

Keynote Address by Professor C. D. Sebastian



Professor C. D. Sebastian is Professor of Indian Philosophy at the Indian Institute of Technology (IIT) Bombay, Mumbai. He holds his MA (first rank with two gold medals) and PhD degrees in Indian Philosophy from Banaras Hindu University, Varanasi, and he did his Postdoctoral Research at the University of Bristol, UK. He was a DAAD Fellow at the University of Erfurt, Germany, and an Erasmus Mundus (of the European Commission) Visiting Scholar at the University of Warsaw, Poland. He was honoured three times with the *Excellence in Teaching Award* of IIT Bombay. He is an established

scholar and researcher in Buddhism, Indian Philosophy, and Comparative Philosophy. He has published more than 130 research papers in refereed international journals, and his books include *Metaphysics and Mysticism in Mahayana Buddhism* (2005, Bibliotheca Indo-Buddhica Series – 238) and *Recent Researches in Buddhist Studies* (2008, Bibliotheca Indo-Buddhica Series – 248). *The Cloud of Nothingness: The Negative Way in Nagarjuna and John of the Cross* (2016, Sophia Studies in Cross-Cultural Philosophy of Traditions and Cultures, Vol. 19 by Springer), and *Buddhism: Essays on Ethics and Religion* (Buddhist World Press – forthcoming). He is a member of the Society for Asian and Comparative Philosophy (SACP) and also a member of the Society for Inter-Cultural Philosophy, Cologne.

Higher Education, AI and Ethics: A Philosophical Reflection on Prospects and Challenges

Artificial Intelligence (AI) is revolutionising Higher Education. AI is transforming the learning of the students, teaching of the faculty, the functioning of the institutions of learning. AI is used in research, administration, and teaching to improve efficiency and access. AI should not be seen as a substitute for human expertise, but rather as a means or way to enhance human skills and efficiency. Machines will never replace humans, but unskilled human-work force could be replaced by machines. So, the need of the hour is training and equipping humans with latest technologies.

Higher Education and AI

Today in AI assisted learning, Generative AI, sometimes called *gen AI* is much used. *Generative AI* is AI that can create original content such as text, images, video, audio or software code in response to a user's prompt or request. It relies on sophisticated Machine Learning models called Deep Learning *models* algorithms that simulate the learning and decision-making processes of the human brain. These models work by identifying and encoding the patterns and relationships in huge amounts of data, and then using that information to understand users' natural language requests or questions and respond with relevant new content. Recent studies on *Generative AI* systems show promising results like assignment platforms have increased student placement efficiency by 20% and improved options for under-assigned students by 38% and so on.¹ In a recent Harvard study found that

¹ Tomás Larroucau, *et al* (2024), "College Application Mistakes and the Design of Information Policies at Scale" at [chrome-extension://efaidnbmnnpbpcajpgclclefindmkaj/https://tlarroucau.github.io/Information_Policies_at_Scale_LRFN_2024.pdf](https://efaidnbmnnpbpcajpgclclefindmkaj/https://tlarroucau.github.io/Information_Policies_at_Scale_LRFN_2024.pdf), accessed on 22 October 2025.

carefully implemented *Generative AI* tools demonstrate meaningful improvements in learning outcomes like students using *AI* tutors learned more than double as much in less time compared to physical and active learning in the classrooms.² In a more recent study by Stanford shows how *AI*-enhanced tutoring could effectively scale expert teaching practices, leading tutors to employ more effective pedagogical strategies while achieving improvements at a cheap cost.³ A recent study report on *Generative AI* in Higher Education shows that *Generative AI* (*GAI*) has quickly gained a significant foothold in academia, and is now used widely for teaching, learning, and research purpose.⁴

However, making use of *AI*'s full potential and efficiency in many of the South and South-east Asian countries requires addressing much more critical impediments and hurdles. Among such barriers we could include the *digital divide*, which limits *AI* accessibility in rural areas of our countries in this region and also very low participation of the institutions of the South and South-east Asian countries in *AI* innovation. In order to address these challenges countries in our region need to invest more on *AI* research, equipment, infrastructure and, above all, faculty training programmes on *AI*. Another important hurdle is ethical concerns surrounding *algorithmic bias* and *data privacy*. I will highlight on this point.

***AI* and Ethics**

Ours is now an age of Artificial Intelligence (*AI*), and it aims at transforming the world through the innovation of *AI*. There is many an advantage of *AI* that makes human life comfortable and easy. They include, *AI*'s precision in diagnosing and treating of diseases, efficiency in production and accounting and banking with the help of machine learning, the development of Micro Aerial Vehicles (*MAVs*) paving the way for innovations in aerial robotics suited for applications like surveillance and environmental monitoring, and agile reconnaissance in terms of safety and security; Wheeled Mobile Robots (*WMR*) in industrial and service robotics. Thus, the boons are manifold with advent of the *AI*.

