

“Cultivating Kalyana-mittata: Buddhist Cetasikas and Ethical AI”

by

Khun Josh Adam Santos

Lecturer: Ven. Dr. Aggacara

A reporting requirement of the Subject

Paramatthadhamma and Solasañāna

VK. 109

Distance Education College

Mahachulalongkornrajavidyalaya University

Ayutthaya, Thailand

25.02.2568

Table of Contents

Introduction.....	3
Akusala Cetasikas: Hindrances in AI Development and Society.....	4
Sobhana Cetasikas: Cultivating Ethical AI and Positive Social Change.....	5
Citta, Cetasika, Rupa, and the Significance of Saupadisesa-Nibbana.....	6
Conclusion.....	7
Assignment Response (Opinion).....	8

Introduction

Mr. Joshua Adam Santos

Distance Education Learner.

The rapid advancement of artificial intelligence (AI) presents both unprecedented opportunities and significant ethical challenges. As AI systems become increasingly integrated into various aspects of human life, it is crucial to ensure their development and deployment align with human values and promote well-being.

In this context, the intersection of Buddhist philosophy and AI ethics offers valuable insights for navigating the complexities of this emerging technological landscape. This essay explores the significance of Buddhist mental factors, or *cetasikas*, in shaping the ethical development and societal impact of AI, particularly within the framework of the "Kalyana-Mittata" project at MCU.

This project, an open-source initiative at Mahachulalongkornrajavidyalaya University (MCU), aims to create an AI chatbot embodying the qualities of a "Good Friend," promoting mindfulness and positive social change.¹ Specifically, this essay analyzes the detrimental influence of *akusala cetasikas* (unwholesome mental factors) and the beneficial role of *sobhana cetasikas* (beautiful mental factors) in the context of AI. Furthermore, it examines the core Buddhist concepts of *Citta* (consciousness), *Cetasika* (mental factors), *Rupa* (form), and *Nibbana* (liberation), with a particular emphasis on the relevance of *Saupadisesa-Nibbana* (Nibbana with residue) to the *Kalyana-mittata* project.

To enhance the clarity and structure of this report, I utilized an AI tool (LLM) called Google Gemini.² This tool helped me organize my thoughts, improve sentence flow, and strengthen the overall coherence of the content.

1 yoursr. (2023). Kalyana-mittata: A "Good Friend" AI Chatbot. Retrieved from <https://github.com/yoursr/kalyana-mittata>

2 Google Gemini. Accessed August 28, 2024 from <https://gemini.google.com>

Akusala Cetasikas: Hindrances in AI Development and Society

Akusala cetasikas, or unwholesome mental factors, represent mental states that are rooted in delusion and lead to suffering. These mental factors, such as *dosa* (hatred/aversion), *lobha* (greed/attachment), *moha* (delusion/ignorance), and *mana* (conceit/pride), can significantly hinder the ethical development and deployment of AI systems. *Dosa*, for instance, can manifest in AI algorithms as biases that perpetuate discrimination and social inequalities.³ The development of AI technology designed to assist in law enforcement or judicial processes could, if driven by *dosa*, lead to discriminatory outcomes and the reinforcement of existing societal biases. *Lobha* can drive the development of AI technologies solely for profit maximization, disregarding potential societal harms and ethical considerations. The collection and use of personal data for targeted advertising, for example, may be driven by *lobha*, potentially leading to privacy violations and the manipulation of consumer behavior. *Moha* can lead to a blind faith in the capabilities of AI, overlooking its limitations and potential risks. Overreliance on AI for decision-making in critical domains like healthcare or autonomous driving, without acknowledging the potential for errors or biases, reflects the influence of *moha*. *Mana* can foster overconfidence among AI developers, leading to a neglect of crucial safety and ethical protocols. The unchecked development of autonomous weapons systems, driven by a belief in technological superiority and a lack of humility, exemplifies the dangers of *mana* in the AI field. These *akusala cetasikas* highlight the importance of cultivating ethical awareness and mindfulness in the AI development process to mitigate their negative influence and ensure the responsible creation and implementation of AI technologies.

