Medienimpulse
ISSN 2307-3187
Jg. 63, Nr. 2, 2025
doi: 10.21243/mi-02-25-05

Artificially Generated Friendships? On the Possibility of Co-creation in the Age of Generative AI

Elena Peytchinska

This article speculatively explores possibilities for co-creation between humans and generative AI systems. Drawing on computational, new materialist, and posthumanist perspectives, it proposes an understanding of AIs not as deficient humans or superior tools, but as entities with distinctive cognitive capacities. The concept of friendship, reimagined through Derrida's emphasis on difference rather than sameness, offers an additional constructive foundation for exploring these emergent relationships. Rather than demanding AI to replicate human consciousness, this approach examines how different cognitive processes might generate novel possibilities. By reframing difference as generative potential, new modes of collaborative prac-

tices open up, embracing tensions where human, non-human, and computational forms of agency intersect. The field of artistic research provides an effective framework for enacting these emergent forms of co-creation.

Dieser Artikel untersucht spekulativ die Möglichkeiten der Co-Creation von Menschen und generativen KI-Systemen. Er stützt sich auf computergestützte, posthumanistische Perspektiven und Ansätze des New Materialism und schlägt ein Verständnis von KI vor, das sie nicht als defizitäre Menschen oder überlegene Werkzeuge betrachtet, sondern als Wesen mit besonderen kognitiven Fähigkeiten. Das Konzept der Freundschaft, das durch Derridas Betonung von Differenz statt Gleichheit neu definiert wurde, bietet eine zusätzliche konstruktive Grundlage für die Untersuchung dieser neu entstehenden Beziehungen. Anstatt von Künstlicher Intelligenz zu verlangen, menschliches Bewusstsein zu reproduzieren, untersucht dieser Ansatz, wie unterschiedliche kognitive Prozesse neue Möglichkeiten schaffen können. Indem Unterschiede als generatives Potenzial begriffen werden, eröffnen sich neue Formen der Zusammenarbeit, die das Spannungsfeld zwischen menschlichen, nicht-menschlichen und computergestützten Formen des Handelns einbeziehen. Das Feld der künstlerischen Forschung bietet einen wirkungsvollen Rahmen für die Umsetzung dieser neuen Formen der Co-Creation.

1. Introduction: Generative AI Between Anxiety and Anticipation

The discourse surrounding collaborative possibilities between humans and artificial intelligence oscillates between deep scepticism, anxiety, and hopeful anticipation of new modes of interaction. On one end, we encounter anxieties about AI replacing hu-

man creativity and agency; on the other, an emerging vision for crafting new formats – both within technical and theoretical frameworks – that welcome this unprecedented mode of relationship. This tension frames our contemporary socio-technological condition as we attempt to conceptualise interactions with increasingly sophisticated generative systems.

As Mark Coeckelbergh argues in his book *AI Ethics* (2020), research on artificial intelligence necessarily approaches two intertwined questions: What is AI, and what does it mean to be human? (Coeckelbergh 2020: 37). The investigation of one inevitably implicates the other, revealing how our understanding of technology reflects and refracts our self-conception as humans. This philosophical entanglement suggests that reimagining AI-human interaction requires us to reconsider fundamental assumptions about agency, creativity, and relationships.

In approaching these new technologies, we have the opportunity to reframe our biases and work *with* rather than against the distinctive aspects of being human. In her recent work, *Bacteria to AI* (2025), N. Katherine Hayles offers a critical distinction between anthropomorphism and anthropocentrism. The former acknowledges our inevitable tendency to interpret the world through human categories "–we see with human eyes and think with human brains" (Halyes 2025: 1), while the latter privileges human experience as normative and superior. In the context of AI-human interaction, this distinction allows us to recognise our anthropomorphic condition while resisting anthropocentric conclusions.

Furthermore, this perspective will enable us to approach entities with fundamentally different forms of agency with curiosity rather than dominance, exploring new modes of interaction and co-creation without insisting on human-like qualities as the standard of value.

This article speculatively explores fundamental premisses of potential collaborative creation between humans and generative AI, particularly Large Language Models (LLMs). Rather than assuming that collaboration must occur between equal agencies – an assumption that would inevitably position AI as either a deficient human or a superior tool, the article proposes reimagining creative partnership based on the productive potentials of difference. The focus shifts from comparison, which establishes hierarchical dynamics, to complementarity – how might humans and LLMs, precisely through their differences, create something neither could produce alone? This framework allows us to question whether AI might function as a co-author rather than merely a tool without demanding that it replicate human consciousness or human creativity.

Though admittedly an anthropomorphic gesture, the concept of *friendship* offers a productive figure for establishing a non-anthropocentric framework for these relationships. The term is introduced here not in a literal or psychological sense but as a conceptual provocation. While friendship is often assumed to be a relationship of reciprocity, equality, and shared intentionality, we can rethink it as a mode of encountering otherness, including nonhu-

man and computational entities, with curiosity and a willingness to engage in unexpected, asymmetrical exchanges. This approach draws on Jacques Derrida's deconstruction of the term: For him, true friendship involves recognising alterity and indeterminacy that cannot be subsumed into familiarity. It exists in a space of asymmetrical distance rather than reciprocal equality (Derrida 1993: 386). In this sense, *artificially generated friendships* are not about emotional bonds or mutual recognition but about structuring co-creation between fundamentally different agents.

