ACTUARIAL SOCIETY
SOUTH AFRICA

From Bias to Black Boxes: Understanding and managing the risks of AI - an actuarial perspective

29 August 2024

Agenda

1. Introduction
   - 1.1. Opening remarks
   - 1.2. Example of AI use in Industry
   - 1.3. AI landscape: Themes in the News
   - 1.4. EU AI Regulation

2. Key Findings from Research
   - 2.1. AI Risk Management and Regulation Themes
   - 2.2. Example: Proxy Discrimination
   - 2.3. Example: Generative AI
   - 2.4. Core Principles of AI Governance Frameworks

3. Conclusion
   - 3.1. What could the Journey forward look like?
   - 3.2. Potential Pathway to AI Readiness

1. Introduction

1.1. Opening Remarks

Actuaries and actuarial teams are seeing the value in new AI/ML techniques and are starting to utilise these techniques in the workplace.

However, actuaries tend to work in a highly regulated environment, but they may be struggling to find **relevant regulation** and **guidance** that is **detailed** enough and keep up with all the technological advancements and **new techniques**.

- Additionally, as society evolves, the definition of fair may change, and this could necessitate **dynamic guidelines.**

IFoA's charter notes that "the objects [of the IFoA] shall be, in the public interest", with this being an implicit consideration underlying the Actuaries Code; **is there a need for public interest and fairness to be explicit in standards such as the Actuaries' Code?**

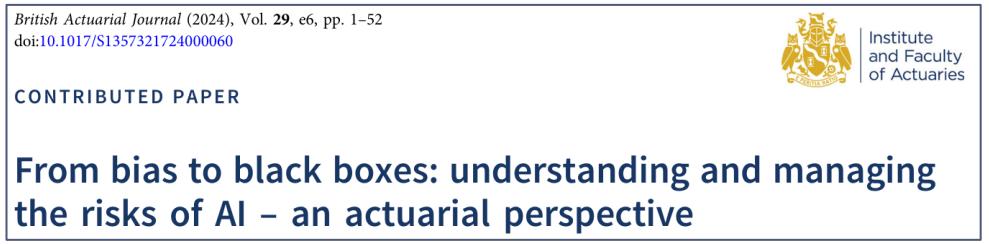
Published regulation (as per our research), does not seem to be detailed enough for the actuary to have clear guidance yet.

We formed a transregional industry group, with the aim of identifying the **challenges** actuaries are facing in **managing AI risk**, and to make **observations** based on collated relevant industry and professional guidelines and regulation.

Our research is ever-evolving, just like technology:

- We are aiming to start a discussion on the topic
- We are discussing our findings with regulators and other professionals
- We would like for industry professionals to comment

Our research report, From Bias to Black Boxes: Understanding and managing the risks of AI – an actuarial perspective, is available online on the *British Actuarial Journal's* website (doi:10.1017/S1357321724000060)



