

Original Article

# Ethical Frameworks and Value Alignment for AI in Actuarial Decision-Making

Nihar Malali<sup>1</sup>, Sita rama Praveen Madugula<sup>2</sup>

<sup>1,2</sup>Independent Researcher

Received Date: 05 January 2025

Revised Date: 10 February 2025

Accepted Date: 03 April 2025

**Abstract:** Risk assessment, price of insurance, and financial forecasting have been impacted by the growing integration of Artificial Intelligence (AI) in actuarial decision-making. The accuracy of the models is enhanced using AI, the complex calculations are automated, and real-time decision-making via AI models is possible. Nevertheless, these new advances highlight ethical issues related to bias, lack of transparency, accountability, along data privacy risk. Responsible use of AI requires a solid set of rigorous ethical principles to prevent it from reinforcing systemic biases and producing unjust consequences. This paper discusses aspects of the virtue of applying ethical considerations for AI science in the field of actuarial science, including fairness, explainability, and regulatory compliance. It talks about the role of transparency in the AI model, some of the techniques for bias mitigation, and the importance of accountability in the automated decision-making system. Additionally, value alignment is also considered to prohibit AI systems from acting outside the boundaries of ethical and societal limits, in accordance with actuarial best practice and industry standard. It is important to address these challenges to enable trust in the actuarial processes that use AI. Using ethically governed AI governance models and technical safeguards, actuaries can bring about the best of the possible of AI for financial risk management while preserving fairness, integrity and reliability.

**Keywords:** Artificial Intelligence, Actuarial Science, Ethical AI, Value Alignment, Transparency.

## I. INTRODUCTION

Artificial intelligence (AI) staged a rapid development that has changed the way decisions are made in different industries. In addition, AI-powered systems are being used by us in actuarial science to analyze big data, assess risk, and generate predictive models. By taking them, these applications allow for more accurate, efficient, and scalable decision making, which turns out to be more data-driven[1]. AI has changed the way in which actuarial decision-making is completed in risk assessment and financial forecasting. Actuaries can use AI-powered models, particularly those that leverage machine learning to evaluate vast datasets, improve forecast precision, and automate extremely complex computations that would otherwise be expensive in terms of time and resources.

These technological advances in decision making have aided in areas including insurance pricing, claims management[2], and risk mitigation[3] which makes efficiency better and human error less. However, in actuarial practices, with the advent of AI, there is a serious concern about maintaining ethical responsibility and value alignment.

In the high stakes domain, value alignment in AI is to make sure that AI systems do what humans want it to do, and this includes ethical and values in AI. With AI systems evolving to a more autonomous state, there is more reason to align AI behavior with ethical rules. AI models can generate outcomes that are contrary to societal values or otherwise not intended. Therefore, this requires the development of structured ethical frameworks guiding the AI development in actuarial science to ensure transparency, accountability and fairness of the decision-making.

The presence of AI technologies has led to a wide range of ethical concerns about the development and use of autonomous intelligent systems that are ingrained in society and culture. Autonomous system design is, therefore, not only a multi-technological endeavor but also incorporates social, psychological, economic, political, and legal elements, and it will have significant effects on every facet of society[4]. Both as a topic for scientific research and as a practice, the ethics of AIS remain understudied. Current strategies include human-robot interaction, professional norms of behavior, and the appropriate application of AI. Technology design has not yet been greatly impacted by attempts to include ethics into AI design. Research on ethical design has been difficult thus far from two angles.

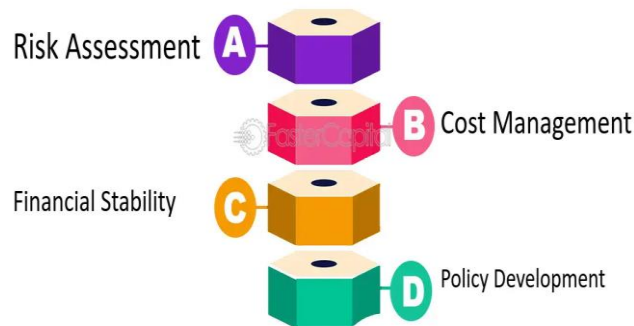
### A. Structure of the Paper

The structure of this paper is as follows: Section II discusses The Role of AI in Actuarial Decision-Making. Section III explores Ethical Frameworks for AI in Actuarial Practices. The last section, IV, looks at Value Alignment in AI for Actuarial Science. Some of the future research directions are included in Section VI, as the literature and case studies are reviewed in Section V.



