

Artificial intelligence governance principles

A summary of the report by EIOPA's Consultative Expert Group on Digital Ethics in Insurance

Charlie Howell, FIA
Kapil Relwani



In June 2021, the European Insurance and Occupational Pensions Authority (EIOPA) published a report from its Consultative Expert Group on Digital Ethics (GDE) setting out six artificial intelligence (AI) governance principles “to promote an ethical and trustworthy Artificial Intelligence in the European insurance sector.” This briefing note summarises the key points and findings of the GDE report.

Against a background of increased adoption of AI across the insurance industry,¹ and a multitude of ethics guidelines published by various public and private institutions, the GDE report sets out an insurance-specific set of considerations. Each principle is followed by guidance on how insurance firms can implement them in practice and with reference to insurance-specific AI use cases.

The six principles covered are:

- Proportionality
- Fairness and nondiscrimination
- Transparency and explainability
- Human oversight
- Data governance and recordkeeping
- Robustness and performance

These are each summarised in the sections which follow.

The report builds on recent European Union (EU) initiatives on AI including the Ethics Guidelines for Trustworthy AI² published by the European Commission's High Level Expert Group on AI (AI HLEG) in April 2019, and the European Commission's white paper on AI³ published 11 months later.

While the GDE recommendations are nonbinding, they provide valuable guidance, and sit alongside the existing legislative framework relevant for the use of AI, such as the General Data Protection Regulation (GDPR), the Solvency II Directive, the Insurance Distribution Directive (IDD) and EU and national antidiscrimination legislation.

The AI governance principles report by the GDE can be found on the EIOPA website:

https://www.eiopa.europa.eu/content/eiopa-publishes-report-artificial-intelligence-governance-principles_en.

Milliman has explored the ethics and ethical guidance around AI use further in our paper "Artificial Intelligence: The ethical use of AI in the life insurance sector." The paper can be found here: <https://uk.milliman.com/en-GB/insight/artificial-intelligence-the-ethical-use-of-ai-in-the-life-insurance-sector>.

Principle of proportionality

EIOPA's GDE stresses the principle of proportionality in the context of AI; insurers should ensure that governance measures implemented are “proportionate to the characteristics (impact) of the specific AI use case at hand.” To assess this, the GDE proposes that insurers undergo its AI use case impact assessment, identifying someone, such as an AI or data protection officer, who will be responsible for developing and documenting the assessment.

The proposed framework is designed to utilise existing assessment mechanisms. The assessment of risk to consumers is designed to follow a process similar to the data protection impact assessment (DPIA) as required under GDPR. The assessment of the impact on insurers is based on insurers already assessing these risks in the Own Risk and Solvency Assessment (ORSA). In addition to these aspects, the framework proposes some additional insurance or AI-specific considerations, such as incorporating the fundamental rights impact assessment (FRIA) proposed in the AI HLEG guidelines.

The AI use case impact assessment entails an investigation into both the severity of any potential harm caused by the use case and the likelihood that harm will occur. The severity assessment will consider factors such as which customers will be affected and to what extent, impacts on business continuity and financial, legal and reputational repercussions. The likelihood for harm will depend on the model purpose, the sensitive nature of any data used and the model complexity, among other factors.

Based on the severity and likelihood, use cases can be assigned as high, medium or low impact. It is expected that this assessment should then inform the level of caution to be taken around each of the other principles, with proportionate governance measures put in place.

Principle of fairness and nondiscrimination

According to Article 17(1) of the IDD, insurance firms have a responsibility to “act honestly, fairly and professionally in accordance with the best interests of their customers.” However, the notion of fairness is difficult to define precisely and can have several meanings. For example, it may be actuarially fair to charge higher premiums to a group because of its risk profile, but may be considered societally unfair given the vulnerabilities of the members of this particular group, and the impact on their ability to access insurance.

The GDE considers two dimensions of fairness: procedural and distributive. Procedural fairness relates to fair conduct and governance, ensuring consumers are treated professionally and can seek redress. Distributive fairness is about ensuring that insurance is accessible, affordable and free from bias and discrimination. Both may be affected by the use of AI. As procedural fairness relies greatly on factors such as transparency, oversight and governance, it will effectively be covered in the later sections of this paper. This section will focus on distributive aspects of fairness.

