ETHICAL AI IN LIFE AND NON-LIFE INSURANCE

A Framework for Mapping Ethical Trade-offs in AI use

October 2024

F&P FORSIKRING OG PENSION / ADD algoritmer data & demokrati

Ethical AI in Life and Non-Life Insurance:
A Framework for Mapping Ethical Trade-Offs in AI Use

Principal Authors
Alexander Gamerdinger, ADD/CBS
Jakob Holm, F&P

Contributors
Sigrid Floor Toft, F&P
Ole Willers, ADD/CBS
Leonard Seabrooke, ADD/CBS
Prins Marcus Lantz, ADD/RUC

Reviewers
Helle Wicksell, CODAN (Alm. Brand)
Alexander Sokol, AP Pension
Mathias Siggard, PKA
Thomas Brenøe, F&P

Layout
Peter P. T. Unset Hommel, F&P

Gamerdinger, A., & Holm, J. (2024). Ethical AI in Life and Non-Life Insurance: A Framework for Mapping Ethical Trade-Offs in AI Use. Copenhagen: Forsikring & Pension.

Table of contents

1. Executive summary

This report builds on previous work on AI ethics governance by bringing ethical governance from the horizontal to the concrete use-case level. It introduces a tool designed to facilitate the ethical adoption of AI systems in the life and non-life insurance sector by addressing ethical dilemmas. The tool can be integrated into an AI use case impact assessment and naturally guide its implementation processes. It highlights three critical insights:

1. The successful adoption of AI systems in insurance demands a combination of ethical reasoning and technical skills

2. The implementation process necessitates a diverse team, including both technical and non-technical members, along with appropriate management engagement

3. Technical solutions and ethical considerations should be anchored in appropriate governance frameworks covering issues on impact, input data, modelling and fairness

The report has been written as part of the *Algorithms, Data and Democracy (ADD)* project in collaboration with the Danish trade association for insurance companies and pension funds *Forsikring & Pension (F&P)*.

2. Towards Ethical AI in insurance

Richer datasets and more sophistical modelling using AI technology have promised insurers a technological transformation of risk assessments, claims processing and customer engagement. Potential benefits include more tailored product offerings, improved prediction and prevention of risks and increased efficiency and lower costs[1].

This hype of technological solutionism has given rise to equally large societal fears of potential bias, discrimination against race and gender and opacity of AI systems[2]. These AI-specific risks have become intermingled with traditional insurance-specific issues including the unaffordability and exclusion of insurance coverage.

Given these complexities, an important task in the adoption of AI systems has been to balance consumer expectations, regulatory demands and business innovation. Consequently, the adoption of fair and ethical AI systems for all stakeholders has emerged as the key priority in the sector.

Adopting ethical AI systems is a challenging task, as it goes beyond the traditional work of legal compliance. An overview over recent regulatory initiatives in the sector can be seen in Annex 1. In recent years, various stakeholders at different levels have developed ethical principles to guide the adoption of AI systems for insurance and pension undertakings. In 2021, the European regulator EIOPA published six ethical principles for life and non-life insurers to navigate the inherent ethical complexities associated with AI use[3]. In Denmark, the trade association for insurers and pension funds F&P, has developed data ethics principles at the sectorial level[4], building on its previous work on data ethics[5].

In 2020, the Danish Annual Accounts Act — and the related financial sector regulation - was amended to require large Danish firms to include a data ethics statement in their annual report, prompting many undertakings to actively engage with data ethics[6]. While the design of ethical guidelines encouraged insurance and pension undertakings to incorporate ethical thinking into their data work, they soon acknowledged that broad ethical principles cannot stand by their own. As the adoption of AI systems matures, insurers have recognized that a principle-based approach needs to be accompanied with explicit governance frameworks, guidelines and ethical considerations on a use case-level to ensure the trustworthy application of AI systems.

2.1. A focus on Ethical trade-off mapping

This report builds on previous work on AI ethics governance by bringing ethical governance from the horizontal, to the concrete use-case level.

We aim to provide a tool that strengthens the capabilities of insurance and pension undertakings in developing ethical AI systems. This is done by shifting the focus away from broad ethical principles to the mapping of ethical trade-offs and dilemmas in concrete AI use cases specific to insurance.

To facilitate a cross-sectorial discussion, F&P in collaboration with the Algorithms, Data and Democracy Project, organized a technical workshop at Copenhagen Business School to carve out crucial parameters that determine the extent to which specific AI use-cases can be considered trustworthy, fair and acceptable.

