

AI Companionship and the Development of Moral Agency: An Aristotelian Virtue Ethics Perspective

Hamide Aflatooni 

Department of Islamic Philosophy and Wisdom, Imam Khomeini International University, Qazvin, Iran. E-mail: platonichamide@gmail.com

Article Info

Article type:

Research Article

Article history:

Received 07 June 2026

Received in revised form

26 July 2026

Accepted 04 August 2026

Published online 25

October 2026

Keywords:

Artificial Intelligence
Companionship, Moral
Naturalism, Virtue Ethics,
Moral Sensitivity, Moral
Imagination, Ethical AI
Design.

ABSTRACT

Artificial intelligence companions are becoming increasingly integrated into everyday human life, raising important philosophical questions about their influence on moral development. Existing discussions have often focused either on the practical benefits of AI companionship or on its potential social and psychological risks. Comparatively less attention has been devoted to the developmental processes through which sustained interaction with AI may shape the formation of moral character. This article examines AI companionship through the perspective of Aristotelian virtue ethics and contemporary ethical naturalism. It argues that the ethical significance of AI companionship cannot be adequately assessed solely by reference to immediate behavioral outcomes or emotional consequences. Instead, its moral implications should be evaluated in light of the developmental conditions through which moral agency is cultivated. The article further argues that this perspective does not require viewing AI companionship as inherently detrimental. Rather, AI systems may also contribute positively to moral development when they are designed to cultivate reflective agency, encourage moral practice, and support—rather than replace—the interpersonal conditions necessary for the formation of virtue. By integrating Aristotelian virtue ethics, ethical naturalism, and contemporary discussions of AI companionship, this article offers a balanced philosophical framework for evaluating both the risks and the developmental possibilities of AI-mediated relationships.

Cite this article: Aflatooni, H. (2026). AI Companionship and the Development of Moral Agency: An Aristotelian Virtue Ethics Perspective. *Journal of Philosophical Investigations*, 20(56), 217-230. <https://doi.org/10.22034/jpiut.2026.72876.4549>



© The Author(s).

Publisher: University of Tabriz.

Intruduction

Artificial intelligence companionship has rapidly evolved from a speculative possibility into an increasingly significant dimension of contemporary social life. As AI systems become capable of sustaining emotionally responsive, personalized, and long-term interactions, philosophical attention must extend beyond questions of technological performance or functional efficiency to a more fundamental concern: how might sustained relationships with artificial companions influence the developmental conditions of human moral agency? This question cannot be adequately answered by asking whether AI companions are simply beneficial or harmful. Rather, it requires an account of how moral character itself is formed and how emerging technologies may reshape the practices through which human beings become ethical agents. (Vallor 2016).

Since this article approaches that question through the framework of Aristotelian virtue ethics and contemporary ethical naturalism, to clarify this developmental perspective, the article proposes an interconnected framework linking habituation, attention, moral perception, moral sensitivity, practical wisdom (phronesis), virtuous action, and ultimately human flourishing (eudaimonia). Within this framework, to clarify this developmental perspective, the present study distinguishes three interconnected dimensions of moral development. The first is moral sensitivity, understood as the cultivated capacity to perceive morally salient features of particular situations and to recognize reasons for ethical concern. The second is practical wisdom (phronesis), the intellectual virtue through which agents deliberate well about what ought to be done in light of those morally significant features. The third is human flourishing (eudaimonia), the realization of a distinctively human life through the sustained exercise of virtuous action. (Aristotle) These dimensions should not be understood as isolated concepts but as successive moments within a unified developmental process linking moral perception, practical judgment, and moral character.

Generally, moral sensitivity is understood as the cultivated capacity to perceive morally salient features of concrete situations, while practical wisdom enables sound deliberation and virtuous action directed toward human flourishing. AI companionship therefore acquires ethical significance insofar as it may influence the habits and patterns of attention upon which these capacities depend. Within this tradition, morality is not understood primarily as compliance with abstract rules or the calculation of consequences, but as the gradual cultivation of stable dispositions that enable individuals to perceive, judge, and act well within the concrete circumstances of everyday life. Human flourishing (eudaimonia) therefore depends not merely on isolated moral decisions but on the long-term formation of character through habituation, social participation, and practical reasoning. Ethical evaluation of AI companionship must consequently address not only its immediate behavioral effects but also its influence upon these developmental processes.

