



# Safeguarding responsible AI and the global initiative of our profession

The actuarial profession is undergoing radical change. We have reached a point where actuarial science, data science, and computer science are beginning to overlap in professional capacities, constantly adapting and being applied to fields previously unimagined. Artificial Intelligence (AI) is no longer an emerging topic; it is now widely discussed and utilized across various industries and professions. However, the responsible use, distribution, and implementation of AI systems have yet to be fully adopted and embraced by the industries we are part of. This article will explore concepts of concern, present the current regulatory landscape for AI in Europe, and share updates on how the global profession is responding, supported by the International Actuarial Association's (IAA) latest initiative.

The responsible use of AI presents a multitude of challenging facets. **Bias** in AI can manifest in forms related to gender or ethnicity. An AI system<sup>1</sup> is trained on large datasets shaped by human interactions, language, and societal feedback. Consequently, an algorithm can unknowingly inherit and perpetuate these biases found in its training materials, as seen in Large Language Models (LLMs). These models do not truly 'understand' the underlying issues but rely on predictive methodologies shaped by their design, inputs, and outputs. Rights and opportunities for women are known to be affected by the way algorithms work, especially in the area of recruitment<sup>2</sup>. In the insurance industry, pricing and risk assessment processes often consider protected characteristics of customers. The consequence of having bias is **direct** and **indirect discrimination**. The former is clear, as it involves unfavorable treatment towards individuals, by misusing one's protected characteristics as a factor. Indirect discrimination might be a bit more subtle. In insurance, proxy variables that relate to non-protected characteristics might be used for pricing. These variables can indirectly reflect protected characteristics, potentially disadvantaging policyholders. Furthermore, sophisticated predictive modeling using AI is subject to feedback loops and dynamic data collection, which can alter its behavior over time. So, if models can influence the existent data, what is once seen as a discrimination free algorithm, can easily change to a discriminatory<sup>3</sup> one. Bias and discrimination are directly linked to the concept of **fairness**. Actuaries<sup>4</sup> should develop and adapt algorithms that are both explainable and transparent to ensure fairness, while avoiding practices that could influence consumer behavior. Additionally, incorporating sensitive information into fairness-unaware algorithms, when available, can effectively identify the features necessary for accurate yet fair modeling frameworks. An extensive explanation and modeled examples are provided by Valerie Du Preez et al<sup>5</sup>.

## ACTUARIES SHOULD DEVELOP AND ADAPT ALGORITHMS THAT ARE BOTH EXPLAINABLE AND TRANSPARENT TO ENSURE FAIRNESS

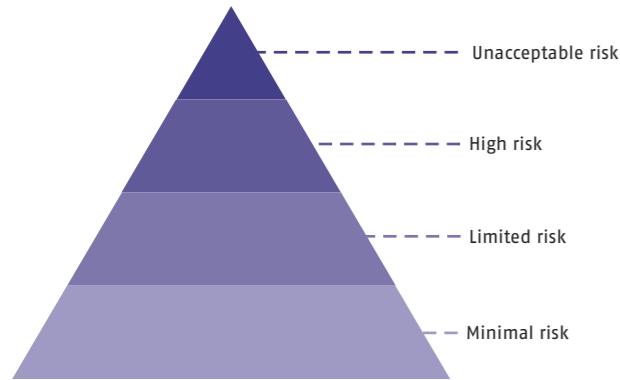
Globally, these concepts are widely discussed at both local and international policy-making levels. Notable efforts include the European Commission's<sup>6</sup> development of the Ethics Guidelines for Trustworthy AI, and the Monetary Authority of Singapore's earlier principles aimed at promoting Fairness, Ethics, Accountability, and Transparency<sup>7</sup>. The Organisation for Economic Co-operation and Development (OECD) has recently published<sup>8</sup> recommendations to further clarify definitions of AI systems, addressing responsible stewardship and international cooperation for trustworthy AI. The following section will delve into more detail on the upcoming European Union legislation, the AI Act.

B.M. Tautan MSc AAG is Reinsurance Actuary and Capital Analyst at Achmea Reinsurance. He is part of the Actuarial Data Science and Kwaliteitszorg / Professionalism commissions at the Royal Dutch Actuarial Association, is co-Vice-Chairing the Actuarial Association of Europe's AI and DS Working Group and is Chair of the AI Governance workstream, part of the International Actuarial Association's AI Task Force initiative.