On June 4, 2024, 36 experts with technical backgrounds from actuarial science, data science and compliance from 16 Danish pension- and insurance sector discussed ethical drawbacks and potential solutions in a variety of AI use cases. In groups of six to seven, the participants discussed four AI use-cases, which were specifically designed by F&P for this workshop. A brief description of each of the cases can be seen below.

Based on the expert discussions, we have designed a tool that can be used side-by-side with the development of an AI-system. The tool maps observed cross-case dy-

[1] Noordhoek, "Regulation of Artificial Intelligence in Insurance: Balancing Consumer Protection and Innovation."

[2] Obermeyer et al., "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations."

[3] EIOPA, Artificial Intelligence Governance Principles, towards Ethical and Trustworthy Artificial Intelligence in the European Insurance Sector.

[4] Forsikring & Pension, "Cool Eller Creepy: Databrug Og Dataetiske Principper i Forsikrings- Og Pensionsbranchen."

[5] Forsikring & Pension, "Towards a Common Data Ethics."

[6] Voldby, "Nyt krav om redegørelse for dataetik i årsrapporten."

namics, helps to identify ethical trade-offs and contributes to define the overall level of governance needed for a specific AI-use case.

The tool, which is attached as an appendix, is targeted at developers, compliance experts, lawyers and everyone else who is actively involved in the implementation of AI systems in insurance and pension undertakings. Specifically, the exercise is intended to inform discussions on:

- The sensitivity of the AI-system based on a variety of parameters
- Trade-offs between parameters that shape the ethical design of the AI system
- The severity of governance and communication measures needed in the AI-use case
- The finetuning of the AI use-cases to reach an acceptable solution to all stakeholders

The two next sections explain how to use the tool in practice.

2.1.1. Step one: Design – Score parameters

Expert discussions highlight that the legitimacy of an AI-system, the severity of governance measures and the chosen communication tools depends on choices made within three core AI governance themes vis-à-vis the specific insurance use case. The three themes are comprised of:

1. **Procedural Fairness & Impact:** The fairness of the AI system situated in a specific use case and its perceived impact on the client, the insurer and overall society.
2. **Data collection:** Choices related to input data used in the algorithm or model
3. **Modelling decision:** Choices related to the design of the algorithm or model used in this use case

Each theme is comprised of a list of parameters that affect the overall riskiness of the AI system. Figure 1 explains the intuition: the farther to the left one scores a parameter, the less risky it is. The opposite is true if the parameter is scored "riskier".

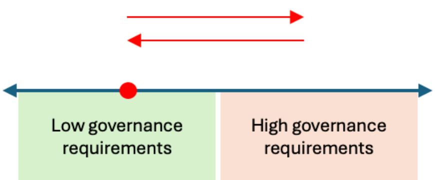


Figure 1: Example of a parameter scale

7 Došilović, Brčić, and Hlupić, "Explainable Artificial Intelligence."

2.1.2. Step two: Discuss - Identify and discuss ethical trade-offs

Once parameters have been assigned a score, the mapping enables the identification of any ethical trade-offs between the parameters.

Figure 2 for example shows how the scoring of two parameters within the *modelling decision* theme reveals a fundamental trade-off: the tension between model accuracy and interpretability. It is a well-established fact that greater predictive accuracy of a model comes at the cost of interpretability; it is not possible to achieve both simultaneously[7]. This trade-off demands developers to make a compromise decision (such as accepting a lower model accuracy) to find an acceptable level of model interpretability.

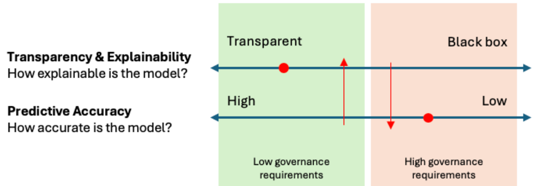


Figure 2: Example of trade-off mapping

Trade-offs are not only possible between parameters within a given theme, but also between different themes. For example, a more accurate model will typically require a greater number of data sources, which might lead to a heightened sense of surveillance among customers.

Group discussions revealed that ethical dilemmas arose in every discussed AI use-case. It was recognized that the nature of these dilemmas often precludes a technical solution that satisfies all interests equally. The successful adoption of an AI system therefore always involves finding compromise positions on identified trade-offs.

2.1.3. Step three: Deploy – Inform fairness, governance and communication requirements

After all parameters within the three themes have been scored, and ethical trade-offs have been identified, the tool will inform developers about decisions regarding *Outcome Fairness, Governance & Communication.*

While the first two steps should be undertaken in the idea phase of adopting an AI system, step three is meant to be scored once the team of developers have gained a more detailed overview of the AI-system and its impacts.