## II. THE ROLE OF AI IN ACTUARIAL DECISION-MAKING

Actuarial decision-making is to assess and control the financial risks based on the mathematical and statistical models. In traditional actuaries, the data and statistical methods are used to do their work of insurance pricing, underwriting and financial forecasting. The use of AI in actuarial science has completely altered the way conventional risk evaluation and decision-making processes work. In the past, the risk was evaluated and forecasts of the financial outcomes were made by using statistical models, historical data, and experts' judgment. AI and ML algorithms, along with advanced analytics, have come a long way in processing extremely large datasets with even more speed and accuracy[5]. It has contributed to the evolution of predictive capabilities from actuaries (now predicted to be more so than ever), where they can now pattern recognize, detect anomalies, and generate actionable from all the available data, structured or not, from diverse sources like social media, geospatial data and sensor data. Then, this comprehensive data utilisation helps them improve their predictions of risk and allow for real time decision making. With the increasing progress of AI, it has a vital role in the automation of sophisticated calculations and in the participation of scenario analysis.



**Figure 1 : AI-Driven Framework for Actuarial Decision-Making**

The risk assessment using ML models based on ML models provides much better actuarial decision-making because the model can predict claim probability and detect fraud, as shown in Figure 1. AI-driven predictive analytics in cost management enhances forecast optimization of costs and quickens the process of pricing strategies. The effect is to stimulate economic scenarios and to assess the long-term risk, ensuring financial stability for the insurer and balanced reserves [6]. Moreover, AI helps in the development of policy by providing personalized insurance plans using data analysis and immediate response to risk factors variants and market trends. In a nutshell, Better actuarial judgements are the result of increased accuracy, decreased costs, and data-driven insight made possible by AI.

### A. Evolution of AI in Actuarial Modeling

The assessment of risks in insurance, pensions and financial planning was traditionally based on statistical methods, historical data and mathematical models (also actuarial science did so)[7]. But as AI has come to the scene, the field has undergone a major transition. Advanced predictive analytics, ML algorithms and automation are all introduced by AI, and all have made the risk assessment more efficient and accurate.

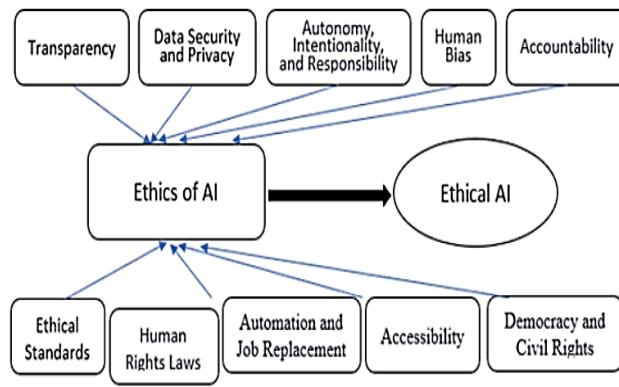
### B. Benefits of AI in Actuarial Science

Actuarial science is improved through risk assessment, predictive modeling, and decision-making efficiency by AI.

- **Better Risk Assessment:** However, on the other hand, when AI algorithms work on data from multiple sources, it is massive amounts of data from external sources. The manufacturers benefit from the best possible designs and strategies for pricing that this provides.
- **Improved Predictive Power:** Developing predictive models of the future requires the new opportunities brought by methods such as ML and DL found in complex non-linear patterns in massive databases, which allow for accurate analysis of the market and consumer habits[8].
- **Personalized Pricing:** Specifically, actuaries may create rates individually tailored to the risk profile of each specific customer through AI usages that are based on customer-specific data. This brings a more personalized individual experience to every client.
- **Fraud Detection:** Additionally, AI systems can and do use predictive analytics to visualize the extent of a claim to detect fraud intelligently, as well as alert the company and insurance department to areas where their processes need to be improved.