One of the major strengths of artificial intelligence is finding patterns and correlations in training data and using them to make predictions based on new data. However, if the data sets are not representative or otherwise contain biases, then the AI system’s outputs may replicate these biases. In supervised machine learning algorithms, another concern is that data sets are labelled in a way that reflects any prejudices of the human developer. Firms should aim to minimise this bias as much as is reasonable.

Additionally, while for traditional models firms may remove protected attributes⁴ and their proxies from data sets or use these variables as control variables to remove bias, this may not be sufficient for AI systems. AI models capture complex relationships between variables, so they may identify unexpected variables, or combinations of variables to serve as proxies for protected attributes. For AI applications assessed as being high impact, the report advises that firms should aim to make their models explainable, use the minimum of data necessary and examine model outputs carefully to identify bias.

The GDE acknowledges that special attention should be given to the impact of AI systems on vulnerable consumers who are more susceptible to harm. For example, vulnerable customers may be “nudged” into buying more by AI systems using price optimisation practices, or customers may struggle to access insurance if they lack a digital profile.

A number of fairness and discrimination metrics are being developed through the research field of “fair machine learning.” While they are not mature solutions, and the applicability of each metric will be case-specific, they provide some ideas of how to assess the fairness of outcomes from an AI model.

The report notes that, while AI and big data bring new risks with respect to fairness, they can also have benefits. For example, telematics devices may improve access to affordable coverage by facilitating more granular risk assessments of young drivers.

For this and each of the subsequent principles, the GDE discusses particular considerations in the context of specific insurance use cases. Fairness considerations are summarised in the table in Figure 1.

FIGURE 1: FAIRNESS USE CASE GUIDANCE

| Use Case | Guidance |
|---------------------------------|--|
| Pricing and underwriting | <ul style="list-style-type: none"> • Should try to mitigate the impact of socioeconomic rating factors, including location and occupation, on vulnerable groups and protected classes even if there is a limited causal link. • Rating factors that can be influenced by the consumer are preferable to rating factors that can’t be changed. • Certain price optimisation practices, such as those that aim to maximise a customer’s “willingness to pay,” should be avoided for essential lines of insurance. |
| Claims management | <ul style="list-style-type: none"> • Need to avoid claims optimisation practices that calculate individuals’ “willingness to accept” when they unfairly harm consumers. |
| Fraud detection | <ul style="list-style-type: none"> • Human oversight necessary for fair and nondiscriminatory outcomes. |

Principle of transparency and explainability

Transparency is defined in the report as “providing information about the use, the nature or design of an AI system and the data variables and parameters used.” Explainability is a related concept about explaining how an AI system got to a particular output by providing information on which variables impacted the decision.

A lack of transparency and explainability makes it harder for consumers to make informed decisions and to challenge decisions made by an AI system, for example if they don’t know what they could do to reduce their premiums or what caused them to be denied cover.

In addition, an opaque AI system may hide weaknesses and biases which thus go unaddressed, potentially resulting in inaccurate predictions and discriminatory outcomes.

The GDE therefore stresses the importance of providing explanations, noting that they should be adapted based on the stakeholder and use case. Regulators and auditors require comprehensive explanations about the inner workings of the AI system and its component parts, while customer communications should focus on providing meaningful, easy-to-understand explanations, covering the variables relevant to the decision.

The report acknowledges that there can sometimes be a trade-off between model explainability and accuracy. For instance, some AI systems, including the use of deep learning techniques, can provide very accurate predictions but they can also be considered as a "black box" due to the difficulty of explaining how the decision was made. In some situations, it may make sense for an insurance firm to compensate a lack of explainability with other governance measures such as an enhanced level of human oversight and data management. However, where a use case has a high impact on consumers or the firm, the report stresses that explainability must be the priority, even at the cost of model performance.

An additional aspect of transparency is in informing consumers of when they are interacting with an AI system (e.g., chatbots or robo-advisors), and what data the system uses.

The GDE report provides further guidance for the use cases shown in Figure 2.