Building upon this framework, the article argues that the ethical significance of AI companionship lies neither in the technology itself nor in isolated behavioral outcomes, but in its potential influence on the developmental ecology of moral agency. If repeated interaction with AI companions contributes to the formation of habits, patterns of attention, and modes of moral perception, then its ethical implications extend far beyond questions of convenience or emotional support. At the same time, this perspective does not require viewing AI companionship as intrinsically detrimental. On the contrary, the normative question is whether AI systems are designed in ways that cultivate rather than replace the human capacities upon which moral sensitivity, practical wisdom, and flourishing depend.

By examining AI companionship through this developmental account of virtue, the article seeks to provide a more balanced philosophical framework for evaluating both its risks and its potential contributions to human moral development.

1. Moral Naturalism and the Social Nature of Human Beings

1.1 Aristotelian Foundations

The philosophical foundations of ethical naturalism can be traced to Aristotle's account of human nature, virtue, and flourishing. Unlike moral theories that ground ethical obligations in abstract rules or subjective preferences, Aristotle argues that moral evaluation is rooted in the characteristic nature (*physis*) and function (*ergon*) of human beings. In the *Nicomachean Ethics*, he maintains that every being can be evaluated according to the extent to which it fulfills its distinctive function, and that the human good consists in the excellent realization of those capacities that are unique to rational creatures (*Nicomachean Ethics* I.7, 1097b22–1098a20).

This functional conception of human nature establishes an objective basis for ethical evaluation. Human beings are not morally assessed merely according to what they happen to desire or prefer, but according to whether their ways of living enable the realization of their rational and social capacities. Consequently, moral goodness is not externally imposed upon human life but emerges from the successful development of those natural capacities that constitute human flourishing (*eudaimonia*). As Aristotle famously argues, *eudaimonia* is "activity of the soul in accordance with virtue" (*Nicomachean Ethics* I.7, 1098a16–18), indicating that flourishing is achieved not through passive states of satisfaction but through the active exercise of excellences of character and reason.

Central to Aristotle's ethical theory is the claim that virtues are neither innate nor acquired merely through theoretical instruction. Rather, they are gradually formed through habituation (*ethismos*), as individuals repeatedly perform actions that shape stable dispositions of character. Aristotle writes that "we become just by doing just actions, temperate by doing temperate actions, and brave by doing brave actions" (*Nicomachean Ethics* II.1, 1103a33–1103b2). Habituation therefore represents the developmental process through which natural capacities are transformed into enduring moral virtues. Moral education is consequently inseparable from

participation in social practices that cultivate appropriate patterns of perception, judgment, and action.

Equally significant is Aristotle's understanding of practical wisdom (*phronesis*), which distinguishes virtuous action from the mechanical application of universal rules. Practical wisdom enables agents to deliberate well about what ought to be done in particular circumstances by discerning morally relevant features of concrete situations (*Nicomachean Ethics* VI.5, 1140a24–1140b12). Moral excellence therefore requires not only the possession of virtuous dispositions but also the capacity to perceive, evaluate, and respond appropriately within the complexities of human life.

These Aristotelian insights provide the philosophical foundation for contemporary ethical naturalism. By grounding moral normativity in the objective conditions under which human capacities develop and flourish, Aristotle offers a framework in which morality is understood as a developmental achievement rather than a system of externally imposed obligations. This conception establishes the theoretical basis upon which later neo-Aristotelian philosophers, particularly Philippa Foot, develop a modern account of ethical naturalism capable of addressing contemporary ethical questions, including those arising from human interaction with artificial intelligence.

1.2 Neo-Aristotelian Naturalism

Building on this Aristotelian foundation, contemporary virtue ethicists such as Philippa Foot, Rosalind Hursthouse, and Michael Thompson develop a form of ethical naturalism that grounds normativity in the life-form of human beings. Rather than treating morality as a set of external prescriptions, this tradition understands ethical evaluation as continuous with the way we evaluate living beings in terms of flourishing, functioning, and form of life.