actuartech

1.2. Example of AI use in Industry: FRC Report

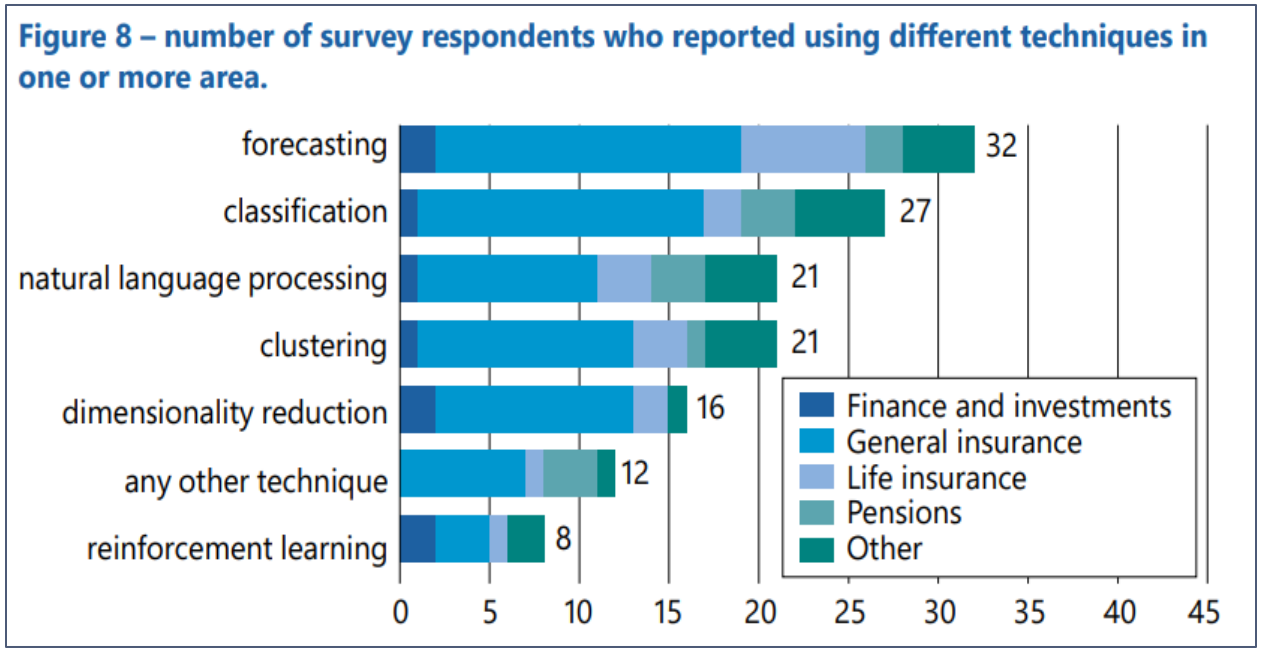
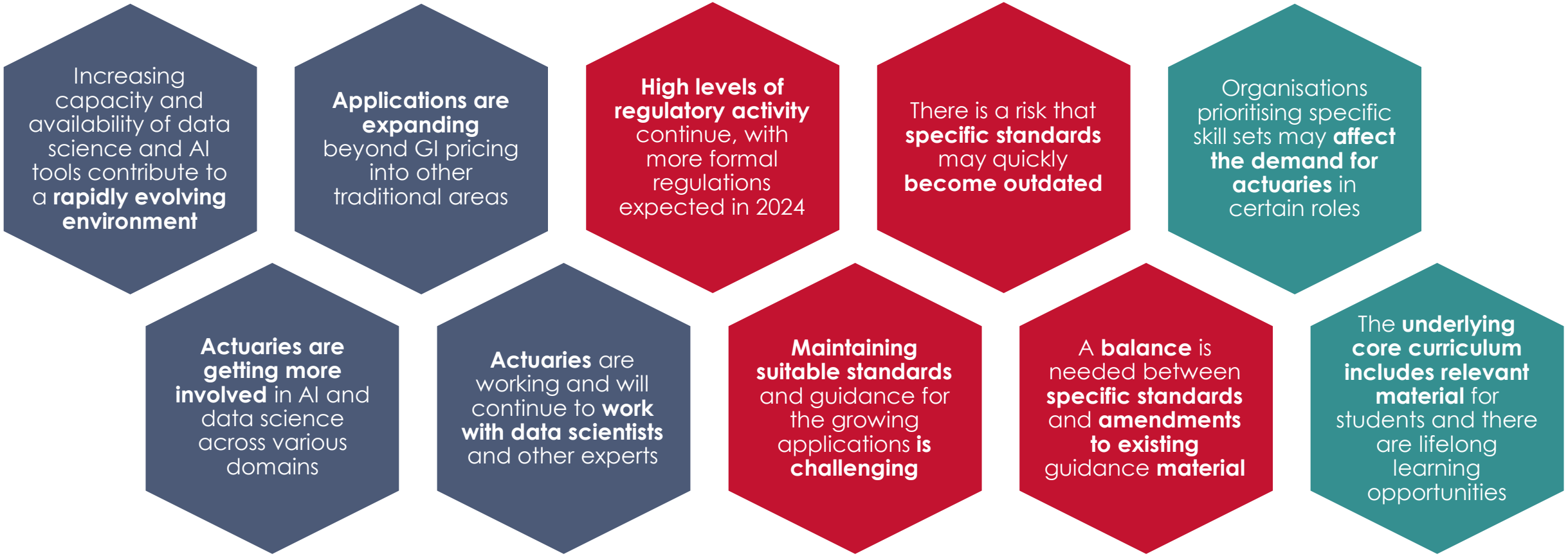


Figure 1: FRC reported usage of AI/ML techniques in actuarial work

Techniques being utilised

- General Insurance and wider fields employ AI/ML the most extensively, with Finance and Investments and Life Insurance expecting a significant increase in the use of AI/ML.
- Various AI/ML techniques are being utilised across the board for different use cases
- Additionally, most respondents in the FRC's 2023 benchmarking mentioned already using LLMs.
  - Most commonly reported use was **helping programmers write code** or assisting novice programmers with coding
  - Other common use included asking ChatGPT to **describe future trends** and to produce **summaries of written code** to help **explain** the code's workings to **non-technical audiences**.
  - Others are trialing the use of a third party LLM to **process transcripts of customer complains**
- Some are developing their own LLMs by fine-tuning third-party models to enable the models to respond to internal policies and information
- The use of LLMs seems to have led to productivity gains in various use cases

1.2. Example of AI use in Industry: IFoA Thematic Review



1.3. AI landscape: Themes in the news

Microsoft lays off team responsible for ethical AI development

Regulation 'critical' to curb risk posed by AI, boss of ChatGPT tells Congress

South Korea, UK to co-host second global AI summit as boom fans risks

Lawsuit claims UnitedHealth AI wrongfully denies elderly extended care

Hope or horror? The great AI debate dividing its pioneers

Humanity at risk from AI 'race to the bottom', says tech expert

In A.I. Race, Microsoft and Google Choose Speed Over Caution

Car insurance quotes higher in ethnically diverse areas

'It's not clear we can control it': what they said at the Bletchley Park AI summit

UK officials use AI to decide on issues from benefits to marriage licences

UK risks scandal over 'bias' in AI tools in use across public sector

The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work

1.4. EU AI Regulation: The EU AI Act (AIA) Risk-based Approach

The AIA takes a **risk-based approach**, setting high standards for model governance.

