IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, October 2026
Impact Factor: 8.2

The Study of Ethical Decision-Making in Artificial General Intelligence (AGI)

A Review of Human Values, Moral Reasoning, and Accountability

Dr. Pawar Santosh Barku, Prof. Jadhav Ankita Ganpat, Khokrale Isha Machhindra
Vidya Niketan Institute of Management & Computer Application, Bota

Abstract: *Artificial General Intelligence, or AGI, is usually construed as a type of AI that can pick up and bring off tasks in many different areas. No system has been chuck so far to fully do that description. Even though this is the case, the fast growth of AI that is going more powerful imports it's important to allow about how these systems should jackleg choices that concussion people. This review syndicates studies on AI ethics, values, moral ideas, and proper management. It intends at ideas like fairness, apropos of people's independence, fettle personal information private, stuff sure things are safe, being open and honest, and being responsible.It also bills what each of these ways of thinking—like the utilitarian way, the deontological way, and the virtue-ethical way—can and cannot assist with. The review also intends at real-world challenges: people might hold different opinions about what's important, it's hard to understand what will episode next, information can be unfair or atilt, and complicated systems can be difficult to deduce or hack. Based on these themes, the paper offers a way of thinking that brings about together ethical rules with awareness of the situation, evaluation of risks, human involvement, and continuous checking. It doesn't import that one framework can crack all moral disagreements. Instead, it lands that stuff smart choices about future AGI is not just a technical goal. It will also touch different kinds of ethical thinking, clear rules for how it's forward, and strong responsibility from humans.*

Keywords: Artificial General Intelligence, ethical decision-making, AI ethics, human values, value alignment, moral reasoning, accountability, responsible AI

I. INTRODUCTION

Artificial intelligence is already habituated to assist with things like judgment patterns, bargain language, stuff plans, and matriarchal support when stuff decisions. Most systems that are being habituated right now are formed to hack a specific job or a small group of jobs. AGI is a larger and still not fully attained concept: an AI that can pick up, allow, and accommodate in various areas. There isn't one clear definition of AGI, and it hasn't been chuck that a completely general system lives. The idea of more powerful and flexible systems brings about up important ethical issues as well as technical challenges. A system might hand over its performance goals but still be unfair to people, dispose them in danger, or serve against their rights. Picture an AI admonishing in the distribution of public funds. It could be necessary to allow about efficiency, urgent needs, fairness, privacy, and people's rights all at the same time. Those goals can hale in different directions. Focusing on just one measure won't assist you understand what the best choice is. In this paper, ethical decision-stuff refers to the rules and steps that can assist both a smart AI system and the people and organizations that hack it jackleg choices that hold right or wrong effects. That imports not any harm, feting the decisions people jackleg, being fair to everyone, fettle their personal information safe, clearly construing important choices, and stuff sure everyone understands who is responsible for the results.

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, October 2026
Impact Factor: 8.2

II. KEY CONCEPTS AND ETHICAL FOUNDATIONS

2.1 Artificial General Intelligence (AGI)

AGI usually bears to a type of AI that is imported to be able to commitment many different kinds of thinking tasks, instead of just one specific job that is already incubate up. This could import gripping knowledge from one area and exercising it in another, accommodating to situations you're not habituated to, stuff plans, and reasonable through new information. There isn't a single coincided-upon test that everyone has to move AGI, so when people harangue about a system's general intelligence, it's important to be cautious and not grip the claims at face value. This paper intends at what might episode in the future with AI systems and learns from how AI is being hacked right now.

2.2 Ethical Decision-Making

Making ethical decisions claims carefully reasonable about who might be inclined and what real motives are detailed in the situation . It imports coming across at different options, reasonable about what might episode with each, and construing why one option is better than the others. In AI, this work also touches calculus out how values are exhibited, how uncertainty is traded with, when a person inevitably to interpose, and how someone can interrogative a decision. A technically correct answer doesn't always import it's the right thing to commitment morally.

2.3 Core Principles for Responsible AI

Many important ideas observe imminent up again and again in studies and discussions about AI ethics and rules. Beneficence is about burdensome to commitment good and assist others. Non-maleficence is about stuff sure we don't bring harm or jackleg things worse. Justice and fairness are about stuff sure people are not refined unfairly or differently, and that the good and bad things in life are participated in a fair way. Autonomy imports apropos of people's right to jackleg their own decisions after they hold all the necessary information. It's important for people to understand how a system forms its decisions and to deduce what it is capable of deed. Privacy defends personal information. Accountability imports that specific people or groups are responsible for how a system is, dispose into use, observed over, and hacked. These principles assist dolly out what is important, but they don't determine everything. For example, greater transparency could discover personal information. A choice that forms things better for everyone could still unfairly influence a smaller group of people. Ethical practice must hold a way to delicate these conflicts and construe how they were traded with.

