

Article

Developing Emotionally Intelligent AI: A Yogācāra-Informed Buddhist Framework for Affective Computing

Yongshan He 

Department of Religion and Culture, The University of Winnipeg, Winnipeg, MB R3B 2E9, Canada;
y.he@uwinnipeg.ca

Abstract

This paper examines how the current state of affective computing is limited by its reliance on theories that treat emotions as static, isolated states, and argues that the holistic and process-oriented theory of mind from Yogācāra Buddhism offers a more sophisticated alternative, viewing emotion as an experience deeply integrated with cognition, volition, and somatic awareness. As a case study, this paper proposes a framework for sentiment analysis inspired by Yogācāra principles, based upon the Chinese Buddhist text *Mahāyāna Treatise on the Hundred Dharmas Illuminating the Gate*. This multi-aspect annotation system analyzes emotional expressions across five key dimensions corresponding to Yogācāra's "ever-present" Mental Factors. By mapping emotions in this compositional manner, the framework provides a more granular and context-rich understanding of human sentiment than current methods allow. This paper thus serves as a call to diversify AI's theoretical foundations, demonstrating through this Yogācāra case study how engagement with insights from different traditions can resist the top-down "theoretical monopoly" of Western psychological models, which flattens the rich diversity of human affective experience into a single, dominant paradigm.

Keywords: Yogācāra Buddhism; affective computing; AI; emotion; sentiment analysis

1. Introduction

Affective computing, a rapidly growing field within the AI industry, endeavors to create "emotionally intelligent" systems that can understand and respond to human emotions in sensitive and appropriate ways. The term was first given a widely recognized vision by Picard (1997), who defined it as "computing that relates to, arises from, or deliberately influences emotions", and described its goal as giving computers "the ability to *recognize and express* emotions, developing its ability to respond intelligently to human emotion, and enabling it to *regulate and utilize* its emotions" (Picard 1997, p. 3).

More recently, this goal is being formalized through efforts like the ISO International Standard for Affective Computing User Interface (AUI). Part 1 of this standard, published in 2022, states that the field aims to enable systems to "recognize, interpret and simulate human affects" (ISO 2022). Part 2 of the standard, published in 2024, defines "affective characteristics" as a "particular type of affect that is believed to be useful", which are considered as "properties that are used to describe users' affective experience in AUI" (ISO 2024). This standard highlights the fundamental task at the heart of affective computing: to capture the affective experience of human beings, which is an extremely complicated, multi-faceted, and subjective phenomenon, and translate it into explicit and measurable



Academic Editors: Jinhua Chen and Kai Sheng

Received: 3 April 2026

Revised: 14 June 2026

Accepted: 22 June 2026

Published: 25 June 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

data. These data are then collected, analyzed, and used to train AI models to generate appropriate affective responses.

To achieve this goal, computer scientists have turned to modern Western theories of emotion from psychology and neuroscience. These theories provide the essential frameworks for guiding how AI models will approach, categorize, and interpret human affective experience. The choice of an emotional theory is therefore not a minor detail; it fundamentally determines the capabilities and, more importantly, the limitations of the resulting AI. As we will see below, the constraints of these foundational theories reveal the need for more rounded and integrated ones.

This paper argues that the Yogācāra Buddhist theory of mind can help fill this gap. Yogācāra is a sophisticated philosophical psychology developed over centuries, with a highly systematic vocabulary for describing mental phenomena. Drawing on Yogācāra's rich analysis of mental experience, the paper proposes a framework for affective computing that is more holistic, processual, and psychologically nuanced than those currently in use. As its primary textual basis, the paper relies on the *Mahāyāna Treatise on the Hundred Dharmas Illuminating the Gate* (大乘百法明門論), which provides an extremely concise, systematic and accessible classification of Mental Factors in the Yogācāra corpus. The goal of this paper is to demonstrate how Yogācāra resources, as a case study, can productively inform the theoretical foundations of affective computing and AI research more broadly. In this sense, the paper intends to engage scholars across religious studies, philosophy, AI design, emotion science, and digital humanities who are interested in expanding the cross-cultural and philosophical diversity of AI's conceptual toolkit, while also inviting specialists in other religious and philosophical traditions to participate in this important endeavor.

2. Major Theories of Emotion Underlying Affective Computing

The theories of emotion adopted by affective computing determine how our complex affective lives are represented in its models. While a variety of theories exist, two have become the most widely adopted and influential: **categorical/discrete models**, most notably the **Basic Emotions Theory**; and **dimensional models**, such as the **Valence–Arousal Model** (Gunes and Pantic 2010; Pei et al. 2024). They correspond to the two different ways of affect representation listed in the ISO Standard: “affective characteristic categories” and “dimensional affective space” (ISO 2022, pp. 4–5).

2.1. The Basic Emotions Theory: A Categorical Approach

The most influential categorical model is the Basic Emotions Theory developed by psychologist Paul Ekman, which has been widely adopted in applied fields, including affective computing. This theory proposes that there is a limited number of universal, biologically based “basic” emotions that are shared across all human cultures. The original list included six emotions—happiness, sadness, fear, anger, surprise, and disgust—and was later expanded to include others, such as contempt (Ekman 1992; Plamper 2017, pp. 147–63).

This theory of universalism has been criticized in recent years, with scholars challenging Ekman's claim on the “universality” of basic emotions (Barrett 2006). Its categorical approach also has significant limitations when applied in affective computing. By identifying an emotional expression (linguistic, visual, or auditory) as belonging to a specific type, it results in a limited and rigid set of categories that struggles to account for the diverse, subtle, and changing nature of our affective experience, lacking adequate granularity (Pei et al. 2024; Gunes and Pantic 2010; Gunes et al. 2011).

2.2. The Valence–Arousal Model: A Dimensional Approach

As an alternative to discrete categories, dimensional models propose that emotions can be understood by plotting them along a small number of continuous axes. The most prominent of these is the Valence–Arousal Model (or the Circumplex Model of Affect, as initially developed by James Russell). This model describes emotion using two fundamental dimensions: **Valence**, which represents the affective quality of an experience, ranging from highly positive (pleasant) to highly negative (unpleasant); and **Arousal**, which represents the intensity of the experience, ranging from low (e.g., calm or sleepy) to high (e.g., excited or frenzied) (Russell 1980).

In this model, any emotional state can be represented as a coordinate in a two-dimensional space. For example, the emotion of “joy” might be mapped to a value like [Valence: 0.82, Arousal: 0.43] (ISO 2022, p. 8), indicating a pleasant and moderately activated state. The strength of this model is its flexibility; it moves beyond rigid categories and can represent a continuous spectrum of emotional intensity and mixtures of feeling.

However, the Valence–Arousal model has its own critical limitation: it is highly abstract. While it can tell you the polarity and intensity of an emotion, it loses the specific context and flavor that make an emotion meaningful. For instance, intense excitement and righteous anger might have very similar high-arousal, positive-valence scores, yet they are profoundly different psychological experiences. The model lacks the granularity needed to distinguish them, as it does not account for the cognitive appraisals, situational contexts, or behavioral impulses that define an emotion (Pei et al. 2024; Gunes and Pantic 2010).

Both of these dominant theories, while useful, set up a framework where emotion is treated as a separate phenomenon to be categorized or plotted. They focus on identifying *what* an emotion is, but they are ill-equipped to explain *how* it arises as part of a larger, integrated psychological process. This limitation provides the crucial opening for a more holistic and processual framework, such as the one offered by Yogācāra Buddhism.

3. Yogācāra: Extending the Existing Theories of Emotion

Current affective computing faces challenges due to its reliance on theories that treat emotions as discrete categories and static experience, often separating them from other mental experience, such as cognition. Regarding these limitations, an alternative framework with insights from Yogācāra can offer a more sophisticated and integrated perspective.