AIA explicitly mentions **all statistical, machine learning, and logic- and knowledge-based approaches** in their definition of AI: It is not limited to ML models, but **may also apply** to all models used by insurers, including **simple regression models**.

Figure 2: Summary of the EU AI Act Risk Tiering (Actuartech based on EY 2023)

2. Key Findings from Research

2.1. AI Risk Management and Regulation Themes

Ethical Risks	Bias, Discrimination, and Fairness	Individualisation of Risk Assessment	Public Interest, incl. Privacy
Professional Challenges	Explainability and Interpretability	Transparency, Replicability, and Reproducibility	Validation and Governance, incl. Accountability
Wider Themes	Lack of Skills	Organisational Strategy and Sustainability, incl. Prudential Risks	Wider Risks related to AGI and LLMs and threat to the profession

FRC research highlighted the following **key risks**:

Bias & discrimination, **data concerns and an overreliance on data**, and **insufficient human-oversight**.

2.2. Example: Proxy Discrimination

Increased social complexity re protected features e.g. gender calls for more consideration around data collection and the potential for proxy discrimination.

As models grow in complexity, they can capture proxies for missing features. This presents the issue of unforeseen discrimination presented in model results. However, if the proxies are removed, the model may lose a lot of its predictive power, creating model risk.

By balancing accuracy and outcomes for protected classes, models may achieve fairer outcomes. This however requires access to the protected class, which presents further data governance considerations.

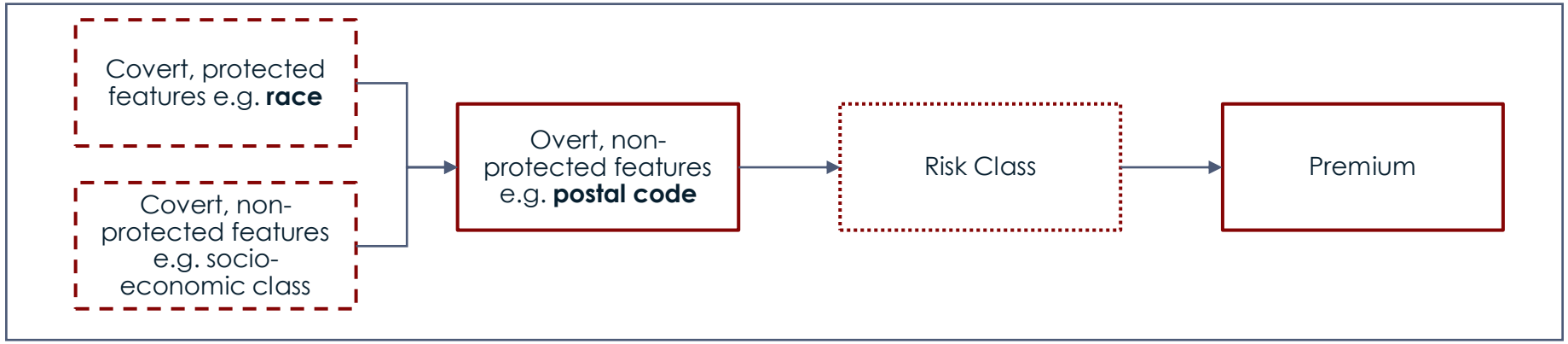


Figure 3: An example showing how overt features could comprise covert features when determining a premium, where some covert features could belong to a protected class.

How do we define fairness and acceptable levels of bias?

2.3. Example: Generative AI:

Example applications of LLMs within an actuarial context

LLMs are a class of AI models trained on large amounts of text data to 'learn' a topic and its language structure to mimic human text. They are a class of generative models that can produce entirely new output based on input.

Examples applications of LLMs within an actuarial context include:

- Reporting
- Summarising documents for:
  - Claims management, including adjusting and compliance
  - Technical reports e.g. for underwriting purposes
  - Identifying emerging risks (e.g. analysing scraped news results)
- Assistance with:
  - Coding
  - Data cleaning and preparation
  - Model development, interpretability, and explainability

Identifying direct applications: Technical assessment

- Data type
- Nature of task (e.g. complex pattern recognition or generating new content)
- Benefit (accuracy, efficiency) vs cost analysis (implementation, maintenance)
- Resource requirement and availability

Identifying direct applications: Risk assessment

- Inherent bias
- Ethical considerations
- Interpretability requirement
- Data leakage or unintentional memorisation
- Variance in results
- Frequency and severity of failures or model errors

