Research Article | Education | ISSN No: 2839-5085 | Volume : 1 | Issue : 3 | July 2026

Artificial Intelligence and Ethics: Philosophical Perspectives on Human-Machine Interaction

Manjula S, Assistant Professor,
B.M.S Govt. First Grade College, Huliyar (Chikkanaikanahalli Tq.),
Tumkur, Karnataka.

Access Online

Abstract

This research paper looks closely at the ethical and philosophical questions that come from the quick spread of Artificial Intelligence into everyday human life. When autonomous systems, predictive models and conversational agents move from theory to use they leave a large gap in who is responsible. Hidden algorithmic choices, built-in bias and the spread of duty across tangled social and technical networks test old ideas about human freedom and control. Using an analytical philosophy approach the study examines Artificial Intelligence through classic moral theories such as Kantian deontology, utilitarianism and virtue ethics and compares them with modern views on machine ethics and shared responsibility. The main results show that Intelligent systems act as moral patients, not true moral agents so all ethical blame stays with humans. The hidden nature of deep learning designs goes against the need for logical interaction. In the end this study says that technology never has no value. Solving these problems means putting oversight from many voices straight into the software development process. Filling the gap, between probability-based computing and human values keeps innovation respectful, free and accountable.

Keywords; *Civil Society Organizations (CSOs), Democratic Institutions, Accountability, Civic Engagement, Political Pluralism, Human Rights, Governance, Democracy.*

Introduction

Artificial intelligence has quickly moved from an academic idea to a key part of modern life. In a short time how people use technology has changed a lot. People do not rely on strict command lines or manual input. Instead they live in a world guided by self-running systems models that predict and chat robots that fit smoothly into life. These changes reshape how society works, talks and works. With these great steps a serious problem remains. As algorithms gain power they often act like "black boxes" that make decisions without clear explanations. This lack of clarity creates problems with bias and who is responsible. When a digital system makes a mistake or

treats people unfairly it is hard to say who should blame. The blame moves between the developers who wrote the code the companies that use the technology and the machines themselves creating a gap in ethics. Relying on technical fixes does not solve this problem. Software updates, security checks and data rules cannot answer human questions about fairness, justice and intention. Machines do not have consciousness. Cannot naturally understand human values. A technical solution is not enough. A deep philosophical look is needed to close this gap. Society must think carefully about how artificial intelligence matches values before giving systems big decisions. Knowing this mix keeps progress in ethics and protects human dignity while keeping the benefits of new ideas.

Objectives of the study

- To explore how artificial intelligence affects human-machine interaction using moral theories.
- To examine if current artificial intelligence systems can have or ever reach real moral responsibility and standing.
- To study how core values such as autonomy, transparency and justice are tested by algorithm structures.
- To suggest a clear philosophical model to judge moral responsibility in high-risk areas where humans and artificial intelligence work together.

Literature Review

Looking at how digital tools meet morality needs us to study new philosophical ideas. Scholars look at automated systems through three historical lenses to see how modern inventions affect people. First classic moral theories give rules. Kantian deontology says that rational beings deserve dignity so algorithms must never use people as tools or data for optimization. At the time utilitarian thinking wants to increase overall happiness weighing the good of automation against harms like job loss or unfair treatment. Aristotelian virtue ethics focuses on character and human growth asking if using tools makes people better or worse and how it changes community bonds. Second modern philosophy of technology expands these ideas into the digital age. Today thinkers debate machine ethics. Whether it is possible to create artificial moral agents that can think about ethics on their own. Researchers study how responsibility spreads across parts of social-technical networks showing that the duty is not only on one person but spread across machines, data collectors, operators and companies. With a lot of study a big gap remains. There is a disconnect between high-level guidelines written by companies and governments and the real philosophical accountability needed in day-to-day life. Many rules are lack enforcement. As hidden feedback loops grow between users and automated programs it is hard to trace responsibility. Bridging this gap needs academic work to keep digital progress safe with human values.

Methodology

To study this topic we use a qualitative, conceptual and analytical method. We do not run lab tests or collect numbers. Instead we use thought, logical reasoning and theory building to look at how software systems and

human morality relate. The main sources are reviews of many materials. We use philosophical writing official reports on algorithm impact and regulations from international bodies. By checking these sources we map how high-level moral theories meet real software deployment and institutional oversight. The analysis uses ways. Conceptual analysis breaks down words like autonomy, fairness and accountability inside digital settings. Thought experiments, specially made for computers, test limits—such as how a machine should solve moral problems in tough situations. Normative critique links ideas straight to real human-machine cases. These real tests focus on areas, like self-driving cars, software that predicts justice sentencing and automated health diagnosis tools making sure theory stays tied to real social effects.

