analytics to process high-dimensional and multimodal information. Machine-learning and AI techniques, including more complex models where justified, support underwriting, pricing, claims handling, fraud detection, prevention and portfolio management. These tools operate within stronger governance frameworks, including testing, monitoring, auditability and human oversight. Technology infrastructure is more scalable, interoperable and real-time, enabling deeper integration across the insurance value chain and with external data providers.

Table 1. Comparative table of scenarios 1 and 2

	Scenario 1: Limited data use scenario	Scenario 2: Extensive data use scenario
Data availability & types	Insurers rely mainly on traditional data, such as demographic information, policy and claims history, exposure data and internal records, complemented by public or aggregated contextual indicators. Emerging data categories (such as telematics, IoT sensor data, behavioural signals, online activity, genetic information and continuous health monitoring) are unavailable or accessible only in aggregated form.	Insurers have broader access to granular, high-frequency and continuous data streams, including telematics, IoT sensor data, health-monitoring data, behavioural and digital interaction signals, subject to strict governance, proportionality, and legal admissibility.
Data processing & analytics	Data processing is mainly based on interpretable and auditable statistical approaches, including generalised linear models, regression techniques, credibility methods and simpler transparent models. Advanced machine-learning models are used selectively, either because data is limited or because governance requirements would be disproportionate.	Advanced analytics and AI are used to process high-dimensional, high-velocity and multimodal data. Models may support underwriting, pricing, claims handling, fraud detection, prevention and portfolio management, with more automation and continuous monitoring/testing.
Technological infrastructure	Technology infrastructure remains largely legacy-based, monolithic and batch-oriented. Interoperability is limited, data exchange is fragmented and API usage is restricted.	Technology infrastructure is scalable, cloud-enabled, interoperable and increasingly real-time, supported by standardised interfaces, event-driven pipelines and stronger operational controls.
People & Skills	Expertise remains centred on actuarial, underwriting, compliance and domain-specific judgement, with stronger reliance on manual review and retrospective analysis.	Data scientists, machine-learning engineers and data-governance experts become embedded across the value chain. Organisation-wide data

		literacy becomes more important as automated pipelines influence business decisions.
Regulation & data protection	The regulatory approach is more restrictive and privacy-oriented, with a narrow operational scope for novel data and complex AI. The stress point for regulation is minimisation and caution.	The regulatory approach needs to enable innovation while imposing strong safeguards on legality, purpose limitation, transparency, fairness, explainability, accountability, security and enforcement. The stress point for regulation is effective governance: broader data use should not outpace supervisory capabilities, consumer protection or the ability to enforce meaningful accountability across the ecosystem.
Customer engagement & acceptance	Consumers show limited willingness to share additional personal or behavioural data. Consent is narrower and easier to understand, and non-data-intensive products remain more visible.	Consumers are more willing to share data where value propositions are clear and safeguards are credible.

Impact assessment dimensions

To compare the scenarios and use cases consistently, the analysis applies five impact assessment dimensions: Consumer, Insurer, Regulator, Data Ethics and Macro effects. The Consumer dimension considers affordability, accessibility, product suitability, transparency, autonomy and consumer behaviour. The Insurer dimension considers business viability, operational feasibility, innovation, risk management and legal or reputational exposure. The Regulator dimension scrutinises the appropriateness and necessity of regulation in the context of new developments. It emphasises protection of policyholder rights and the public interest, effectiveness of compliance and enforcement mechanisms, and alignment with global regulatory trends to avoid regulatory arbitrage, especially within the EU market. The Data Ethics dimension focuses on solidarity, fairness, inclusion, autonomy, privacy, transparency and accountability. Finally, the Macro dimension considers wider market and societal effects, including competition, financial stability, innovation, environmental impacts, social inclusion and overall welfare.

SCENARIO IMPACT ASSESMENT

Table 2. Scenario comparison across impact dimensions

	Scenario 1: Limited data use scenario		Scenario 2: Extensive data use scenario	
	▼ Negative	▲ Positive	▼ Negative	▲ Positive
Consumer	Less risk-based pricing may result in lower-risk consumers subsidising higher-risk consumers. Broader risk categories can reduce incentives for prevention and may increase adverse selection. Some emerging or complex risks may remain difficult to insure.	Simpler and easier-to-understand products. Greater affordability and accessibility for some higher-risk or data-poor consumers due to broader pooling.	More granular pricing can improve risk differentiation but may reduce affordability or access for some higher-risk consumers. Greater risk of exclusion linked to "data poverty". Behavioural monitoring and nudging may affect autonomy, while the use of behavioural data to personalise policy conditions or exclusions may increase product complexity and lead to reduced claim pay-outs for some consumers. Willingness-to-pay pricing and customer segmentation practices may create unfair outcomes.	More accurate risk assessment and increased affordability for some customers. More tailored products and services. Support for risk prevention and mitigation, helping consumers reduce risk exposure and improve access to coverage. Lower-risk consumers may benefit from premiums that better reflect their individual risk characteristics.
Insurer	Greater underwriting uncertainty. Broader pooling may lead to stronger cross-subsidisation. Certain emerging risks may remain difficult to price or insure. Product innovation may develop more slowly. Certain products may become unviable due to a lack of necessary data for risk assessment.	Lower dependence on complex data ecosystems. Lower data governance, cybersecurity and AI-related compliance burdens. Existing infrastructure remains sufficient.	Significant investment in data, technology, AI and specialised skills. Greater exposure to model risk, data quality issues, cyber events, operational dependencies and reputational risk. Reliance on third parties may increase concentration and resilience concerns.	Competitive advantage for insurers with advanced analytics capabilities. Improved data availability enables more differentiated and dynamic risk management with fine-tuned risk steering, more targeted reinsurance and fraud detection. Richer data allows insurers to find causal factors behind risks and improve loss prevention.

