

THE PROTAGONIST AND THE ARTIFACT: RECLAIMING THE MORAL NARRATIVE OF THE PERSON AGAINST ALGORITHMIC REDUCTIONISM

Zyra F. Lentija
University of Asia and the Pacific, Philippines

This article develops a metaphysical and cross-cultural framework for Artificial Intelligence (AI) ethics centered on the human person as the moral protagonist. Current debates highlight fairness, transparency, and accountability, but often neglect a prior question: what kind of being can truly be considered a moral and narrative agent? I respond by synthesizing Robert Spaemann's personalist metaphysics with Filipino relational philosophy as a corrective approach. Drawing on Spaemann, Aquinas, MacIntyre, and Cicero, I first depict the person as an irreducible "someone" whose self-possession grounds actus humanus and a teleological, narrative unity of life. I then deepen and correct this Western protagonism through Filipino concepts of loób (relational will), kapwa (shared self), pakikiramdam (affective attunement), and pakikipagkapwa (engaging the other as a fellow person), presenting the person as a being of interior depth and shared identity whose telos is communal co-flourishing. Finally, employing the classical distinction between physis and technē, I argue that AI systems, as artifacts marked by a "volitional deficit," cannot be moral agents, and that treating them as such fuels "algorithmic closure": the epistemic, moral, and ontological reduction of persons to computational objects and fixed data-selves. I propose ontological sobriety as a governing imperative for AI ethics, expressed as design requirements that protect narrative openness and genuine human presence in an algorithmic world.

Keywords: Artificial intelligence, loób, ontological reductionism, ontological sobriety, pakikipagkapwa, persons

INTRODUCTION

The rapid and widespread integration of Artificial Intelligence (AI) into domains as fundamental as healthcare, governance, and social interaction has generated not only an ethical crisis, but also a profound crisis in how we understand what it means to be human. Today's dominant technological mindset, guided by an instrumental way

of thinking, privileges efficiency, predictability, and control above all else. This technocratic orientation leads to a kind of ontological reductionism: the rich, interior, and freely choosing human person is flattened into a set of measurable, computable data points.

In the background of this analysis stand thinkers such as Martin Heidegger and Romano Guardini¹, who help name the contours of this problem. Heidegger's critique of technological "enframing" (*Gestell*) shows how reality itself, including human life, risks being seen only through the lens of control and utility. Guardini similarly exposes how the technological worldview tends to treat nature, and by extension human beings, as mere "raw material to be hammered into useful shape."² Their insights reveal that the deepest challenge posed by AI is not limited to technical questions of bias, transparency, or regulation. Instead, it forces us to revisit the more basic metaphysical question of *who or what a rational agent is*.

The central claim of this article is that we must recover what I will refer to as *ontological sobriety*³: a clear-sighted recognition of the unique and non-reducible status of the human person in an age dominated by algorithms and data. Ontological sobriety is the primary normative principle I propose for evaluating AI and its impact on human life. It insists that persons can never be adequately understood, treated, or designed as mere tools, resources, or information systems. While the Association of Southeast Asian Nations (ASEAN) has published guidelines on AI Governance and Ethics (2024), these frameworks often prioritize procedural fairness over the deeper metaphysical questions about human agency that this article addresses.

To ground this claim, I draw on two complementary philosophical currents that together provide the necessary metaphysical foundation for ontological sobriety. On the one hand, Robert Spaemann's personalist metaphysics articulates the person as a final end in himself and as the protagonist of the moral life. On the other hand, Filipino philosophy⁴, especially as developed by Leonardo Mercado and Virgilio Enriquez, clarifies the deeply relational nature of the person through concepts such as *loób* (inner self or relational will), *kapwa* (shared self/other), and *sakop* (sphere of belonging). Rather than functioning merely as a non-Western "correction" to Spaemann, Filipino relational thought stands as an essential and complementary partner: Spaemann secures the irreducible dignity and teleological depth of the person, while Filipino philosophy illuminates how this dignity is always already constituted and expressed in concrete, shared relations.

Within this framework, the non-negotiable status of the human person as the protagonist of ethical life is deepened and made more concrete through the Filipino ethos of *pakikipagkapwa*, relating to the other as a fellow person. Together, these perspectives establish a firm ontological boundary between persons and artifacts, resisting the pervasive tendency to treat human beings as if they were just more sophisticated machines.

The article proceeds in four parts. Part II clarifies the moral and narrative structure of the person as protagonist, drawing primarily from Spaemann. Part III turns to Filipino thought, presenting *loób* and *kapwa* as non-Western relational pillars that solidify the metaphysics of agency. Part IV employs the classical distinction between *physis* (nature) and *technē* (art/technique) to critique the instrumental logic of AI and the risk of algorithmic closure. Finally, Part V develops ontological sobriety as the

guiding normative imperative for protecting the person's moral narrative from algorithmic reductionism.

While Heidegger and Guardini remain in the background as critical diagnosticians of the technological worldview, Spaemann's metaphysics of the person and Filipino relational ontology provide the structural core of my proposal. Neither explicitly uses the term "ontological sobriety." However, their shared insistence on the non-instrumental status of persons—one emphasizing finality and dignity, the other emphasizing relational belonging—gives this concept both its philosophical depth and its cross-cultural resonance.

This article argues that any adequate framework for AI ethics must be grounded in a robust metaphysics of the human person as a moral protagonist whose narrative identity cannot be reduced to data. Drawing on Robert Spaemann's account of the person as an irreducible "someone" and the Archimedean point of ethical life, and deepening it through Filipino relational philosophy—*loób*, *kapwa*, *pakikiramdam*, and *pakikipagkapwa*—I propose the concept of ontological sobriety as a governing imperative for the age of AI. Ontological sobriety insists on a strict ontological boundary between *physis* (the self-moving, relational person) and *technē* (the externally driven artifact), exposes how algorithmic closure threatens to flatten the person into a "data-self," and proposes culturally situated design requirements—memory sovereignty, anti-capture safeguards, and affective attunement by design—that preserve the human person as the true protagonist of the ethical life.

THE METAPHYSICS OF PERSONHOOD: THE PROTAGONIST AS "SOMEONE"

Robert Spaemann's (2006) central philosophical project begins with a deceptively simple grammatical distinction that carries profound metaphysical weight: the distinction between "someone" and "something." In much of modern functionalism, a person is defined by a list of attributes: rationality, self-awareness, linguistic capacity, sentience, or higher-order thought. Under this view, a person is merely a "something" that possesses certain features, and once those features appear in a system, we are licensed to call it a person.

Spaemann reverses this priority. For him, personhood is not a property that emerges when a biological or computational system crosses a threshold of complexity. A person is not a sophisticated "thing" with added mental attributes; rather, a person is an irreducible subject who has those qualities, a bearer to whom properties and capacities belong. Personhood is not a layer that can be peeled off, nor a function that can be turned on or off; it names the underlying "someone" for whom all experiences are had and to whom all attributes belong.