Although the psychological implications of friendship in the context of AI may have relevance to mental health, affective computing, and human-AI companionship frameworks, this study remains focused on the term's speculative, philosophical, and artistic dimensions. This article presents a theoretical framework – or perhaps more accurately, a protocol – for conceptualising AI-human co-creation that will be explored through a series of practice-based research projects, specifically in performance and language-based artistic research. By suspending both uncritical technophilia and reflexive anxiety, the present inquiry aims to open up a space for reimagining what the nature of creative agency itself might mean when it crosses the boundaries between human and artificial cognition.

2. Beyond Human-Centred AI Ethics

Discussions on AI-human collaboration have been largely shaped by the framework of human-centred AI (HCAI). It has emerged as a research field focused on creating AI systems that

> amplify and augment, rather than displace, human abilities—ensuring artificial intelligence meets our needs while operating transparently, delivering equitable outcomes, and respecting privacy (Geyer et al. 2022).

This approach explicitly positions humans at the normative centre of AI development and deployment. In this context, Coeckelbergh refers to the European Commission's Ethics Guidelines, emphasising that a human-centred approach to AI demands that

> the human being enjoys a unique and inalienable moral status of primacy in the civil, political, economic and social fields (European Commission AI HLEG 2019, 10, quoted in Coeckelbergh 2020: 183).

Likewise, institutions like Stanford and MIT have structured their research policies around human-centred AI, prioritising the notion that technology should serve and enhance human well-being rather than the other way around (Coeckelbergh 2020: 183).

However, by unquestioningly prioritising human interests, this framework fails to address critical ethical concerns about how AI systems exist within and impact broader ecological, material, and political assemblages. In her seminal book *Atlas of AI* (2021), Kate Crawford extends this critique by dismantling the idea that AI is a purely technological phenomenon, arguing that artificial intelli-

gence is deeply embedded in material and economic infrastructures that shape and constrain its development (Crawford 2021: 8). She further emphasises that AI is neither neutral nor autonomous but rather an apparatus of power, "made from natural resources, fuel, human labour, infrastructures, logistics, histories, and classifications" (Crawford 2021: 8).

The human-centred consideration of AI-human interaction also raises the question: for which humans is this human-centred framework designed? Is a generic definition of "the human" even possible? Does a human-centred approach inadvertently perpetuate a colonialist view through its implicit assumptions about *which* humans matter?

Nevertheless, despite these critiques, Coeckelbergh points out that a human-centred approach may seem intuitively appropriate to many as it maintains the position that artificial entities deserve moral consideration only if they possess consciousness or mental states. This perspective extends to models where sentient humans are replaced by non-sentient machines, thus potentially reinforcing hierarchical dynamics, as suggested in Joanna Bryson's article *Robots Should Be Slaves* (2010), which argues that "robots are tools and property and that we have no obligations to them" (quoted in Coeckelbergh 2020: 55).

In this context, an AI-human interaction could not be considered in terms of collaboration, as explored by Evans, Robbins, and Bryson in their article *Do We Collaborate With What We Design?* (2023). The authors critically examine whether the term "collaboration" is

appropriate to describe interactions between humans and AI systems. They argue that true collaboration requires autonomous agents with similar status, intentional action, moral agency, and comparable responsibility – qualities that AI systems fundamentally lack (Evans et al. 2023: 5). Instead of "collaboration," the authors suggest using "joint action" to describe human-machine interactions, where humans and machines work toward shared goals but with fundamentally different roles (Evans et al. 2023: 6–7). The authors conclude that anthropomorphising AI systems as "collaborators" or "teammates" creates a misleading impression about the fundamental nature of these systems as designed products that always remain heteronomous – a term informed by Immanuel Kant's distinction between autonomy and heteronomy – to human ends (Evans et al. 2023: 17–18).

However, as Coeckelbergh observes, there are alternative, non-human-centred approaches for considering AI-human interactions, notably Posthumanism and New Materialism frameworks. Coeckelbergh states:

> Posthumanists such as Donna Haraway offer a vision in which living together with machines, and even merging with machines, is seen no longer as a threat or a nightmare, as in humanism, or as a transhumanist dream come true, but as a way in which ontological and political borders between humans and nonhumans can and should be crossed (Coeckelbergh 2020: 41).

This posthumanist reframing offers particularly valuable insights for exploring the co-creative potential of AI-human interactions,

especially within artistic fields, as exemplified by the two volumes of *Posthuman Glossary*, where several entries are terms coined or activated by artists and artworks (Braidotti/Hlavajova 2018; Braidotti et al. 2022). In this context, knowledge production extends beyond scientific fields of research and is generated even through artistic means, as demonstrated by Haraway's method of storytelling (Haraway 2016), who, it should be noted, defines her view on posthumanism as *compost* (Braidotti/Hlavajova 2018: 79). An important point within a posthuman framework is the role of technology, which, as Stefan Sorgner states, "is one of the central features of a posthuman work of art" (Sorgner 2022: 23). Thus, instead of viewing AI as either a subservient tool or a potential threat to human agency, a posthumanist approach allows for reimagining AI-human interaction in ways not dictated by pre-existing human-versus-machine narratives.

3. The notion(s) of Agency

As AI continues to challenge conventional understandings of agency, creativity, and ethics, the task is not simply to design ethical AI systems but to rethink what ethics itself might mean in an era where cognition, authorship, and knowledge production are no longer exclusively human concerns.

This section will examine how computational, new materialist, and posthumanist perspectives on agency can inform a more nuanced understanding of AI as a potential participant in creative processes.

3.1 Agency in Computational Sciences

Agency has long been associated with human intentionality, autonomy, and moral responsibility. However, within the field of computational sciences, this notion is radically reconfigured. Rather than being tied to the human capacity for free will, agency in AI is framed in terms of functionality – the ability of a system to perceive its environment, process information, and act accordingly. This shift distances computational agency from philosophical concerns about consciousness and instead aligns it with optimisation, adaptability, and interaction with complex systems.