Regulator	Regulation regarding solvency and risk management is appropriate, but limits on prohibited data/sources may be needed. Risk of covert data practices (scraping, opaque enrichment) increases enforcement needs.	Clear bans/allowances for specific variables and sources; solvency oversight remains stable. Lower complexity of governance and model oversight requirements.	Increased need for oversight of AI, external data providers, data quality, explainability, fairness, cybersecurity and operational resilience. Greater risk of regulatory arbitrage without global alignment. Compliance becomes, in general, more demanding.	More data forces regulation itself to become data driven. Monitoring and supervision of the sector overall improve and become more complete, e.g. not only financial performance but also societal outcomes.
Data Ethics	Less precise risk assessment may weaken risk-based pricing and, in some cases, limit insurability. Unintended subsidies leading to increased adverse selection.	Higher levels of subsidising solidarity. Lower privacy and surveillance risks. Easier explainability and oversight. Simpler models make discriminatory outcomes easier to identify and challenge.	Higher privacy/cyber exposure; needs strong consent, minimisation and security. Access to high-frequency and personal information can result in loss of human autonomy and privacy. Greater risk of opaque or discriminatory models; stronger transparency and explainability demands. Danger of excessive segmentation and reduced subsidising solidarity.	Opportunities to potentially reduce current biases and detect discrimination. Greater opportunities to expand insurability and support inclusion for some consumers through better understanding of risks and consumer needs.
Macro	More commoditised market; scale and brand matter. Slower innovation of the sector. Limited access to granular hazard data undermines catastrophe risk, loss prevention becomes more challenging in the absence of available information. Lower risk of uninsurable population but higher risk for high premiums/limited coverage overall. Smaller players and insurtechs struggle without granular data.	Less data favours incumbents and standardisation. Larger incumbents benefit from scale and capital buffers. Lower computational footprint and reduced negative environmental impact.	Potential consolidation around a few large insurers and risk of monopolisation. Greater segmentation may reduce solidarity and exclude vulnerable groups. Societal debate on human autonomy and surveillance.	Potential to increase welfare through better risk mitigation. Lower loss volatility and improved capital efficiency. Stronger economic incentives for loss prevention and sustainable behaviour where possible. More accurate pricing of environmental externalities. Potential for cost reduction and productivity gains and the rise of new business models.

This scenario analysis compares limited (low) with extensive (high) data use in private insurance and suggests that neither extreme is optimal. As highlighted throughout this report, insurance relies on a balance between risk mutualisation and risk differentiation. Data can strengthen insurers' understanding of risk, support prevention, improve consumer outcomes and expand the boundaries of insurability. At the same time, increasingly granular data may affect affordability, inclusion, privacy, autonomy and the degree of solidarity within insurance pools. A calibrated middle ground is essential—one that rewards low-risk behaviour while ensuring inclusion of high-risk individuals and those lacking digital access.

Limited data use supports simpler products, lower privacy risks and higher levels of subsidising solidarity, but it also creates blind spots, leading to increased volatility, imprecise risk differentiation and adverse selection, and limiting the insurability of certain risks. This undermines market stability and solvency, risks product stagnation, and reduces innovation—particularly affecting vulnerable or niche segments and emerging risks. Overly restrictive data policies may also drive insurers towards opaque proxies, increasing governance complexity without improving fairness.

Conversely, extensive data can improve risk assessment, prevention, operational efficiency and product personalisation, while potentially expanding insurability through a better understanding of risks. However, it may also increase privacy, cybersecurity and governance risks, contribute to greater segmentation of risk pools and limit solidarity, reduce affordability or access for certain consumers, and raise concerns regarding autonomy, exclusion and discrimination.

At the same time, more extensive data use can contribute to greater operational efficiency, lower claims costs, lower loss volatility, a more efficient allocation of capital, and more efficient insurance markets that may reduce insurance costs. However, the overall effect on prices and on consumer welfare is a priori uncertain. More granular information could also enable greater segmentation of consumers and facilitate forms of differential pricing based on behavioural characteristics, customer inertia or willingness to pay. While some consumers may benefit from lower premiums that more accurately reflect their risk profile, others may face higher prices or reduced affordability.

The effect on competition is similarly ambiguous. Granular data can intensify competition by enabling multiple insurers to identify and compete more effectively for low-risk and potentially more profitable customer segments. Digitally savvy consumers may also benefit from greater access to information, digital comparison tools and data-driven services that facilitate price comparison and switching between providers. On the other hand, competitive advantages linked to access to data, analytical capabilities and technological infrastructure may favour larger, more advanced firms and incumbent firms, potentially leading to concentration.

To make the scenarios more tangible, this subsection applies them to two illustrative use cases: telematics in motor insurance and wearables in health insurance. These examples were selected

because they are familiar, directly linked to core insurance activities and illustrate different levels of sensitivity, consumer choice and ecosystem complexity. The use cases do not aim to provide a complete mapping of data-driven insurance models. Rather, they illustrate typical trade-offs across the scenarios and highlight implications for consumers, insurers and supervisors, including key ethical considerations such as privacy, fairness, transparency, and accountability.

4.3. USE CASE A: TELEMATICS IN MOTOR INSURANCE

Telematics-based motor insurance uses vehicle- and driving-related data to refine risk assessment, tailor premiums and support safer driving behaviour. It is typically offered either as an optional add-on to traditional products or, in more mature settings, as an integrated component of motor insurance and mobility services.

Telematics can serve several purposes. It can measure exposure, for example through mileage or time of day. It can capture behavioural indicators, such as braking and acceleration patterns. It can also support prevention and service features, including driving feedback, alerts or incident detection. Data can be collected through embedded vehicle systems, aftermarket devices or smartphone applications. It may also be enriched with contextual information such as road type, weather or time bands. Depending on the business model, data may be collected directly by insurers, obtained through vehicle manufacturers, processed by insurtech partners or accessed through third-party platforms, with each option raising different dependencies and transparency challenges.