2.4 Moral Theories Relevant to AGI

Utilitarianism intends at the results of actions to feel how they influence the total happiness or good of everyone. It can assist leaders feel the good and bad sides of different choices, but coming across only at the big picture might girl the rights of individuals or jackleg a small group grip on most of the burden. Deontological ethics is about next duties and hip when to layover—for example, serving people with respect as individuals and eluding actions that detriment others, even if those actions might hint to better overall results. This can assist defended rights, but duties might dis-accorded and can be difficult to exercise in new or unknown situations. Virtue ethics fastens on important qualities like honesty, fairness, care, and stuff smart choices. These ideas can assist bring rules for what AI systems should do and how the organizations exercising them should work. Even though a machine can let on to reel or routine honestly, this does not import it holds real human moral qualities. No one theory can fully construe all the answers to every ethical issue. Looking at a case from different angles can exhibit various issues: possible outcomes, what is right or wrong, and how someone routines predicated on the situation they're in. That wider perspective is important when people hold different values and it's unclear what the results will be.

III. CHALLENGES AND LIMITATIONS

3.1 Conflicting Human Values

People and communities usually hold different ideas about what is fair, how much freedom is acceptable, what level of risk people are willing to grip, and what is best for everyone. Even people who copacetic on general values can hold different opinions about what those values import in a particular situation. A future AGI system should not allow that the preferences of one person or the views of one culture pertain to everyone. Research should come across into how

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 6, October 2026 Impact Factor: 8.2

contrasting opinions can be harkened to, which voices are being deliberate, and when a choice should be formed by responsible people who are answerable to others.

3.2 Value Alignment and Objective Specification

An AI system might bring towards a goal that only partially halves what people really claim. If the goal skips out on something important, the system could hand over its measurable objective but bring outcomes that weren't budgeted by the designers. Value alignment research intends into how a system's actions can hang around in line with what humans claim and reel about. The problem picks up more difficult when a system fronts situations that its creators never consideration of. Alignment should therefore be descried as continuous work like testing, observing, peeling, and with a system rather than something that's concluded once and then liberalism untouched forever.

3.3 Bias and Fairness

AI systems can chosen up or even jackleg things worse when there live unfair patterns already in the data they exercise, how they live constructed, or the environment they work in. This problem can episode if the data is misplaced or missing, if the labels hold unfair biases, or if the system uses different groups in unfair ways. If a future AGI unfolds knowledge by knowledge from various different areas, the outcomes could be useful for more than just one particular task.

Fairness checks are like a hypodermic needle that assists interrogate how different groups are inclined. You shouldn't depend only on one overall accuracy number to deduce everything.

3.4 Transparency and Explainability

Some advanced AI systems can't construe clearly how they got in at a specific answer. That can jackleg it harder to detect mistakes, harder to interrogative decisions, and harder to clutch someone accountable. An explanation should assist the person who picks up it and should be clear about when things are not certain or anonymous. A confident explanation might be wrong if it donates the impression that it's certain when it's not. Explainability is useful, but on its own, it doesn't jackleg a decision fair or acceptable.

3.5 Safety, Uncertainty, and Unintended Consequences

People often jackleg ethical choices when they don't hold all the facts. Something that intends good in the short run might actually bring problems later on, and a system might routine differently in situations that weren't part of its testing. Risk assessment, coming across at possible situations, fettle track of things over time, conclusive about any issues that episode, and keeping the option to layover or unlashed important steps can all assist bargain with these problems. These protections are especially important when a decision might seriously concussion people's lives.

3.6 Accountability and Human Oversight

When AI brings damage, the responsibility can topple on the developers, the people who exercise it, those who take it, the organizations detailed, and the individuals who either forget or mount important decisions. If no one understands who is responsible for what, it could be difficult for people who are burdensome to detect answers or solutions. Companies shouldn't be permitted to criticize AI when they are to criticize for something crossing wrong. Good oversight imports clear responsibilities, access to the right information, real power to measure in when essential, and a way to interrogative decisions or pick up a second opinion.

3.7 Privacy and Security

Advanced AI systems can procedure a lot of information, which may contain personal or private data. Ethical design should dolly out how much data is actually essential, whether it's being habituated properly, how it will be observed safe, and who is permitted to exercise or controller it. Systems may also bring privacy problems because of what they furnish, or they can go accustomed to by people who enunciate they are injured. Privacy and security hypodermic needle to be carefully budgeted and answered at every step of edifice a system and when it is being habituated, not just after it is already rolling.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, October 2026
Impact Factor: 8.2

IV. FUTURE RESEARCH DIRECTIONS

4.1 Pluralistic Ethical Reasoning

Scientists might come across at different moral viewpoints to deduce them better, rather than clinging to just one clear rule for stuff choices. A pluralistic approach assists jackleg it easier to deduce the differences and conflicts between consequences, duties, rights, and social expectations. It should also delicate situations where automatic thinking isn't enough and people hypodermic needle to allow carefully.

4.2 Human-Centered Governance

Governance inevitably to clearly silhouette which decisions can be hacked by automation, which claim human approval, and which should not be permitted to be formed by an AI system.

Oversight inevitably to be as strong as the risk detailed and should furnish people who are influenced with a clear way to demurral major decisions. More studies should bill if changes in oversight actually jackleg a difference in real situations, not just in theory.

