Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 46 (2025), No. 7, pp. 2243-2252
DOI : 10.47974/JIOS-2126

Understanding importance of probability-driven ethical frameworks in navigating cybersecurity and AI ethics

Supreet Kaur Sahi [†]
Vandana Kalra *
Sri Guru Gobind Singh College of Commerce
University of Delhi
Delhi
India

Amanpreet Kaur [§]
Bow Valley College
Calgary
Alberta
Canada

Abstract

Probability driven frameworks are necessary for implementation of ethics in cybersecurity with Artificial intelligence (AI). By having these framework more informed decisions can be taken that will balance interest of stakeholders as these models will quantify risks and uncertainties. Fairness, accountability, and transparency are prioritized in this probability driven framework, that will ensure ethical norms and societal values are aligned with decisions. Structured approach these mathematical based frameworks provides valuable tool for developing fair practices in cybersecurity and Artificial Intelligence development. In the literature survey of this paper gaps are identified based on available work that further highlights different problems involved in ethical implementation of AI. The framework proposed in paper will provide way to different stakeholders for including risk involved, ethical implementation and promoting the responsible and unbiased use of Artificial Intelligence technologies in order to protect digital privacy.

Subject Classification: *Primary 68M25, Secondary 68Q87.*

Keywords: *Artificial intelligence (AI), Fairness, Cybersecurity, Probability, Risk mitigation.*

[†] *E-mail:* **supreetkaursahi@sggscc.du.ac.in**
* *E-mail:* **Vandana.kalra@sggscc.du.ac.in (Corresponding Author)**
[§] *E-mail:* **akaur@bowvalleycollege.ca**

1. Introduction

Ethical implementation of Cybersecurity and artificial intelligence have become important area of concern in today's digital world[1]. Digital infrastructure needs to be protected from malicious actors who seek to exploit vulnerabilities for personal gain or wrong intention. Different threats to cybersecurity are constantly evolving from identity theft and cyber spying to data breaches and ransomware attacks [2] [3]. The ethical implications of AI technologies is one of important way to deal with these threats as it has the potential to transform industries, revolutionize workflows, and improve human decision-making processes[4]. While implementing different algorithms individual behavior and personality should not interfere with outcome [5]. The use of different probability-driven ethical frameworks, to enhance cybersecurity measures and promote ethical behavior in AI systems is one of way to ensure fairness in algorithm. Unethical working in the era of digitization will have different risks involved [6].

Representation and manipulation of uncertainty can be done by more realistic mathematical framework that can further analyses existing AI ethics guidelines and implements key principles which will have more transparency and accountability. This paper highlights that ethical AI based support system This article analyses how the rise of unbiased AI based decision support systems will affect interactions between different agents by investigating existing literature and identifying gaps.

2. Literature Survey

Different contributions and gaps based on traditional framework of AI and cybersecurity is presented in literature which further highlights need of mathematical framework that can give unbiased results [6]. It is

Table 1
Gaps in Literature

S. No.	Identified Gap	Description
1	Lack of integrated frameworks combining ethics, AI, probability, and cybersecurity [5][7][8]	Most studies treat ethics, probability models, and cybersecurity as separate domains rather than building interdisciplinary, unified frameworks.

Contd...

2	Limited real-world implementation case studies [9][10]	There is a shortage of documented case studies showing how probability-driven ethical models are applied in real cybersecurity/AI systems.
3	Insufficient focus on emerging technologies (e.g., IoT, edge AI) [11][12]	Current literature focuses mostly on traditional AI systems, with limited attention to ethical frameworks for newer tech like edge computing or IoT.
4	Underrepresentation of global and cultural diversity in ethical models [13][14]	Many existing models lack adaptability to different legal, social, or cultural contexts, limiting global applicability.
5	Incomplete bias detection techniques using probabilistic methods [14]	While several fairness metrics exist, comprehensive probabilistic tools for detecting and mitigating hidden and intersectional bias are lacking.
6	Scarcity of probabilistic tools for continuous ethical auditing [15]	Literature often focuses on initial model development, ignoring the need for ongoing probabilistic auditing during deployment phases.
7	Over-reliance on theoretical models without practical validation [16][17]	Many proposed methods remain untested in real-time or industrial environments, leading to questions about scalability and efficiency.
8	Ethical uncertainty under adversarial conditions is underexplored [18]	There is minimal exploration of how probabilistic ethics frameworks perform under adversarial attacks or manipulation.
9	Lack of dynamic decision-making models under uncertainty [19][20]	Existing frameworks often assume static ethical choices, without incorporating probabilistic models for real-time, adaptive decision-making.
10	Limited public and stakeholder engagement in framework design[21]	Frameworks rarely include public perspectives or stakeholder feedback in their probabilistic ethical modelling.