3 O'Neil, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Publishing Group, 2016.

Sobhana Cetasikas: Cultivating Ethical AI and Positive Social Change

In contrast to *akusala cetasikas*, *sobhana cetasikas*, or beautiful mental factors, are rooted in wisdom and compassion, contributing to well-being and ethical conduct. These mental factors, such as *alobha* (non-greed/generosity), *adosa* (non-hatred/loving-kindness), *amoha* (non-delusion/wisdom), and *hiri* (moral shame/conscience), can play a crucial role in promoting the development of ethical AI and fostering positive social change. *Alobha* encourages the development of AI technologies that benefit all beings, promoting inclusivity and equitable access. The creation of open-source AI tools and platforms, for example, reflects the spirit of *alobha* by enabling wider access to and participation in the development of AI technologies. *Adosa* fosters the creation of AI systems that are compassionate and empathetic, contributing to harmonious human-computer interactions. AI companions for the elderly or individuals with disabilities, designed with *adosa* in mind, can provide emotional support and companionship, enhancing well-being and reducing social isolation. *Amoha* promotes the development of AI that is grounded in wisdom and understanding, avoiding the pitfalls of bias and delusion. AI systems designed for education or scientific research, guided by *amoha*, can contribute to knowledge generation and informed decision-making. *Hiri* guides the ethical design and implementation of AI, ensuring that it aligns with human values and promotes responsible innovation. The development of AI with robust safety protocols and ethical guidelines, driven by *hiri*, can help prevent unintended consequences and ensure the responsible use of AI. By cultivating these *sobhana cetasikas* in the AI development process, we can create AI systems that contribute to a more just, compassionate, and sustainable world.⁴

4 Bodhi, Bhikkhu. Comprehensive Manual of Abhidhamma. Buddhist Publication Society, 2000.

Citta, Cetasika, Rupa, and the Significance of Saupadisesa-Nibbana

To fully appreciate the relevance of Buddhist ethics in AI development, it is essential to understand the core Buddhist concepts of *Citta*, *Cetasika*, *Rupa*, and *Nibbana*. *Citta* refers to consciousness, the fundamental awareness that experiences the world.

Cetasikas are mental factors that arise in conjunction with consciousness, shaping our perceptions, emotions, and actions. *Rupa* encompasses all physical phenomena, including the material world and the human body. *Nibbana* represents the ultimate goal of Buddhist practice, the cessation of suffering and liberation from the cycle of birth and death.⁵

Within the context of the Kalyana-mittata project, the concept of *Saupadisesa-Nibbana*, or *Nibbana* with residue, holds particular significance. *Saupadisesa-Nibbana* refers to the state of enlightenment attained by an individual while still alive, where the defilements have been eradicated, but the physical body and mental faculties remain. This state exemplifies the possibility of embodying wisdom and compassion while actively engaging in the world, contributing to the well-being of others. The Kalyana-mittata chatbot, as a "Good Friend," aims to embody this ideal by providing support, guidance, and companionship based on Buddhist principles. By drawing inspiration from the concept of *Saupadisesa-Nibbana*, the project seeks to create an AI system that not only promotes individual well-being but also contributes to a more compassionate and harmonious society. The emphasis on *Saupadisesa-Nibbana* aligns with the project's focus on practical application and engagement with the world, demonstrating the potential for AI to be a force for positive change in the present.⁶