FIGURE 2: TRANSPARENCY USE CASE GUIDANCE

| Use Case | Notes |
|--|--|
| Pricing and underwriting | <ul style="list-style-type: none"> • Need to demonstrate accurate pricing to regulator. • Need to ensure consumer understanding of the rating factors affecting them. |
| "Next Best Action" modelling⁵ | <ul style="list-style-type: none"> • More indirect consumer impact as only providing a suggested product. • Should consider "likelihood to need" not "likelihood to buy" when identifying target customers. |
| Fraud detection | <ul style="list-style-type: none"> • Explainability is less key as there will always be further investigation by a human before a customer is affected, but bias should be investigated and minimised. |
| Optical character recognition (OCR) and image processing techniques | <ul style="list-style-type: none"> • A high level of human oversight may not be needed for small claims but should be in place above certain thresholds. • Use in sensitive contexts (e.g., identifying faces, use in pricing) requires additional governance. |

Principle of human oversight

The report defines human oversight as "some form of direct human involvement in the design, operation, maintenance, adaptation or application of AI systems." Incorporating some form of oversight enables humans to challenge model outputs and intervene if the system starts producing unintended or unethical outcomes.

In the design phase, the GDE advises that specific high-risk AI applications (including in pricing and underwriting) should have guardrails embedded into the system. These guardrails would restrict the extent to which a system could autonomously make a decision, for example by capping the price which could be charged. When the system is in production, the report proposes that "oversight shall focus on the day-to-day operations/processes of the AI system, monitoring and controls of the AI system, adjustments and incident handling according to the previously established procedures, such as in the case of the guardrails." In this phase, the GDE's prescribed oversight measures include reviewing the system's outputs before they become effective and monitoring the system's impact on protected and vulnerable groups.

The report also emphasises that it is important for insurers to establish effective accountability frameworks, documenting the roles and responsibilities of all the employees involved in the implementation of AI systems within the firm. The management board should have an adequate understanding of the use of AI in its organisation and the potential risks. Developers of AI systems should receive frequent training so that they are aware of the ethical issues relating to AI. There may also be value in creating a multidisciplinary ethics committee and/or an AI officer position, responsible for overseeing and advising the firm on the use of AI.

The GDE report provides further guidance for the specific use cases shown in Figure 3.

FIGURE 3: OVERSIGHT USE CASE GUIDANCE

| Use Case | Guidance |
|------------------------------------|---|
| Pricing and underwriting | <ul style="list-style-type: none"> • A relatively high level of human oversight is needed in the design and testing stages. In the production stage, however, a more autonomous system may be fairer as it can help ensure that similar risks are priced consistently. |
| Claims management | <ul style="list-style-type: none"> • Human oversight should be proportionate to impact of claims on consumers. • Useful to have triggers (e.g., claim size, customer vulnerability) that define when oversight is needed. |
| Loss prevention⁶ | <ul style="list-style-type: none"> • Human oversight should be proportionate to potential impact on consumers and should check whether advice provided is practical.⁷ |

Principle of data governance and recordkeeping

Although the insurance sector is no stranger to data analytics, the report notes that "there is an increasing availability of new sources and types of data (e.g. Internet of Things (IoT) data, image data or social media data), which can be processed by increasing powerful and complex AI systems, bringing several opportunities, but also some challenges." The report states that it is essential, therefore, that only high-quality data is used and privacy and data protection are respected throughout all stages of the AI system's life cycle.

The GDE stresses that the GDPR, which came into force in May 2018, should act as the basis for sound data governance. This legislation imposes strict requirements on the processing of personal data by insurers. Firms, for instance, must obtain informed consent from consumers and ensure that they adhere to the principles of purpose limitation,⁸ data minimisation⁹ and confidentiality.¹⁰

In addition, the report notes the importance of ensuring that all data sets used to build AI models are accurate, complete and appropriate so that the systems are robust and without bias. This includes data purchased from third-party vendors as well as that available internally. Lastly, the report proposes that insurers need to keep records of the data used in AI systems including the modelling methodologies. This should enable insurers to trace back and verify decisions. Having these records in place is particularly important for "black box" AI models where there may be limited other information about how a result was derived.

The GDE report provides further guidance for the use cases shown in Figure 4.