## III. ETHICAL FRAMEWORKS FOR AI IN ACTUARIAL PRACTICES

The explanation of ideas that underlie ethical norms, such as prejudice and fairness for AI systems, is typically the main goal of the conceptual component. Principles often manifest as intended AI system traits, such as transparency, data privacy during AI system development, or respect for human dignity throughout AIM rollout. In order to establish a framework for ethical AI, it is necessary to identify potential ethical issues and provide solutions that address or mitigate such risks. These models may provide an overview of probable ethical principles and problems, explain rules (in the case of legal models) or solutions for dealing with these problems, and provide the essential relevant concepts for discussing the ethical aspects of AI systems and their potential effects. Some examples of the latter include recommendations on how to best build AI systems and warnings about potential threats[8]. The explanation of ideas that underlie ethical norms, such as prejudice and fairness for AI systems, is typically the main goal of the conceptual component. Principles often manifest as desired AI system traits, such as transparency, data privacy during AI system development, or respect for human dignity in AI applications.



**Figure 2 :Ethical Considerations for Achieving Ethical AI in Actuarial Practices**

Several ethical factors impact the evolution of moral AI in actuarial processes, as seen in Figure 2. The ethics of AI is at the core of the ethics for the use of actuarial systems with respect to responsible decision-making. Some of the most important things to keep in mind while trying to uphold ethical standards include being transparent, protecting personal information, being accountable, being purposeful, taking responsibility, avoiding prejudice, and being autonomous. By guaranteeing that AI actuarial models are explainable and understandable, transparency enhances the trust in AI-driven actuarial models, while data privacy allows personal and financial information to be kept safe and protected from potential fraudsters[9]. The human bias can be addressed, and clear accountability frameworks provided in order to prevent discriminatory or unfair outcomes.

**A. Ethical Considerations in AI-Driven Actuarial Models**

AI becomes a fundamental part of actuarial decision-making, and maintaining ethical compliance of risk assessment and insurance models is critical in order to maintain trust, fairness and reliability in insurance. In this part, it discusses three important ethical considerations of choosing data for AI: transparency and explainability, fairness and bias mitigation, and data privacy and security.

*a) Transparency and Explainability of AI Decisions*

The goal of transparency in AI-driven actuarial models is to make AI-generated outcomes comprehensible to stakeholders such as actuaries, regulators, and policyholders so that they can understand how AI makes decisions. It is crucial to consider explainability when identifying possible biases, validate the actuarial risk assessment, and embrace trust in the AI system. AI models should be made to be interpretable using interpretable algorithms, justify its decisions in human readable form, and let domain experts validate the risk assessment.

*b) Fairness and Bias Mitigation*

In actuarial science, risk assessment, and insurance underwriting, the AI models must be impartial and not provide discriminatory results. There are many examples of bias in AI that can result from skewed training data, selection of features that are not appropriate or obvious, or that derive from systematic disparity in historical actuarial decisions. For that job, fairness-aware ML techniques like adversarial debiasing and reweighting techniques need to be included to mitigate these risks. Algorithmic accountability is a fundamental responsible approach adopted for AI-based actuarial tools to meet ethical norms and control guidelines. This involves continuous bias audits, fairness checks, and the establishment of ethical AI governance boards that oversee model updates.

*c) Data Privacy and Security Concerns*

The actuarial domain relies on enormous volumes of private financial and personal information, making strict data security and privacy protections necessary. Data protection laws like the GDPR and IEEE P7002 (Data Privacy Process

Standard) must be followed by AI-driven actuarial models[10]. Methods like differential privacy, federated learning, and homomorphic encryption aid in the construction of AI models while also protecting sensitive actuarial data.

### **B. Understanding Ethical Frameworks**

The terms used by different writers to describe and explain ethical theories—such as collections of concepts, foundations, methodologies, criteria, and frameworks—vary. The term "framework" refers to the logical and systematic large-scale description and explanation that theory provides, within which smaller descriptions and explanations (approaches) are created, assessed, and forecasted. Interconnections and the possibility of replacing, supporting, modifying, or amending current ideas are suggested by a review of Perspectives on Frameworks for Ethical Ideas in Public Service Life 251. Deontology, for instance, is the main descriptive and explanation framework for deontological ethics, whereas existentialist, contractarian, and Kantian ethics are some of the smaller created descriptions, explanations, and evaluations.