Analytics typically translate raw trip data into indicators or scores, for example through trip classification, driving-score models, detection of anomalous events (including potential crash signals), and portfolio-level monitoring. Depending on governance expectations and data richness, methods range from relatively simple, auditable scoring approaches to more complex modelling with ongoing monitoring and controls.

Across all scenarios, three governance themes are central. First, privacy and proportionality become more sensitive as data moves from coarse exposure measures to continuous behavioural monitoring. Second, fairness risks increase where behavioural indicators correlate with socio-economic circumstances or where participation is not equally feasible; the availability of meaningful non-telematics alternatives matters for genuine choice. Third, accountability must be clear along the value chain—especially where vehicle manufacturers, platforms or analytics vendors are involved—alongside robust security and operational resilience given the increased data flows and ecosystem dependencies.

Table 3. Use case A: Scenario comparison across impact dimensions

	Scenario 1: Limited data use scenario	Scenario 2: Extensive data use scenario
Consumer	Telematics remains an optional add-on to traditional motor pricing, primarily to reward basic usage indicators and support limited prevention services. Low willingness to share continuous behavioural data means participation remains niche. Consumers retain clearer non-telematics alternatives and face less pressure to participate.	Telematics becomes mainstream and embedded in the mobility ecosystem, supporting dynamic risk management, prevention and more automated claims handling. Participation is higher, driven by perceived value and ecosystem integration, but this may create social or market pressure to participate and reduce the practical availability of non-telematics alternatives.
Insurer	The use case remains focused on non-life motor insurance as an optional telematics add-on or discount product. Insurers mainly use mileage/usage data and coarse contextual information, supported by interpretable scoring and statistical models with conservative feature sets and limited automation. Operational integration into underwriting and claims remains partial and often batch-based.	Telematics is embedded as a standard offering within motor insurance, with increasing differentiation across product tiers and service bundles. Insurers can use high-frequency and continuous driving and contextual data, supported by advanced AI for multimodal and high-velocity data, continuous monitoring and more automated decision-making with auditability controls. Real-time pipelines enable deeper integration across underwriting, dynamic pricing, prevention and claims automation.
Regulator	Supervisory focus is on monitoring alternative data and proxy use, overseeing limited but sometimes opaque third-party arrangements, ensuring transparency, and preventing circumvention of protections.	Supervisory focus shifts to more challenging system-wide oversight of data ecosystems, including auditability of automated decisions, contestability, concentration risks, cyber risks and chain-wide accountability across OEMs, platforms, data brokers and AI vendors.
Data Ethics	Privacy and proportionality concerns are more limited because telematics is restricted to basic usage indicators and limited behavioural detail. Fairness risks remain where participation is uneven or where opaque enrichment substitutes for direct consent-based sharing, but the overall level of continuous monitoring is lower.	Continuous behavioural monitoring raises significantly stronger concerns around privacy, proportionality and surveillance. Fairness risks increase where behavioural indicators correlate with socio-economic conditions or where participation is not equally feasible. Strong safeguards are needed on transparency, meaningful choice, security, accountability and the use of non-telematics alternatives.
Macro	Third-party roles remain limited, though sometimes insufficiently transparent. Telematics stays peripheral to the wider insurance and mobility ecosystem, and adoption is more likely in the short to mid-term depending on market maturity and the governance environment.	Telematics becomes part of a broader mobility data ecosystem, creating stronger dependencies on manufacturers, platforms and analytics providers and raising wider issues of operational resilience and concentration. This is more plausibly a mid- to long-term scenario, contingent on standardised access, widespread adoption and operationally enforceable accountability frameworks.

4.4. USE CASE B: WEARABLES IN HEALTH INSURANCE

Wearables-based health insurance uses data from consumer devices (e.g. activity trackers, smartwatches) to support prevention programmes, wellness engagement, and optional incentives. In some arrangements—subject to strict legal admissibility and strong safeguards—device-derived data may also affect limited elements of pricing or underwriting. The use case highlights the particular challenges of consumer-generated, health-adjacent data: it is sensitive, often noisy or uneven in quality, and can enable secondary inferences about health status or lifestyle.

In practice, the data typically includes activity measures (e.g. step counts or activity bands), heart rate and sleep indicators, exercise sessions, and app-based lifestyle inputs; in some setups, location-derived activity patterns may also be used. Access and sharing are usually mediated by device manufacturers and health platforms, sometimes via wellness programme providers; in certain contexts, employers may be involved, which can raise additional concerns around voluntariness, power imbalances, and purpose limitation. Data access often depends on platform permissions and contractual arrangements, creating dependencies that can be difficult for consumers—and supervisors—to see and assess.

Across all scenarios, three governance themes are central. First, sensitivity and purpose limitation: seemingly "benign" behavioural data can reveal or infer medical conditions, requiring strict minimisation, retention limits and clear separation of purposes under GDPR mandates. Second, fairness and meaningful choice: the ability to participate varies widely (health status, disability, age, pregnancy, mental health, socio-economic constraints), so programme design must avoid penalising non-participation or creating de facto pressure to share data, as such pressure would render consent invalid. Third, accountability across the value chain: device platforms and wellness providers become critical actors, so oversight must cover data quality, security, model governance and contractual clarity—especially where multiple parties touch the data.