In sum, the transition from broad ethical governance principles towards a focus on ethical trade-offs at the use case level highlights the importance of debate and achieving compromise between key stakeholders to understand the dynamics between data collection, AI-model choices and fairness considerations. The adoption of successful AI solutions demands ethical frameworks on the organizational level, the engagement of a diverse team and the focus on tensions.

2.1.4. AI Use-Cases discussed in the workshop

Within the workshop different groups of actors discussed the following AI use cases. The cases are fictive but based in observed trends and use cases observed in the industry.

The use of more granular data in housing insurance underwriting: Using specific addresses instead of zip code as a risk factor for premium setting. Combined with the use of machine learning to create more granular risk groups.

The use of sensor data and AI in car insurance underwriting: Using telematics sensor data from IoT technology to differentiate insurance premiums and provide customer information for risk prevention.

The use of AI in the proactive prevention of stress-related diseases: Using a combination of in-house data and a machine learning algorithm to conduct individual risk assessments to prevent stress-related diseases through human oversight.

The use of voice data and AI in the detection of fraudulent behavior: Using a combination of voice data and machine learning to identify fraudulent behavior through the detection of voice data during customer claims calls.

3. AI Use Case Impact

The development of an AI system should begin with an AI Use Case Impact Assessment, which helps developers and compliance professionals think about the potential impact on key stakeholders. The tool presented in this report can be used from the design- to the evaluation phase of an AI system, as well as whenever it is updated.

An AI Use Case Impact Assessment can help to evaluate how comprehensive the governance ambitions of the specific AI-system should be. The idea behind AI use-case assessment is not new, but firmly integrated in the GDPR, EIOPA's AI ethics principles and the AI-Act . In practice, this tool can be used as part of risk assessments, such as the Data Protection Impact Assessment (DPIA).

The focus of this report is to incorporate an ethical trade-off mapping tool into an impact assessment to highlight areas where parameters are opposing or conflicting, and to facilitate discussion about which parameters should be weighted most heavily for a given use case.

3.1.1. How to use the ethical trade-off mapping?

In the following section, we will briefly describe the different themes and parameters. The overview is kept minimal – reflecting on only selected key parameters identified during the workshop. It is possible to add numerable sub-metrics or other overarching themes to customize the overview.

3.2 Procedural Fairness & Impact

The first theme deals with the *procedural fairness & impact* of an AI system on core stakeholders. In the academic literature, there is a distinction between *procedural fairness* and *outcome fairness*. The former refers to the evaluation of the fairness of the procedures and methods used to reach decisions, while the latter pertains to the fairness of the results or decisions themselves. . For example, procedural fairness in underwriting emphasizes the importance of transparency, consistency, and impartial treatment of customers in the procedure, while outcome fairness would mean that the final decisions are not discriminatory or biased against vulnerable groups of people.

3.2.1. Fairness alignments between core stakeholders

The workshop discussions highlighted the critical role of fairness considerations in all AI use cases. A key topic – and thus key part of the mapping - was the extent to which the AI system aligns the interests of users, the pool of customers, the firm, and society at large. The participants agreed that if an AI system serves the interests of all stakeholders, it will be perceived as more legitimate than if it only caters to two or only one of the core stakeholders.

The **core stakeholders** include:

- The affected client
- The pool of clients
- The insurance or pension company
- The larger society

Experts agreed that if an AI system demonstrably benefits all core stakeholders, the use of AI in more sensitive use cases and the collection of more extensive input data was deemed acceptable.

We can illustrate this logic with two examples from the group discussions. Using a policyholder's address instead of their postal code as a rating factor in housing insurance would allow the insurer to assess risks more precisely. It would benefit the company by allowing it to charge more competitive premiums to lower-risk customers. A lower price would also be beneficial for some customers, but not necessarily for the overall pool of customers as high-risk customers could face higher premiums. The impact on the larger society is thus uncertain.

While some argue that risk individualization could lead to a higher coverage rate, others point out that higher insurance prices for high-risk customers would reduce their coverage rates. All experts agreed that the deployment of AI in this context fails to align the interests of core stakeholders.

The use of AI to identify clients at risk of developing long-term diseases like stress or anxiety is viewed differently. By identifying potential individuals and contacting them proactively, health professionals can prevent the onset of long-term diseases. This is beneficial to customers, especially because there is no penalty for clients who are not identified. The AI system serves also the interest of the company, as it can reduces long-term payouts if disability is prevented. Finally, society benefits from the use of AI, as the system has the potential to reduce long-term diseases on a broader scale.