But there can be dangers as well if the persons those who deal with *AI* is unethical. Machines will never replace humans, but the *AI*-users will be replaced with non-*AI* users. It calls for an *integrity* of *AI*. All intelligence, including that of systems and controls, can be disastrous if it is *sans* ethics and integrity. As *AI* becomes more and more intertwined with human beings, societal functions and environmental challenges, the lack of a comprehensive framework for embedding integrity is a dangerous omission. The main challenge for us now is to configure and form a future where the collaboration between human insight and *AI* enhances value of human dignity and efficiency. It calls for an ethics of *AI* integrity.

² Gregory Kestin, *et al* (2024), "AI Tutoring Outperforms Active Learning", at <https://www.researchsquare.com/article/rs-4243877/v1>, accessed on 24 October 2025.

³ Rose E. Wang, *et al* (2025), "TutorCoPilot: A Human-AI Approach for Scaling Real-Time Expertise", at chrome-extension://efaidnbmnnnibpcajpgclclefindmkaj/<https://arxiv.org/pdf/2410.03017>, accessed on 24 October 2025.

⁴ Claire Baytas and Dylan Ruediger (2024), "Generative AI in Higher Education: The Product Landscape" available at Jstor: <https://www.jstor.org/stable/resrep61192> accessed on 24 October 2025.

Human Dignity and AI Ethics

Given the persistent drive for human dignity at the forefront of all human action, computer ethics has been around almost as long as computers themselves. Beginning with Norbert Wiener's *Cybernetics* (1948)⁵ and *The Human Use of Human Beings* (1950),⁶ Wiener saw in the emerging technology of cybernetics an opportunity, or a destiny, to affect every major aspect of life. "We are already in a position to construct artificial machines of almost any degree of elaborateness of performance. Long before Nagasaki and the public awareness of the atomic bomb, it had occurred to me that we were here in the presence of another social potentiality of unheard-of importance for good and for evil." (Wiener 1948, pp. 27-28).

Merve Hickok who teaches Data Ethics at School of Information at University of Michigan, and President and Policy Director at Centre for AI & Digital Policy (Center for AI and Digital Policy is an independent research organization examining national AI policies and practices, based in Washington, DC).

Algor-Ethics

In order to be able to speak correctly of an ethics of artificial intelligence, it will therefore be necessary that the development of every algorithm always draws on an ethical vision, that is, 'algor-ethics.'

Algorithm is a set of instructions to be followed in calculations or other operations. This applies to both mathematics and computer science. When we speak of algorithm in AI, it refers to the programming that instructs the computer how to learn to operate on its own. It is a set of instructions that is designed to accomplish a task. Here what is important to note is that the algorithms are written or made by human beings.⁷

And hence, the Algor-ethics is the field of ethics that focuses on the development and deployment of algorithms. It aims to ensure that algorithms are transparent, accountable and fair. Algorithms are not neutral or objective. They learn from the values, biases, and

⁵ *Cybernetics: Or Control and Communication in the Animal and the Machine* is a book written by Norbert Wiener (an American computer scientist, mathematician and philosopher who was professor maths at MIT) and published in 1948. It is the first public usage of the term "cybernetics" to refer to self-regulating mechanisms. The book laid the theoretical foundation for **servomechanisms** (whether electrical, mechanical or hydraulic), automatic navigation, analogue computing, AI, neuroscience, and reliable communications.

⁶ *The Human Use of Human Beings* is a book by Norbert Wiener, the founding thinker of cybernetics theory and an influential advocate of automation (**Automation** describes a wide range of technologies that reduce human intervention in processes, mainly by predetermining decision criteria, subprocess relationships, and related actions, as well as embodying those predeterminations in machines). The book argues for the benefits of automation to society; it analyzes the meaning of productive communication and discusses ways for humans and machines to cooperate, with the potential to amplify human power and release people from the repetitive drudgery of manual labour, in favour of more creative pursuits in knowledge work and the arts. The risk that such changes might harm society (through dehumanization or subordination of our species) is explored, and suggestions are offered on how to avoid such risk.