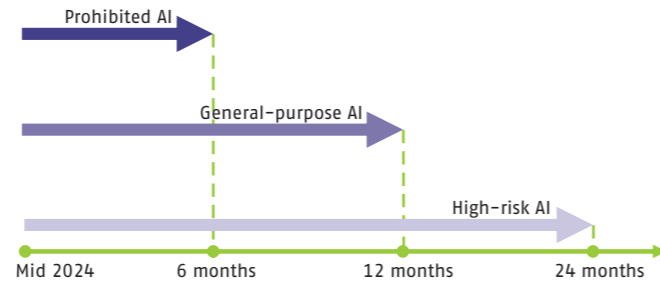
## THE EUROPEAN PERSPECTIVE

One of the most impactful developments for the actuarial profession<sup>9</sup> is the AI Act, a piece of horizontal legislation. The initial draft was published in April 2021, and after undergoing numerous amendments and revisions, it was finally agreed upon during the European Parliament's last legislative voting on March 13, 2024. The act governs AI use, ensures human oversight, and is notable for being the first legislation crafted with consumer rights in mind. It introduces a risk-based approach to the market, including an impact assessment on fundamental rights. Below are the main risk categories highlighted by the regulation:



The risk-based framework is designed to foster innovation while supporting research and scientific activities that facilitate the introduction of AI systems into the European market. Starting with the lower risk categories<sup>10</sup>, **minimal risk** allows for the free use of systems, including examples such as spam filters and video games. **Limited risk** relates to systems that may lack transparency, and it introduces obligations to ensure humans are informed about interactions with machines. AI-generated text must be labeled as such when public interest is concerned. The **high-risk** category, perhaps the most crucial and intriguing, is more complex. Systems classified under this category pose significant risks to the safety and fundamental rights of individuals. Some examples of applicable areas include education, recruitment, public legal systems, employment, and creditworthiness assessments. Pricing and risk assessment activities involving natural persons will be classified as high risk by default for all life and health insurers. Regarding requirements, risk management systems must be established, including technical documentation and regular monitoring of AI systems. Before a high-risk system is released to the market, it must undergo a conformity assessment and be registered in a central EU database. Systems that fall into the unacceptable or **prohibited risk** category, which includes those that involve biometric categorization, threaten human safety and rights, or engage in social scoring or behavior manipulation, will be banned. In addition, a fifth category supports the existing risk classifications and is designated for **general-purpose AI models**. Those involving parallel tasks, such as foundation models (e.g., ChatGPT, Gemini), most large language models (LLMs), and code generators, must undergo thorough examination and be well-documented and transparent before being classified into one of the other categories. Systems that exceed a threshold of  $10^{25}$  FLOPS (Floating Point Operations Per Second—a measure of computer performance) are considered to pose a 'systemic risk' under the act and will be subject to additional obligations. To establish a harmonized governance framework at the European level, the Commission is setting

up an AI Office, which will include a board composed of representatives from member states, an advisory forum, and a scientific panel. Different timelines are set for the underlying risk categories of AI systems, with the AI Act expected to come into force in mid-2024, 20 days after its publication in the EU Official Journal. Below is the implementation period of the act, categorized by risk level:



As a multi-sector legislation, significant work remains on how the AI Act interacts with other directives and regulatory frameworks, such as the Product Liability Directive, AI Liability Directive, and Data Governance Act, among others. Nevertheless, it is clear that actuaries must exercise great care in designing and implementing AI models. Practices in underwriting, pricing, risk assessment, and data management must all align with risk management frameworks that promote responsible AI use.

## INTERNATIONAL ACTUARIAL ASSOCIATION'S INITIATIVE ON AI

Governance frameworks are designed to ensure that AI systems are implemented responsibly across societies, focusing on fairness, reducing bias, and addressing discrimination. The integration of AI adds complexity to actuaries' roles, necessitating interdisciplinary collaboration and an expanded skill set. This evolution is refining professional guidance on the subject. To keep pace with this rapidly evolving landscape, the International Actuarial Association (IAA) launched an AI Task Force at the end of 2023. This initiative<sup>11</sup> will promote the responsible use of AI within the profession, contributing to societal well-being. The IAA will facilitate open dialogue globally with relevant stakeholders and supranational institutions, and advance professional competencies through education and knowledge sharing. The IAA's AI Task Force remains committed to ongoing engagement with Full Member Associations. Since the initiative's inception, over 100 actuaries worldwide have volunteered as delegates. The initiative launched successfully with a two-day AI Summit in Singapore in April this year, attended by more than 50 actuaries in person. The AI Task Force is addressing AI through five workstreams<sup>12</sup>: **Professionalism and Ethics, Education, Changing Role of Actuaries, Governance and Innovation**.