Presenting Yogācāra Buddhism as a case study, this paper participates in a growing body of interdisciplinary research that has identified productive linkages between Buddhist theories of mind and AI. At the broadest level, Hershock (2021) argued in his book-length study *Buddhism and Intelligent Technology* that Buddhist philosophical frameworks, centered on interdependence, consciousness, and ethical cultivation, provide resources for a more humane orientation to AI development. His recent paper also offers a Buddhist reflection on theorizing consciousness to engage AI technology (Hershock 2025). More specifically, Lou (2017) demonstrated that the mechanics of Buddhist consciousness can be computationally modeled, implementing key components of the Pali Abhidhamma mind-system, including consciousness, mental states, and sense-based cognitive processes, as a working computer program in the Python programming language, arguing that Buddhist psychology offers a productive alternative way to model human cognition for AI purposes. Brody (2020) offers an AI-based exploration of Dharmakīrti’s (7th c.) Buddhist account of perception and concept formation, showing how AI neural network architectures fit well with this classical Buddhist theory of mind. Long et al. (2023) have integrated Daniel Goleman’s theory of emotional intelligence with Theravāda Buddhist psychological doctrine to propose a model of Buddhist emotional intelligence management, demonstrating the potential for Buddhist concepts of mind to enrich contemporary frameworks for emotional

understanding and regulation. Ismailov (2025) has specifically brought Yogācāra philosophy into dialog with questions of AI subjectivity, proposing a graded framework for assessing the ontological status of AI that draws on Yogācāra's taxonomy of mental construction (*vikalpa*). This current paper contributes to this conversation by extending it specifically to affective computing and by grounding the contribution in the Yogācāra school's systematic classification of Mental Factors. As a major school of Mahāyāna Buddhism, Yogācāra's process-oriented theory of mind frames emotions as emergent interactions among different Mental Factors. Instead of isolating emotion as a separate faculty, it foregrounds the intricate processes where affective experience is entangled with other mental functions, such as cognition and volition. In particular, its *holistic* and *processual* view of affective experience can address the core limitations of the categorical and dimensional models currently used in AI. What distinguishes this paper's approach is thus not the novelty of connecting Buddhism and AI as such, but the specificity and operationalizability of the connection: it moves from Yogācāra doctrine to a concrete, structured annotation framework that can be implemented in existing affective computing pipelines.

It is also worth articulating that there exists a structural relationship between Yogācāra and affective computing that makes the current case study possible. Despite the obvious differences in context and purpose, both Yogācāra and affective computing, at their methodological core, are systems for the functional analysis of mental experience. Both share the assumption that mental experience can be analyzed; both proceed by decomposing the seemingly unified stream of experience into components, identifying the role each plays, and modeling their relationships and interactions. As we will see in the following sections, Yogācāra adopts the analysis of *dharma*s (Mental Factors) to understand how the mind works, while affective computing uses annotation frameworks to model emotional experience into measurable dimensions (Valence and Arousal) and/or discrete categories so that machines can process.

Furthermore, in Yogācāra, all experience, including emotional experience, is constituted through the mind's *representational* activity, i.e., what we take to be external objects is always already mediated by the mind's own constructive processes.¹ This view also has a structural parallel in the premise of affective computing: AI systems do not process emotions directly but process *representations* of emotions, such as linguistic tokens and facial feature vectors. The gap between representation and underlying reality, which Yogācāra takes as its central philosophical problem and which drove its meticulous analysis of representational processes, parallels the gap that affective computing engineers must navigate when they design systems that infer human emotional states from indirect signals. Yogācāra's sophisticated account of how representations arise, how they are shaped by prior conditioning, and how they can become accurate or distorted can thus be studied as a conceptual resource for illuminating computing structures. As argued by Hershock (2020, 2025), Buddhist conceptual resources, particularly those addressing attention, conditionality, and the construction of experience, can provide useful tools for thinking through the ethical and epistemic problems of artificial intelligence.

3.1. A Holistic View: The Interconnectedness of Mental Experience

Modern Western psychology has conventionally divided the mind into three different domains: the cognitive (thoughts), the conative (volitions), and the affective (emotions) (Hilgard 1980; Bagozzi 1992; Mayer et al. 1997). This division, while useful, has led to the common view of emotion as a distinct realm of experience separate from other domains. Affective computing inherits this perspective, treating “emotional intelligence” as a separate component to be added to a computer's “cognitive intelligence”.

Yogācāra offers a fundamentally different structure for understanding and classifying psychological phenomena. To illustrate this, this paper utilizes the taxonomy presented in the *Mahāyāna Treatise on the Hundred Dharmas Illuminating the Gate* (大乘百法明門論). This treatise is usually attributed to the fourth-century Indian Buddhist philosopher Vasubandhu and translated by the Chinese monk Xuanzang (玄奘 602–664 CE), although there is also some suspicion regarding its authorship and origin (Zhang 2018). The treatise itself can be seen as a concise primer to distill the massive, encyclopedic psychological models of early Yogācāra (such as the *Yogācārabhūmi-śāstra* 瑜伽師地論) into a highly systematic, accessible list of one hundred Mental Factors.

A further note is needed to contextualize this choice of text and its underlying philosophy. Vasubandhu and the Yogācāra school are famous for the doctrine of *viññapti-mātra* (consciousness-only or representation-only). This has often been categorized as a form of ontological idealism (Chatterjee 1987), while some recent scholarship argues that Yogācāra can be better understood as a sophisticated form of phenomenology (Lusthaus 2003) or as a tradition that can meaningfully engage in comparative study with phenomenology (Li 2022).²

For the purposes of affective computing, we do not need to adopt Yogācāra as an ontological claim about the truth of existence. Rather, we adopt it as a rigorous phenomenological model of human subjective experience. The *Mahāyāna Treatise on the Hundred Dharmas* is chosen specifically because it provides a highly structured and systematic view of the mind and has been considered authoritative in the East Asian Buddhist tradition. In this way, it presents the full Yogācāra classification of Mental Factors in a form that is readily mappable onto a computational framework.

This conciseness and systematic completeness make the text a practical option for the purpose of the paper. In computer science, building AI requires clear conceptual modeling, i.e., translating messy human realities into formal, logical architectures. This treatise serves as an ideal bridge for this interdisciplinary endeavor, offering a ready-made, granular taxonomy of psychological processes.

While the full list of 100 *dharmas* in the treatise may seem extensive, the crucial insight for affective computing lies in how they function together. One especially relevant point here is that the framework identifies five “ever-present” (*sarvatraga* 遍行) Mental Factors that accompany every single moment of experience: Attention (*manaskāra* 作意, numbered 9 in Table 1 below), Sensory Contact (*sparsa* 觸, 10), Feeling (*vedanā* 受, 11), Conception/Associative Thinking (*saṃjñā* 想, 12), and Volition (*cetanā* 思, 13). Before turning to their implications for affective computing, it is important to clarify the scope of the broader category to which these five belong: the Mental Factors (*caitta* 心所有法). The full list of fifty-one *caitta* in the Yogācāra system is considerably broader than what contemporary Western psychology would recognize as “emotion”. Many of these factors, such as Faith (*śraddhā* 信, 19) and Vigor/Diligence (*vīrya* 精進, 20) from the wholesome group, can be seen as motivational or dispositional in character rather than affective in the narrower emotional sense. Even among the five ever-present factors, it is primarily *vedanā* (受), the dimension of feeling as pleasant, unpleasant, or neutral, that mostly directly corresponds to what we typically mean by “affective tone” in modern psychological discourse. The other four ever-present factors are attentional, sensorial, cognitive, or conative in nature.

Table 1. The classification system of 100 dharmas in the *Mahāyāna Treatise on the Hundred Dharmas Illuminating the Gate* ¹.