3.2.2. Impact on the data subject

Compared to other actors, the final customers are seen as the most important stakeholder as they are subjected towards the decision of AI models. The workshop revealed three overarching metrics that have a decisive influence on how a customer will perceive the procedural fairness of an AI system:

1. **Sensitivity:** How sensitive is the AI use case? Is the system used in health or bike insurance? Does it have a low or high financial impact?
2. **Client experience:** Is the input data and the model in general seen as intrusive or non-intrusive?
3. **Penalty of "opting out":** Will the customer face a penalty – directly or indirectly -, if she says no to participate?

While the parameters address the procedural fairness of models in relation to clients, it is important to recognize possible trade-offs between parameters within forthcoming themes.

The use of AI to combat fraud was the subject of discussion in the workshop. Experts reached an initial consensus that the AI system has a positive impact on the company, society, and the entire customer base. By reducing fraud, the system will reduce fraudulent payouts, which could be used to reduce premiums. However, non-fraudulent customers may still bear the cost of enhanced fraud detection by "paying" with more data and increased fear of surveillance. This presents a trade-off between comprehensive data collection and fairness for individual customers. Resolving this ethical dilemma by reducing data collection could potentially diminish model accuracy and, consequently, the overall benefits.

The final topic of discussion was about the unfair treatment of customers who opt out of digital insurance services that require expanded data collection. The experts expressed concern that customers who opt out may be perceived as high-risk due to their reluctance to share more data, i.e. sensitive health or driving information.

However, it is unclear if their concerns are specifically related to risk, not wanting to share more data by principle or if they face broader challenges in navigating digital use cases, such as limited digital literacy.

In essence, this parameter should also show if offering some customers better insurance terms due to 'opting in' on enhanced data collection might negatively affect the residual group.

3.3. Data collection

The second theme of the mapping is *data collection*. In group discussions, experts revealed that decisions on data collection, security and storage takes up most of their time. In relation to the other themes, this theme partly overlaps with GDPR requirements on data minimization, data processing and storage rules.

This theme is comprised of four parameters:

1. **Amount of data:** Are 2 datapoints/variables or 1000 collected?"
2. **Type of data:** Is the data based on behavior or demographics?
3. **Frequency of data collection:** Is data collected 24/7 or once a year?
4. **Data predictability:** To what extend would the customer expect the data to be collected?

All four parameters have a decisive impact on the trustworthiness and governance severity of the AI system.

Data collection was widely discussed in the case of using AI to identify customers at risk of developing long-term health conditions. The likelihood of successfully offering proactive help is proportional to the amount and frequency of data collected. However, this can have a negative impact on the customer experience, as large-scale data collection can be perceived as data surveillance.

This led to discussions about how to adjust data collection intervals and ensure clear communication to align expectations between the company and the customer. Interval (monthly/annual) data collection was seen as less intrusive, but the delay in data collection would reduce the ability to respond quickly and limit the ability to offer preventive measures in a timely manner, i.e. in the case of emerging diseases.

Across cases, demographic data such as address, occupation, and age were considered less intrusive than behavioral data such as health, credit card spending, or exercise levels. However, in the case of car insurance, some argued that the use of behavioral data was considered fairer than demographic data because customers can more easily change their driving style than their address, occupation, or age.

Customers' expectations about what data insurers use in their models also figured prominently in the discussions. In general, the use of unexpected data (such as a customer's health data to price home insurance) was considered unfair and would require a clear explanation and alignment of expectations with the customer to avoid a negative impact on fairness.

3.4. Modelling Decisions

The third theme includes parameters on *modelling decisions*. AI systems are comprised of a variety of models including tree-based methods, multivariate regression models, and neural networks.

Each of these models operate differently and afford different use-cases. Most importantly, each modelling choice can involve a trade-off between explainability and accuracy. While some models are more explainable but less accurate (such as linear regression), others are less explainable and more accurate (such as neural networks). Thus, the modelling choice itself has important implications for the procedural and outcome fairness of an AI use case, as well as the appropriate levels of governance and communication.

The expert group discussed modelling decisions in regard to five parameters:

1. **Model purpose:** Can the model function as purely predictive — using only correlations — or are causal explanations needed?
2. **Transparency and explainability:** Is the model truly transparent, can it be explained or is the model a pure black box?
3. **Accuracy:** How much accuracy is needed for the AI system to generate value?
4. **Human oversight:** Is the process fully automatic or to which degree is a human involved?
5. **Granularity:** At what level of granularity is the "final" data used? Is data aggregated for groups or are single individuals identifiable in the data?

A key discussion concerned the use of predictive vs. causal models. Developers need to match models to appropriate use cases. For example, in pricing and underwriting, generalized linear models (GLMs) are appropriate because of their interpretability of why certain customers can be considered high or low risk. A highly predictive model may not fit the use case because it would reduce the transparency needed to explain the model and its specific results to customers, management, and regulators.