Analysis

The study looks at the ways that digital intelligence works and how it affects people. This kind of thinking looks at three areas: what machines want, how people lose their own freedom and who is responsible when something goes wrong. First looking at how machines act needs a line between real thinking and just following math. Big machines use a lot of information to make answers but these answers come from numbers, not real thoughts. Machines do not think, feel or know. They just guess what comes next based on what they have seen. Knowing this helps people not to trust machines much because what they do is not real thinking. Second when people use machines a lot it can take away their choices. The world around people uses tricks to get attention like showing things that match what someone likes. This makes people do things without realizing it. Also when people use machines for thinking they stop using their brains. This weakens their ability to make decisions. These things together take away peoples freedom without them knowing. Third when machines make mistakes that hurt people it is hard to say who is to blame. This is called the responsibility gap. If a machine misses a disease or a car crashes it is not easy to find who is at fault. Machines can't be blamed because they don't understand right or wrong. So blame spreads among people like the people who made the code the people who trained the machine the bosses who pushed the product out and the people who trusted the machine. Fixing this problem needs looking at everything. Society can't think of machines as tools. Understanding how machines work helps protect peoples choices making sure responsibility is where it should be.

Discussion

Connecting ethics with computer work means looking at how ideas work with actual code. This part looks at balancing rules with machine learning making sure people know how computers decide and looking at different ways of thinking around the world. First mixing ideas with new machines brings problems. Rules that say things must be done in a way do not fit with machines that make choices based on numbers. A machine does not know right from wrong it just guesses what is right. Making these fit needs turning ideas into numbers without losing what makes human rights important. Second, making sure people can understand how computers work is a part of being fair. When computers hide how they decide people can't question their choices. Making sure people can see how machines work helps them keep their choices and say no when needed. Third looking at ways of thinking from around the world helps make sure new technology works for all people. Some ways focus on the person, others on the group. Including ideas makes sure technology helps everyone, not just a few.

Result

Looking at the study shows some points about how machines work and what people must do. First smart machines are not people with choices. They do not think or feel so they can't be blamed for things. All the responsibility is on people. Second the way some machines work is not fair. They are hard to understand, which makes it hard for people to make choices. Finally making sure machines work right needs people to be involved all the time. Fixing things after the fact does not work. Developers must think about ethics all the time not after. Keeping things fair needs people to watch every step of making machines.

Conclusion

Ending this study means going to the main ideas about how people and machines work together who is responsible and what is right. Throughout the study old ideas have shown how machines change the world challenging what people think about choices, goals and blame. A main idea is that machines are not neutral. Every computer program shows the choices of the people who made it. Letting machines be made without help is risky. Making sure machines work right needs people to guide them with ideas being open and thinking clearly. Without help machines might take away rights and choices. Moving forward solving these problems needs people, from many areas to work together. Experts must not work alone. Future work must bring together people who think about ideas, computers, laws and rules to make plans. By talking all the time society can make sure new machines help everyone while protecting what is right and making sure people stay in charge.

References

Brey, Philip. "Disclosive Computer Ethics." *Computers and Society*, vol. 30, no. 4, 2000.

Coeckelbergh, Mark. *AI Ethics*. MIT Press, 2021.

Floridi, Luciano. *The Ethics of Information*. Oxford University Press, 2013.

Garfinkel, Simson. *Database Nation: The Death of Privacy in the 21st Century*. O'Reilly Media, 2000.

Introna, Lucas D. "The Enframing of Code: Agency, Authenticity, and the Other." *Theory, Culture & Society*, vol. 24, no. 6, 2007.

Jobin, Anna, Marcello Ienca, and Effy Vayena. "The Global Landscape of AI Ethics Guidelines." *Nature Machine Intelligence*, vol. 1, no. 9, 2019.

Johnson, Deborah G. *Computer Ethics*. 4th ed., Pearson Prentice Hall, 2009.

Moor, James H. "The Nature, Importance, and Difficulty of Computer Ethics." *Computers and Society*, vol. 15, no. 1, 1985.

Pasquale, Frank. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press, 2015.

Russell, Stuart, and Peter Norvig. *Artificial Intelligence: A Modern Approach*. 4th ed., Pearson, 2020.

Taddeo, Mariarosaria, and Luciano Floridi. "Regulating Artificial Intelligence Governance: Federal Approaches to Public Policy in the EU and the US." *Philosophy & Technology*, vol. 31, no. 1, 2018.

Vallor, Shannon. *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. MIT Press, 2016.

Wallach, Wendell, and Colin Allen. *Moral Machines: Teaching Robots Right from Wrong*. Oxford University Press, 2009.

Winner, Langdon. *The Whale and the Reactor: A Search for Limits in an Age of High Technology*. University of Chicago Press, 1986.

Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs, 2019.