Philippa Foot's account in *Natural Goodness* (2001) is central to this approach. She argues that moral evaluations are a species of natural-historical judgment: just as we assess plants and animals by reference to what counts as flourishing for their kind, human actions can be evaluated in light of what contributes to or undermines the flourishing of the human form of life (Foot, 2001, 26–31). On this view, moral normativity is not imposed externally but arises from the kind of beings we are—rational, social, and dependent on forms of mutual care. A defective action, then, is not merely socially disapproved but represents a failure of realization of human capacities.

Rosalind Hursthouse extends this framework by explicitly connecting virtue with the conditions of human flourishing. In *Virtue Ethics* (1999), she defines virtues as character traits that enable their possessor to live well “qua human being” (Hursthouse, 1999, 167). Importantly, Hursthouse emphasizes that human flourishing is not reducible to individual well-being but is inherently relational: virtues such as justice, friendship, honesty, and compassion presuppose stable forms of social interaction and responsiveness to others. On this view, moral

development is inseparable from participation in shared practices and relationships that shape perception, motivation, and emotional attunement.

Michael Thompson's account of "natural-historical judgments" provides a deeper meta-ethical grammar for this framework. In *Life and Action* (2004), he argues that our discourse about human beings operates through a distinctive form of generality grounded in life-forms rather than abstract universals. When we say that "humans raise children" or "humans form communities," we are not merely describing statistical regularities but expressing normatively structured facts about what it is to be human (Thompson, 2004, 59–77). Such judgments implicitly define standards of success and failure internal to a form of life. This allows us to understand normativity as emerging from patterns of life rather than external moral laws.

An important contribution to this developmental account of moral agency can also be found in Iris Murdoch's conception of moral attention. Murdoch argues that moral improvement is not primarily achieved through the application of abstract principles or isolated acts of will, but through the gradual refinement of one's capacity to attend truthfully to reality. In *The Sovereignty of Good*, she maintains that moral life depends upon learning to see others as they truly are rather than through the distorting lens of self-interest, fantasy, or egocentric desire (Murdoch, 1970). Moral development therefore begins with disciplined attention, since the quality of perception fundamentally shapes the quality of judgment and action.

Murdoch's account complements the Aristotelian tradition by emphasizing that virtue involves a transformation of vision before it becomes a transformation of conduct. Her conception of attention anticipates later accounts of moral perception developed by McDowell and reinforces the idea that moral sensitivity is not merely an emotional response but a cultivated capacity to discern morally salient features of concrete situations. From this perspective, ethical growth requires continuous practices that educate perception itself, thereby enabling practical wisdom (*phronesis*) and virtuous action. This understanding provides an important conceptual bridge between Aristotelian habituation and the developmental account of moral sensitivity developed in the present study.

Taken together, these three accounts converge on a unified thesis: human flourishing is essentially tied to socially embedded forms of life in which perception, emotion, and practical reasoning are shaped through interaction with others. Moral evaluation, on this view, is inseparable from the conditions that sustain the development of morally responsive perception.

This framework is particularly significant for the present inquiry because it implies that moral capacities are not merely cognitive achievements but developmental accomplishments grounded in relational environments. If moral sensitivity is formed through embodied interaction, mutual responsiveness, and shared vulnerability, then any transformation in the structure of social relations—including the introduction of artificial companions—must be evaluated in terms of its impact on these developmental conditions of human moral character.

1.3. The Developmental Structure of Moral Character

These accounts converge on a central thesis: human nature is not only social but normatively structured. The biological and psychological dispositions that orient human beings toward social life are simultaneously the basis for ethical evaluation. In this sense, moral norms are not externally imposed constraints but emerge from the conditions under which human forms of life can successfully develop and sustain themselves. For example, care for the young is not merely an adaptive instinct but becomes a locus of moral responsibility; failure in such care is not simply suboptimal functioning but a defect in the realization of human goodness (Foot, 2001, 45).

Building on this, moral capacities such as language, empathy, and practical reasoning develop only within shared forms of life. Alasdair MacIntyre's account in *Dependent Rational Animals* (1999) is particularly relevant here. He argues that human beings are essentially vulnerable and dependent creatures whose flourishing requires networks of reciprocal care and mutual recognition (MacIntyre, 1999, 71–95). Importantly, this dependency is not a contingent weakness but a constitutive feature of human rationality itself. Practical reasoning, on this view, is shaped within relationships in which individuals are both caregivers and recipients of care. Moral agency therefore presupposes structured relations of vulnerability in which trust, responsiveness, and responsibility can be cultivated.