Another key metric is accuracy. Besides the explainability-accuracy trade-off, there is also a fairness-accuracy trade-off. Typically, accuracy is positively related to the amount of data collected. However, given the isolated negative effect of more data collection on procedural fairness, higher accuracy has a fairness cost. Moreover, less explainable models face higher risks of undetected bias and discrimination, which further adds to the fairness-accuracy trade-off.

Transparency, which is the ability to explain each parameter of a model and its effect, becomes unattainable for complex models such as neural networks. However, there are certain other ways to determine explainability, which is the ability to explain how the model works, the importance of the input variables, but not their true weight and relationship due to model complexity.

One way to counter low explainability is human oversight, which can be used with complex AI models to improve their explainability. For a fair model, "true" transparency is not always critical if an expert is involved to evaluate the data and qualify a model's prediction. Thus, for a given use case, increased human oversight could dampen the impact of low transparency/explainability.

Expert discussions highlighted a dilemma in the case of stress prevention. This was whether to use a causality-based model - with less accuracy to achieve "true" causality and transparency, or a more complex model with higher accuracy but less explainability. A more explainable causality-based model would be less accurate but would be able to give customers a good explanation of why they were selected. The more accurate model, on the other hand, would not be able to do this, but could potentially reach more relevant customers.

A third option is to combine a highly accurate model with human oversight, so that humans can verify the algorithmic suggestions. Such a third option was preferred by experts in both the prevention and fraud cases. In both cases, it is not considered crucial to fully understand the logic of the model when a case worker is involved to qualify the model's prediction.

However, in contexts where human colleagues directly engage with an accurate but unexplainable AI systems, there is a threat that they accept the machine's answer without using their own professional judgment. It is therefore important to balance the new possibilities to analyze various types of data without compromising the ultimate control of humans.

Across all cases, experts emphasized that simple models should be prioritized in highly sensitive contexts, as this would enhance developers and caseworkers' ability to critically evaluate a model-decision.

The level of granularity was also an important point of discussion. In sensitive contexts like pricing and underwriting, experts emphasized that fair AI systems should process anonymized, pooled, and untraceable input data. They also considered group-based models less intrusive than using data at the individual level.

In summary, the discussions revealed several dilemmas and trade-offs between accuracy, transparency, and human oversight in relation to fairness, highlighting the need for a discussion on the trade-offs between these metrics for AI use cases.

3.5. Outcome fairnesss, Governance & Communication

As a third step, the parameter scoring and trade-off mapping informs the appropriate decisions on *Outcome Fairness, Governance & Communication.*

After developers - through the first two steps of the tool have considered the expected impact on core stakeholders, input data and the model used in the specific use-case – they can train their model and evaluate the *outcome fairness* of their AI system. The consideration of bias primarily becomes relevant in the third step as the former two are focusing on considerations regarding data and modelling. Model evaluation also involves the use of fairness metrics, which developers can use to test the model for potential negative bias and discrimination.

While bias is often perceived as a purely negative phenomenon by lay people, it should not always be considered as negative by default. It is common practice for developers to strike a balance between acceptable bias and negative bias. There are several ways in which this can be achieved, including tweaking the data, adjusting the model or adapting how it is evaluated.

However, decisions on how bias should be corrected demand ethical considerations, as they have implications for e.g. the insurability or treatment of certain groups of individuals. Decisions on outcome fairness thus include ethical considerations on which fairness metric is preferred for a specific use case. Given that it is not possible to select *all* fairness metrics simultaneously, this choice demands an ethical decision on which technical solution represents the best compromise for core stakeholders.

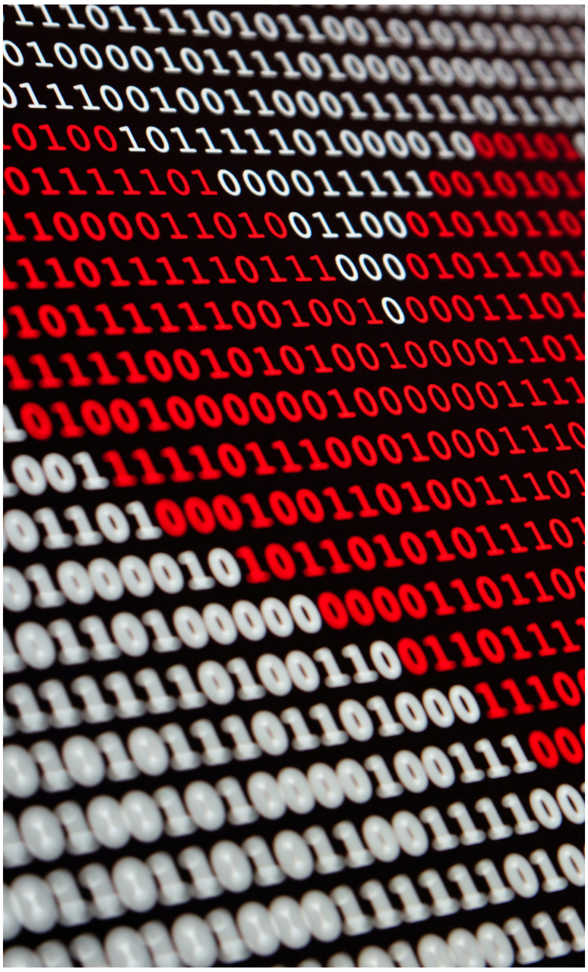
Related to the theme of bias, the experts discussed the following issues. There was consensus that the issue of proxy discrimination becomes more challenging as the complexity of AI models rises. The removal of sensitive variables such as gender, race, or other vulnerable variables might not ensure that other variables carry these signals with them.