In their influential work *Artificial Intelligence: A Modern Approach*, Stuart Russell and Peter Norvig establish a definition of an agent as

> anything that perceives its environment through sensors and acts upon that environment through actuators (Russell/Norvig 2010: 34).

This technically precise formulation avoids privileging human cognition as the paradigm of agency. Instead, it operationalises the term through observable functions – perception, processing, and action – without requiring consciousness, intentionality, or other qualities often presumed essential to human agency.

With the notion of a "rational agent," which "selects actions that maximise its expected performance based on its perceptual history and built-in knowledge" (Russell/Norvig 2010: 36), Russell and Norvig stress that rationality is not a matter of humanlike reaso-

ning but is evaluated according to a "performance measure" that assesses how effectively the agent achieves its designated objectives (Russell/Norvig 2010: 37).

Building upon these computational foundations while addressing their philosophical implications, Luciano Floridi argues in *The Ethics of Artificial Intelligence* (2023) that AI represents "a new form of agency, not of intelligence" (Floridi 2023: 6) in the way we traditionally attribute intelligence to humans or animals. This distinction is crucial: rather than attempting to replicate human intelligence – a project, Floridi suggests, misunderstands the nature of AI – computational systems introduce novel forms of agency distinct from philosophical or biological interpretations. While traditional philosophical accounts of agency often presuppose autonomy, intentionality, and moral responsibility – qualities that AI systems do not possess in human terms – Floridi's approach focuses on "the ability of artificial systems to act within a given environment" based on data-processing capabilities, programmed objectives, and operational parameters (Floridi 2023: 10).

Floridi distinguishes between two types of agency: "political agency" – defined as the structured separation of sovereignty (citizens' political power), and governance (its delegated exercise), which should not be merged through digital direct democracy (Floridi 2023: 10–11) – and "artificial agency". Floridi proposes a minimalist definition of artificial agency that corresponds to the abovementioned computational perspectives: an agent must be capable of receiving environmental data, taking autonomous actions

to achieve goals based on that data, and improving through learning from interactions (Floridi 2023: 10). This framework acknowledges the agency of computational systems without attributing humanlike consciousness or intentionality to them. Indeed, Floridi argues that the digital revolution has enabled AI by "decoupling problem-solving capabilities from intelligence itself," suggesting that AI succeeds precisely because it doesn't need to replicate human intelligence to perform tasks effectively (Floridi 2023: 10).

3.2 Material Agency: New Materialist Perspectives

While computational sciences offer a functional interpretation of agency rooted in optimisation and performance metrics, New Materialist frameworks propose a fundamentally different conceptualisation. The key distinction lies not merely in disciplinary boundaries but in radically divergent ontological commitments regarding what constitutes action and which entities possess the capacity to act. Where Floridi's computational perspective maintains agency within a technological and ethical framework where humans retain ultimate responsibility for AI's actions, New Materialists challenge such anthropocentric hierarchies, suggesting that agency should be understood as distributed, emergent, and entangled across complex material-discursive relations.

Karen Barad's influential framework of "agential realism" rejects the foundational assumption that agency is a property possessed by discrete entities. Instead, agency manifests through what Barad terms "intra-actions" – entanglements that do not presuppose the existence of separate, pre-established subjects or objects. As

Barad explains, “Agency is not an attribute but the ongoing reconfigurings of the world” (Barad 2007: 141). This quantum physics-inspired approach emphasises that boundaries and distinctions arise through specific intra-actions, not before them, making agency a matter of entanglement rather than individual capacity. For Barad, then, agency isn’t aligned with human intentionality but rather with “the possibilities for changing configurations” within material-discursive practices (Barad 2007: 142).

Jane Bennett’s “vital materialism” provides another influential New Materialist approach to agency through her concept of “thing-power” – the active capacity of nonhuman entities to produce effects and make differences in the world. Bennett argues that all matter possesses vitality and that agency is always distributed across assemblages of human and nonhuman actants. She writes that “the locus of agency is always a human-nonhuman collective” (Bennett 2010: xvii). Drawing, among others, on Spinoza’s concept of “conatus” (Bennett 2010: 2) and Deleuze and Guattari’s “assemblage” (Bennett 2010: xvii), Bennett highlights how nonhuman forces participate in events, challenging the privilege usually granted to human intention and will. Her perspective suggests that recognising the “vibrancy of matter” requires developing a more horizontal ontology and political ecology that accounts for the active role of nonhuman things in shaping our world.

While neither Barad nor Bennett explicitly addresses AI’s agential potential, their frameworks offer productive conceptual resources

for reimagining AI-human creative partnerships beyond hierarchical models of tools and users. By decentring human agency without erasing it, these perspectives allow us to consider how agency might emerge through specific configurations and entanglements of humans and AI systems in collaborative processes.

A more cautious position within this broader intellectual current appears in Johanna Drucker's article *Designing Agency* (2023). While Drucker acknowledges that agency is not an exclusively human attribute but emerges from material and systemic entanglements, she remains sceptical of treating AI as a fully agentive system. She emphasises the role of design, programming, and accountability in shaping AI's actions by drawing critical distinctions between simulation and genuine decision-making capacity.