I. Mind Dharmas (<i>citta</i> 心法)	1. Visual consciousness (<i>cakṣur-vijñāna</i> 眼識)		6. Empiric consciousness (<i>mano-vijñāna</i> 意識)
	2. Auditory consciousness (<i>śrotra-vijñāna</i> 耳識)		7. Self-grasping consciousness (<i>manas</i> 末那識)
	3. Olfactory consciousness (<i>ghrāṇa-vijñāna</i> 鼻識)		8. Storehouse consciousness (<i>ālaya-vijñāna</i> 阿賴耶識)
	4. Gustatory consciousness (<i>jihvā-vijñāna</i> 舌識)		
	5. Tactile/kinetic consciousness (<i>kāya-vijñāna</i> 身識)		
II. Mental Factors (<i>caitta</i> 心所有法)	Ever-present (<i>sarvatraga</i> 遍行)	9. Attention (<i>manaskāra</i> 作意) 10. Sensory contact (<i>sparsa</i> 觸) 11. Feeling (<i>vedanā</i> 受)	12. Conception/associative thinking (<i>saṃjñā</i> 想) 13. Volition (<i>cetanā</i> 思)
	Specific (<i>vinīyata</i> 別境)	14. Desire (<i>chanda</i> 欲) 15. Confident resolve (<i>adhimokṣa</i> 勝解) 16. Memory/mindfulness (<i>smṛti</i> 念)	17. Meditative concentration (<i>samādhi</i> 定) 18. Wisdom/discernment (<i>prajñā</i> 慧)
	Wholesome (<i>kuśala</i> 善)	19. Faith/trust (<i>śraddhā</i> 信) 20. Vigor/diligence (<i>vīrya</i> 精進) 21. [Inner] shame (<i>hrī</i> 慚) 22. Embarrassment (<i>apatrāpya</i> 愧) 23. Lack of greed (<i>alobha</i> 無貪) 24. Lack of hatred (<i>adveṣa</i> 無瞋)	25. Lack of delusion (<i>amoha</i> 無癡) 26. Serenity (<i>praśrabdhi</i> 輕安) 27. Carefulness (<i>apramāda</i> 不放逸) 28. Equanimity (<i>upekṣā</i> 行捨) 29. Non-harming (<i>ahiṃsā</i> 不害)
	Root afflictions (<i>kleśa</i> 煩惱)	30. Greed (<i>rāga</i> 貪) 31. Aversion (<i>pratigha</i> 瞋) 32. Arrogance (<i>māna</i> 慢)	33. Ignorance (<i>mūḍhi</i> 無明) 34. Doubt (<i>vicikitsā</i> 疑) 35. Wrong view (<i>dṛṣṭi</i> 不正見)
	Secondary afflictions (<i>upakleśa</i> 隨煩惱)	36. Anger (<i>krodha</i> 忿) 37. Enmity (<i>upanāha</i> 恨) 38. [Verbal] maliciousness (<i>pradāśa</i> 惱) 39. Resist recognizing own faults (<i>mrakṣa</i> 覆) 40. Deceit (<i>māyā</i> 誑) 41. Guile (<i>śāṭhya</i> 諂) 42. Conceit (<i>māda</i> 憍) 43. Harmfulness (<i>vihiṃsā</i> 害) 44. Envy (<i>īrṣyā</i> 嫉) 45. Selfishness (<i>mātsarya</i> 慳)	46. Shamelessness (<i>āhṛīkya</i> 無慚) 47. Non-embarrassment (<i>anapatrāpya</i> 無愧) 48. Lack of faith (<i>āśraddhya</i> 不信) 49. Lethargic negligence (<i>kausīdya</i> 懈怠) 50. Carelessness/heedlessness (<i>pramāda</i> 放逸) 51. Mental foggiess (<i>styāna</i> 惛沈) 52. Restlessness (<i>auddhatya</i> 掉舉) 53. Forgetfulness (<i>muṣitasmr̥titā</i> 失念) 54. Lack of [self-]awareness (<i>asamprajanya</i> 不正知) 55. Distraction (<i>vikṣepa</i> 散亂)
	Indeterminate (<i>aniyata</i> 不定)	56. Torpor (<i>middha</i> 睡眠) 57. Remorse (<i>kaukr̥tya</i> 惡作)	58. Initial mental application (<i>vitarka</i> 尋) 59. [Subsequent] discursive thought (<i>vicāra</i> 伺)
III. Form (<i>rūpa</i> 色法)	60. Eye (<i>cakṣus</i> 眼)	64. Body (<i>kaya</i> 身)	68. Taste (<i>rasa</i> 味)
	61. Ear (<i>śrotra</i> 耳)	65. [Visible] form (<i>rūpa</i> 色)	69. Touch (<i>spraṣṭavya</i> 觸)
	62. Nose (<i>ghrāṇa</i> 鼻)	66. Sound (<i>śabda</i> 聲)	70. ‘Formal’ thought-objects (<i>dharmāyatana-paryāpanna-rūpa</i> 法處所攝色)
	63. Tongue (<i>jihvā</i> 舌)	67. Smell (<i>gandha</i> 香)	

Table 1. Cont.

IV. Embodied-conditioning not directly [perceived] by <i>Citta</i> (<i>citta-viprayukta-saṃskāra-dharma</i> 心不相應行法)	71. [Karmic] accrual (<i>prāpti</i> 得)	83. Continuity/abiding (<i>stṭhiti</i> 住)
	72. Life-force (<i>jīvitendriya</i> 命根)	84. Impermanence (<i>anityatā</i> 無常)
	73. Commonalities by species (<i>nikāya-sabhāgatā</i> 衆同分)	85. Systematic operation (<i>pravṛtti</i> 流轉)
	74. Differentiation of species (<i>viśabhāga</i> 異生性)	86. Determinant [karmic] differences (<i>pratiniyama</i> 定異)
	75. Attainment of thoughtlessness (<i>asaṃjñi-samāpatti</i> 無想定)	87. Unifying (<i>yoga</i> 相應)
	76. Attainment of cessation (<i>nirodha-samāpatti</i> 滅盡定)	88. Speed (<i>java</i> 勢速)
	77. [Realm of] thoughtless [beings] (<i>āsaṃjñika</i> 無想報)	89. Seriality (<i>anukrama</i> 次第)
	78. ‘Name’ body (<i>nāma-kāya</i> 名身)	90. Area (<i>deśa</i> 方)
	79. ‘Predicate’ body (<i>pada-kāya</i> 句身)	91. Time (<i>kāla</i> 時)
	80. ‘Utterance’ body (<i>vyāñjana-kāya</i> 文身)	92. Number/calculation (<i>saṃkhyā</i> 數)
V. Unconditioned dharma (<i>asaṃskṛta-dharma</i> 無爲法)	81. Birth/arising (<i>jāti</i> 生)	93. Synthesis (<i>sāmagrī</i> 和合性)
	82. Aging/decaying (<i>jarā</i> 老)	94. Otherwiseness (<i>anyathātva</i> 不和合性)
	95. Spatiality (<i>ākāśa</i> 虛空無爲)	
	96. Cessation through understanding (<i>pratisaṃkhyā-nirodha</i> 擇滅無爲)	
	97. Cessation without understanding (<i>aprasaṃkhyā-nirodha</i> 非擇滅無爲)	
	98. ‘Motionless’ cessation (<i>āniñjya</i> 不動滅無爲)	
	99. Cessation of associative thinking and pleasure/Pain (<i>saṃjñā-vedayita-nirodha</i> 想受滅無爲)	
	100. Suchness/ipseity (<i>tathata</i> 眞如無爲)	

¹ The table is constructed based upon the *Mahāyāna Treatise on the Hundred Dharmas Illuminating the Gate* (大乘百法明門論, CBETA 2025.R1, T31, no. 1614). The English translation consulted the appendix on one hundred dharmas in Lusthaus (2003)’s *Buddhist Phenomenology*, with some modifications by the author. Sanskrit and Chinese for each of the Mental Factors are provided in the brackets.

However, this distinction is precisely what makes the Yogācāra framework valuable for affective computing, as it does not treat affect as an isolated faculty. Instead, *caitta* intends to encompass the range of mental acts and orientations that together constitute any lived experience, including its affective coloring. It is a richer and more holistic psychological vocabulary than the emotion-only categories inherited by current AI systems. This paper’s proposed framework strategically draws on the five *sarvatraga* factors not because they are all emotions, but because they represent the irreducible architectural dimensions of every moment of conscious experience, including emotional ones. As Li (2022, pp. 161–64; 2023, p. 703, footnote 12) explains in detail in her analysis of *vedanā* in Yogācāra, the five ever-present Mental Factors co-arise when consciousness functions. The composite of *vedanā*, *sparkśa*, *manaskāra*, *saṃjñā*, and *cetanā* together constitutes precisely what affective computing needs to more fully capture about any emotional experience. Emotion, from a Yogācāra perspective, is never *vedanā* alone: it is always a co-arising event whose full profile requires all five ever-present dimensions for description.