One way to identify proxy discrimination is to actively use sensitive personal data to test whether vulnerable groups are adversely affected by the model. However, this requires the collection of sensitive data in the first place, which customers may be reluctant to provide to insurance companies. Specifically, experts discussed collecting racial characteristics of customers to test for racial discrimination of AI models. However, it requires a high degree of trust from customers if they are to provide and consent to the use of their sensitive data – even if it is only for testing.

Outcome fairness combined with all the remaining themes, will provide a comprehensive view on the level of proportionality of AI use. In other words, the greater the number of points allocated to the red areas, the stronger the indication that higher governance requirements are needed for successful use of AI in the specific use case.

Finally, this will also affect the communication to the affected customers. As a ground rule, the parameters that garner the most discussions in a trade-off mapping could indicate areas, that require clear and transparent communication about the model and data use to the affected customer. Here, undertakings might consider the "Frontpage"-test. What are the chances that use of the system would result in a negative media headline?

Being transparent, communicating about data and model design to align expectations is a powerful tool in terms of fairness, and might increase the overall acceptability of AI solutions among customers.



4. Final remarks

The development of ethical AI has achieved significant maturity in the Danish insurance and pensions sector. Undertakings are guided by a variety of resources, including industry-wide ethical principles, recommendations from financial authorities, and AI-specific regulations.

Concurrently, insurers have successfully implemented numerous AI systems across a multitude of use cases. They have acknowledged the necessity for ethical governance that extends well beyond the establishment of overarching guidelines for responsible AI and ethical data use.

Going forward, the governance on ethical AI requires new methods and new ways of thinking when adopting responsible AI on a larger scale. AI Governance needs to move from the horizontal- to the concrete AI use case level.

This report contributes to ongoing work on AI ethics governance by presenting a tool that integrates the ethical trade-off mapping into the existing AI use case impact assessment. Specifically, this report makes three important remarks:

1. The successful adoption of AI systems in insurance demands a combination of ethical reasoning and technical skills
2. The implementation process necessitates a diverse team, including technical and non-technical members, along with appropriate levels of management involvement
3. Technical solutions and ethical considerations should be anchored in appropriate governance frameworks covering issues on modelling, data features and fairness

First, the ethical trade-off mapping has shown that the development of AI is a not a purely technical exercise. The development of algorithms, while technical in nature, always brings forth ethical dilemmas that cannot be addressed neatly through technical solutions. Discussions are needed to highlight dilemmas and define compromise solutions.

In fact, we argue that the successful adoption of AI lies in the recognition that there is no single 'right' solution. Changes in one parameter can always leads to negative effects on other parameters. Decisions on how to build an AI system can only be made after understanding these potential variations and how they affect each other.

Second, AI solutions cannot be isolated from the social context they are embedded in. AI systems affect users, working practices and organizational processes. Thus, it is important to involve a larger, diverse group of professionals with different expertise to find ethically acceptable AI solutions that are deemed legitimate by the relevant stakeholders.

This includes being critical in assessing the purpose and expected value of using AI in the first place in relation to whether it makes sense and what is ethically permissible. Just because AI can provide a solution, it might does not mean, that AI is the best answer.

Third, it has shown that insurance and pension undertakings can benefit from a comprehensive governance framework to guide implementation processes. The tool presented in this report is to help facilitate these discussions, highlight dilemmas, and find acceptable solutions among all the affected stakeholders. In this process, it might also be worthwhile to consider involving customers in the mapping process.

Last, we would like to stress that this overview does not represent an exhaustive list of ethical dilemmas, but merely summarizes key discussions that experts had during the workshop at Copenhagen Business School. We would like to thank the workshop participants for their contributions to discussions and input to create this first version of the mapping. Any input or potential improvements of the mapping are very welcome.