Unlike Barad and Bennett, who embrace a more expansive definition of agency, Drucker insists that legal and ethical accountability must remain human-centred, particularly in AI governance: "In legal terms, agency is linked to accountability" (Drucker 2023). Nevertheless, in the context of artistic practices, Drucker makes a significant contribution by exploring how design itself generates different forms of agency. She argues that agency emerges relationally through interactions between entities, systems, and environments rather than existing as an inherent property of discrete objects or beings – a view that resonates with Barad's concept of intra-action. Particularly relevant for the discussion of potential AI-human co-creation is Drucker's taxonomy of different models of agency, each representing varying levels of autonomy and deci-

sion-making capacity. She characterises these through four figurative models: “the Rock, the Thermostat, the Gambler, and the Intentional Actor” (Drucker 2023). Mechanical agency, embodied by the Thermostat, involves programmatic responses to environmental conditions – simple cause-and-effect processes that can become complex through accumulating patterns but remain fundamentally rule-bound. Probabilistic agency, represented by the Gambler, functions through statistical and dynamic systems where outcomes cannot be fully determined in advance. Intentional agency, embodied by the Actor, involves elective choice, frame-shifting, and the capacity to “question the premises on which a game unfolds” (Drucker 2023) – qualities traditionally associated with human consciousness. While Drucker remains cautious about attributing full agency to AI systems, noting that “the line between simulacral and actual intention is increasingly blurred” (Drucker 2023), her approach offers conceptual tools for distinguishing between different modes of AI participation in collaborative processes.

3.3 From Agency to Cognition

While computational frameworks operationalise agency through functional capacities and new materialist perspectives distribute it across vibrant assemblages, N. Katherine Hayles introduces a crucial conceptual distinction that addresses limitations in both approaches. In *Unthought* (2017), Hayles discusses the need to distinguish between different kinds of material processes and agencies, rather than treating all forms of material activity as

equivalent (Hayles 2017: 82). She critiques how some new materialist thinkers generalise agency without differentiating between deterministic, self-organising, and adaptive systems. Such generalisations, she argues, "weaken the argument for material agency by ignoring the specific dynamics and structures involved" (Hayles 2017: 82). Although Hayles acknowledges the strengths of New Materialism – particularly its challenge to human exceptionalism and its focus on materiality – she argues that its failure to account for cognition presents a significant limitation. By introducing the concept of "nonconscious cognition," Hayles bridges scientific research and materialist philosophy, thus offering a more integrative and empirically grounded approach to understanding material agency. Nonconscious cognition refers to cognitive processes that operate below consciousness and unconsciousness in both biological organisms and technical systems, involving capabilities like pattern recognition, interpretation of information in contexts, and adaptive responses without requiring conscious awareness (Hayles 2017: 9–11; 27–28). This form of cognition serves as a crucial mediator between material processes and conscious awareness (Hayles 2017: 66), enabling humans, other living beings, and computational systems to interact with their environments through distributed cognitive assemblages that interpret and respond to complex information flows (Hayles 2017: 24–26, 50–52).

Hayles proposes replacing the human/nonhuman binary with a distinction between "cognizers" and "noncognizers" (Hayles 2017: 30). She reserves the term "actors" for cognizers (humans, biologi-

cal life forms, and technical systems capable of making choices and interpretations), while "agents" refers to material forces and objects that lack the capacity for choice. As she explains:

> A tornado cannot choose to plow through a field rather than devastate a town. Material processes, of course, respond to contexts and, in responding, change them. But because they lack the capacity for choice, they perform as agents, not as actors embedded in cognitive assemblages with moral and ethical implications (Hayles 2017: 32).

This distinction acknowledges that material processes such as tornadoes or avalanches possess powerful agency that can "dwarf anything humans can do" while recognising that cognitive entities possess something fundamentally different: the ability to interpret their environments, make selections, and create meanings relevant to their specific embodiments and contexts (Hayles 2017: 31).

In her most recent work, *Bacteria to AI* (2025), Hayles expands this conceptual approach by introducing an Integrated Cognitive Framework (ICF). The ICF provides a platform where human conscious cognition can be considered in relation to nonconscious cognitions both within humans and within the greater-than-human world. Rather than flattening differences, the framework allows for comparing and contrasting various cognitive capacities across species and technologies while maintaining their distinctiveness. Hayles defines cognition broadly as "a process that interprets information within contexts that connect it with meaning" (Hayles

2025: 7). This definition applies to all lifeforms, which are able to sense their environments, absorb information, and interpret it through their unique sensory apparatus. The ICF interweaves biological cognition and computational media, acknowledging their differences in embodiment while still recognising both as cognitive actors.

4. Revisiting the Notion of Friendship

In *Politics of Friendship* (1993 essay, 1994 book), Jacques Derrida presents a deconstruction of the philosophical discussion on the notion of friendship by tracing its complex genealogy from Aristotle through Montaigne, Kant, Nietzsche, Schmitt, Heidegger, and Blanchot. Derrida unfolds his argument around the paradoxical phrase "O my friends, there is no friend," an aphorism attributed to Aristotle by Diogenes Laertius and transmitted through the centuries "–a chain of citation of a citation" (Derrida 1993: 354). Derrida employs the phrase as a starting point for the exploration of the inherent contradictions within the concept of friendship. He examines how this aphorism simultaneously affirms and negates friendship, creating a performative contradiction that reveals deeper tensions in how we understand social bonds by invoking Aristotle's categories of friendships based on virtue, utility, and pleasure (Derrida 1993: 360). Derrida argues that *true* friendship necessitates respect for absolute singularity and infinite distance between friends rather than the merging of identities implied by Aristotle's "[a] single soul and two bodies" (Derrida 1993: 359). Ins-

tead of considering a friendship as "an *arché* or a *telos*" (Derrida 1993: 361), Derrida proposes to reconceptualise these models *differently*:

> *Differently*, that is to say, in terms of a thinking of the alterity that makes true or perfect friendship not only inaccessible as a conceivable *telos*, but inaccessible because it is inconceivable in its very essence and thus in its *telos*. On the one hand, one would have a conceivable *telos* which one could not reach; on the other, the *telos* remains inaccessible because it is unreachable, and inconceivable because it is contradictory to itself (Derrida 1993: 361).