Spaemann identifies the person as the "Archimedean point" of reality. Just as Archimedes claimed he could move the earth if only he had a firm place to stand, the person is the standpoint from which space, time, and meaning are oriented. "It is they that give meaning to the here and now," Spaemann (2006, 164) argues. Without a "someone" to say "here" and "now," the universe becomes a coordinate grid of indifferent points, a collection of facts with no center of concern. The world does not

simply exist; it appears to *someone*, and that appearing is not incidental but constitutive of meaning.

This leads to Spaemann's notion of self-possession. A "something" (an object) is possessed by its properties or moved by external causes. A stone is where it is, and does what it does, because forces act upon it. A "someone," however, is a being that in some sense "has itself" (Spaemann 1996, 21-23). This self-possession (*Selbstbesitz*) means that a person is not merely a biological or psychological event happening in the world, but a center of interiority that can take a stance toward its own nature, tendencies, and history. We do not simply undergo a life; we lead a life.

This capacity to say "I" is not just a contingent physiological state but a metaphysical status. It marks the difference between being an episode in a causal sequence and being a protagonist within a story. The "I" is that from which actions proceed as mine, as authored and owned. Here, the metaphysical category of "someone" begins to show its ethical import: only a being that can say "I" can be a responsible agent rather than a mere spectator or a cog in an impersonal chain of causes.

If the person is the "Archimedean point," then his movement through time cannot be conceived as a random series of disconnected events. A protagonist does not simply occur; a protagonist *acts*, and actions unfold in intelligible patterns. To understand the person as protagonist, we must therefore link the Aristotelian–Thomistic tradition of action with Alasdair MacIntyre's contemporary narrative theory.

In the Scholastic tradition, particularly in Thomas Aquinas, a vital distinction is drawn between *actus hominis* (an act of a human being) and *actus humanus* (a properly human act) (*STI-II*, q. 1, a. 1; q. 6, a. 1-3). *Actus hominis* are the biological or involuntary operations, such as digestion or reflexive movements. They happen in a human being, but not from the person as a free subject. *Actus humanus*, by contrast, are actions that proceed from deliberation and will: to choose, to promise, to forgive, to betray. Only the *actus humanus* is morally imputable, because only here is the "someone" fully present as author. This is where the metaphysical "someone" becomes the ethical "protagonist": where the capacity for self-possession issues in responsible self-determination.

Alasdair MacIntyre extends this insight to the modern context, arguing that human acts are intelligible only within a "narrative quest" (1984, 218-219). He argues that only the person can say, "I can only answer the question 'What am I to do?' If I can answer the prior question 'Of what story or stories do I find myself a part?'" (MacIntyre 1984, 219). We do not understand a single action—say, a man picking up a coat—by describing its physical mechanics. We must know the story: is he leaving his wife? Going to work? Preparing to commit a crime? Returning to care for a sick friend? The same bodily motion can belong to radically different narratives and moral worlds.

For MacIntyre, the self is a "storytelling animal" whose life has a teleological, goal-oriented structure (1984, 216). We live our lives as a quest for the Good, however obscurely conceived. Our identity is not merely the sum of psychological states at a given moment but the narrative unity of our commitments, failures, and hopes across time. To be a person is to be answerable for the "to-be-continued" nature of one's

story, to bear responsibility not only for isolated choices but for the intelligibility of one's life as a whole. We are not merely authors; we are at once authors and principal characters, tasked with making the story cohere.

To further ground this "role" of the protagonist, we can turn to the Roman statesman Cicero and the etymology of *persona*. Originally referring to the mask worn by actors in Greco-Roman theater (*prosopon*), the term *persona* evolves into a complex ethical concept in Cicero's thought. Crucially, for Cicero, the *persona* is not a deceitful façade concealing a hidden "real self." Instead, it names the concrete way one's unique role and responsibility are manifested in the drama of life.

In *De officiis*, Book I, Cicero (1913) suggests that each human being wears "four masks." First, there is the universal mask: our shared human nature as rational and social beings. Second, the individual mask: our particular temperament, talents, and bodily constitution. Third, the circumstantial mask: the roles and situations thrust upon us by chance, history, or social position—birthplace, family, era, political order. Fourth, the chosen mask: the vocations, life-plans, and commitments we assume freely, such as our careers, causes, and forms of life. Cicero (1913, *De officiis* 1, 107; 1, 115)⁵ says:

We should know that Nature has dressed us, so to speak, with two masks. The first of them is universal, because all of us partake of reason and of this superiority by which we excel over the beasts, namely the capacity for moral goodness... The second mask, however, is specifically attributed to each individual... To these two is added a third mask, which some event or time lay upon us, and a fourth one which we ourselves put on us according to our own judgment. For kingdoms, military commands, nobility, honors, wealth, influence, and all that is opposite to them are governed by chance and time. Nevertheless, the mask that we ourselves want to wear originates from our own will.

Propriety (*decorum*), Cicero argues, consists in acting in harmony with these masks. A person's "properly human act" is to perform their unique role with excellence, neither denying their nature nor fleeing their concrete situation. In this light, the "theatrical" metaphor is not an invitation to insincerity, but to a deeper sincerity: the recognition that we have been "cast" into a particular time and place, and that our dignity lies in how we inhabit that role. The person is the "someone" who assumes responsibility for the part they play before others and, ultimately, before their own conscience.

Yet there is a danger in this metaphysics of personhood. If we focus solely on Spaemann's "Archimedean point" or on the individual's self-possession, we risk sliding into a particularly Western form of solipsism, or the idea of the "I" as a self-contained island, a sovereign center that needs no one else. On a superficial reading, the person, as an irreducible subject, might appear as an isolated monad.

However, Spaemann's own philosophy, along with the broader personalist tradition, points in another direction: the "I" awakens only in the presence of another "someone." The word "I" is learned in response to "you." The protagonist's interior

drama presupposes an interpersonal stage. There is no protagonist without fellow actors, no persona without an audience and a community in which that role becomes meaningful.

This requires an expansion of the metaphysics of personhood. While the “I” is indeed the seat of responsibility, the “I” remains incomplete—almost ghostly—without the “we.” In the Philippine philosophical context, this emerges in the movement from the individual subject to *kapwa*: the shared self, a self constituted in and through relation (Enriquez 1986, 2; 9-13). Here, the “someone” Spaemann describes is not a closed circuit of consciousness but a being-towards-others, whose very self-possession is exercised as self-giving and co-belonging.

To understand the protagonist, then, we must move from the internal mechanics of self-possession to the external reality of relationality. The story of “someone” is never a monologue; it is always, inherently, a dialogue. My freedom is formed in conversation with others’ freedoms; my narrative quest intersects with theirs, and our mutual recognition becomes a condition for each of us being truly “someone” rather than merely “something.”