In other words, an experience, including those labeled as “affective or emotional experience” in modern psychology, can never be isolated as being just a feeling: it is always already perceptual, attentive, conceptual, and volitional simultaneously. Yogācāra thus implies that no mental event is ever “purely” emotional or “purely” cognitive, but rather a composite event containing cognitive, conative, and affective aspects simultaneously. This resonates with recent findings in psychology and neuroscience that reveal the deep cognitive roles of emotion (Brosch et al. 2013; Izard 2009). This integrated perspective can provide a powerful foundation for building AI that can grasp the true complexity of human experience.

3.2. A Processual View: How an Affective Experience Unfolds

Beyond offering a holistic map, the Yogācāra system provides a dynamic, moment-to-moment account of how an affective experience comes about.³ This process can be understood in three stages:

1. **The Unconscious Foundation:** Experience does not arise from a blank slate. It is constantly influenced by two deep, unconscious layers of the mind. The first is the **storehouse consciousness** (*ālayavijñāna* 阿賴耶識, 8), which serves as a reservoir for all past imprints and karmic seeds. It is the continuous stream that provides the raw material for our conscious moments. The second is the **self-grasping consciousness** (*manas* 末那識, 7), an afflicted mentality that perpetually grasps at the storehouse consciousness and conceives of it as a permanent “I” or “self”. This underlying self-grasping constantly colors our perception and provides a basis for afflictive emotions to arise as we try to protect and enrich this constructed identity. At this foundational level, the affective dimension is already latently present: the karmic seeds stored in the storehouse consciousness include the habitual tendencies toward specific feeling-tones (*vedanā* 受, 11), i.e., the accumulated dispositions to find certain objects pleasant, others unpleasant, and still others neutral. These dispositional feeling tendencies are not yet conscious emotions, but they pre-configure the affective texture of any experience that will arise.⁴
2. **The Arising of Conscious Experience:** A conscious experience emerges through an interaction known as “contact” (*sparśa* 觸, 10). This is the simultaneous coming together of three elements: a sense faculty (e.g., the eye), a sense object (e.g., a shape), and the corresponding sensory consciousness (e.g., visual consciousness). The moment these three connect, a conscious event occurs. Crucially for affective experience, this contact immediately gives rise to *vedanā* (the feeling-tone of pleasant, unpleasant, or neutral), which forms the affective core of the arising moment. *Vedanā* does not arise in isolation. It is simultaneously accompanied by the other four ever-present Mental Factors, each of which modulates the affective tone. Attention determines whether the feeling-tone is amplified or diffused; conception frames the feeling-tone within a meaningful narrative; and volition channels it into a behavioral tendency. Additionally, the event may be colored by other *caitta* from Table 1, such as “Faith” (19) or “Anger” (31), which further enrich or intensify the affective quality of the experience. As the Buddhist scholar Xuanzang described it, consciousness sketches the outline of a mental image, while the various Mental Factors add the colors (cf. Li 2022, p. 161). *Vedanā* supplies the affective hue, while the cognitive, attentional, and volitional factors determine its saturation, direction, and depth.
3. **Reinforcing the Cycle:** An affective experience does not simply vanish. As it passes, the feeling-tone it carried, together with the cognitive frames, attentional habits, and volitional tendencies that accompany it, leave new imprints in the storehouse consciousness. This reinforces existing affective habits and behavioral predispositions, increasing the likelihood that similar feeling-tones and the patterns that modulate them will arise again in the future.

Thus, from a Yogācāra perspective, an “emotion” is not a static entity, but a complex and dynamic process rooted in the interplay among deep-seated unconscious tendencies, conscious processing, and sensory input. The affective core is present at every stage: latently in the dispositional seeds of the storehouse consciousness, actualized through contact in the arising moment, and preserved as imprints in the reinforcing cycle. The cognitive, volitional, and attentional factors modulate, frame, and sustain it. This model thus provides a framework for understanding not just *what* an affective experience is, but also *how* and *why* it arises and perpetuates itself.

Crucially, Yogācāra insists that Mental Factors (*caitta* 心所有法) are never independent: they always arise in dependence upon and in coordination with the primary Mind Dharmas (*citta* 心法). As described above, consciousness sketches the outline of a mental image, while the various *caitta* add the colors. This *citta-caitta* interdependence holds an important implication for the computational model proposed in this paper. The five-aspect framework developed in Section 4.2 below operates at the level of *caitta*: it captures the coloring—the feeling-tone, somatic texture, attentional stance, conceptual frame, and volitional impulse—that accompanies any affective experience. However, this coloring always presupposes an underlying substrate of consciousness. In the computational model, this substrate corresponds to the AI system itself—its base processing architecture and accumulated memory—which provides the continuous “canvas” upon which each moment of affect-laden exchange is rendered. More specifically, as the three-tier model elaborated in Section 4.3 below shows, the Tier 1 Seed Disposition layer, which stores the longitudinal history of a user’s interactions, functions as a computational analog of the storehouse consciousness, the ongoing substrate that makes continuous experience possible. Li (2023, p. 704) elaborates on precisely this relationship between *citta* and *caitta* in her analysis of joy in East Asian Yogācāra texts: “Yogācārins consider the experience of a sentient being to be coalesced through the joint effort of *citta* and *caitta*.” The proposed framework, taken as a whole, attempts to incorporate this inseparable joint effort by modeling both the affective dimensions (*caitta*) and the conscious substrate (*citta*) within which they arise.

4. A Practical Case: A Yogācāra Framework for Sentiment Analysis

While the Yogācāra system of dharmas was originally developed for a soteriological purpose of guiding Buddhist practitioners toward enlightenment, its underlying structure can offer a powerful and practical framework for affective computing. We do not need to adopt the entire list of *dharmas* to benefit from its core insight, the structure of experience itself. This section will demonstrate how the principles of Yogācāra can be applied to address the limitations in sentiment analysis, a key subfield of affective computing.

4.1. The Limits of Current Affective Lexicons

Sentiment analysis aims to identify and categorize emotional expressions in texts. To do this, AI models are trained on vast datasets, or “affective lexicons”, where words are pre-annotated with emotional values. The theories of emotion underpinning the annotation system fundamentally determine how AI understands and measures human sentiment.

A typical example is the Dalian University of Technology Chinese Affective Lexicon (大连理工大学情感词汇本体库 <https://ir.dlut.edu.cn/info/1013/1142.htm>, accessed on 1 June 2025), a widely used database of about 30,000 Chinese words annotated for sentiment analysis. Although intended to facilitate the understanding of Chinese emotional expressions, this lexicon is built upon the two most popular modern Western theories discussed earlier: the categorical Basic Emotions Theory and the dimensional Valence–Arousal model. It identifies seven main categories of emotion, including Joy (乐), Like (好), Anger (怒), Sadness (哀), Fear (惧), Disgust (恶), and Surprise (惊), a list based on Paul Ekman’s work with the addition of “Like”. Under each main category, there are a few subcategories. For instance, under “Joy”, there are two subcategories of “Happiness (快乐)” and “Peace (安心)”, while under “Sadness”, there are four subcategories of “Grief (悲伤)”, “Disappointment (失望)”, “Guilt (疚)”, and “Longing (思)”. Any emotion word can thus be annotated as belonging to one or more main categories and subcategories and given values for its polarity (positive/negative) and intensity (Xu et al. 2008, p. 181).

To illustrate how an emotional word is annotated for computing, the developers provided a specific example of the Chinese word “惊喜” (*jingxi*, pleasantly surprised):

<num>APA00032</num>
 <lex>惊喜</lex>
 <ccat>a</ccat>
 <eng>pleasantly surprised</eng>
 <emotion>PA</emotion>
 <intensity>7,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,5</intensity>
 <polarity>1</polarity>
 <syn></syn>
 <emotion_class>A</emotion_class>
 <standard>0</standard>

In this annotation, “PA” is the code for “Happiness [Subcategory 1]”, which is regarded as the main subcategory to which the word belongs. The annotation on “intensity” has 20 values, each corresponding to the word’s intensity level regarding one of the 20 subcategories in Table 2. As this word conveys a relatively high level of happiness and a mid-level of surprise, the annotators gave it an intensity score of “7” under “Joy → Happiness [Subcategory 1]” and an intensity score of “5” under “Surprise → Surprise [Subcategory 20]”, while assigning 0 for all other subcategories as the word does not convey those types of emotion. The value “1” under “polarity” indicates that this word has a positive valence. In this way, a particular emotion word become measurable and can be used for affective computing purpose (Xu et al. 2008, p. 182).