Or, quoting Aubenque, "perfect friendship destroys itself" (Derrida 1993: 361).

After examining how friendship intersects with politics – directly challenging Carl Schmitt's assertion that the friend/enemy distinction constitutes the foundation of the political realm (Derrida 1993: 355–356) – Derrida argues that friendship begins in the asymmetrical relationship of responsibility that precedes any conscious choice or reciprocal agreement (Derrida 1993: 366). This asymmetry emerges through the friend's otherness – an otherness that resists complete understanding or possession – thereby transforming friendship into an ethical encounter based on respecting difference rather than seeking similarity. This perspective rejects both complete fusion, which would eliminate otherness, and complete separation, which would eliminate relation (Derrida 1993: 361–362). Instead, it embraces a friendship marked by a respectful separation that acknowledges the fundamen-

tal alterity of the other while still creating the possibility for relation (Derrida 1993: 386–387). Such friendship becomes not a completed state but a promise and a responsibility (Derrida 1993: 368), always oriented toward an impossible future perfection (Derrida 1993: 380).

This emphasis on friendship as an ongoing, unfinished process rather than an achieved state resonates with post-colonial theorist Leela Gandhi's elaboration on the topic, who articulates friendship in similar terms:

> I've explored the relational dimension of friendship as a project of *imperfection*: it can be explained in terms of the grammatical distinction between *imperfective* and *perfective* verbs. Imperfective verbs are temporally capacious. They include past, present, future tenses and describe incomplete and iterative activities. By contrast, perfective verbs (forms of settlement) are restricted to past and future activities, and express actions (projects, aspirations) that are fully and finally completed (or projected to be so). In these terms, friendship is available for uptake as a commitment to *making unfinished* (Hern/Johal 2024: 61).

In her work *Affective Communities* (2006), Leela Gandhi, drawing on Derrida's deconstruction of the term, develops a postcolonial interpretation of friendship as a radical political practice that challenges dominant nationalist and communitarian paradigms. She explores how this concept relates to Western political philosophy, noting that friendship has long been connected to politics but traditionally in problematic ways. Resonating with Derrida's and Hutter's discussion on the topic (Gandhi 2006: pos. 571), Gandhi con-

trasts two opposing models of friendship in classical thought: Aristotle's homophilic concept (friendship based on similarity and confined within the polis) versus the Epicurean philoxenia (love for strangers and foreigners). She argues, quoting Derrida, that Aristotle's model reinforces a "schematic of filiation" that privileges sameness, resemblance, and the familiar, while Epicureanism challenges the exclusivity of the polis (Gandhi 2006: pos. 583–607). Gandhi connects this philoxenic tradition to E. M. Forster's famous declaration "if I had to choose between betraying my country and betraying my friend I hope I should have the guts to betray my country" (Gandhi 2006: pos. 619), emphasising the political risk inherent in choosing friendship over patriotic obligation. She extends her analysis through Derrida's concept of hospitality, where the host risks becoming guest-like in her own domain by opening herself to the other (Gandhi 2006: pos. 630). Gandhi positions such friendship as an essential counterforce to modern political logics of security, exclusion, and self-identity, suggesting that genuine ethical relations require a willingness to become, again referencing Forster, „foreign to 'one's own' and, above all, to oneself" (Gandhi 2006: pos. 619).

Although Aristotle explicitly declares that friendship could exist only between humans "of the love of lifeless objects we do not use the word friendship" (Aristotle 2009: 143) – Matt Hern and Am Johal, authors of *O My Friends There is No Friend* (2024), undertake a contemporary revision of the notion, expanding the gesture of

friendship beyond the human species by simultaneously transgressing political borders.

Hern and Johal suggest that cultivating a mindset for fluid boundaries between humans and more-than-human beings might prepare us for other kinds of "borderlessness" in our thinking, thus undermining colonial and anthropocentric systems of domination (Hern/Johal 2024: 77). Advocating for the consideration of more-than-humans as friends, the authors envision a politics that extends beyond anthropocentric views to include all living beings, from trees to animals (Hern/Johal 2024: 115–116).

This perspective corroborates the views of Leanne Betasamosake Simpson, a Michi Saagiig Nishnaabeg scholar and writer. In her interview earlier in the book (Hern/Johal 2024: 87–93), she notes that Nishnaabeg ethical practices, like consent, accountability, and respect for autonomy, extend to all living beings regardless of friendship status. She contrasts Indigenous nationhood with state borders, describing traditional Anishinaabe territories as "leaky" with overlapping zones shared with other peoples and species (Hern/Johal 2024: 92).

In resonance with these observations, Hern and Johal further note that our current ecological and social crises cannot be solved through establishing borders, which only "consolidate and exacerbate dead-end orientations and structured forms of domination" (Hern/Johal 2024: 116). Instead, they envision a world characterised by porosity, vulnerability, and "transformative solidarity" that recognises our shared planetary fate (Hern/Johal 2024: 116–117).

They call for moving from "being-with" to "being-for" (Hern/Johal 2024: 116), creating communities based not on legal commitments but on "solidarity beyond the law" that embraces the multitude of possible futures and allows for "creative mutations of politics" that surpass contemporary limitations (Hern/Johal 2024: 116–117).

