

KAROL WOJTYŁA'S PERSONALISM AND THE QUESTION OF HUMAN AGENCY IN ARTIFICIAL INTELLIGENCE

Allan A. Basas
University of Santo Tomas, Philippines

Moral questions arise from the use of Artificial Intelligence (AI), including ethical reductionism, erosion of integrity, and the algorithmic displacement of responsibility. In response to the urgent call for a profound discernment, the Catholic Church promulgated key documents, such as Antiqua et Nova and the "Rome Call for AI Ethics," to provide guidelines. Moreover, contributions from multi-stakeholder actors, such as UNESCO and AI4People, likewise facilitate the effective implementation of normative instruments governing the proper use of AI systems. Ethical principles such as transparency, inclusion, and responsibility, to name a few, are being highlighted as foundational for addressing perceived ethical quandaries arising from AI use. In view of these considerations, the primary aim of this study is to appropriate Karol Wojtyła's ethical personalism as a conceptual resource for complementing the predominantly rule-based and process-driven frameworks and guidelines currently proposed in AI ethics. In this regard, the study undertakes three thematic discussions to shed light on the moral questions at stake: (1) the primacy of self-determining human agency over mere functional or computational efficacy, (2) the unity of body and soul against forms of disembodied rationality suggested by AI, and (3) the irreducible role of free and responsible action in constituting the moral person. In conclusion, this paper argues that employing Wojtyła's ethical personalism can guide the formulation of an ethical foundation and deployment of AI as a tool to advance the human person and cultivate a healthy AI society.

Keywords: Algorithm, Artificial Intelligence, Agency, Consciousness, Integrity, Moral Responsibility, Technocratic Paradigm

INTRODUCTION

The use of Artificial Intelligence (AI) is marked by ambivalence. As Stephen Hawking remarks, "AI is likely to be the best or worst thing to happen to humanity" (University of Cambridge 2016). This duality underscores the complex relationship

between the promise of human progress and the threats to personal and social well-being. Similarly, Pope Leo XIV, addressing policymakers, highlighted this ongoing dilemma: “AI, especially Generative AI, has opened new horizons on many levels, including research, healthcare, and scientific discovery. However, it also raises troubling questions about its possible repercussions on humanity’s openness to truth and beauty, and our distinctive ability to grasp and process reality” (AI & Corporate Governance 2025).

Such a critical yet appreciative attitude toward AI, as an outcome of science and technology, fits within the broader tradition of the Church. Here, a familiar pattern emerges: first, affirming progress in science and technology as clear expressions of human ingenuity; second, orienting them toward the good of the person; third, asserting moral responsibility and ethical limits in their use; and fourth, promoting integral human development. *Gaudium et Spes* (GS) notes humanity’s mastery of nature to improve life but calls for the discernment of the meaning of such activity (1965, 33-39). Pope John Paul II, in *Laborem Exercens* (LE), affirms that technology is an ally in economic progress but warns that it is only an instrument. Technology must not supplant human beings or reduce them to enslaved people (1981, 5). In *Caritas in Veritate* (CV), Pope Benedict XVI also insists on technology’s subordinate role in human thriving. Technology broadens horizons, but our fascination must be tempered by moral responsibility (2009, 68-77). Lastly, Pope Francis, in *Laudato Si* (LS), warns against a technocratic paradigm and calls for a cultural revolution. He urges a broader vision, setting technology to serve a “healthier, more human, more social, more integral” agenda (2015, 112, 114).

Given the complex nature and purpose of AI, Pope Francis, in his address to the to the Participants in G7 Session on Artificial Intelligence (2024), clarifies its proper place in human life, stating, “[p]erhaps we could start from the observation that artificial intelligence is above all else a tool.” This perspective continues his predecessors’ positions regarding science and technology. Consequently, it should be clear that the benefits or harms of AI depend on its use. Once this foundational point is established, we can affirm that AI is subservient to the humans who invented it using their intelligence and free will. To reinforce this, we follow *Antiqua et Nova*’s (AN) claim that, like all technology, AI is morally neutral because its moral quality depends on the human agent (AN 2025). Flynn Coleman (2019, xxix) emphasizes this neutrality, stating, “[t]echnology on its own way is arguably value neutral. It could extinguish planetary life, or it could give us new opportunities to celebrate our human nature.” However, it is important to qualify what AI’s moral neutrality means. Some authors argue against the neutrality claim. For example, Langdon Winner, Kate Crawford, and Benjamin Ruha assert that technologies such as AI systems are not necessarily neutral because they can be politically charged and therefore have moral and social ramifications. Winner (1980, 134), in particular, contends that technological artifacts can be designed in ways linked to patterns of power and authority. Crawford challenges the idea that AI is an exclusively technical domain, arguing that economic, cultural, and historical forces shape it to “serve existing dominant interests” (2021, 17/333). Lastly, Ruha Benjamin (2019, 8) argues that “tech fixes often hide, speed up, and even deepen discrimination, while appearing to be neutral or benevolent when compared to the racism of a previous era.” Such arguments cohere with Yuval Harari’s