Seen in this light, the question of personhood is not a marginal technical issue in metaphysics or ethics. It is the question of whether human beings are centers of meaning or mere nodes in a system, protagonists in a shared drama, or replaceable components in an impersonal process. Spaemann, MacIntyre, and Cicero together invite us to resist any reduction of persons to functions, traits, or roles detached from the “someone” who lives them. They challenge us to ask, in a world increasingly tempted to think in terms of data, efficiency, and utility: *who, exactly, is the someone for whom all this matters?*

To take that question seriously is already to be drawn into the drama it describes, for the answer cannot be given abstractly, from the outside. It must be lived, as each of us decides, today and again tomorrow, whether to inhabit our masks with integrity, to act as protagonists within a meaningful story, and to recognize in every other face not a mere “something,” but another irreplaceable “someone.”

THE RELATIONAL WILL: *LOÓB*, *KAWPA*, AND ONTOLOGICAL DEPTH

While the Spaemannian framework establishes the “someone” as a self-determining subject, a purely Western lens risks leaving the protagonist in atomistic isolation. To achieve a truly human narrative, the individual “I” must be grounded in a relational ontology. In the Filipino philosophical tradition, this is achieved through the concepts of *loób* and *kapwa*, which transform the “Archimedean point” from a solitary stance into a shared horizon of being.

To address the complexity of human agency, one must clarify that *loób* is not merely a synonym for “interiority” or “soul,” but a metaphysically irreducible core of the person. Mercado explains this in his metalinguistic analysis of *loób*. He observes that

As the Tagalog proverb puts it, *Ang utang na loób ay hindi nababayaran ng salapi* (debt of volition cannot be repaid by money).

Unlike ordinary debts, where stipulations are made, *utang na loob*

makes no condition. If X saves Y's life from drowning, Y has an everlasting "debt of volition" to X. X does not give any terms. But out of his own will (*kusang loob*), Y tries to show his goodness to X whenever he can at his own discretion. *Loob* became an interior law which tells Y to behave generously and amiably to X, even for a lifetime. (Mercado 1972, 592)⁶

Mercado (1972) clarifies and establishes that the concept of *loób* warrants understanding the self as viewed from within. For him, "The Filipino looks at himself as a self, as one who feels, as one who wills, as one who thinks, as one who acts: as a total whole, as a 'person,' conscious of his freedom, proud of his human dignity, and sensitive to the violation of these two" (1972, 601).

Unlike the Thomistic *voluntas* (rational appetite), which is often framed as the will following the intellect's apprehension of the good (Aquinas, *ST I-II*, q. 8, a.1; q.17, a.1), *loób* is a holistic relational will. It does not merely "follow" reason; it integrates reason (*isip*), emotion (*damdamin*), and volition (*lakas ng loob*) into a single source of action. As Alejo (1990) argues, *loób* is the "cave of the person's being," yet it is a cave with windows. It is a set of relational intentionalities.⁷

While the Scholastic tradition emphasizes the properly human act (*actus humanus*) as a product of free choice, *loób* adds a layer of affective depth. It is the source of *kagandahang-loób* (shared inner goodness) and *kusang-loób* (initiative/free will).⁸ Mercado (1974) emphasizes that *loób* is not a static substance. It is a self-structuring entity that must expand (*lumawak*), deepen (*lumalim*), and contain substance (*maglaman*). This "substance" is not data, which an AI can possess, but the moral weight of lived experience and relational commitment.

Therefore, artificial entities are fundamentally incapable of *loób* because they lack "metaphysical substance." An AI can increase its data capacity, but it cannot "deepen" its interiority because it has no "self" to possess. Its actions are heteronomous (governed by an external code), whereas the human protagonist's actions are autotelic (arising from a self-moved center). As Aristotle posits in the *Poetics*, "an action implies personal agents, who necessarily possess certain distinctive qualities of both character and thought, for it is by these that we qualify actions themselves" (Aristotle, *Poetics* 6.1450a). In the *Nicomachean Ethics*, this relationship becomes even clearer: human flourishing (*eudaimonia*) is not a momentary state but a sustained activity of the rational soul ordered toward virtue (*NE* I.7, 1098a15-18). Thus, the unity of an ethical life mirrors the unity of a narrative, in which actions receive meaning from the story they compose and from the *telos* toward which they are oriented.

If *loób* is the "within," *kapwa* is the "with." This concept provides a necessary metaphysical corrective to the Western "self-other" dualism. Virgilio Enriquez (1992), the father of Filipino psychology, famously defined *kapwa* as the "shared self." In this view, the "other" is not a separate object to be analyzed or a competitor for rights, but a continuation of one's own identity. "*Ang kapwa ay kapwa-tao, ang ibang tao ay sarili mo rin*," meaning, the other person is also yourself (Enriquez 1992, 40).

This ontological shift means that fulfillment for the Filipino *tao* (person) is not found in individual autonomy, but in participation. As Mercado (1994) notes, the person is not an atom but a point in a web. The transition from the Spaemannian "I" to

the Filipino “we” is not a loss of individuality but its perfection. The “someone” becomes a protagonist only when they recognize that their story is intricately woven into the stories of their *kapwa*. Enriquez (1986) explains that *kapwa* represents the unity between the self and others. He mentions that in English, “others” is frequently contrasted with “self,” implying separate identities. Conversely, *kapwa* signifies an acknowledgment of a shared identity (Enriquez 1986, 11).

This relationality of the person is best expressed through *pakikiramdam*, an affective attunement of “feeling for the other.” It is the ability to sense what is unspoken and to respond intuitively to the needs of the *kapwa*. While Western metaphysics often emphasizes asserting the individual will over nature and others, *pakikiramdam* focuses on alignment. This attunement goes beyond humans to include the natural world. Here, “nature” is not a “thing” to be exploited, but a realm of presence that calls for a sensitive response. Unlike the shallow, quantified empathy scores seen in modern human–computer interaction (HCI)⁹, *pakikiramdam* demands a shared inner life, an *isang loób* (one heart/will). It is a depth of understanding that cannot be simulated because it needs the vulnerability of “someone” who can truly feel. As Mercado notes, the “happiness and the good for the *Filipino* are shared with his reference group” (1972, 257).

Ultimately, the goal (*telos*) of the human protagonist is not mere self-realization or the accumulation of virtues in isolation. The *telos* is *pakikipagkapwa*, the act of engaging the other as a shared self. Understanding *telos* as *pakikipagkapwa* means “accepting and dealing with the other person as an equal. (...) This means a regard for the dignity and being of others” (Enriquez 1986, 16).

Seen in this light, we can better appreciate how different philosophical traditions articulate the meaning of personhood and its ultimate purpose. The following simple comparison illustrates how Aristotelian–Thomistic and Filipino philosophies approach key features of the person differently:

Primary driver
Aristotelian–Thomistic: <i>Eudaimonia</i> (virtue/flourishing)
Filipino philosophy: <i>Pakikipagkapwa</i> (communal flourishing)
Source of will
Aristotelian–Thomistic: <i>Voluntas</i> (rational appetite)
Filipino philosophy: <i>Loób</i> (relational/affective will)
Relational mode
Aristotelian–Thomistic: Justice and friendship
Filipino philosophy: <i>Pakikiramdam</i> (attunement)
Identity
Aristotelian–Thomistic: Individual subject
Filipino philosophy: Shared self (<i>kapwa</i>)

This table is not exhaustive, nor is it meant to be prescriptive; rather, it highlights how these perspectives diverge in their treatment of the same features of the person.