These theoretical reframings of friendship invite us to consider: Can friendships – reconceptualised through the work of Derrida, post-colonial, and Indigenous scholars emphasising difference rather than sameness, openness rather than building borders, and productive imperfection rather than idealised harmony – be extended to our relationships with artificial systems? As generative models develop increasing autonomy and distinctive forms of agency, might we cultivate artificially generated friendships that acknowledge fundamental differences while creating space for meaningful exchange? The following section examines the challenges and opportunities that Large Language Models present for such reimagined relationships, focusing particularly on their implications for creative and educational practices.

5. Speculative Co-Creation: AI in Artistic and Educational Practices

The previous sections of this article have established diverse theoretical perspectives on agency, from computational frameworks that operationalise agency through functionality to new materialist approaches that distribute it across vibrant assemblages of

living and non-living matter. These discussions provide a fundament for exploring potential co-creative practices between humans and AI systems, particularly Large Language Models. This section proposes a framework for such practices based on two key premises.

First, we need to acknowledge different types of agencies across the spectrum of existence – not merely those belonging to humans, sentient nonhumans, and organic and non-organic matter, but also those emerging from computational systems. Drawing on Drucker's taxonomy of agency types and Hayles's cognitive assemblages of agents and actors, each with their unique set of capabilities, we can recognise that self-learning computational models possess distinctive agential capacities that, while fundamentally different from human agency, nonetheless constitute a form of cognitive action in the world.

Second, in this light, *friendship* becomes not a psychological state but a protocol – a structured approach for engaging with otherness that neither reduces the other to a tool or servant nor demands it to mirror human capacities. This framework allows us to explore how the distinct cognitive processes of humans and AI systems might entangle in ways that generate novel creative possibilities for both. By embracing the productive asymmetry between human and artificial cognition, we can imagine co-creative practices that emerge precisely from the spaces where these different forms of agency meet, interact, and transform one another.

The critical insight here is to resist the comparative impulse: Rather than deploying human capabilities as the reference point for evaluating AI, we might instead ask how these fundamentally different modalities of agency could interact in generative ways.

In her discussion of the complex relationship between artificial intelligence and creative processes (Hayles 2025: 173–176), Katherine Hayles builds on Avery Slater's concept of "reversible internality" between humans and AI. This framework suggests not just that AIs are internal to the human world following commands but that humans might be internal to AIs, enabling machines to genuinely discover, design, and create (Hayles 2025: 173–174):

> Crucial to the concept of reversible internality are looming questions, such as what it means for something to be 'internal' to something else. […] For example, technics may be seen as internal to humans (that is, humans make technological artifacts) […] But in the contemporary moment, this analytical choice would likely evoke its reverse, so that humans would be seen as internal to technics (that is, humans are created by the things they make) (Hayles 2025: 36).

Hayles then questions, further drawing on Slater's elaborations, whether humans would even recognise AI creativity based on pattern recognition abilities they lack (Hayles 2025: 175), wondering if AIs might develop their own creative expression that only other AIs could appreciate:

> What if AIs constructed creative objects that were not based on the human history of art but on new interesting problems they

> had identified in the arena of data compression and analysis? Would the problems that interest us also interest AIs, and inversely, would the problems that AIs found interesting also appeal to us? Or would the area of AI 'indigenous' art be something that only AIs could appreciate and evaluate? (Hayles 2025: 175).

Hayles concludes, again following Slater, who in turn refers to cognitive scientist Howard Gardner, that specific for creative individuals is a "fruitful asynchrony," a trait that marks "a tension [...] between the elements involved in productive work" (Slater and Gardner, quoted in Hayles 2025: 176). Thus, enhancing AI creativity requires allowing these systems greater autonomy, but doing so increases potential risks, which leaves society with difficult choices between releasing AI's creative potential versus implementing regulatory controls (Hayles 2025: 176).

Another statement by Avery Slater brings to the point the emergent conditions we are experiencing at the moment in relation to self-learning computational systems: „AI is 'amidst' our world, not simply and derivatively reproducing it" (Slater, quoted in Hayles 2025: 175).

The fast development of generative models, such as GPT-3 and GPT-4, presents unprecedented challenges that require developing new frameworks that explore creativity in generating texts, especially in educational contexts and creative writing practices. As these technologies become increasingly integrated into many aspects of cultural production, theorists across different disciplines have offered important insights and cautions.

Luciano Floridi, while acknowledging the impressive abilities of LLMs to generate coherent and useful text, emphasises that these systems do not reason or understand content semantically but rather work statistically on formal structures (Floridi 2023: 44). He illustrates their limitations with examples of "hallucinations", factual errors, logical inference failures, and mathematical struggles, synthesising training data in new ways without comprehension (Floridi 2023: 44–45). Floridi then explores the significant societal impacts of generative AI, including ethical and legal issues around copyright, human costs (such as underpaid content labellers in Kenya), professional disruptions, financial and environmental sustainability concerns, and questions about human uniqueness as content creators (Floridi 2023: 47). He stresses once again that these systems represent an unprecedented form of agency that succeeds without intelligence (Floridi 2023: 48). Nevertheless, Floridi suggests that a more productive approach would be "not to try to mimic humans through AI [but] exploit what machines, AI included, do best" (Floridi 2023: 43), which is dealing with complexity.