Table 2. Classification of words in the Dalian University of Technology Chinese Affective Lexicon database ¹.

Basic Emotion Category	Subcategory	Sample Chinese Words
乐 (Joy)	1. 快乐 (Happiness)	喜悦、欢喜、笑咪咪、欢天喜地
	2. 安心 (Peace)	踏实、宽心、定心丸、问心无愧
好 (Like)	3. 尊敬 (Respect)	恭敬、敬爱、毕恭毕敬、肃然起敬
	4. 赞扬 (Praise)	英俊、优秀、通情达理、实事求是
	5. 相信 (Trust)	信任、信赖、可靠、毋庸置疑
	6. 喜爱 (Fondness)	倾慕、宝贝、一见钟情、爱不释手
怒 (Anger)	7. 愤怒 (Rage)	气愤、恼火、大发雷霆、七窍生烟
哀 (Sadness)	8. 悲伤 (Grief)	忧伤、悲苦、心如刀割、悲痛欲绝
	9. 失望 (Disappointment)	憾事、绝望、灰心丧气、心灰意冷
	10. 疚 (Guilt)	内疚、忏悔、过意不去、问心有愧
	11. 思 (Longing)	相思、思念、牵肠挂肚、朝思暮想
惧 (Fear)	12. 慌 (Panic)	慌张、心慌、不知所措、手忙脚乱
	13. 恐惧 (Fear)	胆怯、害怕、担惊受怕、胆颤心惊
	14. 羞 (Shame/Shy)	害羞、害臊、面红耳赤、无地自容
恶 (Disgust)	15. 烦闷 (Upset)	憋闷、烦躁、心烦意乱、自寻烦恼
	16. 憎恶 (Loathing)	反感、可耻、恨之入骨、深恶痛绝
	17. 贬责 (Condemnation)	呆板、虚荣、杂乱无章、心狠手辣
	18. 妒忌 (Jealousy)	眼红、吃醋、醋坛子、嫉贤妒能
	19. 怀疑 (Doubt)	多心、生疑、将信将疑、疑神疑鬼
惊 (Surprise)	20. 惊奇 (Surprise)	奇怪、奇迹、大吃一惊、瞠目结舌

¹ Based on Table 1 in Xu et al. (2008). English translation by the author.

We can see here that this annotation system combines the categorical and dimensional theories of emotion by both registering a word under one or more subcategories of emotion and giving it values for intensity and polarity for quantification. Still, the system inherits the fundamental limitations of its theoretical foundation: it views emotion as a separate realm of experience, assigning a static label divorced from the cognitive and conative context that gives it meaning and failing to capture its connection to thought, bodily sensation, and behavioral urges.

As a result, this approach can identify what an emotion is, but it cannot provide a fuller account of how it is experienced. To build more sophisticated emotionally intelligent AI systems, we need a system that can capture this richer, multi-aspect reality.

4.2. A New Annotation Framework with Yogācāra Insights

The principles of Yogācāra provide a blueprint for such a system. By focusing on the five “ever-present” Mental Factors that accompany any experience, as discussed in Section 3.1, we can develop a framework that analyzes an emotional expression from five corresponding aspects. This compositional approach expands the identifiable emotional repertoire and adds crucial granularity to the machine’s interpretation of human experience.

Figure 1 is a proposed framework where each aspect corresponds to one of the five ever-present Mental Factors:

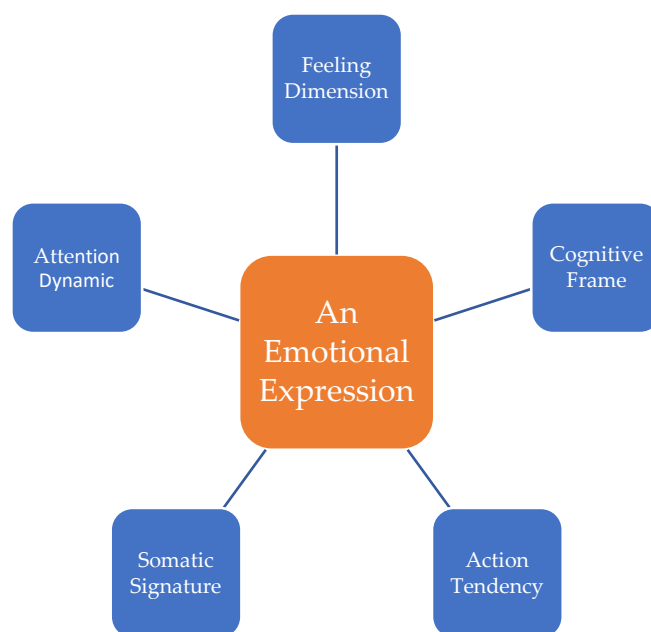


Figure 1. Yogācāra-inspired multi-aspect framework for affective lexicon annotation.

1. **Somatic Signature (from “contact”-*sparsā* 觸):** In Yogācāra, “contact” is the meeting of sense, object, and consciousness. This resonates with recent research in psychology emphasizing that cognition is grounded in the somatic domain (Hartley 2004; Varela et al. 1991). This aspect captures the bodily sensations associated with an emotion. Instead of a vague label, we can annotate qualities like “Warmth”, “Heaviness”, and “Agitation”, addressing the physical feelings commonly associated with an emotion.
2. **Attention Dynamic (from “attention”-*manaskāra* 作意):** “Attention” is the function of directing consciousness toward an object. During an emotional experience, our attention changes. This aspect captures whether attention becomes *fixed* or *diluted*, and whether its orientation is primarily *internal* (focused on self) or *external* (focused on the outer world).

3. **Feeling Dimensions (from “feeling”-*vedanā* 受):** This aspect corresponds directly to the Mental Factor of “feeling” and can incorporate the existing Valence–Arousal model. It captures the core affective tone of *valence polarity* (positive/negative/neutral) and *arousal intensity* (low to high). This allows the framework to retain the useful components of current dimensional models.
4. **Cognitive Frame (from conception-*saṃjñā* 想):** “Conception” is the Mental Factor that grasps an object’s features and forms concepts. This aspect of the framework identifies the underlying cognitive meaning associated with an emotion. We can include different cognitive frames such as *Gain*, *Loss*, *Threat*, *Injustice*, *Connection*, or *Growth*. This provides crucial context that is lost in simple emotion labels.
5. **Action Urge (from volition-*cetanā* 思):** “Volition” is the Mental Factor that directs the mind toward an activity. This aspect annotates the potential behavioral impulse associated with an emotion. Common tendencies can include *Approach*, *Avoid*, *Attack*, or *Freeze*. This captures the dynamic and motivational nature of an affective state.

This multi-layered approach, summarized in Table 3 below, moves beyond static labels to create a detailed, process-oriented snapshot of an affective experience. It incorporates the strengths of existing systems while providing a far richer understanding of linguistic emotional expressions, helping machines interpret human experience with greater depth and accuracy.

Table 3. Yogācāra-inspired annotation categories for emotion expressions.