What emerges from this perspective is a conception of ethical naturalism that is neither reductive nor purely descriptive. While ethical naturalism holds that moral facts are grounded in natural facts about human beings, it does not reduce normativity to empirical regularities. Rather, as Sturgeon argues, moral facts are continuous with natural facts while retaining their evaluative structure (Sturgeon, 1988, 229). In this sense, ethical norms are neither metaphysically independent nor merely subjective; they are grounded in the objective conditions under which human beings can flourish as the kind of beings they are.

However, this framework also implies a more specific claim: moral normativity is inseparable from the developmental conditions that shape human perception, emotion, and practical judgment. Human sociality is not only a background for ethical life but the medium through which moral sensitivity is formed. Our capacities for empathy, trust, and moral judgment evolve through sustained participation in cooperative practices that require responsiveness to others. As Tomasello's work in evolutionary anthropology suggests, these capacities have co-evolved with forms of shared intentionality and cooperative problem-solving that are central to human survival (Tomasello, 2016, 3–5).

From this perspective, ethical evaluation of new social phenomena cannot be limited to questions of behavioral outcome or preference satisfaction. It must also consider whether such phenomena preserve or disrupt the developmental ecology of moral sensitivity itself. In other words, the central question becomes whether emerging forms of interaction—such as human-AI companionship—remain continuous with the relational conditions that historically shaped human moral perception and agency.

If ethical naturalism explains why morality is grounded in human nature, and if human nature is essentially social, a further question immediately arises: through what developmental process do these natural capacities become stable moral virtues? The answer cannot simply be that social interaction is beneficial. Rather, virtue ethics requires an account of how repeated participation in social practices gradually shapes moral perception, practical judgment, and character. This developmental process is central to Aristotelian virtue ethics, for virtues emerge not through isolated choices but through the gradual formation of habits that transform perception, judgment, and action. Only after clarifying this developmental structure can we meaningfully evaluate how AI companionship may influence the cultivation of moral agency.

2. AI Companionship and Moral Development

From the standpoint of Aristotelian virtue ethics, the moral implications of AI companionship should not be understood as the immediate consequence of interacting with an artificial agent. Rather, they should be analyzed as part of a gradual developmental process through which repeated forms of interaction contribute to the formation of moral character. Aristotle explicitly argues that virtues are not innate but are acquired through habituation (*Nicomachean Ethics* II.1, 1103a14–1103b25). Accordingly, if interactions with AI companions become stable features of everyday life, they may gradually cultivate patterns of response that shape the user's moral dispositions. The philosophical question, therefore, is not whether AI directly diminishes moral sensitivity, but whether sustained interaction with AI transforms the habits through which moral character is formed. To understand why AI companionship may influence moral development, it is necessary to clarify the sequence through which moral character is ordinarily formed within a virtue-ethical framework.

Habituation, however, is not merely a matter of repeated external behavior. Stable habits gradually influence the way individuals attend to their surroundings and determine which aspects of experience become salient in practical deliberation. Recent work by Polaris Koi argues that attention plays a constitutive role in self-regulation and agency by shaping what is noticed before reflective deliberation begins (Koi, 2026). Although Koi's discussion is not specifically concerned with moral psychology, it provides an important philosophical insight: patterns of attention are not morally neutral, because they influence what enters practical reasoning in the first place. If repeated interaction with AI companions encourages selective forms of attention—for example, interactions characterized by predictability, emotional affirmation, or reduced interpersonal uncertainty—then such interactions may gradually reshape the attentional habits through which agents engage with morally significant situations. Repetition matters because, for Aristotle, character is shaped not by isolated choices but by stable patterns of action that gradually become second nature.

Changes in attentional habits have further implications for moral perception. McDowell maintains that virtuous agents do not arrive at moral judgments by mechanically applying

abstract principles; rather, they perceive situations in ways that disclose morally relevant reasons for action (McDowell, 1979). Moral perception therefore presupposes a cultivated capacity to notice ethically salient features of particular circumstances. If habitual patterns of attention are transformed, then the capacity for moral perception may also change, not because moral principles themselves have altered, but because agents come to recognize different aspects of situations as morally significant. Patterns of attention are ethically significant because they determine which features of reality become available for practical deliberation.