Table 4. Use case B: Scenario comparison across impact dimensions

	Scenario 1: Limited data use scenario	Scenario 2: Extensive data use scenario
Consumer	Wearables are used, if at all, for voluntary wellness engagement with a strict separation from underwriting and pricing decisions. Low willingness to share health-adjacent data means participation remains small-scale and explicitly voluntary. The consumer experience is focused on optional engagement rather than differentiated insurance outcomes.	Wearables data is integrated into broader health-data ecosystems, enabling highly personalised interventions and, where legally admissible, more granular differentiation. Participation is higher because of stronger integration and clearer value propositions, but there is a greater risk of "soft coercion" or de facto pressure to share data.
Insurer	The use case remains focused on health insurance and health-related services or riders. Insurers use minimal or aggregated activity	Health insurance products include deeply embedded prevention and service bundles with increasing differentiation across product tiers

	indicators and rely on descriptive analytics, simple engagement metrics and limited personalisation. Device and platform providers act as relevant but relatively contained intermediaries.	and service levels. Insurers use higher-frequency and broader device/app signals, potentially linked to richer datasets where legally admissible, and apply more advanced AI for behavioural data with greater automation in personalisation and programme management. Ecosystem dependencies become much stronger, including platforms, analytics vendors, wellness providers and potentially care partners.
Regulator	Supervisory focus is on ensuring that wellness programmes do not become indirect underwriting, and on preventing misleading marketing, opaque data flows and weak contractual accountability.	Supervisory focus expands to system-level impacts on solidarity and access, contestability of automated decisions, ecosystem and platform dependencies, and the prevention of discriminatory outcomes at scale. Oversight must operate across multiple actors rather than only the insurer.
Data Ethics	Health-related data is treated as highly sensitive. Strict purpose limitation, data minimisation, retention limits and separation of uses are central. Fairness concerns focus on ensuring meaningful choice and avoiding the penalisation of non-participation for individuals whose ability to participate is limited by health status, disability, age, pregnancy, mental health or socio-economic circumstances.	Ethical risks increase materially because seemingly benign device data can support inferences about medical conditions, health status or lifestyle. Strong proportionality constraints, clear boundaries on permissible use, and robust accountability across the value chain are essential. Fairness concerns are heightened by the risk that participation capacity differs across consumer groups, potentially leading to exclusionary effects.
Macro	Adoption remains relatively limited, reflecting legal sensitivity, variable device-data quality and cautious market development. This is primarily a short- to mid-term scenario depending on legal interpretation and market maturity.	The use case sits within a broader health-data ecosystem and may become systemically important if adopted at scale. Wider ecosystem dependencies increase third-party and systemic risks. This is more plausibly a mid- to long-term scenario, contingent on enforceable safeguards across platforms and sustained societal acceptance.

The scenario analysis shows that neither extensive nor limited data use delivers uniformly positive outcomes. Instead, increasing data use involves a series of trade-offs that affect consumers, insurers and society in different ways. These trade-offs reflect the broader themes discussed throughout this report, including the balance between risk mutualisation and risk differentiation, the role of solidarity in insurance, the implications for inclusion and affordability, and the importance of maintaining trust in the use of data. The question is therefore not whether data use should increase or decrease, but how it should be governed in a way that supports desirable outcomes while managing associated risks.

5. CONCLUSIONS

The analysis presented in this report suggests that the central challenge is not whether insurance should use more or less data, but how that data is collected, used and governed. As discussed in Chapter 2, insurance relies on a balance between risk mutualisation and risk differentiation. Advances in data collection, analytics and AI are shuffling that balance, creating opportunities to improve risk assessment, prevention, operational efficiency and product design, while potentially expanding the boundaries of insurability. At the same time, as seen in Chapter 3, increasingly granular data raises important questions regarding solidarity, inclusion, discrimination, privacy, autonomy and fairness. In fact, Chapter 4 contributes to illustrate that different approaches involve different trade-offs affecting consumers, insurers and society. These approaches and the attendant trade-offs will vary by national and societal context, by business line, and as technology and data availability evolve.

5.1. RECOMMENDATIONS FOR POLICYMAKERS

Against this background, the policy objective should not be to maximise or minimise data use, but to ensure that data-driven innovation translates into better outcomes for consumers and society without compromising trust, fairness, insurability and financial resilience. Achieving this balance calls for calibrated approaches that recognise both the opportunities and the risks associated with increasing data use, and that remain open to recalibration as markets, technologies and societal expectations evolve. At the same time, the report consistently highlights that many of the challenges associated with data use cannot be resolved by insurers alone. Questions relating to affordability, inclusion, solidarity and protection against difficult-to-insure risks ultimately reflect broader societal choices, and the appropriate balance will differ across jurisdictions, lines of business and social protection systems. Policymakers therefore need to consider not only how insurers use data, but also the broader public policy objectives that insurance markets are expected to serve. Drawing on the analysis developed throughout this report, five interrelated principles emerge to guide this approach.

The principles below focus on the outcomes that data use should support. They should not be interpreted as implying that less data use is necessarily preferable. Rather, the objective is to distinguish harmful or inappropriate uses of data from legitimate and beneficial ones. As data use and analytical capabilities evolve, achieving these outcomes will increasingly depend on effective governance arrangements, including transparency, accountability, human oversight, explainability, performance monitoring and appropriate controls over third-party providers.

