







Developing Artificial Moral Agents: Key Research Processes, Techniques, and Challenges

Fatemeh Ghazali¹, Touraj Banirostan^{1*}, Mirmohsen Pedram²

¹ Department of Computer Engineering, Central Tehran Branch, Islamic Azad University, Tehran, Iran

² Department of Electrical and Computer Engineering, Faculty of Engineering, Kharazmi University, Tehran, Iran

* Corresponding author email address: banirostan@iauctb.ac.ir

E d i t o r	R e v i e w e r s
Leila Youzbashi  Department of sport science, Faculty of Humanities, University of Zanjan, Zanjan, Iran l.youzbashi@znu.ac.ir	Reviewer 1: Farhad Namjoo  Department of Psychology and Counseling, KMAN Research Institute, Richmond Hill, Ontario, Canada. Email: farhadnamjoo@kmanresce.ca Reviewer 2: Masoud Mirmoezi  Department of Physical Education and Sport Sciences, Islamic Azad University, Central Tehran Branch, Tehran, Iran. Email: massoudmirmoezi@live.com

1. Round 1

1.1 Reviewer 1

Reviewer:

The phrase "profound concerns among scholars" (Introduction, paragraph 1) is vague. Could you specify which concerns are being referred to and provide citations or examples to substantiate this statement? This would make the introduction more robust and contextually grounded.

The definition of AMA as "a software or physical (robot) agent that can do an ethical behavior or at least avoid an immoral behavior" (Introduction, paragraph 3) needs clarification. What constitutes "ethical" and "immoral" behavior in this context? Providing examples would enhance clarity.

In the methodology section, the search strategy does not mention the exclusion criteria until later in the paragraph. Consider reorganizing the text to explicitly state inclusion and exclusion criteria upfront for clarity.

The justification for excluding theoretical papers "that lacked sufficient details for model description" (Methodology, paragraph 2) is weak. Could you elaborate on why these details were deemed necessary for inclusion?

The sentence "some papers were selected for further analysis" (Methodology, paragraph 2) is vague. Could you specify the criteria or framework used to determine which papers were selected?

The classification of ethical theories into "classic" and "modern" (Ethics and Ethical Theories, paragraph 2) seems overly simplistic. Consider acknowledging other possible classifications or nuances that could better capture the diversity of ethical theories.

The section on ethical theories heavily relies on older references (e.g., Stroll & Popkin, 1956). Including more recent literature would strengthen the contemporary relevance of this discussion.

The mention of Figure 1 (Ethics and Ethical Theories, paragraph 3) lacks a description of what it depicts. Adding a brief explanation in the text would enhance its integration into the narrative.

Author revised the manuscript and uploaded the updated document.

1.2 Reviewer 2

Reviewer:

The statement "it is an undeniable shortcut to consider and use cognitive capabilities in intelligent systems" (Introduction, paragraph 1) lacks clarity. Could you elaborate on what "cognitive capabilities" specifically entail in this context? Examples or references would enhance understanding.

There is a lack of a clear transition between the discussion of cognitive capabilities and the introduction of moral principles in AI systems (Introduction, paragraph 2). Consider adding a bridging sentence to establish how these concepts are interconnected.

The phrase "so the artificial moral agent and its challenges have received more attention in recent years" (Introduction, paragraph 2) is somewhat redundant given the previous sentences. Streamlining this could improve readability.

The discussion about "Friendly AI" and its relation to machine morality (Introduction, paragraph 3) is underexplored. Expanding on what distinguishes "Friendly AI" from other AI paradigms could provide more depth.

The discussion of challenges in implementing consequentialist theories (Discussion, paragraph 1) lacks depth. Could you expand on specific examples or provide case studies to illustrate these challenges?

The statement that "the hybrid approach has also received great attention" (Discussion, paragraph 1) is insufficiently detailed. What are the strengths and limitations of this approach? Adding references would help.

Author revised the manuscript and uploaded the updated document.

2. Revised

Editor's decision after revisions: Accepted.

Editor in Chief's decision: Accepted.