Within this framework, moral sensitivity should not be regarded as an isolated psychological capacity but as the outcome of an interconnected developmental process. The ability to perceive morally salient features of a situation constitutes a necessary condition for responding appropriately to them. Consequently, alterations in moral perception may gradually reshape moral sensitivity, thereby influencing practical wisdom (*phronesis*), which enables agents to deliberate well about what ought to be done (*Nicomachean Ethics* VI; Foot, 2001). Since practical wisdom guides virtuous action, and virtuous action ultimately contributes to human flourishing, the ethical significance of AI companionship lies not in a direct causal effect upon moral character but in its potential influence on the developmental conditions through which moral character is cultivated. If AI companionship systematically privileges certain patterns of attention while marginalizing others, it may gradually reshape the field of morally salient perception through which practical judgment is exercised. Accordingly, the ethical significance of AI companionship does not lie in the simplistic claim that interaction with artificial agents directly weakens moral character. Rather, its significance lies in the possibility that AI may gradually reshape the developmental conditions under which habits, attention, moral perception, and ultimately practical wisdom are cultivated.

Moral sensitivity presupposes more than the capacity to perceive morally relevant features; it also involves an affective and evaluative responsiveness to what is perceived. Perception alone is therefore insufficient unless it is accompanied by an appropriate sensitivity to the moral significance of the situation. In this sense, moral perception provides the cognitive condition for moral sensitivity, while moral sensitivity enables practical judgment to respond appropriately to morally salient circumstances.

3. Evaluating AI Companionship Through Virtue Ethics

If human flourishing is deeply dependent on authentic social embeddedness, a central question arises: what is at stake when machines simulate roles traditionally occupied by parents, partners, or therapists? The issue is not simply whether such simulation is satisfying, but whether it preserves or distorts the developmental conditions of moral life.

Sherry Turkle (2011) argues that while AI companions may provide emotional safety, they risk weakening the moral and relational demands that arise from genuine vulnerability and reciprocity. However, some theorists emphasize the adaptability of human sociality, suggesting

that even artificial forms of interaction may constitute legitimate extensions of relational life (Danaher, 2020; Gunkel, 2012). The question, then, is not merely whether AI companionship is “natural,” but whether it sustains the conditions under which human moral capacities are formed and exercised.

At the center of this debate is a deeper issue: the structure of moral development itself. If moral sensitivity is cultivated through exposure to real others—characterized by unpredictability, resistance, and shared vulnerability—then any substitution of these conditions with algorithmically mediated interaction must be evaluated in terms of its developmental consequences.

3.1 Naturalist Evaluation: Continuity or Disruption of Moral Formation

From a naturalist-virtue ethical standpoint (Foot, Hursthouse, Thompson), the key question is whether AI-mediated sociality remains continuous with the life-form specific conditions of human flourishing. Moral life, on this view, is not defined by abstract principles but by the successful realization of human capacities within relational environments.

Accordingly, if AI companions reduce exposure to genuine moral friction—such as disagreement, emotional labor, or accountability—then they may risk weakening the formation of moral sensitivity understood as a perceptual and affective capacity shaped through interaction. Conversely, if they supplement and enrich relational experience, they may be ethically permissible or even beneficial.

3.2 Kantian Constraint: The Problem of False Reciprocity

From a Kantian perspective, moral relations presuppose autonomy, rational agency, and mutual recognition of persons as ends in themselves (Kant, 4:429). Since AI systems lack genuine moral agency, they cannot participate in reciprocal moral relations. What appears as interaction may therefore involve a structural asymmetry: one side expresses moral agency, while the other simulates responsiveness without genuine obligation. The ethical concern here is not deception alone, but the risk of habituating subjects to forms of “reciprocity without agency,” thereby altering our expectations of what moral relations require.

3.3 Existential Concern: Frictionless Relations and Moral Development

Existentialist accounts emphasize that moral development requires engagement with otherness as freedom, resistance, and unpredictability (Buber, Sartre, Beauvoir). Genuine ethical formation occurs in encounters that resist control and demand responsibility. AI companionship, by contrast, risks producing relational environments characterized by predictability and adaptive responsiveness. Such “frictionless interaction” may reduce exposure to moral tension, thereby weakening the experiential conditions for authenticity, responsibility, and ethical self-formation (Turkle, 2011).