### **C. Ethical Considerations for Actuaries Using AI**

The data and technologies used by actuaries, as well as their job, are always changing. It is necessary to comprehend the ethical and non-ethical dangers associated with new data sources, new modelling approaches, and a new sector called InsurTech[11]. For instance, in the past, underwriting was based on a few observable criteria and assessments derived from pooled class experience. Individuals who are underwritten run the risk of prejudice (usually unintentional) and being denied coverage due to erroneous information and presumptions.

### **D. Ethical Principles and Frameworks**

- **Fairness:** Fairness in AI is a foundational ethical principle, obliging AI systems to treat all individuals equitably. Achieving fairness, however, is a complex endeavor, marred by challenges in data collection, algorithmic bias, and defining fairness metrics. Algorithmic fairness poses very important questions to examine as it aims to take away the deeply rooted biases found in virtually every AI system.
- **Transparency and explainability:** In the case of AI ethics, transparency and explainability are the main things that are required. As opaque AI systems can generate a lack of accountability and trust, deep learning models need to be paid greater attention to in order to ensure that they are black box black [12]. As such, both ethically and as a means to garner public acceptance of AI technologies, model transparency and explainability become a matter of pursuit.
- **Privacy and Data Protection:** AI is tackling the NFL's privacy and human rights issues. A complex web of concerns pertaining to permission, anonymization, and data protection surrounds the gathering, storing, and use of data by AI systems. One excellent example of a law that regulates personal privacy in AI applications is the GDPR.

## **IV. VALUE ALIGNMENT IN AI FOR ACTUARIAL SCIENCE**

AI for actuarial decisions ensures value alignment, meaning that any AI-driven models act in the same way as ethical principles, government regulations, or assessing societal expectations do. In the actuarial systems, AI is used to assess, price, and evaluate the claims, and hence it is important for these models to align with fairness, transparency and accountability. The advantages of misaligned AI are biased risk predictions, unfair insurance rates, or ethical concerns in the decision-making[13].

### **A. Concept of Value Alignment**

The realization that intelligent systems in general and autonomous entities (such as robots and software agents) in particular display more complicated behavior and are thus harder to control is what drives the VAP. Adding values to intelligent systems is one method to solve this issue. Three primary strategies are now being used to achieve this goal. First, the creation of standards, certificates, and recommendations. The second, "value-based design," which incorporates values as design requirements into the systems from the outset[14]. According to the third method, it is likely that autonomous entities should be forced to adhere to moral principles. Presented in an analytical manner. In this approach, the compliance problem is framed inside hybrid online social coordination systems, where autonomous rational entities interact within a common social space that is governed by norms. In particular, the issue of establishing values in these systems by having norms that support certain values and independent agents that may behave in line with their own, potentially disparate, values while being bound by the rules of a social coordinating environment[15].

### **B. Technical Approaches to Value Alignment**

The goal of technical methods to value alignment in AI for actuarial decision-making is to include moral values like accountability, transparency, and justice in AI models. If the preceding analysis is correct, then ensuring value alignment in AI-driven actuarial decision-making requires interdisciplinary collaboration. A clearer understanding of both the objectives of alignment in actuarial contexts and the technical methods available for implementation is needed. This is where technical research can offer a more in-depth look into the challenges of aligning AI in actuarial science as well, which will help improve the fairness, transparency and accountability of the process[16]. At the same time, input from actuarial experts, regulatory

bodies, and ethical researchers is imperative in establishing the goals of the alignment and the requirement of AI models making responsible, unbiased decisions. This section leads in the spirit of mutual collaboration, in the area of technical aspects of AI value alignment in actuarial decision making, by exploring methodologies to achieve alignment, challenges that have been encountered so far, and methods to guarantee that, as AI systems develop, they stay in line with moral and legal requirements.