⁷ How to write an AI algorithm? 1. Identify the problem you want to address/solve with AI. 2. Gather Data. 3. Clean and prepare the Data for model training. 4. Choose an AI technology. 5. Build and train the model. 6. Test the model. 7. Finally deploy the model.

assumptions of their creators and users.⁸ This can lead to unintended and harmful consequence for individuals, marginalised groups, and society at large. For example, algorithms can be written to unknowingly perpetuate discrimination against vulnerable populations based on their caste, race, gender, age, or other characteristics. They can also affect people's privacy, security, and autonomy. But most alarming, algorithms can learn to manipulate people's behaviour, preferences, and opinions.

Algor-ethics is so important that through the understanding and applying of algor-ethics, we can ensure that all AI, ML, and DP systems are aligned with human values, respect for human dignity and human rights. Through this we can also foster trust and confidence in technology and its benefits for society. That's what it is essential that all technocrats, computer wizards, engineers have to learn and study Arts, Humanities, Social Sciences, particularly, philosophy, that bring critical thinking and value based approach in one's endeavours. Human dignity and autonomy are to be safeguarded for common good. AI systems can affect human dignity and autonomy by manipulating, deceiving, or coercing users, or even by replacing human roles and functions in various domains.

Conclusion

The rapid rise in artificial intelligence (*AI*) has created many opportunities globally, from facilitating healthcare diagnoses to enabling human connections through social media and creating labour efficiencies through automated tasks.⁹ However, these rapid changes also raise profound ethical concerns. These arise from the potential AI systems have to embed biases, contribute to climate degradation, threaten human rights and more. Such risks associated with AI have already begun to compound on top of existing inequalities, resulting in further harm to already marginalised groups.¹⁰

Before I conclude, let me make a point on the concept of *Artificial Integrity*: The concept of artificial integrity¹¹ emphasizes the development and deployment of *AI* systems that uphold and reinforce human-centered values, ensuring that AI's integration into society enhances rather than undermines the human condition. Through a holistic approach that combines external guidelines, internal operational consistency, and a commitment to a synergistic human-*AI* relationship, artificial integrity aims to ensure that *AI* systems contribute positively to society, marking a pivotal evolution in the field's trajectory.

**

⁸ Paul Wagle, "What is Algorithethics?", at <https://paulwagle.com/what-is-algorithethics/> accessed on 24 October 2025.

⁹ UNESCO Recommendation: "Ethics of Artificial Intelligence" <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics> accessed on 24 October 2025.

¹⁰ UNESCO Recommendation: "Ethics of Artificial Intelligence" <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics> accessed on 24 October 2025.

¹¹ Hamilton Mann, "Introducing the Concept of Artificial Integrity: The Path for the Future of AI" in *European Financial Review*, 22 December 2023: <https://www.europeanfinancialreview.com/introducing-the-concept-of-artificial-integrity-the-path-for-the-future-of-ai/> accessed on 24 October 2025.

Appendix

A Brief Historical Overview of AI

John McCarthy, a pioneer in computer science, coined the term artificial intelligence (AI) in 1955, setting the stage for the domain's formal inception at the 1956 Dartmouth Conference. Inspired by Alan Turing's¹² foundational work on machine intelligence, McCarthy envisioned AI as the science and engineering of making intelligent machines capable of performing tasks requiring human-like intelligence, such as reasoning, learning and autonomous action.

August 31, 1955: The term "artificial intelligence" is coined in proposal for a "2 month, 10 man study of artificial intelligence" submitted by John McCarthy (Dartmouth College), Marvin Minsky (Harvard University), Nathaniel Rochester (IBM), and Claude Shannon (Bell Telephone Laboratories). The workshop, which took place a year later, in July and August 1956, is generally considered as the official birthdate of the new field.

December 1955: Herbert Simon and Allen Newell develop the **Logic Theorist**,¹³ the first artificial intelligence program, which eventually would prove 38 of the first 52 theorems in Whitehead and Russell's *Principia Mathematica* (in chapter 2).

1958: John McCarthy develops programming language **Lisp**¹⁴ which becomes the most popular programming language used in artificial intelligence research.

1959: Arthur Samuel coins the term "**machine learning**," (ML) reporting¹⁵ on programming a computer "so that it will learn to play a better game of checkers than can be played by the person who wrote the program."