Guiding principles

- **Principle of access**: As discussed, data use can both expand and restrict access to insurance. Better risk information may enable insurers to cover risks that were previously considered uninsurable, while increasingly granular risk differentiation may make cover unavailable or unaffordable for some consumers. Innovation should improve reserving, prevention and service quality, but not at the cost of unfairly excluding high-risk, vulnerable or data-poor consumers. Data use should therefore support access to insurance products and services rather than unnecessarily restrict it. Particular attention should be paid to situations where insurance plays an important social or economic function and where exclusion may have wider societal consequences.
- **Principle of fair and affordable outcomes**: The report highlights that more accurate risk assessment does not automatically translate into outcomes that are perceived as fair. Increasingly granular data may improve actuarial precision while also affecting affordability and the degree of solidarity embedded within insurance pools.
  This highlights the important distinction between risk differentiation that can be justified by underlying risk and pricing practices that rely on willingness to pay, customer inertia or proxies for personal characteristics deemed discriminatory. The key issue is not actuarial accuracy alone, but whether outcomes remain fair in social terms. In some cases, ensuring affordability may require a combination of public and private measures.
- **Principle of integrity, transparency and meaningful choice**: Trust is central to the functioning of insurance. As data collection increases and the use of analytics becomes more complex, preserving trust requires ensuring that consumers understand how their data is used and retain meaningful choice over participation. Data use must not come at the expense of integrity or consumer confidence. This implies strong expectations around transparency, explainability, accountability and proportionality. It also means that consent should not become a mere formality. Where refusal to share behavioural or sensitive data leads in practice to material penalties or loss of realistic alternatives, policymakers should be cautious about treating any resulting consent as genuinely meaningful.
- **Principle of prudential soundness**: Restrictions on data use should not prevent insurers from carrying out accurate risk assessment where this is necessary for underwriting, solvency and risk management. An overly restrictive framework can create blind spots, increase volatility, weaken insurability and push firms towards cruder or more opaque proxies. Data use to drive better risk sensitivity should not be treated the same as data use for individual pricing, underwriting decisions or targeting of customers.

- **Principle of safe progress**: The report's scenario analysis highlights the potential benefits of innovation in improving prevention, resilience and some consumer and societal outcomes. At the same time, innovation can generate new ethical, operational and supervisory challenges. Insurance markets should therefore remain capable of innovation and adaptation, while ensuring that new uses of data are subject to appropriate governance, oversight and accountability, considering the principles listed here and how they are met.

Main trade-offs policymakers need to navigate

The five principles above reflect a series of recurring trade-offs identified throughout the report:

- **Risk differentiation versus solidarity**: Better data can improve underwriting precision and reduce adverse selection, but more granular pricing can weaken cross-subsidisation and make some risks uninsurable or unaffordable.
- **Fairness and predictive accuracy:** Throughout the discussion of causality, correlation and discrimination, the report highlights that variables that improve predictive performance do not always produce outcomes that are perceived as fair. Less data may reduce some direct discrimination risks, but it can also increase mispricing, other potential discrimination risks, and cross-subsidisation. More data may reduce some broad-brush proxies, yet it can also create indirect discrimination, “data poverty” effects and systematic exclusion. Where there was less data, insurance could rely on simpler kinds of risk pooling, yet the scope and impact of insurance were reduced by the inherent inefficiency of this broad pooling.
- **Autonomy and prevention**: The report identifies different ways in which data can support prevention and loss mitigation. Telematics, wearables and other forms of behavioural data can encourage safer behaviours and earlier interventions. However, these same developments may also lead to intrusive monitoring, overly strong nudging and practical pressure to participate in data-sharing schemes. This is especially sensitive where individuals' ability to participate differs due to, for example, age, disability, mental health, pregnancy or socio-economic circumstances.
- **Innovation and the status quo**: Extensive data flows may drive new business models, productivity and better risk mitigation. However, they may also strengthen the position of firms with privileged access to data, large incumbents, platform operators, vehicle manufacturers or other actors controlling a data ecosystem, leading to changes in relative power between different players. Limiting data flows may strengthen those who are already data-rich. Policy choices may therefore need to balance the benefits of innovation against potential risks relating to concentration, competition and market resilience.

Recommended policy approach

Putting these dimensions together, policymakers should consider taking an approach that reflects the purpose of data use and its impacts or outcomes. A one-size-fits-all approach is unlikely to be effective: answers will be different for different lines of business.

- **First, policymakers should assess the purpose of the data use.** Uses linked to risk assessment, prevention, fraud control or claims efficiency may be viewed differently from a consumer and societal perspective than uses primarily designed to optimise prices based on behavioural factors. While differentiated pricing can reflect legitimate commercial objectives and risk-based underwriting, a clear 'red flag' arises where pricing strategies are designed to exploit consumer vulnerabilities, such as inertia, low financial literacy or willingness to pay. It should be considered that the purpose of data use can change as analytics and technologies evolve: it is not a fixed picture known from the outset. Moreover, not all data is equally material, and the principle of data minimisation can be a strong basis for also using anonymised data or synthetic data where possible.
- **Second, policymakers should assess the likely impacts or outcomes at three levels at once: the consumer, the insurer and society as a whole.** For consumers, the key questions are access, affordability, autonomy and clarity. For insurers, they are viability, risk management, compliance burdens and innovation. For society, they include solidarity, competition, inclusion, resilience and concentration and market structure risks. The report's own scenario analysis is built around these kinds of dimensions and shows that different choices shift costs and benefits across these groups. Policymakers should take a clear stance on the balance of costs and benefits. This focus on impacts or outcomes would need to take a holistic view. A given restriction might reduce one risk while worsening another, for example by increasing average prices or pushing firms towards weaker proxies. Equally, richer data may improve certain forms of fairness while worsening affordability or autonomy. The question should therefore be whether the overall practice produces acceptable outcomes under the five principles above.
- **Third, policymakers should differentiate by line of business.** Safeguards should be tailored to the context. For essential insurance such as life and health, the threshold for intrusive, behavioural or sensitive-data uses should be higher, given the essential character of the products and heightened fairness concerns. In motor insurance, telematics may be acceptable and beneficial, but only where meaningful alternatives remain available, and the governance of transparency, accountability and proportionality is robust. In health-related and wellness-related business, policymakers should be particularly attentive to indirect underwriting, inference of health status from apparently benign device data, and exclusionary effects linked to participation capacity.