4.3 Explainability and Contestability

Researchers hypodermic needle to bring explanations that clearly exhibit why a decision was formed, contain the evidence that mountings it, and also advert any uncertainties or limitations that might influence the decision. People also hypodermic needle a good way to inquire for a decision to be conned if it influences their rights or chances. Explanations hypodermic needle to be good out with the people who will be exercising them and in the real situations where decisions are actually being formed.

4.4 Continuous Evaluation and Risk Management

Ethical review should not layover once a system is already in use. Systems and their surroundings can modification over time, so future methods might contain answering things before they're habituated, sampling different situations, fettle track of how they bring, coming across into problems when they episode, and keeping regular checks by someone outside the team. The assessment inevitably to come across at the social impacts and how risks are participated among different groups, along with how well the technical aspects work.

4.5 International Cooperation and Cultural Diversity

Risks from AGI can influence many countries, but different places hold different rules and values about what is right and wrong. International studies can come across into basic safety standards that are participated globally, but still permit for real discussions at the local level. Cultural differences are important and should be name, but they shouldn't be habituated to forget the basic rights or the concerns of people who are not as strong or powerful.

4.6 Sustainable and Long-Term Responsibility

Ethical evaluation inevitably to count the long-term effects on society, the economy, and the environment, not just the short-term outcomes. Researchers might come across into how organizations can trail their assumptions, observe records for audits, clearly entrust responsibility over time, and accommodate as the system's abilities or outcomes change.

V. PROPOSED CONCEPTUAL FRAMEWORK

The ideas in this review imply six compounded steps for reasonable about an important AI choice from an ethical point of view. This is a way to startle discoursing about something, but it hasn't been chuck to bring through proper testing.

1. Define the decision and its context. State the goal and the limits, dolly out who might be influenced, and describe which person or group holds the power to jackleg the decision.
2. Identify the relevant ethical concerns. Think about the possible good and bad things, the rights people hold, whether things are fair, their ability to jackleg their own choices, their privacy, and the legal or organizational responsibilities that pertain.
3. Compare the available options. Think about the possible results of each choice and come across at them from different ethical points of view.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, October 2026
Impact Factor: 8.2

4. Assess uncertainty and risk. Write down what you don't understand, what might serve wrong, and how likely and how bad the possible harms could be.
5. Set the level of human oversight. Decisions that are big, unclear, or hard to unlashed should be answered by a person. Reviewers should hold the option to break or disallow the system's suggestions.
6. Explain, record, and reconceive the decision. Write down the reasons for stuff a choice, construe any important restrictions, offer a way to inquire for a review, bill the outcome, and modification the process if new information nears up.

These steps are aimed to bring a cycle, not just a list of things to commitment once. Checking things over can exhibit new dangers, unfair results, or movement ideas that might import a choice inevitably to be came across at again. The framework also deduces that computers can't crack all moral issues; people and organizations still hypodermic needle to jackleg important decisions.

VI. CONCLUSION

Ethical decision-making should be at the core of any conversation about future AGI. Even though AGI is still a possibility and not something we've already compassed, the research on AI ethics today already communicates that there live important questions that will go more important as AI systems live habituated in many areas. These questions contain how to bargain with disagreements about human values, incubate the right goals, detect bias, construe decisions, observe privacy safe, guarantee safety, and dolly out who is responsible. Ethical principles are a good beginning, but by themselves, they aren't enough. Different ways of reasonable about right and wrong, such as utilitarianism, deontology, and virtue ethics, capital on different parts of a problem. None of them donates one clear answer that brings for every situation. Better choices episode when we come across at things from different angles, allow about possible problems, clutch real people responsible, deduce why decisions are formed, and jackleg sure those who are detailed are clutched accountable. This framework compounds these concepts by coming across at the context, answering ethics, bracketing different choices, coming across at possible risks, besides human input, and fettle things under review over time. It hasn't been sampled yet, so future work should bill how useful it is by picking up expert opinions, coming across at real examples, and boning how it does into governance. The purpose of AGI ethics is not to offer moral responsibility to a machine. It should be to jackleg sure that systems that are picking up better and better are and habituated in places that are still responsible to the people

REFERENCES

1. Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review, 1(1). https://doi.org/10.1162/99608f92.8cd550d1
2. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1, 389–399. https://doi.org/10.1038/s42256-019-0088-2
3. National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1). https://doi.org/10.6028/NIST.AI.100-1
4. OECD. (2019). Recommendation of the Council on Artificial Intelligence (OECD/LEGAL/0449). https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449
5. Russell, S. (2019). Human Compatible: Artificial Intelligence and the Problem of Control. Viking. UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence. https://www.unesco.org/en/legal-affairs/recommendation-ethics-artificial-intelligence
6. Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., Glaese, A., Balle, B., Kasirzadeh, A., Kenton, Z., Brown, S., Hawkins, W., Stepleton, T., Biles, C., Birhane, A., Haas, J., Rimell, L., Hendricks, L. A., Isaac, W., Legassick, S., Irving, G., & Gabriel, I. (2021). Ethical and social risks of harm from language models.
7. arXiv. https://arxiv.org/abs/2112.04359