¹² **Alan Mathison Turing** (1912-1954) was an English mathematician, computer scientist, logician, cryptanalyst, philosopher and theoretical biologist. Often considered the father of modern computer science, Alan Turing was famous for his work developing the first modern computers, decoding the encryption of German Enigma machines during the second world war, and detailing a procedure known as the Turing Test, forming the basis for artificial intelligence. (The **Enigma** was a type of enciphering machine used by the German armed forces to send messages securely. Although Polish mathematicians had worked out how to read Enigma messages and had shared this information with the British, the Germans increased its security at the outbreak of war by changing the cipher system daily.) Turing Test is a thought-experiment that measures a machine's ability to mimic human-like communication. The test is a fundamental tool in the development of AI. Turing Test involves a game between a human, a machine, and an interrogator. The interrogator's goal is to determine which of the other two is the person and which is the machine. The machine's goal is to cause the interrogator to mistakenly conclude that the machine is the person. The test is conducted using three terminals , each physically separated from the other two – one is for computer, one is for human who functions as respondent and one is the interrogator. The interrogator asks the respondents a series of questions within a specific subject area. After a preset length of time, the questioner tries to determine which player is the human and which is the computer. If the interrogator fails to determine which player is which, the computer is declared the winner. Today's CAPTCHA system which is commonly used online to tell humans and bots apart, is based on the Turing Test.

¹³ It was the first program deliberately engineered to perform **automated reasoning**, and has been described as "the first artificial intelligence program. (The study of **automated reasoning** helps produce computer programme that allow computers to reason completely, or nearly completely, automatically. Although automated reasoning is considered a sub-field of artificial intelligence, it also has connections with theoretical computer science and philosophy).

¹⁴ Lisp an abbreviation of "list processing") is a family of programming languages, a practical mathematical notation for computer programmes.

¹⁵ Computer-assisted **reporting** describes the use of computers to gather and analyze the data necessary to write news stories. The spread of computers, software and the Internet changed how reporters work.

Deep learning: Deep learning (DL) is a method in artificial intelligence (AI) that teaches computers to process data in a way that is inspired by the human brain. Deep learning models can recognize complex patterns in pictures, text, sounds, and other data to produce accurate insights and predictions.¹⁶ A large language model called **ChatGPT**¹⁷ is based on deep learning, specifically a type of neural network called a transformer. ChatGPT's transformer architecture uses attention mechanisms to focus on the most important parts of the input, allowing it to process and comprehend a large amount of text data.

Generative AI or *gen AI*: Generative *AI* creates new content, like text or images, based on patterns in data. Generative *AI* tools have become more common since 2020s.¹⁸ In Generative AI, Large Language Models (LLMs) are a powerful form, generating human-like text, and Small Language Models (SLMs) focus on specialised and confined tasks with less data. Adding to LLMs and SLMs, Retrieval-Augmented Generation (RAG) enhances these models by bringing external information for more accurate results. It must be also stated here that today, *AI* Agents use Generative *AI* to autonomously perform tasks such as writing or research which could be good and at the same time bad (when it comes to academic honesty and integrity). They, in fact, represent cutting-edge advancements in automation and creativity.

C. D. Sebastian, PhD

Professor of Philosophy

Humanities and Social Sciences

Indian Institute of Technology Bombay (IIT Bombay)

Mumbai – 400076

sebastian@iitb.ac.in

¹⁶ The origins of deep learning and neural networks date back to the 1950s, when British mathematician and computer scientist Alan Turing predicted the future existence of a supercomputer with human-like intelligence and scientists began trying to rudimentarily simulate the human brain.

¹⁷ Chat GPT stands for Chat Generative Pre-Trained Transformer and was developed by an AI research company, Open AI. It is an artificial intelligence (AI) chatbot technology that can process our natural human language and generate a response.

¹⁸ Generative *AI* has a long history. In the 1960s the first notable example of a Generative *AI* system was the ELIZA chatbot, created by Joseph Weizenbaum at MIT which used pattern matching to simulate a conversation but had no true understanding. In the late 1970s and early 1980s, Rule-based expert systems continued to get developed, but there was also a period known as the “*AI* Winter” where progress slowed considerably. In the 1990s *AI* research continued with programmes showing advances in computational problem-solving, like IBM's *Deep Blue* beating a chess grandmaster. It was, in fact, in 2010s Generative *AI* saw a significant resurgence due to breakthroughs in *Deep Learning*, specifically with the development of models like *Generative Adversarial Networks* (GANs) and the *Transformer Architecture*, which powered models like Open AI's GPT. After 2020 there was boom in *AI*. The widespread commercial use and popularization of Generative *AI* began in the early 2020s augmented the accessibility of *Cloud Computing* and the launch of user-friendly interfaces like ChatGPT in late 2022 (For a detailed historical trajectory of Generative *AI*, see the sixth chapter of the Volume by Liza Long *Cyborgs and Centaurs: Academic Writing in the Age of Generative Artificial Intelligence*, published College of Western Idaho by titled “Where Did Generative AI Come From? An Exercise in Co-Writing” available at: <https://cwi.pressbooks.pub/longenglish102/chapter/writing-and-artificial-intelligence/> accessed on 24 October 2025.)