
AI AND ETHICS: POTENTIAL DANGER OR MANAGEABLE RISK

Prin. Dr. R. S. Salunkhe

Arts, Commerce and Science College, Palus Dist: Sangli (MS)

prindrssalunkhe@gmail.com

Abstract

Artificial Intelligence (AI) is more and more influencing almost every portion of human life, becoming extremely integrated into communication, work, healthcare, education, transportation, shopping, and personal well-being. Working through machine learning algorithms, natural language processing, and computer vision, AI systems often function unnoticeably within digital platforms and daily devices, determining decisions and behaviours in subtle yet powerful ways. While these advancements offer significant opportunities for innovation, efficiency, and communal progress, they also present ethical challenges such as privacy concerns, algorithmic unfairness, job displacement, and the risk of exploitation. This paper significantly examines the dual nature of AI whether it should be supposed as a potential danger or a manageable risk. The paper further emphasizes the importance of ethical governance, regulation, and human oversight in ensuring the responsible development and application of AI.

Keywords: Artificial Intelligence, Ethics, Risk Management, Governance

Introduction

Artificial Intelligence has quickly evolved into one of the most transformative technologies of the 21st century, influencing various sectors such as healthcare, finance, education, transportation, and entertainment (Brynjolfsson & McAfee, 2017). Its extensive applications, ranging from facial recognition in hospitals to predictive tools used by law enforcement, demonstrate its rising role in present society. Beside these benefits, however, the increase of AI has sparked strong ethical debates regarding privacy, bias, accountability, and even existential risks.

The vital question that arises is whether AI poses a significant danger to humanity or whether, with proper regulation and governance, it can remain a safe and beneficial tool. Ensuring that AI systems stay under human control and are developed responsibly is key to preventing potential harm. While the complete impact of these technologies is not so far entirely understood, it is evident that society must proactively deal with both their promises and their risks in order to safe a sustainable and ethical future.

Statement of the Problem

Artificial Intelligence has quickly emerged as a transformative force, influencing significant aspects of human life including healthcare, education, finance, and transportation (Bryson, 2019; Floridi & Cowls, 2021). While its capabilities propose extraordinary opportunities for efficiency, innovation, and societal progress, they at the same time raise serious concerns about potential fear to human independence, safety, and social structures (Bostrom, 2014). The main ethical challenge is to set up effective frameworks, policies, and technological

safeguards that can ensure AI systems stay associated with human values, care for fundamental rights, and promote long-term well-being. Without such dealings, the unchecked development of AI could degenerate inequalities, compromise privacy, and create risks to public stability.

Objectives of the Study

1. To examine the Artificial Intelligence poses potential dangers to human life and society.
2. To analyze the extent to which the risks associated with AI can be managed through ethical governance and regulation.
3. To draw a balanced conclusion regarding the dual nature of Artificial Intelligence as both a potential danger and a manageable risk.

Methodology of the Study

This study is based on secondary data, representation information from a wide range of credible sources including books, research articles, newspapers, online databases, and other published materials. The collected data has been critically reviewed and analyzed to explore the ethical scope of Artificial Intelligence and its potential risks and benefits. The researcher has synthesized these insights to present interpretations and perspectives on whether AI should be considered a potential danger or a manageable risk.

Limitations of the Study

As the study relies exclusively on secondary data, its findings are dependent on the availability, accuracy, and reliability of existing literature and published sources. The absence of primary data may limit the scope for capturing immediate perspectives and emerging trends in the field of AI and ethics.

Discussion and Interpretation

In **personal life**, AI is present in digital assistants such as Siri, Alexa, and Google Assistant, which interpret voice commands, schedule reminders, and control smart home devices (Hoy, 2018). Streaming platforms like Netflix and YouTube employ recommendation algorithms that analyze viewing history to suggest relevant content, while e-commerce platforms like Amazon use similar systems to personalize product offerings (Gomez-Urbe & Hunt, 2016). Wearable devices such as Apple Watch and Fitbit leverage AI to monitor physical activity and detect potential health anomalies (Piwek et al., 2016).

In **communication and information access**, AI powers automated translation services like Google Translate, enabling cross-linguistic interaction in real time (Wu et al., 2016). Social media platforms apply AI to curate news feeds, recommend connections, and detect harmful content (Tarleton, 2020).

In the **workplace**, AI tools assist in automating repetitive administrative tasks through Robotic Process Automation (RPA), provide data-driven decision support, and enhance productivity through predictive text and advanced grammar suggestions in office software (Van der Aalst et al., 2018). In **public life**, AI systems enhance navigation (e.g., Google Maps traffic prediction), enable ride-hailing services (e.g., Uber's dynamic routing), and support healthcare through diagnostic imaging and predictive analytics (Esteva et al., 2019).

Although these applications express considerable convenience and efficiency, they also raise ethical considerations, including concerns over privacy, algorithmic bias, and the public implications of automation. Ensuring that AI systems are transparent, fair, and accountable remains a critical priority for researchers, policymakers, and industry practitioners (Jobin et al., 2019).