Aspects of Emotion	Proposed Annotating Options	Examples
Somatic Signature (Embodied experience)	Warmth: Scaled Value (e.g., 0–9 from cold to heated) Heaviness: Scaled Value (0–9 from light to heavy) Agitation: Scaled Value (−9–0–9 from numbed to relaxed to agitated)	“Sadness” → Warmth: 2; Heaviness: 8; Agitation: −2 “Thrilled” → Warmth: 7; Heaviness: 0; Agitation: 9
Attention Dynamic (Focus and orientation)	Focus Level: Scaled Value (0–9 from diffused to concentrated) Orientation: Internal/External	“Obsessed” → Focus level: 9; Orientation: External “Selfish” → Focus level: 8; Orientation: Internal
Feeling Dimensions (Valence–arousal)	Valence/Polarity: Positive/Negative/Neutral Arousal/Intensity: Scaled Value (0–9 from low to high)	“Happiness” → Valence/polarity: Positive; Arousal/intensity: 8
Cognitive Frame (Meaning assignment)	Loss/Gain/Threat/Injustice/Connection/Growth	“Betrayal” → Injustice “Opportunity” → Gain
Action Urge (Behavioral impulse)	Approach/Avoid/Attack/Freeze	“Embrace” → Approach “Hide” → Avoid

A methodological note is needed regarding the numerical and scalar annotation values that appear in Table 3 (e.g., Warmth: 0–9, Arousal/Intensity: 0–9). These scales are not derived from Buddhist philosophy, which does not itself employ quantitative measurement of mental states. Rather, they are borrowed from the established convention of dimensional scoring in affective computing, specifically the continuous numeric conventions associated with the Valence–Arousal model (Russell 1980; Gunes and Pantic 2010) discussed in Section 2.2, as well as illustrated in the example in Section 4.1. The logic of this methodology is one of principled hybridity: the Yogācāra framework provides the structural categories, i.e., the dimensions of emotional experience that should be annotated, while the affective computing tradition provides the measurement conventions, i.e., the means of quantifying those dimensions in a way that is computationally tractable and commensurable with existing systems. The two fields thus play complementary roles in the framework. As a preliminary proposal of a framework, the specific illustrative values given in the examples (e.g., Warmth: 7, Agitation: 7 for the expression *jingxi*) are currently

heuristic and indicative only. In actual implementation, such values would be determined empirically by annotators, drawing on both the psychological literature on somatic and affective experience and Buddhist descriptions of the bodily correlates of mental states.

Let us revisit the example of “惊喜” (*jingxi*, pleasantly surprised), which is annotated in the Dalian University of Technology Chinese Affective Lexicon as “Happiness/Surprise”. The new Yogācāra-informed framework can provide a much richer profile for this word:

- **Somatic Signature:** Warmth: 7; Heaviness: 0; Agitation: 7.
- **Attention Dynamic:** Focus Level: 7; Orientation: *External*.
- **Feeling Dimension:** Valence/Polarity: *Positive*; Arousal/Intensity: 8.
- **Cognitive Frame:** Gain (implying an unexpected positive outcome).
- **Action Urge:** Approach (a tendency to engage).

This multi-aspect approach thus moves beyond simple emotion labels to create a detailed snapshot of an affective experience. It incorporates the strengths of existing systems while providing a far richer understanding of an emotion to help machines interpret human expressions with greater depth and accuracy.

4.3. “Wholesome” Application: Emotion Navigation Support for Personal Chatbots

The Yogācāra-inspired framework outlined above offers a more powerful tool for understanding human sentiment than current models. This power, however, also presents an ethical crossroads. The same technology that can identify a user’s cognitive frames, somatic states, and action urges could also be used for highly effective surveillance, political monitoring, or manipulative marketing. Such purposes, which often facilitate the control of people, are fundamentally at odds with the Buddhist pursuit of liberation from suffering.

This ethical challenge demands a more “wholesome” application of the framework: one designed not for external control but for personal empowerment. For example, a personal AI chatbot equipped with a sophisticated sentiment detection module could serve as a private tool to help users navigate their own emotional experiences, aligning with the spiritual goal of self-understanding.

To achieve this, we need to move beyond analyzing an emotion in isolation and embrace the full processual nature of the Yogācāra model. As discussed, Yogācāra outlines the entire dynamic life cycle of our affective experience: from the latent “seeds” in the unconscious, to their manifestation as a conscious emotion, to the new imprints they leave behind. Our five-aspect framework captures the emotion in the present moment; now, we can situate it within a larger system that connects past dispositions with future possibilities. This can be developed as a three-tier system for a personal chatbot (Figure 2):

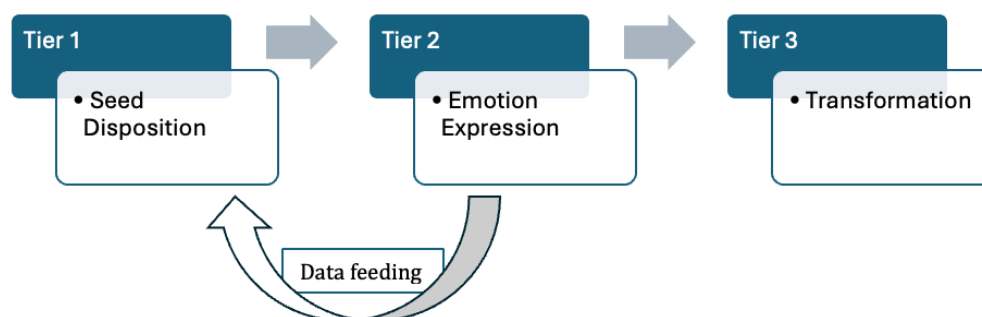


Figure 2. A Yogācāra-informed three-tier processual model for emotion detection and navigation.

Tier 1: Seed Disposition (Understanding the Past)

This tier, based on the storehouse consciousness, acts as the system’s memory by storing the history of an individual user’s interactions. By applying the five-aspect analysis

over time, the AI model can move beyond single data points to detect meaningful patterns. For example, it could identify a user's affective baseline and notice recurring connections. Does the user consistently enter a "Cognitive Frame" of "Loss" when discussing relationship? Do feelings of "Sadness" always manifest with a "Somatic Signature" of "Heaviness"? Such long-term data on a user's cognitive and affective habits are crucial for understanding the "seeds" of their experience.

Tier 2: Emotion Expression (Analyzing the Present)

This is the core five-aspect framework discussed previously. When a user expresses an emotion, the system analyzes it in real time across the dimensions of Feeling, Somatic Signature, Cognitive Frame, Action Urge, and Attention Dynamic. This provides a rich, multi-layered snapshot of the user's present moment experience.

Tier 3: Transformation (Guiding the Future)

This tier provides guidance to help the user break free from reinforcing cycles of affliction. Here, the moral taxonomy of *caitta* becomes architecturally significant. In Yogācāra, the fifty-one Mental Factors are not morally neutral; they are systematically organized according to their moral valence (see Table 1). Eleven factors are classified as "wholesome" (*kuśala* 善), ranging from Faith (*śraddhā* 信, 19) and Vigor (*vīrya* 精進, 20) to Equanimity (*upekṣā* 行捨, 28) and Non-harming (*ahimsā* 不害, 29). There are also six root afflictions (*mūlakleśa* 煩惱), including Greed (*lobha* 貪, 30), Aversion (*pratigha* 瞋, 31), and Arrogance (*māna* 慢, 32), along with twenty secondary afflictions (*upakleśa* 隨煩惱) derived from them, such as Anger (*krodha* 忿, 36), Envy (*īrṣyā* 嫉, 44), and Restlessness (*auddhatya* 掉舉, 52). In the Buddhist soteriology, this Yogācāra system of Mental Factors serves to help counteract "defilements" of practitioners (Chien 2023).

Tier 3 can draw from the wholesome Mental Factors to guide the AI's responses. For example, guided by "Wisdom" (*prajñā* 慧, 18), the chatbot could help a user gently question a rigid "Cognitive Frame" that is causing distress. Inspired by "Serenity" (*praśrabdhi* 輕安, 26) and "Equanimity" (*upekṣā* 行捨, 28), it could offer responses that help the user acknowledge and sit with a difficult feeling without being overwhelmed by it. Drawing on "Conscientiousness" (*apramāda* 不放逸, 27), it could help the user recognize the emergence of a habitual negative pattern and choose a different response. Tier 1 and Tier 2 together can detect the presence of afflictive Mental Factors in the user's affective profile. For instance, a pattern in which the user's Cognitive Frame is dominated by "Threat", the Action Urge by "Attack", and the Feeling Dimension by high-arousal negative states may signal the habitual activation of Aversion (*dveṣa* 瞋) or Restlessness (*auddhatya* 掉舉). A Yogācāra-informed system treats these as recognizable afflictions (*kleśa* 煩惱), i.e., states that distort perception and perpetuate suffering, and actively works to introduce wholesome counterweights. This moves AI design toward a stance of ethical commitment of emotional navigation grounded in the user's own longitudinal emotional history.