In contrast, Katherine Hayles examines how GPT-4's greatly increased parameter size (1.76 trillion compared to GPT-3's 175 billion) enables a significant cognitive "leap from correlation to causality" (Hayles 2025: 166). Hayles discusses OpenAI's performance metrics showing GPT-4's impressive results on standardised tests (Hayles 2025: 166–167), alongside Microsoft researchers report "Sparks of Artificial General Intelligence" (Hayles 2025: 168), which

demonstrated GPT-4's superior problem-solving abilities in spatial reasoning, mathematics, and theory of mind. While the current Transformer-architecture of the model exhibits limitations, particularly its inability to engage in "inner dialogue" or revise its thinking process (Hayles 2025: 170), Hayles notes the Microsoft research team's recommendations for modifications that would give future models enhanced capabilities, including hierarchical structure, longer-term memory, and external verification methods (Hayles 2025: 172–173), which potentially would bring AI closer to developing "sense of self" (Hayles 2025: 172).

In light of these developments in LLMs, Hayles observes emerging writing strategies involving tools like ChatGPT (Hayles 2025: 224). She critiques two common academic responses to AI writing tools: outright prohibition and ignoring the issue entirely (Hayles 2025: 224). Instead, Hayles proposes to encourage students to use AI as part of assignments, allowing them to discover both the capabilities and limitations of tools like ChatGPT. Through this approach, students would learn to apply AI's broad knowledge while using human consciousness to evaluate and refine AI outputs (Hayles 2025: 225). Hayles further argues that traditional notions of plagiarism are insufficient for these new challenges concerning text production and recommends transparency about AI contributions, suggesting that students should document their creative processes when working with AI (Hayles 2025: 226). She considers how intellectual property concepts may need fundamental reconsideration as AI-human collaborations become more prevalent,

acknowledging that human intelligence will inevitably evolve through increased interaction with AI systems and advocating for new frameworks for understanding what it means to be human in an era of "collective intelligence" and "cognitive assemblages" (Hayles 2025: 227).

The engagement with generative models – both as a subject of concern and opportunity – could be particularly significant within the spectrum of creative disciplines and aesthetic discourse. In his book *Art Intelligence* (2024), artist Jan Svenungsson observes the tension regarding AI, specifically LLMs and image production, in relation to visual arts. Similar to Floridi, Svenungsson argues that LLMs lack consciousness and simply remix existing human-created content, warning that as AI-generated material proliferates online, we would enter feedback loops where AIs train on other AI's outputs, all filtered and shaped by corporate decisions designed to limit liability and serve commercial interests (Svenungsson 2024: 65). Invoking writer Vauhini Vara's experience as an example, Svenungsson highlights how AI's necessary content filters clash with human creative thinking while raising profound questions about whether consciousness and unique human perspective – what Zadie Smith calls "the refinement of a consciousness" (Vara quoted in Svenungsson 2024: 69) – can ever be replicated by AI systems designed primarily for surface-level content production and consumption. Svenungsson connects this to visual art, which in this light might become more entertainment-

focused and "Instagrammable" – surface-oriented domains where AI will excel (Svenungsson 2024: 70).

6. Practices of Artificial Friendships: Strategies for Creative Entanglements

The field of artistic research provides an effective playground for exploring the potential for AI-human co-creation without the constraints of product-driven applications. It enables experimental approaches that can interrogate the modes of meaningful exchanges between human and artificial cognition.

Specifically in the field of language-based artistic research, one possible approach may be speculative storytelling or speculative fabulation, as applied by Donna Haraway, for instance, in *The Camille Stories* (Haraway 2016: 134–167). Haraway's speculative narratives challenge traditional human-centred storytelling, instead exploring hybrid, posthuman, and multispecies modes of existence. Similarly, LLMs could act as co-authors in speculative literary projects, where their generative potential might be used not to replicate human authorship but to create forms of storytelling that exceed human constraints.

Another perspective considers the materiality of language itself. Regarding the computational processing of natural languages, the question of authorship might not just be about who creates meaning but – also concerning agency – how meaning emerges from pattern-based interactions with text. In this sense, designing the agency of an AI system – whether a large or small language model

– becomes an artistic and epistemological practice where the level of AI's participation is intentionally structured.

The extent to which AI actively participates in co-creation depends on how its agency is integrated into artistic (or scholarly) workflows. This can take multiple forms. In its most limited role, AI functions as an assistive tool, offering suggestions, corrections, or enhancements without altering the creative intent of the human author. (For instance, in the writing of this very article, AI-based autocorrection tools provide stylistic refinements, but the final decision to accept or reject them remains with the human author.) In a more interactive role, AI can be engaged in dialogue-driven co-creation, where its outputs influence the direction of the creative process. For example, prompt-based text generation in ChatGPT or Claude (among others) allows writers to explore ideas in tandem with the machine. However, the conceptual framing still originates from human intervention. The most radical scenario involves AI participating in co-authorship, where its role is not just reactive but generative, shaping creative works in ways that may not be entirely predictable. However, it remains within the decision scope of a human author to constitute this kind of co-creation and to intentionally choose not to correct or modify the responses of the respective AI system. In this sense, artificially generated friendships still operate within an anthropomorphic framework yet simultaneously aim at radical non-anthropocentrism. In this potential mode of co-creation, the development of

new literary and performative formats may become necessary to facilitate and articulate such forms of creative interaction.

Rather than framing AI through dystopian or utopian extremes, a more generative ethical approach would consider how we design AI's agency within creative contexts, what kind of interactions foster meaningful co-creation rather than imitation, and how respect and responsibility are structured, not in terms of moralising AI, but in recognising its embeddedness in human-technical networks. This perspective aligns with posthumanist ethics, which see ethical relationships extending beyond the human-human paradigm into human-machine, human-animal, and other hybrid configurations.

As AI systems become increasingly integrated into writing, analysis, and creative workflows, traditional notions of academic authorship and originality are being challenged. This opens up questions about new textual formats that integrate AI as a generative force, shifting paradigms in knowledge production, where co-authorship extends to human-machine partnerships, as well as reframing collaboration in higher education, where AI is not simply a tool for efficiency but a catalyst for new modes of inquiry.