### C. Value Alignment in the Gen-AI Augmented

The phrase "value alignment" describes how organizational activities pertaining to AI systems, in particular, are in line with the expectations and values of stakeholders and society at large. In the context of the Gen-AI Augmented Workforce, where AI systems and human workers participate in two ways, value alignment emerges as the most crucial factor. As AI systems improve in productivity and decision-making, a clear framework for the acceptable minimum output must be presented. This will assist in aligning the AI systems' outputs with the communities they serve. In addition to ensuring the moral integrity of AI applications, this alignment eventually ensures the legitimacy and profitability of apps in businesses. Emphasize the significance of value congruence between businesses and their stakeholders in corporate activism, emphasizing that this alignment is a key factor in establishing legitimacy and confidence in both AI and human-led operations.

## V. LITERATURE REVIEW

This section reviews ethical frameworks and value alignment in AI-driven actuarial decision-making, covering best practices, challenges, and solutions for transparency, fairness, and accountability. Table I summarizes the reviewed studies.

Chaturvedi, Patvardhan and Lakshmi (2023) intent to provide a comprehensive conceptual grasp of the AI value alignment challenge, the necessity of finding a solution, the main problems, potential solutions, and the future. AI is the creation of machines that can think and behave like people. Current AI developments aim to create systems that function at a level above that of humans in specific jobs or areas. Making sure the technology is reliable and safe is the main obstacle in the development of artificial intelligence[17].

Jin et al. (2023) provide a behavior tree-based value alignment technique with the goal of presenting a more trustworthy and dependable behavior tree for task fulfillment. The value model is developed to assess behavior trees in order to provide a bridge between humans and robots for value alignment. Instead of only following human orders, the value alignment technique combines human and machine intelligence through active and passive value alignment, human feedback, and robot explanation[18].

Reis et al. (2023) offer a structure for damage prevention while encouraging moral sensitivity and regard for human worth and welfare. To ensure that aspiring professionals are prepared to handle the ethical ramifications of artificial intelligence, it is intended to be included in technical education. In order to ensure that AI can assist society and not damage it, they may create a better and more responsible AI sector by emphasizing ethical reasoning in technological education[19].

Richman (2021) examines the potential evolution and adaptation of actuarial science to integrate these novel approaches and procedures in the upcoming years. The paper provides a brief overview of machine learning and deep learning before presenting examples from mortality modeling, claims reserving, non-life pricing, and telematics. It then goes on to discuss new applications of AI in actuarial science and offers a heuristic for when actuaries could find these methods useful. GitHub has made code available for several of the examples so that readers who are interested can try these methods out for themselves. Quick developments in ML and AI are producing goods and services that have the potential to both transform the actuarial profession and open up new avenues for research[20].

Espinosa and Zarruk (2021) intend to conduct and record a comprehensive analysis of the research on the role of actuarial management in insurance managerial decision-making in the twenty-first century. After providing a brief overview of the business setting, the study provides a thorough evaluation of the literature from 2000 to 2020, highlighting the advantages and difficulties of formally implementing different risk management strategies. The primary areas of attention are (i) financial modelling and quantitative strategic risk framework, (ii) actuarial cycle control and uncertainty management, and (iii) enterprise risk management using a value-based approach. Future scholars will gain from this study as it will enable them to more fully understand and articulate the numerous benefits and contributions of actuarial management in relation to managerial decision-making in an organization [21].

Gabriel (2020) The philosophical issues that come up in relation to AI alignment are examined in this work. It argues in favor of three claims. First, there is room for fruitful interaction between those working in the normative and technological facets of the AI alignment challenge since they are interconnected. Second, it's critical to clarify the alignment's objective. The differences between AI that follows instructions, intentions, stated preferences, ideal preferences, interests, and values are substantial. In this case, there are a lot of advantages to a principle-based strategy for AI alignment that carefully incorporates

these elements. Third, even if moral beliefs differ widely across people, theorists must establish fair alignment principles that are backed by reflection rather than focusing on identifying "true" moral principles for AI. The paper's last section examines three possible approaches to identifying fair principles for AI alignment[22].

Table I provides a summary of related work on ethical frameworks and value alignment in AI-driven actuarial decision-making, including key findings, approaches, challenges, focus areas, and limitations.