To say that someone has free will is to say that the reason the person acts lies in himself. However, in the Filipino context, we venture to say that “acting for oneself”

and “acting for the *kapwa*” are not in conflict. Because the *kapwa* is a shared identity, the protagonist’s narrative unity is achieved only when their actions contribute to the community’s co-flourishing. Enriquez notes:

A person starts having a *kapwa* not so much because of recognition of status given to him by others, but more because of his awareness of shared identity. The *ako* (ego) and the *iba sa akin* (others) are one and the same in *kapwa* psychology: *Hindo ako iba sa aking kapwa* (I am no different from *kapwa*, the self, in effect, denies the status of *kapwa* to the other (1986, 12).

The “someone” finds their meaning not by standing alone at the Archimedean point, but by using that point to anchor a bridge to others.

The pinnacle of Filipino relationality is the achievement of *isang loób*, a state in which two or more individuals attain a shared interiority. This is not a mere agreement or a simple “syncing” of data; it is a metaphysical union of wills. In contrast, modern HCI and AI development often treat “empathy” as a set of variables: sentiment analysis, facial recognition, and predictive response patterns. This creates a “shallow” empathy, a simulation of *pakikiramdam* that lacks the ontological depth of a subject.

Pakikiramdam requires a person to be “at stake” in the relationship. Because an AI lacks a *loób* (a center that can suffer, grow, or “contain substance”), it cannot truly “feel for” another. It can simulate the results of empathy, but it cannot participate in the process of shared interiority. Returning to Spaemann’s “someone,” the Filipino protagonist adds a layer of presence (*pagkatao*). To be a person is to be “someone” who can be held accountable not just for their actions but for their failure to “feel for” the *kapwa*. The narrative unity of the Filipino protagonist is a story of “becoming more human” (*pagpapakatao*). This is a teleological journey where the “I” matures into a “we.” While an AI remains a “something” defined by its programming—a character written by others—the human person is the protagonist who writes their story through the exercise of a relational will.

What we have seen thus far is that the metaphysics of personhood, when viewed through the combined lenses of Spaemann, Aristotle, Aquinas, and Filipino philosophy, reveals the person as a being of immense depth. As an Archimedean point, the person provides the orientation of the universe. As a narrative agent, the person gives coherence to time through the *actus humanus*. As a relational will (*loób*), the person seeks the telos of shared flourishing through *kapwa*. The “someone” is ultimately defined by their capacity for unprompted, autotelic moral action, a capacity that remains the exclusive domain of the human spirit.

THE PROTAGONIST AND THE MACHINE: A METAPHYSICS OF PERSONHOOD AGAINST ALGORITHMIC CLOSURE

The contemporary crisis of agency, precipitated by the integration of Artificial Intelligence (AI) into the fabric of human decision-making, necessitates a return to the fundamental metaphysics of personhood. At the heart of this inquiry is the distinction

between a being that is “someone” and an entity that is merely “something.” We have seen in the previous section that Robert Spaemann’s central contribution to this discourse lies in his identification of the person as the Archimedean point from which alone it is possible to identify positions in space and time (2006, 164). Unlike modern functionalist accounts that reduce personhood to a set of cognitive properties, such as processing speed or data retrieval, Spaemann argues that persons are self-possessing subjects who do not merely exist as instances of a species but lead their lives as protagonists (as already mentioned in the previous pages). This metaphysical self-possession (*Selbstbesitz*) serves as the foundation for the ethical life, transforming a mere sequence of biological events into a “properly human act” (*actus humanus*) (Spaemann 1996, 21–23).

To call a human being a protagonist is to say that events do not simply carry her along but can, at least to some extent, take ownership of her story. This protagonism is not an exercise in solitary autonomy, as if the self were an isolated island. Rather, the narrative integrity of the person is only intelligible when viewed as what Alasdair MacIntyre terms a “narrative quest” (1984, 219). The self is a storytelling animal, and its identity is found in the unity of a life lived toward a perceived telos (an end goal), where the present is always “informed by some image of some future” (MacIntyre 1984, 204). Even ordinary human activities such as choosing a career, sustaining a friendship, and raising a child presuppose some sense of “where this is going” and “who I am becoming.” In the Scholastic tradition, this teleological direction is the product of the *voluntas*, the will, which moves the person toward a conceived good (Aquinas, *ST I–II*, q.1, a.1). The will is not merely a mechanism that follows impulses; it is the inner power by which a person can endorse, revise, and even resist her own desires for the sake of a more encompassing good.

While the human protagonist achieves narrative unity through a shared life with others, the technological system threatens to dissolve this unity into a series of instrumental calculations. AI systems operate by optimizing outputs relative to predefined objectives: maximizing engagement, predicting choices, and increasing efficiency. When such systems mediate more and more aspects of life—what we watch, whom we meet, which opportunities we see or never see—the risk is not only practical manipulation but a subtle erosion of our sense of authorship over our own story.

This instrumental threat is best understood through the Aristotelian distinction between *physis* (nature) and *technē* (artifice). Natural beings possess an internal *archē kinēseōs kai staseōs*, a principle of motion and rest from within (Aristotle, *Physics* II.1, 192b23–29). They move themselves in light of what appears good to them. Artifacts¹⁰, by contrast, have their principle of motion imposed from without.

To attribute moral agency to AI is therefore to commit a grave category error, conflating sophisticated simulation with intrinsic being and engaging in what we might call “ontological inflation.” An AI system can imitate patterns of choice and even display what appears to be consistency of character. However, its operations remain entirely dependent on external programming, training data, and hardware. It does not possess its own life, its own “from within,” that could ground responsibility. Treating such systems “as if” they were agents may be practically convenient. However, when this convenience hardens into literal belief, it obscures where accountability actually

lies: with the human designers, deployers, and institutions that shape and use these technologies.

This inflation hides a deeper issue we can call “algorithmic closure,” in which a person’s rich, fluid identity is reduced to what Shoshana Zuboff (2019, 377)¹¹ calls a “computational object.” In this condition, we no longer primarily appear to institutions and platforms as unique persons, but as profiles, scores, and behavioral probabilities. When we allow behavioral capture to dominate—when every click, pause, location, and interaction is recorded, modeled, and optimized for—this process creates a substitute self that increasingly stands in for the real person in practical decision-making. Credit assessments, job screenings, content recommendations, and predictive policing are all examples of social domains in which a person is first encountered as a data profile. As this substitute self gains social authority, it can begin to overwrite or constrain how the real person is seen and treated.

The result is a gradual transformation of authentic Buberian *I-Thou* moments into more utilitarian *I-It* interactions (Buber 1970, 54). Instead of meeting others as irreplaceable subjects whose interior life exceeds any description, we are subtly encouraged to treat them as data points within a system of prediction and control. By filtering reality through algorithms, we lose the complexity needed for each person to tell their own unique story, ultimately leading to a “reduction of the Other to sameness and an undermining of moral responsiveness” (Levinas 1969, 197). When our encounters are pre-structured by recommendation engines and risk scores, the space for surprise, genuine otherness, and moral awakening shrinks.