clear from that rise of AI based support systems will affect interactions between different agents ethically and fairly. While considering ethical implementation probabilistic reasoning allows decision-makers to assess

the likelihood of various outcomes and risk involved. At the bottom of these probability-driven ethical frameworks lies the concept of expected utility, which combines the probability of an outcome with its associated utility or value. Ethical decisions are evaluated based on their expected consequences, taking into account both the likelihood of different outcomes and their ethical implications[7]. Several gaps and area for future work as emerged after literature review are summarized in Table 1.

3. Suggestive Probability Methods for Ethical Framework

As discussed in literature probability-based method plays important role for having fairness in different AI system. This session provides brief overview of different methods [20][21][22][23].

I. **Disparate Impact Analysis:** In order to have different outcomes for different groups of people disparate impact analysis can be used. Imagine there is AI system that evaluates job applicants for hiring. Disparate impact analysis will check if there are unfair differences by looking at the probability of getting hired for different demographic groups, like men and women.

II. **Statistical Parity Testing:** The chances of a positive outcome are the same for everyone or not are examined by statistical parity testing. For example, in order to evaluate approval of loan applications by AI system are similar or not for different racial groups. statistical tests are used to determine to evaluate significant differences.

III. **Equalized Odds Testing:** It is used for ensuring that the AI system makes accurate predictions for everyone. As for example equalized odds testing evaluates whether the system predicts loan defaults with the same accuracy for different groups like young and old people.

IV. **Calibration Analysis:** In order to examine whether the probabilities predicted by the AI system match the actual outcomes **calibration analysis is done**. Suppose AI system is used to predict medical condition. In order to check predicted probabilities, match the actual chances of having the condition for different groups, like males and females' calibration analysis can be used.

V. **Conditional Demographic Parity: Individuals with similar relevant characteristics should have same probability of receiving positive outcome from AI system. If personalized ads**

are recommended by AI system, then people with similar interests but from different age groups receive similar recommendations is ensured by conditional demographic parity.

VI. **Counterfactual Fairness:** Personal characteristics of people should not interfere with outcome is determined by this method. This plays very important role in admissions by checking whether admission decisions with respect to change in certain attributed like gender or ethnicity.

VII. **Kernel Density Estimation:** This model will check outcomes for different groups. The probability distribution of approval rates for various demographic groups, helping identify any disparities can be done by Kernel Density Estimation.

VIII. **Probabilistic Fairness Metrics:** In this method probabilities are used to measure fairness. The chances of making prediction errors for different groups or the probability of reaching certain performance levels can be done by this method.

IX. **Bayesian Fairness** Analysis: In this method uncertainty when evaluating fairness is considered. Bayesian methods help in estimating confident level when in fairness assessments, taking into consideration uncertainty of AI predictions.

X. **Fairness-aware Machine Learning Models:** This model includes fairness in training process ensure high level of fairness in predictions. These algorithms might penalize unfair predictions.

These models result in more equitable and ethical AI applications. There are many challenges associated with these models. Investment is needed in data analytics and modelling capability for probability driven framework implementation. Resistance to probabilistic thinking is another problem. Even there are challenges still probability-driven frameworks present numerous opportunities for organizations to enhance their ethical decision-making processes. By accepting uncertainty and using probabilistic reasoning, organizations can navigate ethical challenges with greater confidence, transparency, and adaptability. In the next session gaps identified in literature will be addressed based on probability models.

3. Proposed Framework for Addressing Gaps

For responsible innovation and ethical decision-making addressing issues identified in literature are very important. There is need to have

comprehensive framework that integrates different gaps like interdisciplinary perspectives, empirical research, and cultural and contextual considerations. Table 2 presents outline of proposed framework and maps it with different gaps identified.

