Review of: "Beyond Algorithms: A Framework for Ethical Assessment of AI Applications in Industry"

Marcel Patalon[1]

1 South Westphalia University of Applied Sciences

Potential competing interests: No potential competing interests to declare.

Dear Researcher, I am grateful for the opportunity to review your research.

The paper, "Beyond Algorithms: A Framework for Ethical Assessment of AI Applications in Industry," proposes a novel framework to assess the ethical implications of AI applications in industrial settings. Grounded in the ethical principles outlined in the European Union's Trustworthy AI guidelines, the framework aims to support organizations in addressing the ethical challenges of AI adoption. It covers multiple dimensions, from technical robustness to human-centric considerations, ensuring AI systems align with ethical standards throughout their deployment and operation.

As AI technologies are adopted at an accelerating pace worldwide, ethical AI practices have become increasingly crucial. This paper addresses the issue in a timely manner, offering a structured approach to evaluating ethical risks in AI implementation. Given the rapid expansion of AI use in various sectors, this framework fills an important gap in current academic and industrial discussions, especially in industrial manufacturing. The framework presented covers a broad range of ethical considerations, from human oversight to data governance, making it holistic and actionable. Its systematic approach, which organizes ethical pillars into themes such as governance, operational procedures, and stakeholder engagement, demonstrates a deep understanding of the complexities involved in AI ethics.
One of the paper’s strongest aspects is its focus on practicality. The framework provides specific guidelines and evaluation criteria that organizations can directly apply to their AI systems. This action-oriented approach makes it a valuable tool for industries seeking to ethically align their AI adoption with their organizational goals and values.

Suggestions & Considerations

Clarity of Objectives: While the framework is described as novel, the paper could benefit from a clearer articulation of how it advances existing ethical AI frameworks, such as the European Union’s ALTAI. Although the paper mentions the framework's adaptability to future technological advancements, it lacks specific details on how this adaptability will be realized. The framework’s uniqueness, particularly in comparison to existing models, could be more clearly defined. Strengthen the paper by including specific case studies or examples that show how this framework operates in a dynamic, evolving technological landscape. Additionally, highlighting explicit contrasts between this framework and existing ones would better showcase its contribution to the field.

Operationalization of Key Concepts: Concepts such as "trustworthy AI" and "ethical AI" remain somewhat abstract in their operationalization. While the paper outlines key criteria, it could further clarify how organizations should practically assess these elements. For example, transparency is identified as important, but the paper does not fully explain how organizations should measure or verify transparency in AI systems. Provide more detailed guidelines on how to measure compliance with ethical principles. Offering specific examples from different sectors could illustrate how the framework is applied across various contexts and environments.

Methodology and Justification: The paper could elaborate further on the methodology used to develop the framework. Although the seven ethical pillars and five thematic areas provide a solid foundation, the process by which these elements were selected and prioritized is not fully explained. Additionally, the reasoning behind the binary scoring system could be clarified, especially regarding the weighting of ethical criteria. Include a section detailing the methodology used to build the framework, which would enhance the paper's credibility. If the ethical pillars and themes were developed in collaboration with industry experts or based on empirical research, this should be made explicit. Additionally, explaining the rationale behind the scoring system would clarify how ethical compliance is quantified.

Validation and Empirical Testing: Although the paper outlines the framework, it lacks empirical evidence or real-world case studies that validate its effectiveness. Without such validation, the framework remains theoretical and untested. Conduct pilot studies or provide examples from industries that have applied the framework. If empirical testing is beyond the scope of this paper, outlining a clear plan for future validation would still strengthen the argument. Discussing potential challenges organizations might face when implementing the framework could also provide a more nuanced understanding of its practical application.

Language and Terminology: The paper occasionally uses technical jargon and repetitive phrases, which may limit its accessibility to a wider audience. For instance, the term "trustworthy AI" is used frequently without further elaboration in certain sections. Streamlining the language and avoiding repetition would improve readability.

Stakeholder Involvement: While the framework emphasizes the importance of stakeholder engagement, it does not fully explain the practicalities of involving stakeholders in the assessment process. For organizations looking to apply this framework, clearer guidance on how to effectively include various stakeholders would be helpful. Expanding on how stakeholder feedback should be gathered and incorporated into the ethical assessment process would make this component more actionable. Offering practical methods for stakeholder inclusion (e.g., surveys, workshops, or collaborative platforms) would operationalize this key aspect of the framework.

Conclusion:

The paper offers a valuable contribution to the field of AI ethics, particularly in industrial settings where AI adoption is rapidly increasing. However, its impact could be enhanced by providing clearer differentiation from existing frameworks, expanding on how the framework can be operationalized, and including empirical examples or case studies to validate its effectiveness. A more detailed explanation of the methodology will

also strengthen the paper's argument. By addressing these areas, the framework proposed in this paper has the potential to serve as an essential tool for organizations seeking to align AI applications with ethical standards and societal values, ultimately contributing to more responsible AI practices.