A SHORT DECISION FRAMEWORK FOR POLICYMAKERS

A practical decision framework could run as follows:

- **Step 1: Are insurance products impacted socially essential?**

  If yes, apply a higher threshold for exclusionary effects, stronger scrutiny of fairness, and/or tighter limits on highly damaging granular personalisation.

- **Step 2: What is the purpose of the data use?**

  If it primarily improves risk assessment, prevention, claims handling or fraud detection, it may be justified more easily. Price optimisations unrelated to risks may warrant separate consideration.

- **Step 3: Is the data use proportionate and plausibly linked to risk?**

  The report stresses the need to distinguish causality from mere correlation. Policymakers should be cautious where data points act as proxies that are difficult to explain, challenge or justify from an actuarial perspective, especially where they may correlate with protected characteristics. Consider how sensitive the relevant data is; can anonymous data work? However, it is also important to remain agile, as data use changes in view of new products, services, consumer needs and technologies. Data can have a dynamic life.

- **Step 4: What are the effects on consumers, insurers and society?** Always develop a data use impact assessment when assessing policies that can impact data use (encourage or restrict data use in practice). The impact assessment needs to be holistic and take into account consumers – individually and collectively -, insurers and society, and be explicit about the balance of outcomes that are being sought or accepted, both positively and negatively.

- **Step 5: Are meaningful alternatives preserved?**

  Where participation depends on telematics, wearables or similar data sources, or on providing consent for access to social media data, consumers should retain realistic non-participation options.

5.2. RECOMMENDATIONS FOR SUPERVISORS

Consistent with the principles and trade-offs discussed throughout this report, and taking into account the need for a proportionate approach that builds, where appropriate, on existing regulatory and supervisory frameworks, avoids unnecessary administrative burden, and preserves insurers' ability to underwrite risks and set premiums freely within the applicable legal framework, supervisors may wish to consider the following:

- **Embedding data use as a dimension in conduct supervision,** with a focus on outcomes rather than data use in itself. Supervisory assessments could consider impacts on access, affordability, fairness, trust, risk management and innovation, including through Product Oversight and Governance (POG) frameworks and AI-related supervisory activities.
- **Developing and monitoring – with market participants – indicators relating to exclusion and affordability**, particularly for vulnerable consumers, underserved groups and individuals who may be disadvantaged by limited data availability or limited ability to participate in data-sharing arrangements.
- **Challenging undertakings on their data-use impact assessments,** including pricing, underwriting, claims handling and distribution practices. Such assessments could consider not only predictive performance and operational benefits, but also implications for consumer outcomes, inclusion and fairness.
- **Following up on EIOPA's 2023 statement on differential pricing practices (non-risk-based pricing)**, particularly where pricing relies on consumers' willingness to pay, behavioural biases, inertia or other factors that may undermine fair and affordable outcomes.
- **As part of AI work, strengthening model governance expectations**, including data use aspects in relation to explainability, bias and fairness testing, model validation, meaningful human oversight, outcome monitoring and redress. As data-driven models become more sophisticated, governance should evolve accordingly, with supervisory attention focusing on ensuring accountability and responsible use.
- **Monitoring the implications of increasing reliance on third-party data providers, AI systems and digital ecosystems**, ensuring insurers remain accountable for the services and products they rely on.

GLOSSARY

(Direct) discrimination: There are different definitions of discrimination (Annex I). The common theme of these definitions is that a group of people is treated less favourably compared to others based on certain attributes shared by the group that is discriminated against. It is important to note that, while a certain practice may not be per se legally classified as discrimination, it does not mean that this practice does not constitute discrimination.

Indirect discrimination involves an apparently neutral practice or policy which puts members of a protected group (e.g., women) at a disproportionate disadvantage compared with members of a cognate group (e.g., men), and which fails to satisfy a means-end justification test[23]. The discriminator is usually entitled to show that the discriminatory policy is a necessary and proportionate means to pursue a sufficiently important objective or some variation of a means-end formula.

Intersectional discrimination: Discrimination arising from a combination of characteristics. A practice may not discriminate against either group considered separately, but may discriminate against individuals who belong to both groups.

Statistical discrimination is based on statistical beliefs about the members of a group rather than on direct evaluation of individuals alone. In other words, decision-makers use statistical generalisations about a group (for example, beliefs about the average likelihood of certain traits or outcomes) to make decisions about individuals who belong to that group.[24]

Fairness in insurance means ensuring that decisions concerning pricing, underwriting and claims are neither arbitrary nor discriminatory, including when they are made by algorithms.

Transparency: Ensuring that all stakeholders have access to clear and relevant information that enables them to evaluate products, pricing, and risk assessments according to their specific needs and objectives.

Explainability: A component of transparency referring to the ability to clearly articulate how decisions or outcomes are derived. For example, it may involve explaining how factors such as age, claims history or location influence underwriting decisions or pricing. It means making the relationship between data inputs and final outputs understandable to the intended audience.

[23] Lippert-Rasmussen, K. (ed.) (2018), The Routledge Handbook of the Ethics of Discrimination, Routledge.
[24] Ibid.

Accuracy and data quality: The obligation to ensure that data is complete, correct, reliable, and representative of the populations or phenomena it describes. Using inaccurate, outdated, or biased data can lead to unfair, discriminatory, or harmful outcomes.

Privacy: Ensuring that individuals' personal data is protected and handled in a secure, lawful, and transparent manner.

Consent requires that individuals provide a freely given, specific, informed, and unambiguous agreement for their data to be processed, with the ability to withdraw that consent at any time.

Purpose limitation means that personal data may only be collected and used for clearly defined, legitimate purposes and cannot be repurposed in ways that are incompatible with the original intent.

ANNEX I – LEGAL FRAMEWORK

REGULATORY BASELINE

There are legal rules on how to differentiate for each use case. There are broad legal rules like the General Data Protection Regulation (GDPR) that apply to all use cases where algorithms/models are used for differentiation. There are more specific legal rules for high-risk use cases such as insurance pricing for life and health insurance.