Figure 4: Key considerations when assessing the suitability of an LLM

2.4. Core Principles of AI Governance Frameworks

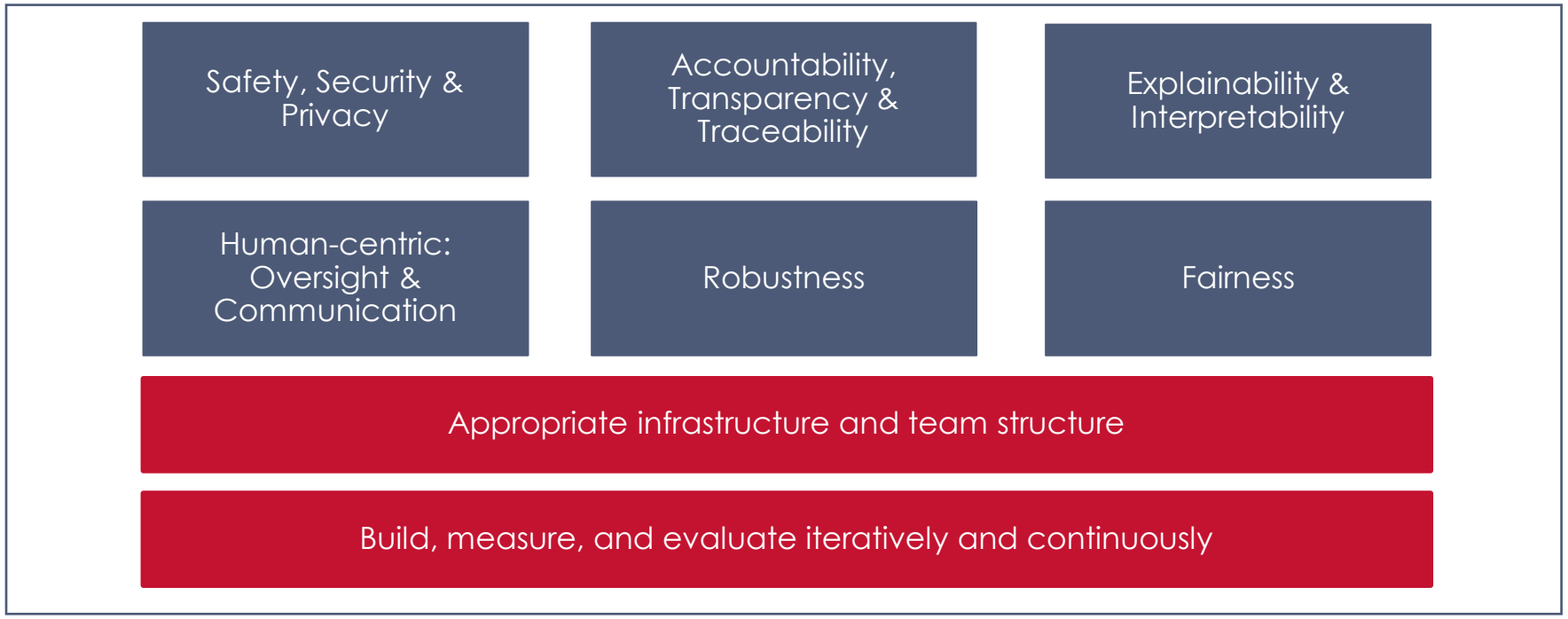


Figure 5: Core principles of AI governance frameworks.
For detailed frameworks, refer to Singapore's Model AI Governance Framework, the Australian Actuaries Institute's Practical Framework for assessing AI models and managing risks, and the US National Institute of Standards and technology's AI Risk Management Framework.

Consider how ML governance frameworks overlap with risk & actuarial frameworks, and how to integrate them

3. Conclusion

3.1. What could the Journey forward look like?

- The industry may need to reach consensus regarding **best and acceptable practices** e.g. acceptable levels of bias, the extent to which other experts could be relied on, etc., through initiatives such as **repositories and frameworks** from professional and industry bodies.
- Professional bodies may need to offer more **detailed guidance** and best practice recommendations.
- Regulation may need to be considered in a way that avoid taking divergent paths and guidance notes should offer **clear interpretations** and not lead to ambiguity.
- Regulation could clarify different levels of risk for AI, provide clear guidelines on **classifying models into risk categories** (not only high risk), and the regulatory **requirements per category**.
- **Upskilling and training** could be offered as part of the qualifying process and incorporated into CPD.
- The **ethical considerations** regarding AI and ML could be part of **requisite examinations**, with additional (optional) specialisations offered in e.g. Data Science for Insurance.
- The **individual may need to** become **more invested** in the new AI/ML landscape and ensure that they incorporate **techniques and guidelines from multiple expert sources** into their everyday work. Actuaries may need to:
  - **emphasise transparency** in processes and models;
  - **include explainability** and **interpretability** as properties of our solution, particularly for black boxes; and
  - **ensure bias and fairness** is considered and **adequately addressed** within our solution.
- In addition, these principles can be used to inform and **enhance our governance and validation frameworks** in the context of AI/ML.