Table 2
Proposed Framework

Component	Description	Gap Addressed
Ethical Risk Assessment Module	Uses probabilistic models (e.g., Bayesian networks) to predict and quantify ethical risks in AI systems.	Gap 1, Gap 9
Bias Detection Engine	Implements statistical parity, equalized odds, and counterfactual analysis to identify group-based biases.	Gap 5, Gap 4
Continuous Ethical Monitoring	Real-time auditing of AI systems using probabilistic metrics to track ethical compliance and fairness.	Gap 6, Gap 7
Contextual Adaptability Layer	Adjusts ethical parameters based on cultural, legal, and sectoral inputs using probabilistic weights.	Gap 4, Gap 3
Adversarial Resilience Check	Uses scenario modelling and Monte Carlo simulations to test ethical decision-making under attacks.	Gap 8
Dynamic Decision-Making Unit	Real-time decision engine that adapts ethical choices based on changing data distributions and risks.	Gap 9
Stakeholder Feedback Loop	Integrates user and stakeholder feedback to reweight probabilities in ethical outcomes.	Gap 10
Case Study Repository	Collects and analyses real-world implementation cases to refine the framework using probabilistic learning.	Gap 2
Modular Interdisciplinary Architecture	Combines ethical theory, AI models, cybersecurity protocols, and probabilistic reasoning into a unified system.	Gap 1

Contd...

Transparency Dashboard	Visual interface to communicate probabilistic fairness, risk scores, and ethical audit outcomes.	Gap 6, Gap 10

These points will help in informed and ethical decisions in cybersecurity and ai ethics. By incorporating these elements into a unified framework, organizations and policymakers can make more informed and ethical decisions in cybersecurity and AI ethics.

4. Discussion and Conclusion

In this paper importance of including probabilistic reasoning into ethical decision making is considered. This is very much important in case of complex and uncertain domains like AI and cybersecurity. But this integration faces several challenges like transparency, bias mitigation and accountability. Despite these challenges future research in this field offers significant opportunities. There is need of collaboration between policymakers, practitioners and researchers to refine existing frameworks and develop new methodologies. Probability driven framework presented in this paper ensures better handling of ethical complexity of digital age.

Based on understanding of review presented policymakers should support integration of probabilistic model by funding research projects, establishing regulatory guidelines, and promoting interdisciplinary collaboration. There is need to adopt continuous evaluation approach in order to assess the ethical implications and effectiveness of probabilistic frameworks over time. Stakeholder engagement also plays key role as they can contribute to the design of more inclusive and effective frameworks.

References

[1] S. Sudha, A. Narmadha, G. Amutha, and N. Ayedee, "Role of emerging technologies in marketing and decision making – PRISMA model," *Journal of Information and Optimization Sciences*, vol. 45, no. 6, pp. 1591–1605 (2024), doi: 10.47974/JIOS-1636.

[2] S. K. Singh, N. K. Bhasin, and D. Kumar, "Unleashing the fintech revolution: Transformative impacts of technology on financial services in the National Capital Region (NCR)," *Journal of Information and Optimization Sciences*, vol. 44, no. 8, pp. 1501–1514 (2023), doi: 10.47974/JIOS-1472.

[3] F. Cremer, F. Cremer, T. Holz, R. Mertens, and A. Risius, "Cyber risk and cybersecurity: a systematic review of data availability," *Geneva Pap Risk Insur Issues Pract*, vol. 47, no. 3, pp. 698–736 (Jul. 2022), doi: 10.1057/s41288-022-00266-6.

[4] A. Trotta, M. Ziosi, and V. Lomonaco, "The future of ethics in AI: challenges and opportunities," *AI & Soc,* vol. 38, no. 2, pp. 439–441 (Apr. 2023), doi: 10.1007/s00146-023-01644-x.

[5] M. F. Dixon, I. Halperin, and P. Bilokon, "Probabilistic Modeling," in Machine Learning in Finance: From Theory to Practice, M. F. Dixon, I. Halperin, and P. Bilokon, Eds., Cham: Springer International Publishing, pp. 47–80 (2020). doi: 10.1007/978-3-030-41068-1_2.

[6] R. Nishant, M. Kennedy, and J. Corbett, "Artificial intelligence for sustainability: Challenges, opportunities, and a research agenda," *International Journal of Information Management,* vol. 53, pp. 102104 (Aug. 2020), doi: 10.1016/j.ijinfomgt.2020.102104.

[7] C. Vidal and F. Heylighen, "Ethics and Complexity: Why Standard Ethical Frameworks Cannot Cope with Socio-Technological Change," in Humanism and its Discontents: The Rise of Transhumanism and Posthumanism, P. Jorion, Ed., Cham: Springer International Publishing, pp. 197–216 (2022). doi: 10.1007/978-3-030-67004-7_12.