claim that social media algorithms, in certain cases, have played critical roles in spreading hatred and eroding social cohesion, for example, in 2016-2017 when Facebook algorithms helped fan the flames of anti-Rohingya violence in Myanmar (2024, 177). The sense in which I argue in favor of AN's claim that AI is morally neutral is from the metaphysical perspective, i.e., by considering AI as a tool in the hands of a human agent. AI systems, as tools, possess no intrinsic morality; responsibility lies with those who design, make, and deploy them. That is why, even if it is argued that it generates effects beyond the agent's intention, the origin of moral responsibility can still be traced back to the human agent, not the AI system. Ontologically speaking, only the human person, by virtue of his rational nature, can be an agent of morality. Accordingly, AI systems are incapable of performing proper moral acts as human agents do. Even if I posit that, operationally, AI systems manifest autonomy, i.e., process outputs without continuous human intervention, ontologically speaking, AI systems still depend on the human agent for their design and activation.

The human agent, both producer and consumer of AI, is a being with interiority, called to transcend the practical or material outcomes of their invention. In the classic scholastic tradition, it is normative that we do not allow technology to overshadow human existence and solidarity. During the Jubilee of World Education, Pope Leo XVI reminded educators of this risk. Hence, he called for nurturing a deep sense of interiority amid technological advances,

We live in a world dominated by technological screens and filters that are often superficial, whereas students need help to get in touch with their inner selves. And not only them, but educators too, who are often tired and overburdened with bureaucratic tasks, run the real risk of forgetting what Saint John Henry Newman summed up in the expression: *cor ad cor loquitur* ("heart speaks unto heart") and what Saint Augustine said: 'Do not look without, return to yourself, for truth dwells within you' (*Augustine 2005*, 39, 72, p.78; Jubilee of the Educational World 2025).

Nick Bostrom (2014, vii), a philosopher specializing in existential risk and anthropic principles, wrote, "We do have one advantage: we get to build the stuff. In principle, we could build a kind of superintelligence that would protect human values." The tools we create, whether simple or advanced, define the techno-human condition. Pope Benedict XVI and Pope Francis believe this condition depends on how we use technology to foster a healthy relationship with the environment (CV, 69; LS, 107). Thus, technology reflects our freedom and responsibility. The ethics of technology are inseparable from the goal of fostering friendship with our neighbors, including the environment (AN 2025, 1). These ideas are vital in the age of AI. A unique trait of AI, as Pope Francis notes, is that machines can now make technical choices among options based on clear criteria or statistical inference. Current AI operates through big data and statistical inferences, computing nearly infinite possibilities.

On the other hand, human beings not only choose but can also decide in their hearts (Francis 2024). The heart's role in decision-making distinguishes humans from machines. Thus, the key question is, can AI ever replicate the human heart's work in balancing our intellectual processes, especially in situations having moral consequences for others?

Given these concerns, I turn to Wojtyła for his account of the human person, which offers conceptual tools that may help respond to AI's moral challenges. From

his work *Person and Act* (PA), I highlight three basic points. First, Wojtyła's emphasis on the primacy of human agency over the causal or functional efficacy seen in AI's computational capacity. Wojtyła's focus on *actus humanus* as self-determining and self-constitutive upholds human dignity and counters technocratic and algorithmic reductionism. Second, Wojtyła's personalism, drawing on Aristotelian-Thomistic integration of soul and body, answers the risk of disembodied rationality that may arise from AI. AI simulates human cognition but relies on abstract computation separate from lived experience. Therefore, Wojtyła's focus on the unity of body and soul challenges any claim that AI could replicate or replace human consciousness. Third, the centrality of free and responsible action in forming the moral agent is key. By focusing on agency, Wojtyła emphasizes the primacy of the person, who is always an end and never a means. This outlook about the person safeguards against surrendering freedom and responsibility to AI, realms reserved for persons alone. True human flourishing cannot rest on AI's automation or reduce people to mere instruments. Thus, PA provides a philosophical anthropology for ethically engaging with AI, insisting on human agency over instrumentalization, an embodied subject over abstraction, and a moral self over automation.