5. Annex

1.1. Regulatory initiatives

The Insurance- and pension sector is subject to a large variety regulatory initiative which limit adoption of AI in certain instances. In the European Union, several regulations target AI-specific issues including bias, discrimination and fairness, transparency and data governance, as well as obligations for human oversight. These include cross-sectoral legislation like the Gender Directive and the Racial Equality Directive that prohibit various kinds of discriminations, and the GDPR that mandates explainability, transparency, and robust data processing. Insurance-specific regulations such as the Insurance Distribution Directive and the Solvency II Directive provide additional requirements such as providing customers with objective information about products, as well as governance and oversight requirements to ensure the solvency of insurance undertakings.

The forthcoming AI Act will provide more specific requirements, including technical documentation and record keeping for high-risk AI applications, as the use of AI for risk assessment and pricing of life- and health insurance is considered high-risk in the act. An impact assessment from the European Commission estimates that only 5-15% of all AI use cases will fall under the high-risk category, meaning that a large share of AI use is considered medium to low risk and therefore not subject to the comprehensive high-risk rules in the act.

Nevertheless, the AI Act encourages all actors to develop and follow codes of conduct based on the act on a voluntary basis. This highlights the continued relevance of AI ethics governance as a crucial strategy towards the digitalization of the sector.

6. References

Council. Commissions Staff Working Document: Impact Assessment Accompanying the Proposal of the AI Act, 2021. https://doi.org/10.5040/9781782258674.

Došilović, Filip Karlo, Mario Brčić, and Nikica Hlupić. "Explainable Artificial Intelligence: A Survey." In 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 0210–15, 2018. https://doi.org/10.23919/MIPRO.2018.8400040.

EIOPA. Artificial Intelligence Governance Principles, towards Ethical and Trustworthy Artificial Intelligence in the European Insurance Sector: A Report from EIOPA´s Consultative Expert Group on Digital Ethics in Insurance. LU: Publications Office, 2021. https://data.europa.eu/doi/10.2854/49874.

European Commission. Proposal for the Artificial Intelligence Act (2021). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206.

Forsikring & Pension. "Cool Eller Creepy: Databrug Og Dataetiske Principper i Forsikrings- Og Pensionsbranchen," 2019.

Jerusalmy, Olivier. "Financial Exclusion: Making the Invisible Visible | Finance Watch," March 2, 2020. https://www.finance-watch.org/publication/financial-exclusion-making-the-invisible-visible/.

Martin, Kirsten, and Ari Waldman. "Are Algorithmic Decisions Legitimate? The Effect of Process and Outcomes on Perceptions of Legitimacy of AI Decisions." Journal of Business Ethics 183, no. 3 (March 1, 2023): 653–70. https://doi.org/10.1007/s10551-021-05032-7.

Noordhoek, Dennis. "Regulation of Artificial Intelligence in Insurance: Balancing Consumer Protection and Innovation." The Geneva Association, 2023.

Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." Science 366, no. 6464 (October 25, 2019): 447–53. https://doi.org/10.1126/science.aax2342.

Tyler, Tom R. "The Relationship of the Outcome and Procedural Fairness: How Does Knowing the Outcome Influence Judgments about the Procedure?" Social Justice Research 9, no. 4 (December 1, 1996): 311–25. https://doi.org/10.1007/BF02196988.

Voldby, Marie. "Nyt krav om redegørelse for dataetik i årsrapporten." Deloitte Denmark, 2020. https://www2.deloitte.com/dk/da/pages/audit/articles/Nyt-krav-om-redegorelse-for-dataetik-i-arsrapporten.html.

Whittlestone, Jess, Rune Nyrup, Anna Alexandrova, and Stephen Cave. "The Role and Limits of Principles in AI Ethics: Towards a Focus on Tensions." In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, 195–200. AIES '19. New York, NY, USA: Association for Computing Machinery, 2019. https://doi.org/10.1145/3306618.3314289.