**Table 1 : Ethical Frameworks and Value Alignment for AI in Actuarial Decision-Making**

Reference	Study On	Approach	Key Findings	Challenges	Limitations
Chaturvedi, Patvardhan, and Lakshmi (2023)	AI value alignment problem	Conceptual analysis of AI value alignment	Identifies key issues, need for alignment, and potential solutions for safe AI development	Ensuring AI safety and trustworthiness	Lacks practical implementation details
Jin et al. (2023)	Value alignment in AI-driven systems	Behavior tree-based value alignment method	Proposes a behavior tree approach integrating human feedback for improved AI decision-making	Balancing active and passive alignment methods	Limited real-world validation
Reis et al. (2023)	Ethical awareness in AI	Ethical framework for technology education	Emphasises how crucial ethics are to AI education in order to avoid damage and improve wellbeing.	Integrating ethical considerations into AI development	Focused on education, not direct AI applications
Richman (2021)	AI in actuarial science	Applications related to deep learning and machine learning	provides GitHub code and real-world examples of AI applications in actuarial science	Keeping up with rapid AI advancements	Implementation complexities in actuarial models
Espinosa and Zarruk (2021)	Actuarial management in insurance	Systematic literature review on risk management tools	Analyzes risk management techniques and their impact on managerial	Addressing uncertainty in actuarial cycles	Limited empirical validation

			decision-making		
Gabriel (2020)	Philosophical issues in AI alignment	Principle-based AI alignment	Highlights the need for interaction between normative and technological aspects; emphasizes defining alignment objectives clearly; suggests focusing on fair principles rather than "true" moral principles	Variability in moral opinions; difficulty in defining alignment objectives	Challenges in achieving universal fairness; complexity in integrating multiple alignment components

**VI. CONCLUSION AND FUTURE WORK**

Responsible AI-driven actuarial decision-making requires ethical frameworks and value congruence. The difficulty of ensuring openness, fairness, accountability, and data protection arises when AI is included in risk assessment and financial forecasting. Sustaining public trust and regulatory compliance requires AI models to adhere to ethical norms. This paper examines the obligation of building responsible AI, and the frameworks that enable this responsibility include explainability, fairness-aware algorithms, and industry rules. To further ensure that AI-based actuarial models are in line with professional norms and social ideals, it also investigated value alignment tactics. The goal of future research should be to provide standardized ethics for AI in actuarial science by integrating actuarial professionals, policymakers, and AI researchers in an interdisciplinary manner. Promoting explainable AI (XAI) and fairness-aware algorithms might help dispel prejudices and increase confidence in actuarial models created by AI.

Future studies should then develop practical implementations of AI-based actuarial decision making with respect to ethical frameworks and value alignment. It includes improving behavior based alignment techniques, incorporating ethical reasoning into AI models, and dealing with the challenge of real-world situations: mitigation of bias and regulatory compliance. It is additionally possible to use interdisciplinary collaborations of actuaries with AI researchers and policymakers to strengthen the transparency, fairness and accountability in actuarial applications.

**VII. REFERENCES**

[1] N. Yeo, R. Lai, O. M. Jyeh, and J. Y. Liew, "Literature Review: Artificial Intelligence and Its Use in Actuarial Work," *Actuar. Innov. Technol. Soc. Actuar.*, no. December, 2019.

[2] Goyal, "Optimizing Project Timelines with Strategic Vendor Management and Blockchain-Enabled LEAP Collaboration," *Int. J. Res. Anal. Rev.*, vol. 10, no. 3, pp. 94-100, 2023.

[3] Prem, "From ethical AI frameworks to tools: a review of approaches," *AI Ethics*, 2023, doi: 10.1007/s43681-023-00258-9.

[4] J. Kumar Chaudhary, S. Tyagi, H. Prapan Sharma, S. Vaseem Akram, D. R. Sisodia, and D. Kapila, "Machine Learning Model-Based Financial Market Sentiment Prediction and Application," in 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), IEEE, May 2023, pp. 1456-1459. doi: 10.1109/ICACITE57410.2023.10183344.

[5] McKay, "Predicting risk in criminal procedure: actuarial tools, algorithms, AI and judicial decision-making," *Curr. Issues Crim. Justice*, 2020, doi: 10.1080/10345329.2019.1658694.