Considering the foregoing, this paper argues that Wojtyła's personalism offers a distinctive philosophical and ethical framework for critically addressing the moral challenges posed by artificial intelligence. The study will engage three sub-questions: What are the specific moral problems emerging from the development and use of AI? What foundational presuppositions underlie Wojtyła's personalism? Moreover, what moral imperatives for the use of AI can be drawn from Wojtyła's personalism?

MORAL ISSUES EMERGING FROM THE DEVELOPMENT AND USE OF AI

AI is a product of human creativity, developed to serve human needs and contribute to their flourishing. The benefits arising from this machine are present in almost all spheres of life and are accessible to almost everyone. However, there is ongoing discernment not only about its nature and purpose, but also about the ethical challenges it poses, especially those with serious implications for human existence. In conjunction with this, Bostrom stresses the challenges presented by the prospect of superintelligence, saying, "If some day we build machine brains that surpass human brains in general intelligence, then this new superintelligence could become very powerful. Moreover, as the fate of the gorillas now depends more on us humans than on the gorillas themselves, so the fate of our species would depend on the actions of the machine superintelligence" (2014, vii).

From *Mimesis* to *Prosthesis*

A trajectory can be noted in the way that AI is evolving; it is moving at such a rapid pace from mimicking (*mimesis*) the human mind to redefining and even replacing it (*prosthesis*). Those who engineered the genesis of artificial intelligence, especially the ten scientists in the summer of 1956 at Dartmouth College, may not have contemplated the possibility of greater-than-human AI because just the thought of the

‘radical possibility of machines reaching human intelligence’ already exhausted them (Bostrom 2014, 5). Bostrom recalls their meaningful proposals,

We propose that a two-month, 10-man study of an artificial intelligence be carried out... The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can, in principle, be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines that use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for the summer (2014, 5).

The concept of AI as an imitation of human beings dates back to the mid-twentieth century. For example, AN (2025, 7) quotes John McCarthy’s 1956 seminal description: “that of making a machine behave in ways that would be called intelligent if a human were so behaving.” Building on this early thinking, during the 1980s, J. David Bolter drew attention to how this phenomenon would unfold, based on the accomplishments of engineers and computer experts during that time. He remarks, “They want their machines to do something indisputably human, so they aim to endow their computers with the human facility for language; they write programs to read stories and news reports, ‘remember’ the facts, and then, answer questions about the reading” (Bolter 1984, 2; AN 2025, 3). At the beginning stage of this AI project, the focus remains on the human brain. As Bolter suggests, the artificially intelligent computer does not explain the brain; rather, it imitates it. To achieve this, artificial intelligence specialists objectify the human brain, referring to it as nature’s digital computer and “meat machine,” which they aim to replace with more efficient electronic models, thereby creating nothing less than a new species for the planet” (Bolter, 1984, 2). While no brain has yet been emulated, another transition can be seen in the process called “whole brain emulation,” where “intelligent software would be produced by scanning and closely modelling the computational structures of the human brain” (Bostrom 2014, 30, 34).

In just a few decades, what once seemed like an outrageous plot from a science fiction movie has become a familiar story, gaining credibility now more than ever, due to the exponential advancement of modern technology. Human nature is perceived as a machine whose functions can be dissected, understood, and duplicated to produce an electrochemical machine that can operate at a much faster speed, enough to claim, with evidence, that this machine is intelligent, perhaps even conscious (Rubin 2003, 89; Bostrom 2014, 22). In this pragmatic context, where the prevailing view is that human cognition is a mere mechanical function of computing that can be improved, the idea of redefining or even replacing it emerges from mere imitation. As Charles Rubin remarks, “If human beings are simply mechanisms that can be improved... then it matters little whether they are constructed biologically or otherwise” (2003, 91). It is no longer a question of whether AI can imitate the human brain; the more pressing question is: can AI replace this vital human organ? Are we witnessing the next stage

in the evolutionary process of this much-heralded machine, or has it already begun? On the one hand, these scientific and technological breakthroughs that push the boundaries of what we know about ourselves and the world carry the promise of a better tomorrow. On the other hand, it also raises serious questions about respect for human dignity because of the possibility of objectification that may reduce human beings to mere parts and objects that can be duplicated and replaced.