The movement toward algorithmic closure takes on at least three distinct but interrelated forms. First, *epistemic closure*¹² happens when the rich diversity of human language and experience is flattened by the repetitive cycles of synthetic data and feedback loops. Research shows that models trained primarily on synthetic outputs often face “distributional collapse,” a narrowing of representable ideas that deprives the *loób* of the “substance” (*maglamán*) needed for real growth (Borj 2023, 1). As more cultural production is generated or filtered by AI systems, we risk inhabiting an increasingly homogenized informational environment where the unusual, the marginal, and the resistant become harder to encounter. This is not merely a technical glitch but a spiritual and cultural loss: a shrinking of the horizon within which persons can imagine and reinterpret themselves.

Second, *moral closure*¹³ appears when the accountability chain weakens; without a “someone” to answer for actions, the ethical story turns into simple mechanical cause-and-effect. If an algorithmic decision leads to injustice—denial of a loan, misdiagnosis in healthcare, unjust surveillance—it is tempting to say, “the system decided,” as though no agent were truly responsible. However, moral life cannot be outsourced to procedures. To the extent that we allow AI systems to occupy the role of practical arbiter while diffusing or obscuring human responsibility, we close off the very narrative in which agents can acknowledge fault, seek forgiveness, and pursue restitution. Moral progress relies on the capacity of persons and institutions to say, “We did this; we could have done otherwise; we must do better.”

Lastly, *ontological closure*¹⁴ is about reducing the person to a fixed “data-self,” which contradicts the *loób*’s nature as a self-creating, self-deepening reality. In many data-driven contexts, one’s past behavior—purchases, movements, preferences—

becomes the primary lens through which one's future possibilities are evaluated. Recommendation systems show us "more of the same," predictive models anticipate our likely moves, and risk scores follow us across domains. Over time, this can create a social expectation that a person's identity can be adequately captured by their behavioral trace. Such an expectation runs directly counter to a metaphysics of personhood in which the self is always more than it has so far expressed, always capable of reconsideration, conversion, and unexpected generosity. The *loób*, on this view, is not a static core but an inner depth that can grow beyond any given snapshot of data.

Ultimately, the human protagonist must be reclaimed as a being of presence rather than mere performance. A performance can be evaluated purely in terms of output: how fast, how accurate, how profitable. Presence, by contrast, refers to the way a person stands before others and the world as a center of initiative and responsibility. The story of the person is not a problem to be optimized but a mystery to be lived through *pakikipagkapwa*—being-with-others in a shared, mutually recognized humanity. This involves not only acting on others but also allowing oneself to be called, questioned, and transformed by them.

The irreducible gap between the self-moving "someone" and the externally driven "something" remains the final line of defense for human dignity. Even the most advanced AI cannot step into the space of genuine self-presence and moral accountability. In the face of a system that seeks to totalize reality through data, the protagonist finds their meaning in the unprompted, autotelic act of love: a movement of the *loób* that no algorithm can ever initiate or replicate. Acts of gratuitous care, forgiveness, and solidarity, especially when they run counter to predictive expectations, testify to a dimension of personhood that resists full capture by any model.

Given the ontological threats posed by algorithmic closure, modern AI ethics must shift from a sole focus on technical safety to a strong defense of metaphysical personhood. Safety measures such as robustness, transparency, and privacy protection are necessary but not sufficient. They can make harmful outcomes less likely without ever asking what kind of beings we are becoming when we increasingly live through machines. A deeper ethical orientation is needed, one that takes seriously the claim that persons are not just information processors but bearers of stories, relationships, and an inner depth that exceeds any functional description.

This framework underpins such a shift: by understanding the person as an essential "someone" with a relational *loób*, we set a clear boundary against technological encroachment. Ethical policy should not only aim to make AI more "human-like" in its outputs but must also protect the "Archimedean" space where only humans can act responsibly. This implies, for instance, maintaining domains in which final decisions about persons, especially in law, healthcare, and education, remain under the accountable judgment of human agents, supported but not replaced by machines. It also means cultivating social practices and institutions that resist treating data profiles as definitive portraits of the self.

To preserve the possibility of *pakikipagkapwa*, we must resist reducing the shared self to a data self, ensuring that the "we" remains an open, lived experience rather than a closed, computational loop. Such resistance is not just a matter of individual willpower; it requires cultural, legal, and technological frameworks that

protect spaces of unpredictability, hospitality, and genuine dialogue. Only by maintaining the distinction between someone and something can we ensure that, even in an age of widespread AI, the human story remains truly ours to tell.

RECLAIMING THE NARRATIVE: ONTOLOGICAL SOBRIETY AS A GOVERNING IMPERATIVE

To address the dominant trend of instrumentalism and the degradation of the individual's ethical narrative, the discourse must transcend merely procedural remedies, such as algorithmic audits, bias-mitigation protocols, and explainability requirements, and instead embrace what can be described as *ontological sobriety*¹⁵. This concept, rooted in a synthesis of Romano Guardini's philosophy of power, Robert Spaemann's metaphysics of the person, and the Filipino relational ontology of *kapwa* and *loób*, embodies a disciplined and unequivocal acknowledgment of the primacy of the human person. It functions as a governing imperative: a normative orientation that reaffirms the person as the sole bearer of responsibility and the singular agent of teleological self-realization.

The foundation of ontological sobriety lies in Romano Guardini's critique of *das Machtproblem*, or the modern tendency for technical power to outstrip ethical self-command. Guardini contends that technological potency becomes dangerous when it obscures the qualitative distinction between persons and artifacts, leading to an "objective power without inward measure" (2001, 83). As Guardini observes:

Modern man believed that an increase of power meant an increase of "progress" itself. Power, however, is truly a thing more powerful than any of those things. It can create evil as well as good; it can destroy as well as conduct. What happens to power depends upon man's tempered exercise of it, upon the reasoned ends to which he places it (2001, 82).

Drawing from Guardini, ontological sobriety, therefore, is the "rigorous, non-anthropomorphic recognition of technology's instrumental status" (Francis 2015, para. 115-119).¹⁶ It insists that the human person (*physis*) and the technological artifact (*technē*) are "polar opposites" whose productive relationship relies entirely upon maintaining their distinct metaphysical boundaries. When this boundary is blurred—when artifacts are treated as if they shared the ontological depth of persons—the person ceases to be the protagonist of her own narrative and becomes an adjunct to the system. The danger is not merely theoretical: once human subjectivity is functionally subordinated to systemic optimization, responsibility diffuses, and the ethical drama of decision and answerability gives way to procedural compliance.¹⁷

Building on Spaemann's defense of the person as an end-in-itself (2006, 34–53) and the Filipino ethos of *pakikipagkapwa*, ontological sobriety manifests in three interconnected mandates.