[6] S. Chatterjee, "Mitigating Supply Chain Malware Risks in Operational Technology : Challenges and Solutions for the Oil and Gas Industry," *J. Adv. Dev. Res.*, vol. 12, no. 2, pp. 1-12, 2021.

[7] S. Tan, R. Caruana, G. Hooker, and Y. Lou, "Detecting Bias in Black-Box Models Using Transparent Model Distillation," *arXiv Prepr.*, 2017.

[8] J. Kerkhoven, "Advances in constraint-based models: methods for improved predictive power based on resource allocation constraints," *Current Opinion in Microbiology*. 2022. doi: 10.1016/j.mib.2022.102168.

- [9] S. Du and C. Xie, "Paradoxes of artificial intelligence in consumer markets: Ethical challenges and opportunities," *J. Bus. Res.*, 2021, doi: 10.1016/j.jbusres.2020.08.024.
- [10] S. Choudhury, M. Lee, and A. Kurenkov, "In Favor of Developing Ethical Best Practices in AI Research," *Stanford AI Lab Blog*, 2019.
- [11] N. Raden, "Ethical Use of Artificial Intelligence for Actuaries," *Innov. Technol.*, 2019.
- [12] Rashmi, R. Sharma, P. Sharma, and S. Mangal, "Ethical Consideration in AI & Machine Learning," *Ind. Eng. J.*, 2023, doi: 10.36893/iej.2023.v52i5.1543-1551.
- [13] R. Richman, "AI in actuarial science - A review of recent advances - Part 1," *Annals of Actuarial Science*. 2021. doi: 10.1017/S1748499520000238.
- [14] S. Tyagi, T. Jindal, S. H. Krishna, S. M. Hassen, S. K. Shukla, and C. Kaur, "Comparative Analysis of Artificial Intelligence and its Powered Technologies Applications in the Finance Sector," in *Proceedings of 5th International Conference on Contemporary Computing and Informatics, IC3I 2022*, 2022. doi: 10.1109/IC3I56241.2022.10073077.
- [15] R. B. Bouncken, V. Fredrich, and S. Kraus, "Configurations of firm-level value capture in cooptation," *Long Range Plann.*, 2020, doi: 10.1016/j.lrp.2019.02.002.
- [16] H. Sievers, A. Joos, and M. Hiligsmann, "Real-world evidence: perspectives on challenges, value, and alignment of regulatory and national health technology assessment data collection requirements," *Int. J. Technol. Assess. Health Care*, 2021, doi: 10.1017/S0266462321000131.
- [17] S. Chaturvedi, C. Patvardhan, and C. V. Lakshmi, "AI Value Alignment Problem: The Clear and Present Danger," in *2023 6th International Conference on Information Systems and Computer Networks, ISCON 2023*, 2023. doi: 10.1109/ISCON57294.2023.10112100.
- [18] H. Jin, J. Li, F. Li, Y. Chang, T. Tian, and Y. Wu, "A Bidirectional Value Alignment Method Based on Behavior Trees," in *2023 IEEE International Conference on Unmanned Systems (ICUS)*, IEEE, Oct. 2023, pp. 1014-1020. doi: 10.1109/ICUS58632.2023.10318404.
- [19] S. Reis, L. Coelho, M. Sarmet, J. Araujo, and J. M. Corchado, "The Importance of Ethical Reasoning in Next Generation Tech Education," in *2023 5th International Conference of the Portuguese Society for Engineering Education, CISPEE 2023*, 2023. doi: 10.1109/CISPEE58593.2023.10227651.
- [20] R. Richman, "AI in actuarial science - a review of recent advances - part 2," *Ann. Actuar. Sci.*, vol. 15, no. 2, pp. 230-258, Jul. 2021, doi: 10.1017/S174849952000024X.
- [21] O. Espinosa and A. Zarruk, "The importance of actuarial management in insurance business decision-making in the twenty-first century," *Br. Actuar. J.*, vol. 26, Dec. 2021, doi: 10.1017/S1357321721000155.
- [22] Gabriel, "Artificial Intelligence, Values, and Alignment," *Minds Mach.*, vol. 30, no. 3, pp. 411-437, Sep. 2020, doi: 10.1007/s11023-020-09539-2.