Laws/recommendations can either provide insights on outcomes (for example):

- EU Gender Directive (2004/113/EC) – bans gender as a pricing variable in insurance.
- EU Charter of Fundamental Rights – protects against discrimination on grounds such as ethnicity.

Or procedures (for example):

- EU Artificial Intelligence Act – treats life and health insurance pricing as high-risk, imposing risk-management, transparency and explainability duties on AI systems.

NON-DISCRIMINATION

Legislation requires fair treatment of customers. The Insurance Distribution Directive (IDD, Article 17) requires insurers and distributors to 'act honestly, fairly and professionally in accordance with the best interests of their customers'. Article 25 of the IDD and the POG Delegated Regulation set out that the product approval process must ensure that products are aligned with the needs, objectives and characteristics of the target market and that products do not adversely affect customers, to prevent customer detriment. The product approval process also needs to be proportional vis-à-vis the products sold and the nature, scale and complexity of the business model of the insurance manufacturer.

The AI Act seeks to address discrimination by imposing obligations on high-risk AI systems, their providers and deployers to ensure that such systems are designed, trained, and monitored in ways that prevent biased or discriminatory outcomes, and uphold fundamental rights throughout the AI lifecycle. This includes banning AI practices that exploit vulnerabilities or lead to materially unjustified or disproportionate treatment of individuals or groups (Article 5). Furthermore, Article 10 requires that high-risk AI systems be developed using training, validation and testing datasets that are relevant, representative, error-free and reflect the specific context, and permits the processing of special categories of personal data only when strictly necessary for bias detection and

correction. Additionally, the AI Act mandates that deployers of high-risk AI systems conduct a Fundamental Rights Impact Assessment (FRIA) under Article 27, which requires them to describe how the system will be used, who it may affect, what risks to rights might materialise, and how human oversight and mitigation will be built in.

The GDPR restricts the use of sensitive personal data—such as race, ethnicity, or health status— to prevent, inter alia, discriminatory effects on natural persons.

In the European Union, discrimination is addressed in the Gender Directive. It addresses both of direct and indirect discrimination:

- Direct discrimination: where one person is treated less favourably, on grounds of sex, than another is, has been or would be treated in a comparable situation.
- Indirect discrimination: where an apparently neutral provision, criterion or practice would put persons of one sex at a particular disadvantage compared with persons of the other sex, unless that provision, criterion or practice is objectively justified by a legitimate aim and the means of achieving that aim is appropriate and necessary.

Very similar definitions apply in the case of race, ethnic origin, religion or belief, disability, age or sexual orientation. When considering the AI Act, these definitions should be used.

Legislation often defines so-called protected attributes (arising, for example, from the UN Declaration of Human Rights) that cannot be used as a basis for categorising people. It is clear from EU rulings that even with these protected attributes there can be cases where indirect discrimination is allowed.

PRIVACY – LEGAL GROUNDS FOR DATA PROCESSING

Under the GDPR, data processing is governed by specific legal bases to ensure that personal data is handled lawfully and ethically. These bases provide a framework that balances the operational needs of organisations with the protection of individuals' rights and privacy. The key legal bases include:

- Fulfilment of Contract: Data processing is permissible when it is necessary for the performance of a contract to which the data subject is a party, or to take steps at the request of the data subject prior to entering into a contract. This basis ensures that organisations can process personal data to deliver services or fulfil contractual obligations.
- Legitimate Interest: Processing is allowed when it is necessary for the legitimate interests of the data controller or a third party, provided these interests are not overridden by the rights and freedoms of the data subject. This basis requires a careful balancing test to ensure that the interests are legitimate and do not infringe on the individual's privacy rights.

- Consent: Organisations can process personal data if the data subject has given clear, informed, and explicit consent. Consent must be freely given, specific, and easily withdrawable, ensuring that individuals have control over their personal data and understand how it will be used. Legal Obligations or Specific Statutory Allowances: Data processing is permissible when it is necessary to comply with a legal obligation to which the data controller is subject, or when specific statutory allowances (e.g. tax laws) authorise such processing. This basis ensures that organisations can fulfil legal requirements, such as tax reporting or regulatory compliance.

These legal bases collectively provide a framework for lawful data processing, balancing the need for organisational functionality with the protection of individuals' rights and privacy.

TRANSPARENCY

Transparency is a cornerstone of both the GDPR and the EU AI Act, reflecting a commitment to safeguarding individuals' rights and fostering trust in digital environments. Under the GDPR, organisations are required to provide clear, concise, and accessible information about how personal data is collected, processed, and used. This includes details such as the identity of the data controller, processing purposes, legal bases, and any third-party disclosures. Privacy notices must inform individuals of their rights, including access, rectification, erasure, and objection to processing. Consent must be specific, informed, and easily withdrawable. Additionally, individuals have the right to access their data and be notified promptly in case of data breaches. Transparency in automated decision-making is crucial, requiring explanations of the logic and potential consequences.

Similarly, the EU AI Act mandates transparency in AI systems, particularly in high-risk applications. Users must be informed when interacting with AI, and clear documentation of the AI system's capabilities, limitations, and decision-making processes is required. Together, these regulations aim to create a transparent digital environment that upholds individuals' rights, promotes ethical AI deployment, and fosters accountability and trust between individuals and organisations.

ACCURACY AND DATA QUALITY

Under the GDPR, data accuracy is a fundamental principle, requiring organisations to ensure that personal data is accurate, up-to-date, and maintained with high quality. Organisations must take reasonable steps to correct inaccurate data and keep it current, reflecting the true state of the information. This is essential for protecting individuals' rights, as inaccurate data can lead to erroneous decisions affecting individuals' lives. The GDPR also provides individuals with the right to rectify inaccurate data, empowering them to ensure their information is correctly represented.