3.2. An Approach to AI Governance & Risk Management

Culture, training, and awareness

Employee training and leadership programmes

Policies and procedures

Description of how AI and AI governance is embedded

Operations and sustenance

Health checks, framework implementation, and due diligence

Strategy and framework

Strategy design and roadmap

Demonstrating accountability

Governance forum, ethical-by-design approach, and impact assessments

Figure 6: AI governance considerations

How are you ensuring effective collaboration between different functions where they may be a slight imbalance between conduct requirements from a professional perspective?

Closing

For further questions, comments, or follow-up discussions, please contact us on
info@actuartech.com

Sources

Sources

- Actuaries Institute. 2022. Artificial intelligence and anti-discrimination: Major new publication. Available at: https://www.actuaries.asn.au/Library/MediaRelease/2022/FINALMRAHRC.pdf
- Actuaries Institute and The Australian Human Rights Commission. 2022. *Artificial intelligence and discrimination in insurance in pricing and underwriting*.
- Association for British Insurers (ABI). 2024. AI Guide: Practical ideas for getting started with responsible AI. Available at: https://www.abi.org.uk/globalassets/files/subject/public/ai/ai-guide.pdf
- Bank of England. 2022. DP5/22 - Artificial Intelligence and Machine Learning. Available at: https://www.bankofengland.co.uk/prudential-regulation/publication/2022/october/artificial-intelligence
- Balona, C. 2023. ActuaryGPT: Applications of Large Language Models to Insurance and Actuarial Work. Available at SSRN: https://ssrn.com/abstract=4543652
- Bird, S., Dudik, M., Horn, B., Lutz, R., Milan, V., Sameki, M., Wallach, H., and Walker, K. 2020. Fairlearn: A toolkit for assessing and improving fairness in AI. Microsoft. Available at: https://www.microsoft.com/en-us/research/publication/fairlearn-a-toolkit-for-assessing-and-improving-fairness-in-ai/
- Bodanis, C. 2023. A responsible approach to using AI in corporate reporting: guidance for boards and management on approach and disclosure. Falcon Windsor. Available at: https://www.falconwindsor.com/s/FW_Guidance_Responsible_Use_AI_in_Reporting_Nov23.pdf
- Cheung, D., Kang, M., Goldfarb, A. 2022. Report: The Use of Predictive Analytics in the Canadian Property and Casualty Insurance Industry. Available at: https://www.cia-ica.ca/docs/default-source/research/2022/rp222067e.pdf
- Casualty Actuarial Society (CAS) Race and Insurance Research Task Force. 2021. Understanding Potential Influences of Racial Bias on P&C Insurance: Four Rating Factors Explored. Available at: https://www.casact.org/sites/default/files/2022-03/Research-Paper_Understanding_Potential_Influences.pdf?utm_source=Landing+Page&utm_medium=Website&utm_campaign=RIP+Series
- Centre for Data Ethics and Innovation. 2021. The roadmap to an effective AI assurance ecosystem. Available at: https://www.gov.uk/government/publications/the-roadmap-to-an-effective-ai-assurance-ecosystem
- Department for Science, Innovation and Technology. 2023. A Pro-innovation approach to AI Regulation. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1146542/a_pro-innovation_approach_to_AI_regulation.pdf
- Department for Science, Innovation and Technology. 2023. Initial £100 million for expert taskforce to help UK build and adopt next generation of safe AI. Available at: https://www.gov.uk/government/news/initial-100-million-for-expert-taskforce-to-help-uk-build-and-adopt-next-generation-of-safe-ai
- Digital Regulation Cooperation Forum. 2022. Auditing algorithms: the existing landscape, role of regulators and future outlook. Available at: https://www.gov.uk/government/publications/findings-from-the-drcf-algorithmic-processing-workstream-spring-2022/auditing-algorithms-the-existing-landscape-role-of-regulators-and-future-outlook
- Department for Science, Innovation and Technology, and the Centre for Data Ethics and Innovation. 2023. Report: Enabling responsible access to demographic data to make AI systems fairer. Available at: https://www.gov.uk/government/publications/enabling-responsible-access-to-demographic-data-to-make-ai-systems-fairer/report-enabling-responsible-access-to-demographic-data-to-make-ai-systems-fairer
- European Council. 2021. Proposal for a Regulation of the European Parliament and of the Council: Laying Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending certain Union Legislative Acts. COM(2021) 206 final. Available at: https://artificialintelligenceact.eu/the-act/
- European Insurance and Occupational Pensions Authority (EIOPA). 2021. Artificial Intelligence Governance Principles: Towards Ethical and Trustworthy Artificial Intelligence European Insurance Industry. A report from EIOPA's Consultative Expert Group on Digital Ethics in insurance. Available at: https://www.eiopa.europa.eu/system/files/2021-06/eiopa-ai-governance-principles-june-2021.pdf