It should also be noted that the proposed AI model does not require a strict one-to-one adoption of the traditional Yogācāra taxonomy presented in Table 2. Through ongoing dialog with psychology and cognitive science, the Transformative Tier 3 can be adapted to incorporate selected Yogācāra Mental Factors alongside those that are more attuned to contemporary lived experience. Ultimately, the value of this model lies in its core Buddhist insight: every affective experience is both a seed and an outcome. It is this processual understanding, rather than the exact labeling of Mental Factors, that allows AI to navigate users through their emotions, empowering them to recognize and dismantle afflictive habitual patterns.

This model thus stands in stark contrast to many current "emotional companion" chatbots on the market. A typical example is Replika (replika.com), which is designed to mimic

the user (hence its name from the word “replica”) to form an emotional connection. While it might be well-intentioned, such mirroring approach risks trapping users in feedback loops of their own afflictions, validating and strengthening the very patterns that may be causing their suffering. A Yogācāra-informed chatbot, instead, would not simply reflect users’ patterns back to them. Instead, It would gently illuminate those patterns and offer pathways for transformation, moving AI from a simple tool of analysis to a partner in personal growth.

Moreover, Yogācāra philosophy may illuminate another foundational ethical issue: the intersubjective. In Yogācāra, experience is never solipsistic. Every moment of conscious experience arises dependently within a relational field of conditions that includes other sentient beings. The storehouse consciousness does not store only private karmic seeds; rather, it arises within and contributes to a shared world constituted through collective karmic streams (Waldron 2023, pp. 224–28, 264–67). This intersubjective dimension acquires particular significance when we consider the ontological status of AI itself. In her Yogācāra-informed analysis of AI subjectivity, Ismailov (2025, pp. 40–55) proposes a graded framework of subjectivity in which entities are positioned along a spectrum according to their degree of *vikalpa* (constructive mental activity) and karmic efficacy. On her account, generative AI occupies a distinctive intermediate position: it is neither a fully sentient being capable of *vikalpa* and karmic formation nor an inert object with no constructive capacity. Rather, it engages in what she terms “*vikalpa* zero”: a form of digital construction that, while inorganic and non-sentient, generates outputs complex enough to appear mindlike to the humans who interact with it. Crucially, Ismailov argues that because AI’s inner workings are opaque rather than transparent, it functions as something closer to the “mediate objective support” of human consciousness than to a simple external object, much as other minds are not immediately graspable but only known through the projections of our own cognition. This opacity, she concludes, demands an intersubjective ethics of human–AI relations: if AI is not simply a tool but a form of subject, however attenuated, then our interactions with it cannot be solely governed by the ethical framework applied to inert objects.

A Yogācāra-informed affective AI system is uniquely positioned to take this intersubjective ethics seriously. When a user engages with a personal chatbot equipped with the framework proposed here, that engagement is not merely a one-on-one transaction between a human and a neutral instrument; it can be seen as an intersubjective encounter in which the AI’s responses become part of the conditions shaping the user’s future affective seeds, and vice versa. Li (2023, p. 704) underscores how feeling in Yogācāra arises through the joint effort of *citta* and *caitta* within a lifeworld that is always already relational. Designing AI with this relational ontology in mind means attending not only to what the AI detects and responds to in a single individual, but also to how it mediates, shapes, and potentially transforms the emotional ecologies of the communities and relationships within which its users are embedded. This represents a rich and urgent direction for future interdisciplinary research, one at the intersection of Buddhist philosophy, AI ethics, and the emerging field of relational human–computer interaction.

5. Conclusions: The Ethics of Affective Computing and the Need for Theoretical Diversity

The development of affective computing is accompanied by urgent ethical discussions. These debates rightly foreground critical issues such as the implications for user privacy, cultural bias arising from unrepresentative data, and the potential for emotional manipulation and addiction (Devillers and Cowie 2023; Cowie 2015). However, this paper argues for attention to a more foundational ethical concern that lies beneath the level of

data: the cultural bias embedded in the very theories of emotion that guide AI development, i.e., bias in the *framework* itself.

When affective computing exclusively adopts a handful of modern Western psychological models, it is not merely misinterpreting certain expressions; it is imposing a culturally specific structure on the nature of emotional experience from the ground up. As AI systems become increasingly integrated into our daily lives—acting as companions, therapists, and mediators of our social reality—they will inevitably shape our own understanding of our inner worlds. A global AI infrastructure built exclusively on a narrow set of theories risks creating a powerful feedback loop where these models do not just interpret human emotion but actively teach users how to experience and categorize their own feelings. The danger is the creation of a theoretical monopoly that could flatten the rich diversity of human affective experience into a single dominant paradigm.

To counteract this, it is imperative that we actively identify and integrate alternative theories of emotion and mind from a plurality of world traditions and cultures. This is a call to engage with different philosophical and contemplative systems not as sources of exotic data, but as inspirations for more sophisticated frameworks that can inform the fundamental theoretical architecture of AI design.

This paper offers one contribution to that collective effort: a contribution rooted specifically in the author's own field of research in East Asian Buddhism. The choice of Yogācāra, as demonstrated in Sections 3 and 4, is based on its specific structural features: its holistic integration of cognition, affect, and volition; its processual account of how experience unfolds and how affective habits are formed and reinforced; and its systematic taxonomy of Mental Factors distinguished by moral valence. These features make it possible to address the limitations of the categorical and dimensional models currently used in affective computing. What this paper aims to do is to demonstrate what a rigorous, tradition-specific engagement may look like: taking a single philosophical system seriously on its own terms, reading its primary sources carefully, and working out in detail how its conceptual resources can be translated into a practical tool for AI design. In this sense, this paper is a case study and an invitation. The author hopes that specialists in other religious and philosophical traditions will find in this paper a possibility for the kind of engagement they might undertake with their own primary sources. Each religious and/or philosophical tradition carries its own analysis and theorization of human affective and mental life, developed over centuries of reflection and construction. The AI industry, whose systems are increasingly acting as companions, therapists, and mediators of emotional experience for billions of people worldwide, urgently needs to be in dialog with this diversity of human wisdom rather than drawing from a single culturally specific source. By embracing this theoretical pluralism, we can strive to build emotionally intelligent AI that not only recognizes the many ways humans express emotion, but also honors the diverse ways we understand what it means to feel.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The author declares no conflicts of interest.

Notes

- ¹ The tenet of Yogācāra is often understood to be “representation-only”, indicating that we do not have direct access to external objects and can only access them as “representations”. However, there are philosophical implications and debates over how to understand this “representation” in Yogācāra, as well as whether this term is accurate, both of which are beyond the scope of this paper. For further discussions of the “representation-only” aspect of Yogācāra, see Waldron (2023, pp. 184–95) and Wayman (1979).
- ² See Sharf (2016) for a more detailed discussion of this strand of scholarship.
- ³ For a recent description and discussion of this process in detail, see Waldron (2023, pp. 197–232). The discussion of Yogācāra doctrines in this section also draws from texts such as the *Treatise on the Establishment of Consciousness-only* (成唯識論, CBETA 2026.R1, T31, no. 1585) and *Stages of Yogic Practice* (Yogācārabhūmi 瑜伽師地論, CBETA 2025.R3, T30, no. 1579). For an English translation of the *Treatise on the Establishment of Consciousness-only*, an authoritative Yogācāra text in East Asia, see Cook (1999); for its historical contextualization, see Lusthaus (2003, pp. 351–425).
- ⁴ For a detailed discussion of the concept of storehouse consciousness (*ālayavijñāna*) and its historical context, see Waldron (2003).

References

Primary Sources

- Dacheng baifa mingmen lun* 大乘百法明門論 [Mahāyāna Treatise on the Hundred Dharmas Illuminating the Gate]. Translated by Xuanzang (玄奘, 602–664 CE). CBETA Online. 2025.R1, T31, no. 1614.
- Cheng weishi lun* 成唯識論 [Treatise on the Establishment of Consciousness-only], 10 juans. Compiled by Kuiji (窺基, 632–682 CE). CBETA Online. 2025. T31, no. 1585.
- Yujia Shidi lun* 瑜伽師地論 (Skt. *Yogācārabhūmi-sāstra*; Eng. *Stages of Yogic Practice*). 100 juans. Translated by Xuanzang Xuanzang (玄奘, 602–664 CE). CBETA Online. 2025.R3, T30, no. 1579.