[8] J. Chaki, "Probability and Bayesian Theory to Handle Uncertainty in Artificial Intelligence," in Handling Uncertainty in Artificial Intelligence, J. Chaki, Ed., Singapore: Springer Nature, pp. 13–24 (2023). doi: 10.1007/978-981-99-5333-2_2.

[9] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of AI ethics guidelines," *Nature Machine Intelligence*, vol. 1 (Sep. 2019), doi: 10.1038/s42256-019-0088-2.

[10] C. Huang, Z. Zhang, B. Mao, and X. Yao, "An Overview of Artificial Intelligence Ethics," *IEEE Transactions on Artificial Intelligence*, vol. 4, no. 4, pp. 799–819 (Aug. 2023), doi: 10.1109/TAI.2022.3194503.

[11] V. Belle, "Tractable Probabilistic Models for Ethical AI," in Graph-Based Representation and Reasoning, T. Braun, D. Cristea, and R. Jäschke, Eds., Cham: Springer International Publishing, pp. 3–8 (2022). doi: 10.1007/978-3-031-16663-1_1.

[12] R. L. Baskerville, J. Kim, and C. Stucke, "The cybersecurity risk estimation engine: A tool for possibility based risk analysis," *Computers & Security,* vol. 120, pp. 102752 (Sep. 2022), doi: 10.1016/j.cose.2022.102752.

[13] R. Kaur, D. Gabrijelčič, and T. Klobučar, "Artificial intelligence for cybersecurity: Literature review and future research directions," *Information Fusion*, vol. 97, pp. 101804 (Sep. 2023), doi: 10.1016/j.inffus.2023.101804.

[14] M. Braun, P. Hummel, S. Beck, and P. Dabrock, "Primer on an ethics of AI-based decision support systems in the clinic," *J Med Ethics,* vol. 47, no. 12, pp. e3 (Dec. 2021), doi: 10.1136/medethics-2019-105860.

[15] T. Miller, "Explanation in artificial intelligence: Insights from the social sciences," *Artificial Intelligence,* vol. 267, pp. 1–38 (Feb. 2019), doi: 10.1016/j.artint.2018.07.007.

[16] L. Floridi, "Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical," *Philosophy & Technology,* vol. 32 (May 2019), doi: 10.1007/s13347-019-00354-x.

[17] D. Dennehy, A. Griva, N. Pouloudi, Y. K. Dwivedi, M. Mäntymäki, and I. O. Pappas, "Artificial Intelligence (AI) and Information Systems: Perspectives to Responsible AI," *Inf Syst Front*, vol. 25, no. 1, pp. 1–7 (Feb. 2023), doi: 10.1007/s10796-022-10365-3.

[18] M. Ashok, R. Madan, A. Joha, and U. Sivarajah, "Ethical framework for Artificial Intelligence and Digital technologies," *International Journal of Information Management,* vol. 62, pp. 102433 (Feb. 2022), doi: 10.1016/j.ijinfomgt.2021.102433.

[19] K. Xivuri and H. Twinomurinzi, "A Systematic Review of Fairness in Artificial Intelligence Algorithms," in Responsible AI and Analytics for an Ethical and Inclusive Digitized Society, D. Dennehy, A. Griva, N. Pouloudi, Y. K. Dwivedi, I. Pappas, and M. Mäntymäki, Eds., Cham: Springer International Publishing, pp. 271–284 (2021). doi: 10.1007/978-3-030-85447-8_24.

[20] Pfeiffer, J., Gutschow, J., Haas, C. et al. Algorithmic Fairness in AI. *Bus Inf Syst Eng* 65, 209–222 (2023). https://doi.org/10.1007/s12599-023-00787-x.

[21] "A Causal Bayesian Networks Viewpoint on Fairness | SpringerLink." Accessed: Apr. 13 (2024). [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-16744-8_1

[22] "Fairness-aware machine learning engineering: how far are we? | Empirical Software Engineering." Accessed: Apr. 13 (2024). [Online]. Available: https://link.springer.com/article/10.1007/s10664-023-10402-y

[23] Z. Irfan, F. McCaffery, and R. Loughran, "Evaluating Fairness Metrics," in Advances in Bias and Fairness in Information Retrieval, L. Boratto, S. Faralli, M. Marras, and G. Stilo, Eds., Cham: Springer Nature Switzerland, pp. 31–41 (2023). doi: 10.1007/978-3-031-37249-0_3

Received March, 2025