Sources

- Even, Y. 2022. A journey into Responsible AI: Veritas Fairness Assessment Methodology & Toolkit for the Financial Industry. Available at: https://www.swissre.com/risk-knowledge/advancing-societal-benefits-digitalisation/a-journey-into-responsible-ai.html
- *Equality Act 2010*. Available at: https://www.legislation.gov.uk/ukpga/2010/15/contents
- Financial Conduct Authority (FCA). 2022. FG22/5 Final non-Handbook Guidance for firms on the Consumer Duty. Available at: https://www.fca.org.uk/publication/finalised-guidance/fg22-5.pdf
- Financial Reporting Council (FRC). 2023. The use of Artificial Intelligence and Machine Learning in UK actuarial work. Available at: https://media.frc.org.uk/documents/Research_on_the_use_of_Artificial_Intelligence_and_Machine_Learning_in_UK_actuarial_work_AK5H1We.pdf
- Institute and Faculty of Actuaries (IFoA). 2010. Charter of the Institute and Faculty of Actuaries. Available at: https://www.actuaries.org.uk/system/files/documents/pdf/new-charter.pdf
- Institute and Faculty of Actuaries (IFoA) and Royal Statistical Society (RSS). 2019. A Guide for Ethical Data Science A collaboration between the Royal Statistical Society (RSS) and the Institute and Faculty of Actuaries (IFoA). Available at: https://www.actuaries.org.uk/system/files/field/document/An%20Ethical%20Charter%20for%20Date%20Science%20WEB%20FINAL.PDF
- *Insurance Act 2015*. Available at: https://www.legislation.gov.uk/ukpga/2015/4/contents/enacted
- International Actuarial Association. 2024. *Task Forces.* Available at: https://www.actuaries.org/iaa/IAA/Committees/Task_Forces/IAA/Committees/Task%20Forces.aspx?hkey=4fcf76f5-ed39-4dec-8564-aa12bebc79f2
- Levene, M. 2023. How ISO/IEC 42001 guides Organisations toward Trustworthy AI. *AI Standards Hub*. Available at: https://aistandardshub.org/iso-iec-42001-trustworthy-ai
- Lindholm, M., Richman, R., Tsanakas, A., and Wüthrich, M.V. 2023. What is Fair? Proxy Discrimination vs. Demographic Disparities in Insurance Pricing. Available at: https://ssrn.com/abstract=4436409
- Marshall, A. 2024. For the IFoA: Thematic review Report: Actuaries using data science and artificial intelligence techniques. Available at: https://actuaries.org.uk/media/g4ymoxep/actuaries-using-data-science-and-artificial-intelligence-techniques.pdf
- Murphy, M., Tully, T. Xiao, D. 2024. The Modern AI Stack: Design Principles for the Future of Enterprise AI Architectures. *Menlo Ventures.* Available at: https://menlovc.com/perspective/the-modern-ai-stack-design-principles-for-the-future-of-enterprise-ai-architectures/
- Peters, S. and Van Den Brink, G. 2023. Can AI Go to Jail? – Consequences for the Corporate Governance. *EAA e-Conference on Data Science & Data Ethics*. Available at: https://actuarial-academy.com/Documents/Abstracts/E0338_Abstract_e-Conference_Recording_Peters+vandenBrink.pdf
- Prince, A.E.R. and Schwarcz, D. 2020. Proxy Discrimination in the Age of Artificial Intelligence and Big Data. *Iowa Law Review*, 105(3): 1257-1318.
- Prudential Regulation Authority. 2023. SS1/23 – Model risk management principles for banks. Available at: https://www.bankofengland.co.uk/prudential-regulation/publication/2023/may/model-risk-management-principles-for-banks-ss
- Smith, L.T., Pirchalski, E., and Golbin I. 2022. Avoiding Unfair Bias in Insurance Applications of AI Models. Society of Actuaries. Society of Actuaries, August 2022. Available at: https://www.soa.org/resources/research-reports/2022/avoid-unfair-bias-ai/
- The European Parliament and the Council of the European Union. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union*, L 119/1. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679
- World Economic Forum. 2020. Empowering AI Leadership: AI C-Suite Toolkit. Availability at: https://www.weforum.org/reports/empowering-ai-leadership-ai-c-suite-toolkit/
- Usher, K. on behalf of the IFoA AI Ethics Risk Working Party. 2023. Man vs Machine. *The Actuary* (May 2023). Available at: https://www.theactuary.com/2023/05/04/man-vs-machine