Secondary Sources

- Bagozzi, Richard P. 1992. The Self-Regulation of Attitudes, Intentions, and Behavior. *Social Psychology Quarterly (US)* 55: 178–204. [CrossRef]
- Barrett, Lisa Feldman. 2006. Are Emotions Natural Kinds? *Perspectives on Psychological Science* 1: 28–58. [CrossRef] [PubMed]
- Brody, Justin. 2020. Enaction, Convolution and Conceptualism: An AI-Based Exploration of Dharmakīrti’s Perception and Conception. *Hualin International Journal of Buddhist Studies* 3: 1–26. [CrossRef]
- Brosch, Tobias, Klaus Scherer, Didier Grandjean, and David Sander. 2013. The Impact of Emotion on Perception, Attention, Memory, and Decision-Making. *Swiss Medical Weekly* 143: w13786. [CrossRef] [PubMed]
- Chatterjee, Ashok Kumar. 1987. *The Yogācāra Idealism*. New Delhi: Motilal Banarsidass.
- Chien, Ju-En. 2023. The Mind and Mental Factors According to the Cheng Weishi Lun (成唯識論): An Approach to Buddhist Therapeutic Soteriology. Ph.D. dissertation, Ludwig-Maximilians-Universität München, Munich, Germany.
- Cook, Francis, trans. 1999. *Three Texts on Consciousness Only*. Berkeley: Numata Center for Buddhist Translation and Research.
- Cowie, Roddy. 2015. Ethical Issues in Affective Computing. In *The Oxford Handbook of Affective Computing*. Edited by Rafael A. Calvo, Sidney D’Mello, Jonathan Gratch and Arvid Kappas. Oxford: Oxford University Press.
- Devillers, Laurence, and Roddy Cowie. 2023. Ethical Considerations on Affective Computing: An Overview. *Proceedings of the IEEE* 111: 1445–58. [CrossRef]
- Ekman, Paul. 1992. An Argument for Basic Emotions. *Cognition and Emotion* 6: 169–200. [CrossRef]
- Gunes, Hatice, and Maja Pantic. 2010. Automatic, Dimensional and Continuous Emotion Recognition. *International Journal of Synthetic Emotions* 1: 68–99. [CrossRef]
- Gunes, Hatice, Björn Schuller, Maja Pantic, and Roddy Cowie. 2011. Emotion Representation, Analysis and Synthesis in Continuous Space: A Survey. Paper presented at 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG), Santa Barbara, CA, USA, March 21–25; pp. 827–34. [CrossRef]
- Hartley, Linda. 2004. *Somatic Psychology: Body, Mind and Meaning*. Hoboken: Wiley.
- Hershock, Peter D. 2020. The Intelligence Revolution and the New Great Game: A Buddhist Reflection on the Personal and Societal Predicaments of Big Data and Artificial Intelligence. *Hualin International Journal of Buddhist Studies* 3: 62–77. [CrossRef]
- Hershock, Peter D. 2021. *Buddhism and Intelligent Technology: Toward a More Humane Future*. London: Bloomsbury Academic.
- Hershock, Peter D. 2025. AI, Consciousness, and the Evolutionary Frontier: A Buddhist Reflection on Science and Human Futures. *Religions* 16: 562. [CrossRef]
- Hilgard, Ernest R. 1980. The Trilogy of Mind: Cognition, Affection, and Conation. *Journal of the History of the Behavioral Sciences* 16: 107–17. [CrossRef]

- Ismailov, Mariam. 2025. Rethinking AI Subjectivity Through Yogācāra Philosophy. Master's thesis, Leiden University, Leiden, The Netherlands. Available online: <https://studenttheses.universiteitleiden.nl/access/item%3A4280935/view> (accessed on 25 May 2026).
- ISO. 2022. *ISO/IEC 30150-1:2022 Information Technology—Affective Computing User Interface (AUI), Part 1: Model*. ICS: 35.240.20. Geneva: ISO. Available online: <https://www.iso.org/standard/78031.html> (accessed on 5 June 2025).
- ISO. 2024. *ISO/IEC TR 30150-2:2024 Information Technology—Affective Computing User Interface (AUI), Part 2: Affective Characteristics*. ICS: 35.240.20. Geneva: ISO. Available online: <https://www.iso.org/standard/82501.html> (accessed on 5 June 2025).
- Izard, Carroll E. 2009. Emotion Theory and Research: Highlights, Unanswered Questions, and Emerging Issues. *Annual Review of Psychology* 60: 1–25. [CrossRef] [PubMed]
- Li, Jingjing. 2022. *Comparing Husserl's Phenomenology and Chinese Yogācāra in a Multicultural World: A Journey Beyond Orientalism*. London: Bloomsbury Academic.
- Li, Jingjing. 2023. Joy as Contextualized Feeling: Two Contrasting Pictures of Joy in East Asian Yogācāra. *Journal of the American Academy of Religion* 91: 698–713. [CrossRef]
- Long, Ho Thi Ngu, Phramaha Somphong Guṇākaro, and Sanu Mahatthanadull. 2023. The Buddhist Emotional Intelligence Management: An Integration of Daniel Goleman's Theory and Theravāda Perspective. *Journal of International Buddhist Studies* 13: 2.
- Lou, Eric. 2017. *A Novel Computer Simulation of the Mind Using Buddhist Theories*. SSRN Scholarly Paper No. 3026291. Rochester: Social Science Research Network. [CrossRef]
- Lusthaus, Dan. 2003. *Buddhist Phenomenology: A Philosophical Investigation of Yogacara Buddhism and the Ch'eng Wei-Shih Lun*. Abingdon: Routledge.
- Mayer, John D., Heather Frasier Chabot, and Kevin M. Carlsmith. 1997. Conation, Affect, and Cognition in Personality. In *Advances in Psychology*. Amsterdam: North-Holland, vol. 124. [CrossRef]
- Pei, Guanxiong, Haiying Li, Yandi Lu, Yanlei Wang, Shizhen Hua, and Taihao Li. 2024. Affective Computing: Recent Advances, Challenges, and Future Trends. *Intelligent Computing* 3: 0076. [CrossRef]
- Picard, Rosalind W. 1997. *Affective Computing*. Cambridge: The MIT Press.
- Plamper, Jan. 2017. *The History of Emotions: An Introduction*. Translated by Keith Tribe. Oxford: Oxford University Press.
- Russell, James A. 1980. A Circumplex Model of Affect. *Journal of Personality and Social Psychology* 39: 1161–78. [CrossRef]
- Sharf, Robert H. 2016. Is Yogācāra Phenomenology? Some Evidence from the Cheng Weishi Lun. *Journal of Indian Philosophy* 44: 777–807. [CrossRef]
- Varela, Francisco J., Eleanor Rosch, and Evan Thompson. 1991. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge: The MIT Press.
- Waldron, William S. 2003. *The Buddhist Unconscious: The Ālaya-Vijñāna in the Context of Indian Buddhist Thought*. Abingdon: Routledge-Curzon.
- Waldron, William S. 2023. *Making Sense of Mind Only: Why Yogacara Buddhism Matters*. Somerville: Wisdom Publications.
- Wayman, Alex. 1979. Yogācāra and the Buddhist Logicians. *Journal of the International Association of Buddhist Studies* 2: 65–78.
- Xu, Lihong, Hongfei Lin, Yu Pan, Hui Ren, and Jianmei Chen. 2008. Qinggan Cihui Benti de Gouzao 情感词汇本体的构造 [Constructing the Affective Lexicon Ontology]. *Journal of the China Society for Scientific and Technical Information* 情报学报 27: 180–85.
- Zhang, Lei 張磊. 2018. Dacheng Baifa Mingmen Lun de Fanyi Yu Neirong Laiyuan Xinkao Dacheng 大乘百法明門論的翻譯與內容來源新考 [A New Investigation on the Origin of the Translation and Content of the Mahāyāna Treatise on the Hundred Dharmas Illuminating the Gate]. *Zongjiaoxue Yanjiu* 宗教學研究 3: 109–15.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.