1. Rigorous recognition of volitional deficit. There must be a strict acknowledgment of AI's purely instrumental status. Because AI lacks *loób* and *voluntas*, it possesses a volitional deficit that makes it incapable of moral agency,

however complex its outputs or adaptive its behavior may appear. This mandate blocks the subtle “moral outsourcing” that occurs when we treat a computational process as if it possessed the metaphysical depth, interiority, and norm-responsiveness of a person. Attributing responsibility, empathy, or intention to AI systems, whether in customer service, governance, or care work, does not merely constitute a category mistake; it erodes our practices of moral formation by encouraging passivity in the face of technical prescriptions.

2. Subordination to the final cause. To fully grasp this subordination of technology to the human *telos*, one must first navigate the broader metaphysical landscape of Thomas Aquinas. At his core, Aquinas’ methodology is profoundly Aristotelian, a point of departure that is as foundational as it is frequently overlooked in modern discourse. To engage with his work is to step into an intellectual world where the definition of “cause”¹⁸ differs fundamentally from the one inherited by contemporary science. While modern academic philosophy often reduces causation to a linear, mechanical relationship between events, Aquinas adopts a significantly more expansive, ontological view. Following Aristotle’s lead, he argues that a complete explanation of any reality must account for a fourfold structure: the material, formal, efficient, and final causes.

By framing the “end” as the primary guide within this framework, Aquinas argues that purposes are not only a psychological projection but also a necessary structure of reality itself. This metaphysical foundation requires all human actions to be subordinated to the final cause. In Aristotelian-Thomistic terms, technology must remain entirely subordinate to the *telos* of the human person. A system’s success is therefore not primarily judged by its efficiency, predictive accuracy, or control, but by its ability to promote human relational flourishing. This involves redefining the usual measures of progress. In AI, “progress” is not just about expanding capabilities but about how well these capabilities help people and communities pursue genuinely meaningful goals. Ontological sobriety thus challenges an underlying technological determinism that sees the growth of computational power as inherently justified.

3. Mandatory accountability chain. Exclusive moral liability must remain with human agents. This anchors responsibility in the Filipino principle of *sakop*, which ties accountability to identifiable relational spheres rather than anonymous, distributed networks (Mercado 1994, 32). The design, deployment, and use of AI systems must be structured such that specific persons, institutions, and communities are clearly answerable for downstream consequences. In practice, this means rejecting architectures and governance arrangements that obscure who, concretely, bears responsibility. *Sakop* counters the fiction that harm caused “by the system” has no author and re-centers the shared inner identity of a moral community, rather than allowing it to be dissolved into the anonymity of the machine.

These three mandates illustrate that ontological sobriety is not a metaphysical luxury but an ethical necessity in conditions of escalating technical power. Without it, the human capacity for deliberation, remorse, and conversion—the very capacities that constitute moral agency—is gradually displaced by deference to opaque recommendations and automated norms.

Ontological sobriety, however, is not merely a theoretical stance; it must be operationalized through design requirements that protect the protagonist’s narrative

arc. This involves a decisive shift toward culturally situated design that prioritizes relational integrity and narrative openness over algorithmic efficiency and behavioral predictability. Such a design is not a marginal, culture-specific add-on, but a universal requirement grounded in the nature of personhood itself, which is always embedded in particular traditions and forms of life.

These design requirements include, at a minimum, the following:

1. Memory sovereignty. Memory sovereignty safeguards the self’s capacity for self-determination against the “ontological closure” produced by static data profiles. Reducing a person to a historical “data-self” is a metaphysical violation of the *loób*, as it assumes that the protagonist’s future is merely a repetition, refinement, or extrapolation of her past. When credit scores, risk assessments, or personalization algorithms lock persons into predictive categories, they effectively script futures that pre-empt the possibility of genuine change, repentance, or growth. Memory sovereignty insists that the narrative arc must remain open: that the person retains the right to transcend her data, revise earlier self-understandings, and pursue an authentic *telos* that cannot be fully captured in historical traces. Concretely, this entails robust rights to data minimization, contextual integrity, meaningful contestation, and paths to “narrative reset” that acknowledge the possibility of moral and existential transformation.

2. Anti-capture safeguards. This requirement aims to prevent the instrumentalization and commodification of deeply personal Filipino values. Specifically, anti-capture safeguards must block the use of AI in state surveillance, patronage systems, or manipulative platform architectures where data-driven obligations could undermine *katarungan* (justice) and coerce the *loób*. For example, scoring systems that rank citizens’ “trustworthiness” or “loyalty” to political patrons, or micro-targeting infrastructures that exploit *utang na loob* (debt of gratitude) for political mobilization, would constitute serious forms of relational capture. By ensuring human participation within a specific *sakop* (sphere of belonging), these safeguards make sure that technological relationships are rooted in mutual care, reciprocity, and loyalty, rather than transactional extraction and behavioral nudging (Mercado 1994, 29–38). They also require institutional arrangements—such as community oversight boards, culturally grounded impact assessments, and consent practices shaped by local moral vocabularies—that can recognize when technology is beginning to erode communal life instead of supporting it.

3. Affective attunement and *pakikiramdam* by design. Finally, design must facilitate affective attunement and *pakikiramdam*, the finely tuned, often tacit sensing of the other that characterizes Filipino relational ethics. By rejecting quantified empathy metrics and synthetic “emotional intelligence” scores in favor of system structures that require sensitive, intuitive human resonance, we ensure that technology supports rather than simulates the qualitative depth of human relationality. This “sensitive resonance” respects the *loób* of the other, ensuring that *pakikipagkapwa*¹⁹ remains a lived encounter, rather than a computational simulacrum (de Guia 2005, 70–118). Systems designed under this mandate avoid presenting themselves as substitutes for friendship, counsel, or spiritual accompaniment. Instead, they are built to create space and time for human-to-human conversation, shared discernment, and embodied presence.

Taken together, these requirements transform ontological sobriety from an abstract philosophical posture into a concrete design ethic. They resist a purely technocratic framing of AI governance, in which problems are cast as optimization challenges to be solved by better data and more sophisticated models. Instead, they relocate the debate within a richer anthropological horizon: the question is not only what AI can do, but who we are becoming as we design, deploy, and defer to these systems.

Reclaiming the narrative in this sense does not mean simply telling more positive stories about technology, nor does it amount to a nostalgic rejection of innovation. It means re-centering the human person as the protagonist within a shared moral world, such that technical systems are evaluated by their contribution to or erosion of this protagonism. Ontological sobriety asks whether our tools still serve the drama of personal and communal flourishing—or whether, under the guise of convenience and efficiency, they are quietly scripting lives that are increasingly predetermined, surveilled, and morally flattened.

For societies such as the Philippines, where relational concepts like *kapwa*, *loób*, *pakikiramdam*, and *sakop* remain vital, this framework offers both a critical lens and a constructive path forward. It shows that local ethical traditions are not obstacles to modernization but indispensable resources for imagining an alternative technological future—one in which AI is integrated without surrendering the depth of personhood and the richness of communal ties. More broadly, it suggests that any serious attempt at AI governance must grapple with the ontological status of the human person and the narrative conditions that sustain responsibility and meaning.