Similarly, the EU AI Act emphasises data quality as a critical factor in the development and deployment of AI systems. High-quality data is essential for training AI models, ensuring they

perform accurately and effectively. The Act requires that data used in AI systems be relevant, representative, and free from errors or biases that could lead to discriminatory or harmful outcomes. This focus on data quality aims to ensure that AI systems make reliable decisions and operate ethically, minimising risks and maximising benefits for users and society.

Together, the GDPR and the EU AI Act underscore the importance of data accuracy and quality, fostering environments where personal data and AI technologies are used responsibly and effectively. By prioritising accurate and high-quality data, these regulations aim to protect individuals' rights, enhance trust in digital systems, and ensure ethical and reliable AI deployment.

While, as mentioned above, the EU AI Act stipulates that the data sets used by high-risk AI systems must be relevant, sufficiently representative, and free of errors, the regulation does not specify which types of data meet these standards for the different use cases of AI systems, including insurance. Similarly, while the GDPR includes the principle of data minimisation, it does not specify the types of data that should and should not be collected and processed for the different use cases under which AI processes personal data in consumer financial services.

HUMAN OVERSIGHT

Under Article 22 of the GDPR, individuals have the right not to be subject to decisions based solely on automated processing, including profiling, if such decisions have significant effects on them. This includes decisions that could affect an individual's legal rights or similarly significant outcomes, such as credit approvals or hiring decisions. The GDPR mandates that individuals must be provided with safeguards, including the right to obtain human intervention, express their point of view, and contest decisions. This ensures that automated systems are not making critical decisions without human oversight, protecting individuals from potential biases or inaccuracies inherent in automated processes.

The EU AI Act similarly emphasises the importance of human oversight in AI systems, particularly in high-risk applications. It requires that AI systems be designed and implemented in ways that allow for human monitoring and intervention. This means that human operators must have the ability to understand and manage the AI system, intervening when necessary to prevent harm or correct errors. The Act aims to ensure that AI systems do not operate in isolation but are subject to human control and accountability, reinforcing ethical standards and mitigating risks associated with autonomous decision-making.

Together, the GDPR and the EU AI Act highlight the necessity of human oversight in automated and AI-driven processes. By integrating human judgement and intervention, these regulations aim to safeguard individuals' rights, enhance accountability, and ensure that technology serves humanity ethically and responsibly.

As laid out in the previous section, the GDPR grants individuals the right not to be subject to automated decision-making where such decisions have significant effects on them. A common misinterpretation is that automated decision-making is generally allowed, while an opt-out mechanism is provided to individuals who wish to object. In fact, a closer look at the text and a review of guidance by the European Commission reveals that the GDPR in general prohibits the use of automated decision-making, while introducing three narrow exceptions under which automated decision-making is permitted:

- The decision based on the algorithm is necessary (i.e. there must be no other way to achieve the same goal) to enter into or to perform a contract with the individual;
- A particular EU or national law allows the use of algorithms and provides for suitable safeguards to protect the individual's rights, freedoms and legitimate interests; or
- The individual has explicitly given their consent to a decision based on the algorithm.

Accordingly, when relying on decisions made by AI which affect individuals, organisations should ensure that they carefully assess whether these decisions have a significant effect on these individuals and, if so, whether an exception allowing reliance on automated decision-making is permissible.

Notably, an individual's right not to be subject to automated decisions relating to them only applies if such decisions have significant effects on them. However, it is recommended that in case an individual complains about any automated decision relating to them, a human review of the decision is granted also if not legally required (i.e., also for decisions that have no significant effect on the individual).

More specifically to the insurance sectors, AI systems cannot make any binding financial decisions taken by insurance companies (i.e. setting of premiums, decisions on insurance claims, etc.) without human oversight and involvement. Moreover, there should be a written documentation of AI outputs which insurers must store so that they can be reviewed by consumers and/or public authorities at any time upon request.

ANNEX II – MEMBERS OF EIOPA'S CONSULTATIVE EXPERT GROUP ON DATA USE IN INSURANCE

Name & Surname	Affiliation
Alejandra Fernandez	VidaCaixa
Alexander Gamerdinger	Aalborg University
Anna Martin	Bureau Européen des Unions de Consommateurs (BEUC)
Chris Holland	Loughborough University & Universität Münster
Claudio Senatore	SAS Institute & Actuarial Association of Europe (AAE)
Davide Casinelli	Crédit Agricole Creditor Insurance (CACI)
Erzsébet Kovács	Corvinus University of Budapest
Esko Antero Kivisaari	Actuarial Association of Europe (AAE)
Felix Wenzel	ERGO Group AG
Florence Bailly	CNP Assurances
Jaakko Mikkonen	If P&C Insurance
Jonas Hirz	Boston Consulting Group & Actuarial Association of Europe (AAE)
Katarzyna Malinowska	Kozminski University & BIPAR
Lee Harris	Zurich Insurance
Melanie Sophie Schlünder	PricewaterhouseCoopers GmbH WPG

Mikko Kuusela	Finance Finland
Nico Lauridsen	European University Institute (EUI)
Nicolas Shire	Independent advisor, Data & AI
Paola Scarabotto	Groupama Assicurazioni
Pedro Écija Serrano	Independent consultant
Peter Norwood	Finance Watch
Philipp Raether	Allianz SE
Rens Garssen	Oliver Wyman
Simon Hatzesberger	Deloitte & Actuarial Association of Europe (AAE)
William Vidonja	Insurance Europe

EIOPA

Westhafen Tower, Westhafenplatz 1
60327 Frankfurt – Germany
Tel. + 49 69-951119-20
info@eiopa.europa.eu
https://www.eiopa.europa.eu