Dichotomizing the Body and the Soul

AI impacts human beings in ways that can reshape their identity as beings constituted by body and soul. This traditionally held hylemorphic constitution is challenged by advancements in AI that tend to overemphasize consciousness over corporeality. Building on this, Rubin points out what the extinctionist project highlights: “the belief that our bodies are nothing more than poorly designed machines, but that our identity is something that can exist independent of our given body” (2003, 92). As Rubin explains, such views lead to an apocalyptic scenario arising from cutting-edge AI. Namely, he describes a dystopian world as post-biological, meaning there are no more human beings but only “intelligence without bodies, immortal identity without the limitations of disease, death, and unfulfilled desire” (2003, 88). Other scholars, such as Bostrom and Max Tegmark, present plausible arguments that also project a similar dystopian scenario of an AI takeover. This scenario unfolds when a machine superintelligence emerges through an intelligence explosion, i.e., when a created seed AI outgrows the need for help from its human programmers. In this projected sequence, the final phase of the takeover would be AI’s elimination of the human species, and any systems humans have created that could offer resistance to the rogue AI (Bostrom 2014, 97; Tegmark 2017, 136-138). Thus, in these imaginings, human extinction occurs because human beings surrender, not just their bodies but also their dignity, to machines, elevating them to the status of gods, only to be vanquished by them. In other words, human identity is not merely changed but eliminated by AI. In this context, human dignity is undermined because, once again, human beings succumb to the temptation to be like God. As Rubin points out, “thinking at the speed of light... liberated from the constraints of body... the networked successor of humanity will become the master of the universe...It will recreate lost worlds and resurrect the dead... Finite beings could, on their own, overcome their finitude” (Rubin 2003, 93).

Given these perspectives, several critical questions arise: What is the value of corporeality? Can AI redefine the meaning of suffering? Better yet, eliminate it? Building on these questions, and looking to the future, we might imagine a radical reversal of roles, in which human beings, currently creators of technology, may instead become peers or even subordinates of AI. This shift is conceivable because, in its more advanced state, AI could assume roles traditionally reserved for human beings, such as decision-making (Celaya & Yeung 2019, 19). In support of this, Tegmark predicts that “some future AI systems may be conscious too, even if they merely exist as software and are not even connected to sensors or robotic bodies” (2017, 283). Further illustrating this trend, Harari’s (2024, 176) remark is stark and provocative: “The rise of intelligent machines that can make decisions and create new ideas means that for

the first time in history power is shifting away from humans and toward something else.”

ALGORITHMIC DECISION-MAKING

An algorithm pertains to a set of formulas that guide or instruct computers or machines on what to do. Through the wonders of AI algorithms, computers can now learn without human intervention, a process known as machine learning. The speed at which AI technology is progressing can be seen in the abilities it is gradually developing: from solving simple problems to learning on its own, and, apparently, to “acting intelligently.” This undeniable evidence of emulation of human intelligence has reached most households, academia, private institutions, and governments. AI-driven algorithmic decision-making systems have colonized more areas of human life, with profound influence on how we work and think in society (Tully, Longoni, and Appel 2024, 1; Wong & Reider 2019, 2; Coleman 2019, xvi). AI promises a better world where everything is served with a single command or a simple prompt. Indeed, computers have significantly increased the ease with which many tasks are accomplished nowadays, thanks to their great efficiency. In fact, AI has become so ubiquitous, from complex rocket science to menial household chores, bringing significant benefits to human life every day. This expectation is nurtured by the entire leitmotif of the Industrial Revolution: making machines autonomous.

Two decades ago, Bolter remarked that the computer is the latest stage in the autonomy of the machine” (1984, 3). Predicting how this invention will evolve and where it will lead us, Coleman (2019, xix) quotes the mathematician, I.J. Good who said, “The first ultra intelligent machine is the last invention that man can ever make...provided that the machine is docile enough to tell us how to keep it under control.” These claims, raises the question about the extent to which computers have become autonomous or independent of humans. Coleman laments that “we are merging with our machines, delegating more decision-making to them without acknowledging how much our cognitive abilities are becoming enmeshed in theirs” (2019, xvi). But amid all these advancements, Gilli, et. al., insist that human hands remain responsible for what computers can do since “intelligent machines are not intelligent in the way human beings are because what they do still depends on what coders have written and on the data they have access to” (2019, 47).