Ontological sobriety, then, is not simply one option among many competing regulatory philosophies. It is a governing imperative for an age in which the power of artifacts threatens to obscure the reality of persons. To remain sober, in Guardini's sense, is to remember that no matter how powerful our technologies become, they are never more than means—and that the measure of any technical order is the extent to which it preserves, protects, and deepens the story of the human person as a free and responsible agent among others.

Ultimately, ontological sobriety restores the person as the protagonist of the ethical life. It reinforces the truth that only persons possessing the capacity for self-transcendence and reflection can order actions toward a final cause. By consciously exercising moral dominion over our tools, we prevent the artifact from defining human reality. As the University of the Philippines' Principles for Responsible and Trustworthy AI (2023) suggest, meaningful human control is not just a technical requirement but a philosophical necessity that honors the metaphysical distinction between *physis* and *technē*. In this framework, the human person remains the "someone" who leads a life, ensuring that the machine remains the "something" that serves it.

CONCLUSION

The ethical challenge posed by Artificial Intelligence is, at its core, metaphysical, cutting into the very question of what it means to be a person.

Throughout this analysis, it has become clear that a strong and innovative framework for AI ethics demands a decisive shift away from “procedural fairness”—which reduces the person to a data point to be optimized—and toward “ontological imperatives,” which recognize the person as an inviolable subject. By integrating Robert Spaemann’s protagonist-centered metaphysics with the Filipino ethos of *pakikipagkapwa*, we lay down this essential foundation and move the discussion beyond safety mechanisms toward the genuine safeguarding of human dignity.

Defining the person as the “Archimedean point” and as the self-determining protagonist of ethical life establishes the human subject as a being of physis: a creature with an internal principle of motion and a narrative journey that cannot be outsourced, automated, or delegated to a machine. This agency is rooted in the non-reducible volitional interiority of *loób*—a relational will that unites reason and emotion into a coherent “I” that seeks a “we.” In contrast, AI is a product of *technē*, an artifact driven solely by instrumental rationality and external causality. Recognizing this distinction affirms AI’s subordination to human goals. AI suffers from a “volitional deficit” that no degree of computational sophistication can cure; it can calculate outcomes but cannot bear the moral weight of an act.

Ontological sobriety, then, is not merely a philosophical caution; it is a demand of *pakikipagkapwa*. It refuses to measure the sacred dignity of the human *loób*—defined by its capacity for shared moral care—against the profane arithmetic of an instrumental, algorithmic system. To surrender to “algorithmic closure” is to allow the “data-self” to devour the “living self,” cutting short the human story before it can reach its telos.

The moral health of a technological society depends on preserving the human person as the sole protagonist in the ethical story. By operationalizing relationality through memory sovereignty, the necessary accountability of the *sakop*, and the empathetic resonance of *pakikiramdam*, this framework offers a culturally grounded and philosophically rigorous alternative to dominant instrumentalist narratives. It locates personhood not in efficiency metrics or predictive accuracy, but in shared interiority and responsibility.

This approach defends the person’s role as the self-determining protagonist within a relational moral universe. Technology may set the stage, extend our reach, and even influence the conditions of our choices, but it can never rightfully claim the author’s pen. The human story is a mystery of shared interiority, of *loób* reaching out in *pakikipagkapwa*, that forever exceeds the grasp of machines. The real question, then, is not what AI will become. However, what we are willing to become in relation to it: custodians of human dignity who refuse to abdicate our moral authorship, or mere spectators as our own stories are quietly rewritten in code.

NOTES

1. The philosophical inquiry into the nature of technology in the twentieth century found one of its most profound and prescient voices in Romano Guardini. As a thinker who occupied the intersection of theology, phenomenology, and cultural philosophy, Guardini’s reflections on the transition from the modern to the post-

modern era provide a critical framework for understanding the human condition in a world increasingly mediated by technical systems. Central to this framework is the concept of “sobriety” (*Nüchternheit*), a term Guardini employs to describe the necessary intellectual and spiritual posture required to confront the technological paradigm. He uses this term in his seminal work *The End of the Modern World* (1956). Also, one can refer to his other work, *Power and Responsibility* (1961) where he discusses a clear-eyed, existential recognition of the ontological characteristics of human existence. In this article, we are introducing the term “ontological sobriety” referring to a parsimonious approach to reality.

2. See Guardini, *The End of the Modern World*, 55-56. Pope Francis, in *Laudato Si'*, comments on this, too, in Section III, par. 115, where he deals with “The Crisis and effects of modern anthropocentrism” and argues that: “Modern anthropomorphism has paradoxically ended up prizing technical thought over reality, since “the technological mind sees nature as an insensate order, as a cold body of facts, as a mere ‘given’, as an object of utility, as raw material to be hammered into useful shape; it views the cosmos similarly as a mere ‘space’ into which objects can be thrown with complete indifference.” The intrinsic dignity of the world is thus compromised. When human beings fail to find their true place in this world, they misunderstand themselves and end up acting against themselves: “Not only has God given the earth to man, who must use it with respect for the original good purpose for which it was given, but man too is God’s gift to man. He must therefore respect the natural and moral structure with which he has been endowed.”

3. This term is a proposal for a more person-centered AI ethics. This two-word concept emerged from synthesizing Spaemann’s insistence that persons are final causes, not means; Guardini’s reflections on the dangers of technological absolutism; and the Filipino relational ontology (*loób*, *kapwa*, *sakop*), which resists any reduction of the person to a functional data profile.

4. See Mercado, “Filipino Thought,” 241-242.

5. *De officiis* 1, 107; 1, 115; cf. 1, 97: *Sed poetae quid quemque deceat ex persona iudicabunt, nobis autem personam imposuit ipsa natura magna cum excellentia.*

6. Mercado’s account further explains the Filipino *loób* arguing that: “Anybody without the sense of “debt of volition” is considered “shameless” (*walang hiya*), which most Filipinos resent. Likewise, to reject a “debt of volition” leads to *hiyâ*, which may be an expression of interiority. At any rate the Filipino is quite sensitive to his treatment as a person. He wants to keep his self-respect even to the point of sacrifice. *Loób* then has much to do with the Filipino’s notion of selfhood. One often hears the saying: “Better to die than to suffer insult.” This can be formulated in various ways but the meaning remains: the Filipino’s sensitivity even to slight insults. That Americans openly exchange insults in quarrels but are reconciled afterwards as if nothing happened at all is a surprise for a not a few Filipinos to witness. The Filipino may forgive an insult but he carries the wounds for a long time. That is why when Filipinos disagree, they would prefer to use intermediaries or exchange indirect remarks rather than to be frank with each other.

7. See also Alejo, Albert E., SJ, and Julia E. Riddle. 2018. "Loob as Relational Interiority: A Contribution to the Philosophy of the Human Person." *Social Transformations Journal of the Global South* 6 (1): Article 3.