Sources
Headlines

- Ahmed, M. 2024. Car insurance quotes higher in ethnically diverse areas. *BBC*. Available at: https://www.bbc.com/news/business-68349396
- Grant, N. and Weise, K. 2023. In A.I. Race, Microsoft and Google choose Speed over Caution. *New York Times*. Available at: https://www.nytimes.com/2023/04/07/technology/ai-chatbots-google-microsoft.html
- Grynbaum, M.M. and Mac, R. 2023. The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work. *New York Times*. Available at: https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html
- Jamieson, A. 2023. Regulation 'critical' to curb risk posed by AI, boss of ChatGPT tells Congress. *The Independent*. Available at: https://www.independent.co.uk/independentpremium/business/ai-regulation-sam-altman-congress-openai-chatgpt-b2340077.html
- Lee, J. 2024. South Korea, UK to host AI summit in Seoul as risks mount. *Reuters*. Available at: https://www.reuters.com/technology/south-korea-uk-co-host-second-global-ai-summit-boom-fans-risks-2024-05-20/
- Milmo, D. 2023. Hope or horror? The great AI debate dividing its pioneers. *The Guardian*. Available at: https://www.theguardian.com/technology/2023/oct/24/ai-risk-climate-crisis-google-deepmind-chief-demis-hassabis-regulation
- Milmo, D. and Helmore, E. 2023. Humanity at risk from AI 'race to the bottom', says tech expert. *The Guardian*. Available at: https://www.theguardian.com/technology/2023/oct/26/ai-artificial-intelligence-investment-boom
- Milmo, D. and Stacey, K. 2023. 'It's not clear we can control it': what they said at the Bletchley Park AI summit. *The Guardian*. Available at: https://www.theguardian.com/technology/2023/nov/01/elon-musk-calls-ai-one-of-the-biggest-threats-to-humanity-at-summit
- Pierson, B. 2023. Lawsuit claims UnitedHealth AI wrongfully denies elderly extended care. Available at: https://www.reuters.com/legal/lawsuit-claims-unitedhealth-ai-wrongfully-denies-elderly-extended-care-2023-11-14/
- Sankaran, V. 2023. Microsoft lays off team responsible for ethical AI development. *The Independent*. Available at: https://www.independent.co.uk/tech/microsoft-layoff-ai-ethics-team-b2300237.html
- Stacey, K. 2023. UK officials use AI to decide on issues from benefits to marriage licences. *The Guardian*. Available at: https://www.theguardian.com/technology/2023/oct/23/uk-officials-use-ai-to-decide-on-issues-from-benefits-to-marriage-licences
- Stacey, K. 2023. UK risks scandal over 'bias' in AI tools in use across public sector. *The Guardian*. Available at: https://www.theguardian.com/technology/2023/oct/23/uk-risks-scandal-over-bias-in-ai-tools-in-use-across-public-sector

Appendix

AI Regulatory and Professional Journey in the UK as an Example

Government

- **UK: Pro-innovation approach** with principles and guidelines, including AI assurance guides and Foundation Model Taskforce
- Additional regulation around specific items e.g. **data protection (GDPR), Equality Act**
- Other legislation being developed and released in **other countries** e.g. **EU AI Act**
- **Development of global AI Standards e.g. through DRCF,** ISO/IEC DIS 42001, and Global Summit on AI Safety
- **Additional guidelines** from entities spanning across regions and industries e.g. from OECD and WEF

Industry

- Regulators **per sector/industry** are required to produce **relevant regulation** within the government's high-level AI framework
  - Initial benefits, risks, and regulation presented in DP5/22 (PRA/FRC); Model risk management for banks in PS6/23 (PRA)
  - The FCA and the Bank of England released updates on their approach to regulating AI in April 2024
- Whilst not AI-specific, **existing guidance and legislation should be applied to AI** e.g. FCA's Consumer Duty, Insurance Act, etc.
- Guidelines released by **EIOPA** and **ABI** supporting the Industry's use of AI; regulation from **other countries** e.g. from OSFI, CBI
- FRC's 2023 Research Report on the use of AI/ML for UK actuarial work

Profession

- Actuaries follow **specific guidelines and standards from professional bodies** e.g. the Actuaries' Code from the IFoA. **Additional, non-mandatory governance** includes:
  - IFoA and RSS's 2019 Guide for Ethical Data Science & Ethical and Professional Guidance on Data Science: A Guide for Members
  - IFoA AI Ethics Risk Working Party's 2023 risk taxonomy (in *The Actuary*)
  - IFoA's 2024 Thematic Review on the use of data science and artificial intelligence techniques by actuaries
  - IAA AI Task Force established 2024 covering education, professionalism, innovation, the future actuary, and governance
- **The concept of AI is fast developing: is there a need for dynamic guidance?**

There are various societal and ethical elements the actuary needs to pay more attention to due to AI, but regulation is not purely aligned on the various risks and how to address them (e.g. discrimination).
Is regulation and guidance sufficient and easily accessible for actuaries?