### THE IAA WILL FACILITATE OPEN DIALOGUE GLOBALLY WITH RELEVANT STAKEHOLDERS AND SUPRANATIONAL INSTITUTIONS

The decision to divide the work into different workstreams underscores the commitment to comprehensively address aspects of professionalism. This approach involves testing the existing code of conduct, developing case studies on ethics, and fostering professional

advancements within the actuarial field. Efforts include addressing governance issues, setting examples through innovation, and updating the educational syllabus with relevant AI topics, among others.

The actuarial field is continually evolving due to advancements in data analytics, predictive modeling, and reporting practices. AI algorithms, which thrive on data, require actuaries to master advanced data modeling, data structures, alternative data sources, and to deal with synthetic data. Moreover, technical modeling is becoming increasingly complex, as seen in deep learning models like convolutional or recurrent neural networks, which process spatial or sequential data. Complying solely to traditional risk management frameworks could lead to underdeveloped practices and work ethics, resulting in the inexplicable, opaque, and irresponsible use of AI. The AI Task Force aims to address some of these challenges and is scheduled to complete its initial work by the end of 2024, with plans to continue its efforts thereafter.

As actuaries, we bear a significant responsibility to ensure that our practices remain **fit and proper** in the context of AI. By actively participating in international activities, such as the IAA's recent initiative, following closely the changing regulatory landscapes, and engaging with other industries, we can maintain our relevance and ensure a sound adaptation of our profession. These efforts are crucial for upholding our societal duties and driving the responsible evolution of actuarial practices. ■

- 1 – OECD Definition of AI: <https://oecd.ai/en/wonk/ai-system-definition-update>
- 2 – UNESCO, IRCAI (2024). "Challenging systematic prejudices: an Investigation into Gender Bias in Large Language Models"
- 3 – European Union Agency for Fundamental Rights, Bias in algorithms: Artificial intelligence and discrimination: [https://fra.europa.eu/sites/default/files/fra\\_uploads/fra-2022-bias-in-algorithms\\_en.pdf](https://fra.europa.eu/sites/default/files/fra_uploads/fra-2022-bias-in-algorithms_en.pdf)
- 4 – Actuarial Association of Europe, January 2024, 'What Should and Actuary Know about Artificial Intelligence?'
- 5 – Valerie Du Preez et al., 2024, British Actuarial Journal, From bias to black boxes: understanding and managing the risks of AI – an actuarial perspective
- 6 – Ethics guidelines for trustworthy AI, 2019, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- 7 – Monitoring Authority of Singapore, Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector
- 8 – OECD, May 2024, Recommendation of the Council on OECD Legal Instruments Artificial Intelligence
- 9 – Actuarial Association of Europe, March 2024, [https://actuary.eu/wp-content/uploads/2024/04/AAE-Note\\_The-AI-Act-sets-out-the-way-Artificial-Intelligence-is-to-be-used-in-the-EU-28-03-2024-FINAL.pdf](https://actuary.eu/wp-content/uploads/2024/04/AAE-Note_The-AI-Act-sets-out-the-way-Artificial-Intelligence-is-to-be-used-in-the-EU-28-03-2024-FINAL.pdf)
- 10 – European Commission, AI Act, <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>
- 11 – International Actuarial Association, March 2024, [https://www.actuaries.org/IAA/Documents/CMTE\\_EXEC/IAI%20Task%20Force/Sol\\_on\\_Artificial%20Intelligence\\_Council\\_approved\\_BMarch2024.pdf](https://www.actuaries.org/IAA/Documents/CMTE_EXEC/IAI%20Task%20Force/Sol_on_Artificial%20Intelligence_Council_approved_BMarch2024.pdf)
- 12 – International Actuarial Association, Artificial Intelligence Initiative, <https://www.actuaries.org/IAA/IAI>

# Summerschool 2024

**19 september** 09.00–21.00 uur  
(aansluitend borrel)

**20 september** 08.30–13.00 uur

**Conferentiecentrum Woudschoten**  
(Zeist)

Aangevraagd voor **13 PE-punten**

**Schrijf nu in!**

Schrijf nu in via [www.actuarielinstituut.nl/permanente-educatie/summerschool-2024-1](http://www.actuarielinstituut.nl/permanente-educatie/summerschool-2024-1)

**Enkele sprekers:**  
Frank van Berkum, Joris Krijger, Gerrit Jan van den Brink, Jan-Huug Lobregt, Elly Smits, Lukas Straathof, Inez Bosch en Laura Grob.

**Actuariel Instituut®**