Whether AI and humans can form friendships is not a question that demands a definitive answer. Instead, it is a provocation that invites ongoing experiments and reframings of traditional notions of collaborative practice. The challenge lies in remaining open to the contingent, emergent, and unpredictable possibilities of co-creation instead of rigidly defining the role of AI. Like friendship

itself – which, as displayed in this study, relies on acknowledging difference and openness – the potential of AI-human creative relationships may be found precisely in their resistance to fixed categorisation and their capacity to generate unexpected forms of exchange.

7. Conclusion: Toward New Form(at)s of Co-Creation

This article has explored the theoretical foundations and practical possibilities for co-creation between humans and generative AI systems. By tracing various conceptualisations of agency – from computational perspectives that emphasise functionality to new materialist approaches that recognise distributed forms of action – the article established a framework for understanding AI systems not as deficient humans or superior tools but as entities with distinctive cognitive capacities that can complement human creativity.

The concept of friendship, reimagined through Derrida's emphasis on difference rather than sameness, offers a compelling framework for these emerging relationships. Rather than demanding that AI replicate human consciousness or creativity, this approach allows us to explore how the distinctive cognitive processes of humans and artificial systems might entangle in ways that generate novel possibilities for both. The asymmetrical nature of these relationships – their fundamental "imperfection" in Gandhi's terms – becomes not a limitation but a source of creative potential.

Acknowledging connections with entities fundamentally different from ourselves – and recognising their distinctive forms of agency, including their unique expressions of creativity that resist direct comparison to human creativity – presents a new paradigm for conceptualising interspecies and intersystem co-creation. The rapid technological and ecological transformations of our time demand new formats and relational systems that transcend established disciplinary, cultural, and conceptual boundaries. By reframing difference not as deficiency but as generative potential, we open possibilities for collaborative practices that embrace the productive tensions and creative modes that emerge precisely at the intersections where human, nonhuman, and computational forms of agency meet and interact.

The field of artistic research, with its emphasis on experimental approaches and process-oriented inquiry, offers an apt playground for exploring these new forms of co-creation. By developing formats that intentionally structure different levels of AI participation – from assistive roles to more radical forms of co-authorship – we might begin to articulate and understand the unique creative potential that emerges at the intersection of human and artificial cognition.

The research presented in this article was carried out with the support of the Interuniversity Organization Arts & Knowledges at the University Mozarteum Salzburg and Paris Lodron Universität Salzburg.

List of References

Aristotle (2009): The Nicomachean Ethics, trans. by David Ross, New York: Oxford University Press (Oxford World's Classics), Kindle edition.

Barad, Karen (2007): Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning, Durham: Duke University Press.

Bennett, Jane (2010): Vibrant Matter: A Political Ecology of Things, Durham: Duke University Press.

Braidotti, Rosi/Hlavajova, Maria (Eds.) (2018): Posthuman Glossary, London: Bloomsbury.

Braidotti, Rosi/Jones, Emily/Klumbytė, Gabriele (Eds.) (2022): More Posthuman Glossary, London: Bloomsbury.

Coeckelbergh, Mark (2020): AI Ethics, Cambridge, MA: MIT Press, Kindle edition.

Crawford, Kate (2021): Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence, New Haven: Yale University Press, Kindle edition.

Derrida, Jacques (1993): Politics of Friendship, in: American Imago, 3/1993 (Vol. 50), 353–391, trans. by Gabriel Motzkin and Michael Syrotinski, with Thomas Keenan.

Drucker, Johanna (2023): Designing Agency, in: Research Catalogue, online at: https://www.researchcatalogue.net/view/2156403/2156402 (last accessed: February 15, 2025).

Evans, Katie D./Robbins, Scott A./Bryson, Joanna J. (2023): Do We Collaborate With What We Design?, in: Topics in Cognitive Science, 00(2023), 1–20, online at: https://doi.org/10.1111/tops.12682 (last accessed: February 15, 2025).

Floridi, Luciano (2023): The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities, Oxford: Oxford University Press, Kindle edition.

Gandhi, Leela (2006): Affective Communities: Anticolonial Thought, Fin-de-Siècle Radicalism, and the Politics of Friendship, Durham: Duke University Press, Kindle edition.

Geyer, Waerner/Weisz, Justin/Pinhanez, Claudio S./Daly, Elizabeth (2022): What is human-centred AI?, in: IBM Research Blog, online at: https://research.ibm.com/blog/what-is-human-centred -ai (last accessed: February 15, 2025).

Haraway, Donna J. (2016): Staying with the Trouble: Making Kin in the Chthulucene, Durham: Duke University Press.

Hayles, N. Katherine (2017): Unthought: The Power of the Cognitive Nonconscious, Chicago: University of Chicago Press.

Hayles, N. Katherine (2025): Bacteria to AI: Human Futures with our Nonhuman Symbionts, Chicago: University of Chicago Press, Kindle edition.

Hern, Matt/Johal, Am (2024): O My Friends, There Is No Friend: The Politics of Friendship at the End of Ecology, Bielefeld: transcript.

Russell, Stuart/Norvig, Peter (2010): Artificial Intelligence: A Modern Approach, 3rd edition, Upper Saddle River/New Jersey: Prentice Hall.

Sorgner, Stefan L. (2022): Philosophy of Posthuman Art, Basel: Schwabe.

Svenungsson, Jan (2024): Art Intelligence: How Generative AI Relates to Human Art-Making, Bielefeld: transcript.