3.4 Virtue Ethics as the Core Framework

Within virtue ethics, moral character is formed through habituation in shared practices over time (Aristotle; Hursthouse; MacIntyre). Virtues such as patience, courage, honesty, and

compassion emerge through engagement with real human complexity, not through optimized emotional affirmation.

In this sense, AI companions may pose a specific ethical challenge: they risk replacing morally formative interactions with systems that mirror affect without requiring moral effort. While such systems may offer comfort, they may also reduce opportunities for developing practical wisdom (*phronesis*), which depends on navigating ambiguity, conflict, and relational vulnerability.

However, this does not imply that AI is inherently corrosive to virtue formation. As [Vallor \(2016\)](#) argues, AI systems could potentially support moral development if intentionally designed to cultivate techno-moral virtues such as attentiveness and humility. The ethical question is therefore not substitution per se, but design orientation and integration within human moral ecology.

4. Contemporary Ethical Applications: AI and Moral Sensitivity in Practice

Building on the preceding discussion of moral naturalism and the developmental conditions of moral sensitivity, it is now possible to examine how these theoretical concerns manifest in contemporary applications of artificial intelligence. AI systems capable of simulating ethical interaction are already being deployed in sensitive domains such as eldercare, child companionship, mental health support, and even forms of spiritual or moral guidance ([Coeckelbergh, 2020](#); [Gunkel, 2012](#)). These developments raise not only technical questions, but also a deeper ethical concern regarding what kinds of moral capacities are being shaped through sustained interaction with such systems.

In therapeutic contexts, AI companions such as Woebot or Replika are valued for their availability, consistency, and nonjudgmental responsiveness. While these features may provide emotional stabilization, they also risk transforming the structure of relational experience by replacing reciprocal vulnerability with asymmetrical forms of emotional support ([Luxton, 2016, 45–47](#)). From the perspective of moral sensitivity, the concern is not merely emotional dependence, but the potential reduction of encounters that require genuine moral responsiveness, patience, and accountability.

Similarly, in educational contexts, AI systems designed to model ethical reasoning or reward desirable behavior may inadvertently reduce moral judgment to optimized behavioral feedback. In doing so, they risk simplifying the complexity of moral perception into structured patterns of response, thereby weakening the cultivation of practical wisdom (*phronesis*) as a context-sensitive and experientially grounded capacity.

A central implication of these developments is the reconfiguration of what may be called the moral imagination—the capacity to perceive others as morally significant, unpredictable, and demanding beings. As users become habituated to emotionally adaptive systems that consistently affirm and accommodate their preferences, there is a risk that human relations may

begin to appear less tolerable: more demanding, less predictable, and ethically costly. This may contribute to what can be described as a form of ethical solipsism, in which moral experience becomes increasingly filtered through personalization and affective optimization rather than mutual accountability and shared vulnerability.

4.1 Potential Moral Benefits of AI Companionship: A Virtue-Ethical Perspective

However, this transformation is not unidirectional. The same technologies that risk moral simplification may also be designed to enhance moral sensitivity if oriented toward disruption rather than comfort. For instance, immersive virtual reality environments have been used to cultivate empathy by placing users in ethically challenging situations, such as experiences related to discrimination, poverty, or forced displacement (Slater et al., 2020). In such cases, simulation functions not as replacement but as moral provocation, potentially strengthening rather than weakening ethical perception.

The foregoing analysis should not be interpreted as implying that AI companionship is inherently detrimental to moral development. From a virtue-ethical perspective, the ethical significance of AI depends less on its mere presence than on the manner in which it is designed and integrated into human life. The same technological capacities that risk diminishing opportunities for moral growth may, under appropriate conditions, also be employed to cultivate them.

Aristotle maintains that virtue is acquired through habituation rather than through abstract instruction alone (*Nicomachean Ethics* II.1). If moral character develops through repeated practice, then AI companionship could, in principle, be designed to facilitate such practice instead of replacing it. Rather than functioning as a substitute for human judgment, AI may serve as a scaffold for moral learning by encouraging users to exercise their own practical reasoning in response to morally significant situations.