Ethical Trade-off Mapping

Step One & Two: Design and Discuss

Procedural Fairness & Impact

Impact on core stakeholders:

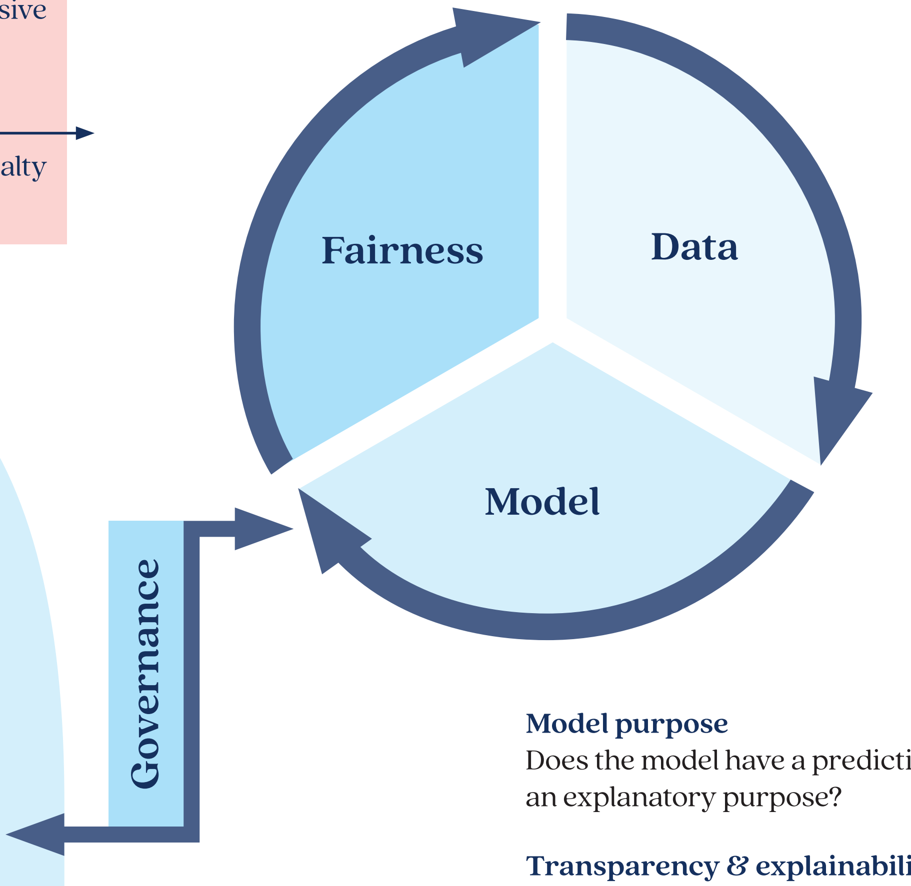
The affected client Financial and social impact	Positive	Negative
The pool of customers Financial and social impact	Positive	Negative
The company Financial and reputational impact	Positive	Negative
Society at large Financial and social impact	Positive	Negative

Client impact:

Sensitivity (in outcome & data) Would data-collection, processing or outcome (impact) be considered as sensitive?	Not sensitive	Sensitive
Experience Level of "intrusiveness". I.e. does the AI system cause feelings of surveillance	Not intrusive	Intrusive
Penalty of "opting out" E.g. reduced of coverage or limited access to resources	No penalty	Penalty

Data Collection

Amount of data Are 2 datapoints/variables or 1.000 collected?	Minimal	Maximal
Type of data Is the data based on behavior or demographics?	Demographic data	Behavioral data
Frequency of data collection Is data collected 24/7 or once per year?	Rarely	Frequently
Data predictability Will/can the client expect this kind of data to be collected?	Expectable data	Unexpectable data



Step Three: Deploy

Outcome fairness, Governance & Communication

Proportionality Level of governance in relation to the expected impact	Governance requirements for data collection and adoption of AI system Low	 High
Outcome fairness Risk of bias and discrimination and the need for deciding on fairness metrics in relation to expected impact	Risk for bias and discrimination Low	 High
Need for communication Transparency and explanations for external stakeholders	Need for communication and alignment of expectations Low	 High

Model Decisions

Model purpose Does the model have a predictive or an explanatory purpose?	Causal explanations		Predictive
Transparency & explainability How explainable is the model?	Transparent	Explainable	Black box
Accuracy How accurate is the model?	High		Low
Human oversight Are decisions made automatically or does it involve a human?	Decision supporting	Review /Verification	Fully automatized
Granularity What is the final level of granularity?	Aggregated in groups		Individual granularity

F&P INSURANCE AND PENSION DENMARK

Jakob Holm
Chefkonsulent,
digitaliseringspolitik
Tlf. 41 91 91 76
jho@fogp.dk

ADD algorithms data & democracy

Alexander Gamerdinger
PhD Fellow, Department
of Organization, CBS
Tlf. 50 32 58 15
Aga.ioa@cbs.dk

F&P — Insurance & Pension Denmark — is the Danish trade association for insurance companies and pension funds. We represent the interests of the industry and ensures that the industry in Denmark is recognized for its important role in helping solve some the largest challenges facing the Danish and international society. This concerns welfare and security for the individual Dane, as well as sustainability and the necessary green transition.