FIGURE 4: DATA GOVERNANCE USE CASE GUIDANCE

| Use Case | Guidance |
|--|---|
| Pricing and underwriting (price optimisation) | <ul style="list-style-type: none"> • A high-impact use case, so data use must be well documented. • Behavioural data is dynamic so models require more retraining and re-estimation. |
| Claims management (image recognition) | <ul style="list-style-type: none"> • Need to specify image requirements so claimants can provide high-quality images which can be correctly interpreted by the model. |
| Loss prevention | <ul style="list-style-type: none"> • Ongoing new data leads to new recommendations. • High-impact use case, so needs thorough documentation and checks on data quality, retraining and data reconciliation. |

Principle of robustness and performance

The GDE stresses that the robustness and performance levels of AI systems is critical to ensure users can be confident that the results are fair.

This is not just a question of the model itself; it depends on the context in which the model is trained and applied. Models rely on representative and high-quality training data and must be used in the context for which they were designed. If there are significant changes to the purpose or characteristics of an AI system (including changes to the input data and the legal/economic environment), then the system should be retrained, recalibrated and revalidated.¹¹

This should also be done on an ongoing basis, assessing any changes, and understanding when the model's limitations may become more relevant.

Monitoring of AI systems should use performance metrics that depend on the intended purpose. The GDE uses the example of classification systems used in fraud detection under which "insurance firms should decide if the objective is to maximise the prediction accuracy (number of fraudulent claims detected), reduce the number of false positives (legitimate claims wrongly labelled as fraudulent) or false negatives (claims labelled as legitimate which in the end are fraudulent)." For each one of these objectives, the performance metrics used may be different. More broadly, the report advises that insurers should assess the stability of model predictions over time, comparing against older outputs or independent models.

A further consideration is security. AI models require more computational power than the traditional models and are therefore more at risk from cyberattacks, making strong information security procedures critical. The GDE recommends that insurers create fallback plans in case the system no longer works as intended or the firm is a victim of a cyberattack, particularly for high-impact systems.

The GDE report provides further guidance for the use cases shown in Figure 5.

FIGURE 5: ROBUSTNESS USE CASE GUIDANCE

| Use Case | Guidance |
|--|--|
| Pricing and underwriting | <ul style="list-style-type: none"> • Should assess how to reduce the uncertainty of the risks being priced. • Lack of data can worsen pricing model performance so should disclose whether data shortcomings will alleviate over time. • Need to understand and disclose the impact of inadequate model performance. • Pricing models should require significant robustness. • Should explore fallback plans if the pricing model doesn't perform adequately. |
| Optical character recognition and image processing techniques | <ul style="list-style-type: none"> • If human oversight is in place, less need for robustness. • Image quality can impact OCR model performance; so should request better quality data. |
| Loss prevention | <ul style="list-style-type: none"> • If loss prevention measures are clearly just used to provide advice, then performance and robustness requirements don't need to be as strict. |

Conclusion and further thoughts

The GDE's guidance aims to set out how insurers can harness the benefits of AI, while taking a proportionate approach to address the risks AI generates and to promote trust in the use of AI. EIOPA notes that it plans to use the report's recommendations to explore potential supervisory initiatives in this field.

The report should be considered in the context of the broader EU developments to promote the adoption of ethical AI. It also provides an insight into what ethical issues are expected to be most relevant for the insurance industry. For example, the six principles outlined above were adapted from the seven "Key Requirements for Trustworthy AI" identified by the AI HLEG.¹² Some of the key requirements are considered less applicable for insurers, such as "Societal and environmental well-being,"¹³ which is folded into the discussion around fairness, while a new principle of proportionality is added.

Firms may still benefit from consulting the overarching guidance by the AI HLEG for a broader perspective, or to make use of tools like the assessment list for trustworthy AI.¹⁴

As the GDE guidance is principle-based, firms will still need to seek out more detailed guidance on best practice techniques to implement AI in line with these principles. For example, firms might explore the use of surrogate models to improve model interpretability, or put together an Adversarial Threat Matrix to assess security issues. It is at this level of detail that firms have the freedom to tailor their controls and governance to address their specific circumstances and uses of AI.