AI Could Become a Danger to Humans

The fast growth of Artificial Intelligence has generated a number of ethical challenges that must be critically examined. While AI promises innovation and efficiency, its unregulated development may create significant social, economic, and existential risks. The major ethical concerns include the following:

1. Loss of Human Control: Advanced AI systems are moving toward autonomous decision-making, reducing direct human error. If their actions move away from human values, such systems may make harmful decisions (Bostrom, 2014). This highlights the risk of humans losing authority over AI outcomes. While autonomy improves efficiency, it increases the danger of unintentional consequences, making strong oversight and ethical safeguards necessary. Tesla's self-driving cars have been involved in accidents where the AI system failed to recognize obstacles, showing the dangers of limited human intervention.

2. Job Displacement and Economic Inequality: Automation reduced millions of jobs, mostly in manufacturing, retail, and customer service, potentially leading to large-scale unemployment and widening wealth gaps (Frey & Osborne, 2017). AI-driven automation can boost productivity but risks displacing human workers, increasing unemployment and income inequality. Re-skilling, up-skilling, and comprehensive policies are crucial to balance technological progress with social equity. Amazon warehouses use AI-powered robots for packaging and inventory, reducing the need for human labour and raising concerns in relation to worker layoffs.

Will AI take your job? Indians say 'Yes!' Survey conducted in 21 countries of the world; Developing countries have the highest fear of losing jobs due to AI, while developed countries have the lowest fear of losing jobs. (Lokmat News Network Mumbai dated 23rd August, 2025) According to a survey called Global Public Opinion on Artificial Intelligence, 54% of people in the world think that their jobs could be threatened by AI or machines in the next 10 years. People are especially afraid of this in India, Pakistan and Indonesia.

What do people from different countries think? Will jobs go away? Figures in %

Country	Definitely Yes	Perhaps, Yes	Perhaps, No	Definitely No
India	36	39	17	8
Pakistan	31	41	18	9
South Africa	23	38	24	14
Mexico	21	44	21	3
America	18	30	28	24
Indonesia	16	60	16	8
Brazil	14	43	29	14
China	13	57	20	10

Spain	11	38	30	21
Global	15	39	27	18

It is seen that the fear is greater in developing countries. India, Pakistan and Indonesia have the highest proportion of people who answered 'yes'. In India, 75% (36%) of people believe that their jobs will be at risk due to AI, compared to 72% in Pakistan and 76% in Indonesia. This may be because these fast-growing economies have a high proportion of low-skilled jobs that are likely to be automated by AI.

Less concern in developed countries: The survey shows that the fear of losing jobs due to AI is low in developed countries like Germany, Japan, Canada, Britain, and Australia. Japan has the lowest proportion of those who say "definitely yes", at just 5%, while Germany has the highest proportion of those who say "definitely no" (34%). In contrast, in France, 41% of people think their job is a paycheck, while Portugal also has a high proportion of those who say "yes". The low fear of losing jobs in these countries may be due to strong labour laws or the occasional use of AI. Also, some research has shown that human interaction jobs are sometimes threatened by AI.

3. Bias and Discrimination: AI models regularly reflect biases present in their training data, resulting in discriminatory outcomes in hiring, policing, or loan sanctions (Buolamwini & Gebru, 2018). Such biases can reinforce existing inequalities and unfair conduct. Ensuring diverse, transparent, and accountable training processes is vital to minimize discrimination in AI systems. A study 2018 showed that Amazon's AI hiring tool discriminated against female applicants because it was trained on male-dominated past hiring data.

4. Privacy Concerns: AI-powered technologies, such as facial recognition and predictive analytics, frequently involve mass observation and large-scale data collection, raising concerns about the violation of personal privacy (Zuboff, 2019). Excessive data collection without proper precautions can remove individual freedoms. Strong data protection laws and ethical use of information are needed to protect human rights. In China, facial recognition technology is widely used for surveillance, sparking global debates about privacy and human rights violations.

5. Weaponization of AI: AI-enabled autonomous weapons could activate without human ethical constraints, leading to terrible consequences in warfare (Russell, 2019). The lack of human control in military AI poses risks of indiscriminate violence and conflict increase. Global cooperation and strict regulations are essential to prevent exploitation of AI in warfare. The development of killer drones, such as the Turkish Kargu-2, which can autonomously select and attack targets, demonstrates the dangers of weaponized AI.

6. Existential Risk: Hypothetically, if Artificial General Intelligence (AGI) surpasses human intelligence, it could act in ways beyond human comprehension or control, posing a threat to human survival (Bostrom, 2014). The concern is that AGI, if misaligned with human values, could prioritize its individual goals over human well-being. International collaboration on AI safety research and alignment strategies is critical to mitigate this long-term risk. Though AGI does not yet exist, leading scientists like Stephen Hawking and Elon Musk have warned that unchecked super intelligent AI could threaten humanity's future.