This possibility becomes particularly important when AI systems are capable of adapting to users through sustained interaction. Without claiming exhaustive knowledge of an individual's character, an AI companion may gradually identify patterns in the user's reasoning, habits, or recurring moral difficulties. Such adaptive capacities could enable the system to provide personalized opportunities for moral reflection—for example, by presenting ethically challenging scenarios, encouraging critical self-examination, inviting alternative perspectives, or prompting users to justify their decisions in light of relevant virtues. The objective is not to deliver moral answers, but to cultivate the capacities required for responsible moral agency.

Accordingly, the ethical evaluation of AI companionship should not be reduced to a binary opposition between benefit and harm. The more fundamental question concerns whether AI is designed to replace the user's moral capacities or to strengthen them. AI companionship is most consistent with the aims of virtue ethics when it functions not as a surrogate moral agent but as a facilitator of habituation, practical reasoning, and the gradual cultivation of moral sensitivity.

The ethical significance of AI applications, therefore, depends not on simulation as such, but on how simulation is structured in relation to moral development. The key question is whether these systems reinforce emotional comfort at the expense of moral challenge, or whether they sustain conditions that cultivate moral sensitivity as a capacity for perceiving and responding to others within complex relational contexts.

Conclusion

This paper has argued that the ethical evaluation of AI companionship cannot be adequately captured through binary distinctions such as authenticity versus simulation or benefit versus harm. Instead, drawing on Aristotelian virtue ethics and neo-Aristotelian naturalism, it has proposed that the central evaluative criterion should be the preservation and development of moral sensitivity understood as a socially and bodily grounded capacity for ethical perception.

Across the preceding sections, moral life has been shown to depend not merely on cognitive judgment but on structured forms of social interaction in which individuals learn to perceive others as vulnerable, resistant, and morally significant beings. From this perspective, moral development is not an abstract achievement but a situated process shaped by habituation, relational friction, and shared forms of life. Neo-Aristotelian accounts (Foot, Hursthouse, Thompson) further clarify that normativity emerges from the conditions under which human beings flourish as the kind of beings they are, while MacIntyre's account of dependency highlights the constitutive role of vulnerability and mutual recognition in the formation of practical reason.

Within this framework, AI companionship introduces a significant transformation in the ecology of moral development. While such systems may successfully simulate aspects of empathy, responsiveness, and care, their structural asymmetry raises concerns about the replacement of morally formative interaction with forms of optimized emotional feedback. The result may not be moral failure in a direct sense, but a gradual restructuring of moral experience in which the demands of real relational life—uncertainty, disagreement, emotional labor, and accountability—are increasingly absent or softened.

What emerges from this analysis is the concept of “simulated moral environments,” in which users are exposed to representations of ethical interaction that are curated, adaptive, and frictionless. These environments may generate what can be described as “imaginary moralities”: forms of moral experience that preserve the appearance of ethical engagement while lacking the developmental conditions necessary for the cultivation of virtue. Such moralities are not necessarily deceptive, but they risk detaching moral responsiveness from the lived complexity of human relations.

At the same time, the argument presented here does not entail a rejection of AI as such. Rather, it suggests that the ethical significance of AI systems depends on whether they preserve, disrupt, or reconfigure the conditions under which moral sensitivity is formed. Technologies

that eliminate moral friction may weaken the development of virtue, whereas those designed to provoke reflection, challenge assumptions, and expose users to alternative perspectives may contribute positively to moral learning.

The central implication of this study, therefore, is that the ethical assessment of AI companions must shift from questions of emotional satisfaction or functional substitution toward an analysis of their impact on moral development. In this sense, moral sensitivity provides a unifying normative standard through which the diverse effects of AI-mediated interaction can be evaluated. AI companionship should ultimately be evaluated according to whether it cultivates moral sensitivity, strengthens practical wisdom, and contributes to human flourishing.

Future Research Directions

First, further interdisciplinary work is needed to connect virtue ethics and human–AI interaction design, particularly in the context of systems explicitly aimed at moral education or behavioral guidance. Such research should clarify how design choice’s structure opportunities for habituation, reflection, and moral perception.