Whilst these European developments might not be directly relevant to UK insurers, there is an overlap in how the UK and EU approach AI ethics. Indeed, discussions^{15,16} by the Financial Conduct Authority (FCA) and the Financial Services Artificial Intelligence Public-Private Forum (AIPPF) of the Bank of England (BoE) have looked to guidance in the AI HLEG's "Ethics guidelines for Trustworthy AI" and the example provided by the European Commission proposal for AI regulation.¹⁷

AIPPF meetings have covered topics similar to those in the EIOPA paper such as the challenge of defining fairness, the development of stronger data governance standards and the need for a risk-based approach, requiring more explainability in high-risk use cases. UK firms may therefore benefit from understanding the EIOPA guidance and considering how their practices align with the proposed principles.

Naturally, this is a rapidly evolving field, with extensive ongoing research into AI applications and new tools for explainability, transparency and fairness being developed. Guidance will evolve along with this and firms should keep up to date to ensure they are applying best practices.



Milliman is among the world's largest providers of actuarial and related products and services. The firm has consulting practices in life insurance and financial services, property & casualty insurance, healthcare, and employee benefits. Founded in 1947, Milliman is an independent firm with offices in major cities around the globe.

[milliman.com](https://www.milliman.com)

CONTACT

Charlie Howell
charlie.howell@milliman.com

Kapil Relwani
kapil.relwani@milliman.com

ENDNOTES

- ¹ European Commission. European Enterprise Survey on the Use of Technologies Based on Artificial Intelligence. Retrieved 29 July 2021 from <https://digital-strategy.ec.europa.eu/en/library/european-enterprise-survey-use-technologies-based-artificial-intelligence>. The European Commission in 2020 found that the financial and insurance sector had one of the highest growth prospects of any industry, with 27% of firms planning to adopt AI technology in the next two years.
- ² European Commission (8 April 2019). Ethics Guidelines for Trustworthy AI. Retrieved 29 July 2021 from <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.
- ³ European Commission (19 February 2020). White Paper: On Artificial Intelligence – A European Approach to Excellence and Trust. Retrieved 29 July 2021 from https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf.
- ⁴ The list of protected characteristics in the EU Charter of Fundamental Rights includes gender, ethnic or racial origin, religion or belief, disability, age, sexual orientation and nationality. Age, disability, religion or belief and sexual orientation are permitted for risk-based pricing and underwriting purposes (with restrictions).
- ⁵ Modelling to identify up-selling or cross-selling opportunities.
- ⁶ Providing customers with advice to help them reduce the risk or severity of a claim
- ⁷ The provided example is that it may be safer to drive earlier in the morning, but a recommendation to leave two hours early for work will not be useful advice.
- ⁸ This is defined as “requiring firms to inform individuals about the specific purposes for processing the data and ensuring that any further processing is compatible with the original purpose.”
- ⁹ This is understood in the report as only collecting and processing the personal data that is necessary for the purposes for which it is to be used.
- ¹⁰ Firms should have protocols in place to ensure that data used in AI models is stored in a safe environment (e.g., data should only be handled by the appropriate staff).
- ¹¹ The validation process should include scenario analysis and stress testing for high-impact AI systems.
- ¹² Ethics Guidelines for Trustworthy AI, op cit.
- ¹³ This covers the impact of AI applications on the environment, social relationships, institutions, democracy and society at large.
- ¹⁴ High-Level Expert Group on Artificial Intelligence. The Assessment List for Trustworthy Artificial Intelligence (ALTAI). Retrieved 29 July 2021 from https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342 (PDF download).
- ¹⁵ Artificial Intelligence Public-Private Forum (26 February 2021). Second Meeting: Minutes. Retrieved 29 July 2021 from <https://www.bankofengland.co.uk/-/media/boe/files/minutes/2021/aippf-minutes-february-2021.pdf?la=en&hash=9D5EA2D09F4D3B8527768345D472D9253906ADA1>.
- ¹⁶ Artificial Intelligence Public-Private Forum (15 June 2021). Third Meeting: Minutes. Retrieved 29 July 2021 from <https://www.bankofengland.co.uk/-/media/boe/files/minutes/2021/aippf-minutes-june-2021.pdf?la=en&hash=783B2A11EA5074CC90B4E8A71F1745F56ACB4377>.
- ¹⁷ Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. Retrieved 29 July 2021 from <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206>.