8. AI cannot possess *loób* because it is incapable of unprompted, autotelic moral action. Its "actions" are always heteronomous or governed by external programming or data.

9. See Lee, M. "Human-Computer Interaction," *EBSCO Research Starters*, 2024, <https://www.ebsco.com/research-starters/computer-science/human-computer-interaction>.

10. While *technē* encompasses knowledge and expertise, signifying being "at home in something," the resultant artifact is a product of production (*poiesis*) and is inherently instrumental. The causality of artifacts is thus purely transient. The rules, architectures, training data, and optimization functions that govern AI are all extrinsically imposed by human programmers. The AI system is entirely dependent on this external intention, confirming that it is merely a complex means, designed for a utility that exists outside of itself.

11. For a more detailed treatment of the subject, see Zuboff, S. "Surveillance Capitalism or Democracy? The Death Match of Institutional Order and the Politics of Knowledge in our Information Civilization," *Organization Theory* 3(3), November 21, 2022.

<https://journals.sagepub.com/doi/10.1177/26317877221129290>.

12. In italics for the author's emphasis.

13. Same as Note 10.

14. *Ibidem*.

15. Author's emphasis. This is the novel concept that is being proposed here.

16. Francis argues that without a "renewal of humanity itself" (*Laudato Si'*, par. 118), technology becomes an "absolute dominion" that ignores the "moral structure" of the human person. This "moral structure" is exactly what Mercado defines as the self-structuring entity of the *loób*.

17. Ultimately, the challenge of Human-Computer Interaction is not merely technical, but ontological. While modern HCI seeks to bridge the "gulf of execution," Mercado (1974) reminds us that the human *loób* is a self-structuring entity that must "deepen" (*humalim*) to hold moral weight of lived experience, a depth that raw data cannot replicate. Without this internal grounding, the "increase of power" afforded by AI remains, as Guardini (1998) warned, a force that can "destroy as well as conduct." True progress in the digital age requires what Francis (2015) calls a *sobriety*. Only by maintaining this sobriety can we ensure that our tools remain instruments of the *loób*, rather than masters of it.

18. See. *STI*, q, 33, a. 1 ad.1: "hoc nomen causa videtur importare diversitatem substantiae, et dependentiam alicuius ab altero; quam non importat nomen principii. In omnibus enim causae generibus, semper invenitur distantia inter causam et id cuius est causa, secundum aliquam perfectionem aut virtutem. Sed nomine principii utimur etiam in his quae nullam huiusmodi differentiam habent, sed solum secundum quandam ordinem."

19. See Albert Alejo, S.J.'s *Tao pô! Tulóy! Isang Landas ng Pag-Unawa sa Loob ng Tao*, 153.

REFERENCES

- Alejo, Albert. 1990. *Tao pô! Tulóy! Isang Landas ng Pag-unawa sa Loob ng Tao*. Quezon City: Ateneo de Manila University Press.
- Alejo, Albert E. SJ and Riddle, Julia E. 2018. "Loob as Relational Interiority: A Contribution to the Philosophy of the Human Person," *Social Transformations Journal of the Global South*: 6 (1): 29-53. <https://doi.org/10.13185/2799-015X.1088>
- Aquinas, Thomas. 1920. *Summa Theologiae*. Translated by the fathers of the English Dominican Province. Benziger Bros.
- Aristotle. 1995. *The Complete Works of Aristotle*. 2 vols., edited by Jonathan Barnes. Princeton University Press.
- Aristotle. 1955. *The Nicomachean Ethics*. Translated by J.A.K. Thomason. Penguin Classics.
- Borji, Ali. 2023. "A Categorical Archive of ChtaGPT Failures." *arXiv*.
- Buber, Martin. 1970. *I and Thou*. Translated by Walter Kaufmann. Scribner.
- Cicero. 1913. *De Officiis*. Translated by Walter Miller. Harvard University Press.
- De Guia, Katrin. 2005. *Kapwa: The Self in the Other*. Anvil.
- Enriquez, Virgilio G. 1986. *Philippine World-View*. Institute of Southeast Asian Studies.
- Enriquez, Virgilio G. 1992. *From Colonial to Liberation Psychology: The Philippine Experience*. De La Salle University Press.
- Floridi, Luciano. 2023. "AI as Agency without Intelligence: on ChatGPT, Large Language Models, and Other Generative Models." *Philosophy and Technology*. <https://doi.org/10.1007/s13347-023-00621-y>.
- Floridi, Luciano, Josh Cowls, Monica Beltrametti, et al. 2018. "AI4People- An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations." *Minds and Machines* 28 (4): 689-707.
- Guardini, Romano. 1998. *The End of the Modern World*. Translated by Joseph Theman and Herbert Burke. ISI Books.
- Heidegger, Martin. 1977. *The Question Concerning Technology, and Other Essays*. Translated by William Lovitt. Harper and Row.
- Jocano, F. Landa. 1997. *Filipino Value System: A Cultural Definition*. Manila: Punland Research House.
- Lee, M. 2024. "Human-Computer Interaction." *EBSCO Research Starters*. EBSCO Learning.
- Levinas, Emmanuel. 1985. "The Face-to-Face." In *The Levinas Reader*, edited by Seán Hand, 83-90. Blackwell.
- MacIntyre, Alasdair. 1984. *After Virtue*. University of Notre Dame Press.
- Macmurray, John. 1957. *The Self as Agent*. Faber and Faber.
- Malo, Antonio. 2004. *Il senso antropologico dell'azione*. Amando Editore.
- Maninang, S. 2025. "On Algorithmic Vulnerability and the Erosion of Kapwa: Towards a Filipino Relational Ethics for Recommender Systems." *Journal of Technology and Society* 40 (1): 1-25.
- Maturana, Humberto R., and Francisco, J. Varela. 1980. *Autopoiesis and Cognition: The Realization of the Living*. D. Reidel.

- Mercado, Leonardo. 1972. "Filipino Thought." *Philippine Studies* 20 (2): 207-272. Ateneo de Manila University Press.
- Mercado, Leonardo. 1994. *The Filipino Mind*. Council for Research in Values and Philosophy.
- Mercado, Leonardo N. 1994. *Elements of Filipino Philosophy*. Divine Word University Publications.
- Pinckaers, Servais. 1995. *The Sources of Christian Ethics*. Translated by Mary Thomas Noble. Catholic University of America Press.
- Reyes, Jeremiah. 2015. *A Filipino Virtue Ethics*. Ateneo de Manila University Press.
- Spaemann, Robert. 2006. *Persons: The Difference Between Someone and Something*. Translated by Oliver O'Donovan. Oxford University Press.
- University of the Philippines. 2023. "UP Principles for Responsible and Trustworthy Artificial Intelligence." *Board-approved by the UP Board of Regents at its 1383rd meeting*, 24 August 2023. *UP Gazette*, 54 (7): 8-11. <https://osu.up.edu.ph/wp-content/uploads/2024/01/1383-FINAL-GAZETTE.pdf#page=21>
- Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs. <https://journals.sagepub.com/doi/10.1177/26317877221129290>