

Peer Review

Review of: "AI Ethics by Design: Implementing Customizable Guardrails for Responsible AI Development"

Saara Ala-Luopa¹

1. Tampere University, Tampere, Finland

The paper presents an ethical guardrail framework for AI systems, emphasizing adaptability to diverse user needs and values. It outlines a structure for promoting responsible AI behavior and compares this new framework to existing guardrail models. The topic itself is interesting and important. Implementing ethical and philosophical concepts into practice (AI design and development) is challenging, and more research is needed on the topic. However, concerning the academic validity and reliability of the article, I have a few comments. I wish this feedback helps the authors to improve the paper.

Already after reading the abstract, I started to wonder about the research methods: What is the data behind the study? How was it collected? What is the background for this new framework the authors propose? For instance, have the authors conducted a literature review to identify the problem? What kind of previous research has been conducted on the topic?

Following this, there is no clear indication of the research problem, such as research questions. The reader can “read between the lines” and conclude the research gap from the abstract, but more clarity concerning the research problem is needed. A research question could be, for example, something like this: “How to develop an ethical guardrail framework for AI systems while emphasizing diverse user needs?” This can be followed by another question such as “How does the developed ethical guardrail framework for AI systems compare with existing guardrail models?”

Most importantly, I would suggest the authors explore previous research on the topic and clarify the research objective. This would help to format the latter part of the paper, for instance, by adding a discussion section in which the authors reflect on their findings with previous research and answer the research question(s). I also recommend that the authors familiarize themselves with academic

paper formatting, which typically includes methods, results, and discussion sections in addition to the abstract, introduction, and references.

Overall, the readability of the article is good, and as already said, the topic is important. I wish the authors all the best with this work!

Declarations

Potential competing interests: No potential competing interests to declare.