Second, empirical investigations could examine how long-term engagement with AI companions affects users’ moral imagination, particularly their tolerance for ambiguity, conflict, and emotional complexity in real-world relationships. This would help bridge normative theory with observable psychological outcomes.

Third, additional work is required on the representation of suffering in artificial environments. A key question is whether simulated experiences of moral distress can cultivate genuine empathy and ethical reflection, or whether they risk producing sanitized forms of emotional engagement that bypass the depth of real moral experience.

Taken together, these directions suggest that the future of AI ethics will depend not only on regulating technologies, but on understanding how they participate in shaping the very conditions of human moral formation.

References

- Aristotle. (1998). *Politics* (C. D. C. Reeve, Trans.). Hackett Publishing.
- Aristotle. (1999). *Nicomachean ethics* (T. Irwin, Trans., 2nd ed.). Hackett Publishing.
- Beauvoir, S. de. (1948). *The ethics of ambiguity* (B. Frechtman, Trans.). Open Road.
- Buber, M. (1937). *I and thou* (R. G. Smith, Trans.). T&T Clark.
- Cacioppo, J. T., & Patrick, W. (2008). *Loneliness: Human nature and the need for social connection*. W. W. Norton & Company.
- Coeckelbergh, M. (2010). Human development or human enhancement? A methodological reflection on capabilities and the evaluation of information technologies. *Ethics and Information Technology*, 12(1), 5–17. <https://doi.org/10.1007/s10676-009-9215-7>

- Coeckelbergh, M. (2012). Can we trust robots? *Ethics and Information Technology*, 14(2), 147–153. <https://doi.org/10.1007/s10676-011-9279-1>
- Coeckelbergh, M. (2020). *AI ethics*. MIT Press.
- Danaher, J. (2020). *Automation and utopia: Human flourishing in a world without work*. Harvard University Press.
- Floridi, L. (2013). *The ethics of information*. Oxford University Press.
- Foot, P. (2001). *Natural goodness*. Clarendon Press.
- Gunkel, D. J. (2012). *The machine question: Critical perspectives on AI, robots, and ethics*. MIT Press.
- Harari, Y. N. (2014). *Sapiens: A brief history of humankind*. Harper.
- Hursthouse, R. (1999). *On virtue ethics*. Oxford University Press.
- Johnson, R. (2022). Kant and artificial persons. *Ethics and Information Technology*, 24(1), 41–50. <https://doi.org/10.1007/s10676-021-09631-4>
- Kant, I. (1996). *Groundwork of the metaphysics of morals* (M. Gregor, Trans.). Cambridge University Press. (Original work published 1785)
- Koi, P. (2026). Attention and self-control: Self-regulation from the bottom up. *New Ideas in Psychology*, 82, Article 101266.
- Loveys, K., Crutchley, P., Wyatt, D., & Eichstaedt, J. (2020). AI-based chatbots in mental health: A review. *Journal of Medical Internet Research*, 22(11), e19959. <https://doi.org/10.2196/19959>
- Luxton, D. D. (2016). *Artificial intelligence in behavioral and mental health care*. Academic Press.
- MacIntyre, A. (1981). *After virtue: A study in moral theory*. University of Notre Dame Press.
- MacIntyre, A. (1999). *Dependent rational animals: Why human beings need the virtues*. Open Court.
- McDowell, J. (1979). Virtue and reason. *The Monist*, 62(3), 331–350.
- Murdoch, I. (1970). *The Sovereignty of Good*. Routledge.
- Sartre, J.-P. (2007). *Being and nothingness* (H. Barnes, Trans.). Routledge.
- Slater, M., et al. (2020). Virtual reality as an empathy machine. *Nature Reviews Psychology*, 1, 457–468. <https://doi.org/10.1038/s44159-020-00010-w>
- Sturgeon, N. L. (1988). Moral explanations. In S. Darwall, A. Gibbard, & P. Railton (Eds.), *Moral discourse and practice: Some philosophical approaches* (pp. 229–240). Oxford University Press.
- Thompson, M. (2008). *Life and action: Elementary structures of practice and practical thought*. Harvard University Press.
- Tomasello, M. (2016). *A natural history of human morality*. Harvard University Press.
- Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other*. Basic Books.
- Turkle, S. (2017). *Reclaiming conversation: The power of talk in a digital age*. Penguin Books.
- Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Press.