In which countries is ChatGPT banned and why

Although ChatGPT is used in many countries, it has been banned in some countries. Some places have banned ChatGPT because they do not want a platform that creates anti-government opinions while providing information freely. (Countries shown in red on the map) Cuba, China, North Korea, Hong Kong, Syria, Russia, Afghanistan, Bhutan, Eritrea, Libya, South Sudan, Sudan, Yemen, Central African Riel., Iswatini, Chad, Burundi, DRC, Iran, Belarus

Why have these countries imposed a ban; Fear of forming anti-government opinion, Government censorship. US/international sanctions, Security threats and conflict-ridden countries, Internet shortage, there are doubts about Open AI's data storage and sending of citizens' data to the US.

Source: Cybernews, OpenAI

AI Risks Are Manageable

Whereas concerns about Artificial Intelligence are valid, growing evidence suggests that with proper governance, ethical design, and human mistake, AI risks can be effectively managed. Key arguments include:

1. Ethical AI Frameworks are Emerging:

Governments, academic institutions, and corporations are developing ethical guidelines such as the EU's *Ethics Guidelines for Trustworthy AI*, which emphasize transparency, accountability, and human mistake (European Commission, 2019) these frameworks give a structured approach to ensure AI is aligned with human values and public needs. They provide as preventive measures against exploitation and help establish global standards. The European Union's AI Act (2021)

categorizes AI systems by risk levels and prohibits harmful practices like social scoring, demonstrating proactive governance.

2. AI as an Augmenter, Not a Replacer: In many fields, AI complements human abilities rather than replacing them. For instance, AI-assisted diagnostics in healthcare improve accuracy though leaving final decisions to doctors (Esteva et al., 2019). AI can improve human judgment rather than remove it, leading to better outcomes when humans and machines work together. IBM Watson assists oncologists by analyzing medical data and suggesting treatment options, but doctors hold the authority to make the final call.

3. Regulation and Governance Can Mitigate Risks: With effective regulations and monitoring, risks such as data exploitation and bias can be significantly reduced (Jobin, Ienca, & Vayena, 2019). Regulatory bodies ensure accountability and avoid AI from being used unethically. Strong governance builds public trust in AI applications. The General Data Protection Regulation (GDPR) in Europe enforces strict data privacy rules, limiting how companies use AI to process personal data.



4. AI for Social Good: AI is already being used to address urgent global issues such as predicting climate change impacts, optimizing energy use, and improving disaster response. (Rolnick et al., 2019) AI has the potential to make solutions for sustainable development and kind efforts, showing its positive public impact. Google's DeepMind AI has helped reduce energy consumption in data centers by 40%, contributing to environmental sustainability.

5. Historical Parallels: New technologies, from electricity to the internet, were initially met with fear. With time, governance, and adaptation, societies have integrated them safely (Tegmark, 2017). AI risks are not extraordinary; history shows that proper adaptation and regulation can convert disruptive technologies into everyday tools. The internet was once criticized for security and misinformation risks, but today it is necessary in communication, education, and commerce.

Balanced Conclusion

Artificial Intelligence has emerged as a transformative strength that permeates almost every aspect of human life, from healthcare and education to commerce, communication, and governance. Its capabilities offer extraordinary opportunities for efficiency, innovation, and problem-solving. However, along with these benefits lie important ethical concerns, including bias in decision-making, privacy violations, job displacement, and potential misuse in surveillance or autonomous weaponry. The debate is not about uncertain AI progress but about steering it responsibly. A balanced approach requires robust regulatory frameworks, transparent algorithms, inclusive stakeholder engagement, and continuous ethical mistake. By harmonizing technological advancement with human values, society can ensure that AI remains a tool for empowerment rather than a source of harm, securing both innovation and human dignity in the digital age (Floridi & Cowls, 2019; Jobin, Ienca & Vayena, 2019).

Conclusion

Artificial Intelligence represents both outstanding opportunities and thoughtful challenges for humanity. On one side, it poses ethical concerns such as loss of human control, job displacement, bias, privacy violations, weaponization, and even existential risks. On the other, AI also holds massive promise when developed responsibly through ethical frameworks, human-centered design, effective regulation, and applications for social superior.

The debate is therefore not about whether AI is inherently dangerous or safe, but regarding how societies decide to govern and apply it. If left unchecked, AI could increase inequalities and threaten social structures; however, with positive oversight, transparency, and international cooperation, it can become a transformative force for sustainable progress.

In conclusion, AI should be viewed as a **dual-edged tool**: potentially dangerous if misused, however highly beneficial if managed with responsibility and foresight. The ultimate outcome depends on striking the right balance between innovation and ethics, ensuring that technology serves humanity rather than undermines it.

References

1. Bostrom, N. (2014). *Super intelligence: Paths, Dangers, Strategies*. Oxford University Press.



-
2. Brynjolfsson, E., & McAfee, A. (2017). *Machine, Platform, Crowd*. W.W. Norton & Company.
 3. European Commission. (2019). Ethics Guidelines for Trustworthy AI.
 4. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399.
 5. Russell, S. (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking.
 6. Tegmark, M. (2017). *Life 3.0: Being Human in the Age of Artificial Intelligence*. Alfred A. Knopf.
 7. Daily news papers
 8. Cybernews OpenAI