Example: Proxy Discrimination

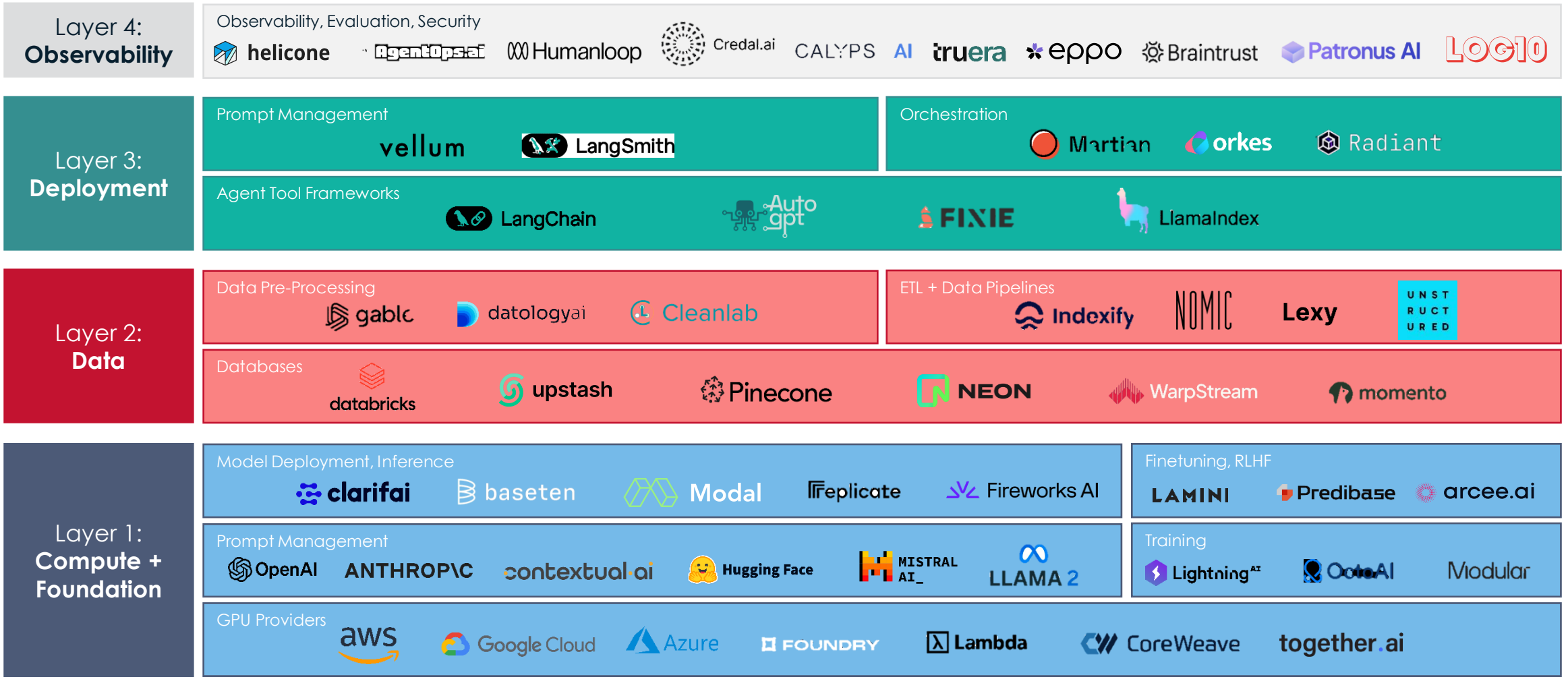
Other considerations

- How do we establish and implement appropriate frameworks* to address fairness?
- How to balance fairness and accuracy with increasingly advanced and complex techniques?
- How does one define an acceptable level of bias?
- Is it not problematic to use simpler, more well-known techniques that may require additional margins and increased prices due to low accuracy?
- The concept of AI is fast developing, perhaps there is a need for the profession to provide more specific guidance?

* See the FEAT Fairness Principles Assessment Methodology developed by MAS, SwissRe, and Accenture (2022)

Example: Generative AI

Building blocks of generative AI: The art of the possible



See "The Modern AI Stack: Design Principles for the Future of Enterprise AI Architectures" published by Menlo Ventures for a more detailed discussion.

Initial Considerations of AI Governance

Big Data and increased computing power has led to more awareness of analytical use-cases which can be employed through AI and ML. This has raised several areas of consideration for actuaries:

- Use of external databases
- Working with other professions not used to the same level of professional standards
- Greater data protection considerations
- Increased awareness of potentially **misleading results and hallucinations**

AI systems are not independently thinking i.e. they require developers to train it, therefore validation and governance metrics can be built into the system and the AI can be trained to be **ethical and self-governed**.*

However, **accountability** for AI systems should be established first.

FRC findings on governance:

- In many cases, organisations have not changed governance process as they were effective, but there were differences in practice, especially with a shift in focus toward explainability
- Actuarial practitioners with the skills to review AI/ML are in short supply
- Reproducibility is not viewed as a high priority
- Explainability is key in the choice of modelling techniques

* See "Can AI go to Jail? – Consequences for the Corporate Governance" by Peters and Van den Brink (2023).