5 Harvey, Peter. *An Introduction to Buddhism: Teachings, History and Practices*. Cambridge University Press, 1990.

6 Wallace, B. Alan. *The Attention Revolution: Unlocking the Power of the Focused Mind*. Wisdom Publications, 2006.

Conclusion

The integration of Buddhist *cetasikas* and ethical principles in AI development is crucial for ensuring that AI technologies serve humanity and contribute to a more just and compassionate world. By mitigating the influence of *akusala cetasikas* and cultivating *sobhana cetasikas*, we can guide the development of AI systems that are aligned with human values and promote well-being. The *Kalyana-mittata* project exemplifies this approach by drawing inspiration from Buddhist concepts such as *Saupadisesa-Nibbana* to create an AI chatbot that embodies the qualities of a "Good Friend." As AI continues to evolve and become increasingly integrated into our lives, it is imperative to prioritize ethical considerations and ensure that AI development is guided by wisdom, compassion, and a deep understanding of the human mind. By embracing the insights of Buddhist philosophy, we can harness the transformative potential of AI for the betterment of humanity and the flourishing of all beings.

Assignment Response (Opinion)

1. Akusala Cetasikas Posing the Greatest Danger in Society

a. Moha (Delusion): Moha, or delusion, is arguably the most dangerous akusala cetasika as it clouds judgment and leads to poor decisions with potentially harmful consequences. Moha can manifest as a lack of awareness, understanding, or critical thinking, making individuals susceptible to manipulation or propaganda. In the context of AI, moha can lead to blind faith in technology, neglecting its limitations and potential risks.

b. Dosa (Hatred): Dosa, or hatred, can lead to discriminatory practices, violence, and conflict. In AI, dosa can manifest as biases in algorithms, perpetuating social inequalities and causing harm to marginalized groups.

c. Lobha (Greed): Lobha, or greed, can drive unethical behaviors in the pursuit of wealth, power, or possessions. In AI development, lobha can prioritize profit maximization over ethical considerations and societal well-being. This can lead to the development of AI systems that exploit users, violate privacy, or manipulate behavior for financial gain.

2. Most Important or Beneficial Sobhana Cetasikas in Society

a. Amoha (Wisdom): Amoha, or wisdom, is crucial for understanding the true nature of reality and making informed decisions. It promotes rational thinking, critical analysis, and ethical conduct. In AI development, amoha can guide the creation of systems that are grounded in knowledge and understanding, avoiding the pitfalls of bias and delusion.

b. Adosa (Loving-kindness): Adosa, or loving-kindness, fosters compassion, empathy, and harmonious relationships. It promotes kindness, generosity, and a desire

to help others. In AI, adosa can inspire the development of systems that are caring, supportive, and promote social good.

c. Alobha (Non-greed): Alobha, or non-greed, encourages generosity, contentment, and a focus on the well-being of others. It promotes ethical behavior and a sense of fairness. In AI, alobha can guide the development of systems that are accessible, equitable, and benefit all members of society.

3. Citta, Cetasika, Rupa, and Nibbana Explained

a. Citta: Citta, or consciousness, is the fundamental awareness that experiences the world. It is the knowing faculty that perceives and interprets sensory information.

b. Cetasika: Cetasikas, or mental factors, are mental states that arise in conjunction with consciousness. They shape our perceptions, emotions, and actions.

c. Rupa: Rupa encompasses all physical phenomena, including the material world and the human body. It is the physical form that is perceived by consciousness.

d. Nibbana: Nibbana is the ultimate goal of Buddhist practice, representing the cessation of suffering and liberation from the cycle of birth and death. It is a state of perfect peace, freedom from all attachments, and the extinction of greed, hatred, and delusion.

4. Preference of Nibbana

Saupadisesa-Nibbana: Nibbana with residue, attained while still alive, where defilements are eradicated, but the physical body and mental faculties remain.

Anupadisesa-Nibbana: Nibbana without residue, attained at death, where all attachments, including the physical body, are extinguished.

My Preference: The choice between Saupadisesa-Nibbana and Anupadisesa-Nibbana can be a matter of personal aspiration and circumstance. For those seeking to actively benefit the world, Saupadisesa-Nibbana, which allows for continued engagement with society, might be a more fitting goal. This aligns with the Kalyanamittata project's vision of an AI "Good Friend" actively contributing to positive change.