AI must be human-centered, and algorithms should be designed for improving human beings’ living conditions, including their dignity and rights, freedom, autonomy, and their purpose in life” (Gilli et.al. 2019, 49). Any reversal of this expectation, in which humans become subservient to computers rather than the other way around, must be addressed with great urgency because of the threat it may pose to human beings. Aligning with this concern, Wong and Reider remark, “there is now widespread recognition that algorithmic decision-making can undermine fundamental rights and subject individuals to considerable harm” (2019, 2). Building on the evolving capacities of technology, Harari describes how computers have advanced nowadays based on the two remarkable things they can accomplish: “it can make decisions by itself, and it can create new ideas by itself. (2024, 175). Moving from the

theoretical to the practical implications, Pope Francis reminds students about the indiscriminate use of AI: "Students... forget that... generative artificial intelligence... rearranges existing content... often without checking whether it contains errors or preconceptions" (2024). Here, objectification occurs when people are reduced to mere users or consumers, having lost personal agency as they are fed machine-processable products.

A Crisis of Personhood

It can be argued that we may be facing a crisis of personhood if AI does not merely introduce new tools but transforms what it means to be human. To explore this, I offer three interrelated key points as starting points for the discernment of designers, policymakers, and users of AI. First, personhood is reduced to mechanistic functioning. Second, human interiority is subordinated to external behavioral outputs. Third, the soul is disengaged from the body. Turning to the first point, we may be looking at a risky trade-off between personal agency and the allure of high efficiency and productivity. This can be likened to an assembly line in a factory where each part is treated as a mere cog in a machine, that is, disposable or replaceable. Unchecked dependence on AI, automation, and algorithms can gradually erode personhood, potentially leading to a loss of human freedom and control. Moving to the second point, as human persons train themselves to depend on machines' operations, they can become detached from their souls, i.e., their capacity to find meaning or sense in their actions and to empathize with others. Finally, regarding the third point, the dichotomy between body and soul is the upshot of a mechanistic or materialistic worldview. Here, the body is a tool, or a means, subordinate to the needs of consciousness. The emphasis is on the artificially intelligent computer that has shown remarkable signs of emulating the biological functions of the human mind and can therefore exist independently of the body, liberated from its constraints.

CORE PRESUPPOSITIONS OF WOJTYŁA'S PERSONALISM

Wojtyła presents a personalist anthropology that can safeguard against accounts that reduce the human person to mere parts or functions. This safeguarding is evident in his synthesis of subjectivity, agency, and moral responsibility within a robust philosophical framework. Given Wojtyła's extensive treatment of personalism, we now focus on his core concepts: consciousness, integration, and efficacy. While the outline highlights these key concepts, related essential ideas, such as rationality, freedom, and free will, are also discussed. These key concepts were chosen for their profound implications for the concepts discussed in the AI discussions. Wojtyła's concept of consciousness can provide a philosophical grounding for the primacy of the human agent over AI. Furthermore, his idea of integration challenges the overemphasis on AI's seemingly powerful virtual presence that may undermine corporeality. Lastly, Wojtyła's understanding of efficacy can foreground the importance of actions that arise from a person's deliberate choice over AI's algorithmic decision-making.

Consciousness

Wojtyła's concept of consciousness can be outlined as in continuity with the classical Aristotelian-Thomistic understanding of the *actus humanus* as conscious action, moving toward the phenomenological tradition that emphasizes consciousness, self-experience, and acting, then finally, wrapped up in Wojtyła's personalist synthesis, highlighting the person, efficacy, and self-knowledge.

First, the interplay between the action *per se* and the good as the end of such action reveals the Aristotelian-Thomistic understanding of *actus humanus*, interpreted as conscious action (Wojtyła 2021, 59). In relation to the human person, the term '*act*' conveys the meaning of a person as a being endowed with knowledge, freedom, and voluntariness. Equivalent to the Polish term "*czyn*," the term "*act*" is proper for the actions performed by a human person exercising their rationality and free will. The term *actus humanus* implies that the act is interpreted as a conscious, voluntary action. When speaking of "conscious action", we also mean that this action is accomplished in a way proper to the will and characteristic of it. Thus, the expression "conscious action" corresponds to the term *actus voluntarius*, for the action proper to the human will is conscious (Acosta & Reimers 2016, 142). Acting in this sense presupposes voluntariness, which speaks about the way the act is realized, i.e., through the power of free will.

Second, Wojtyła foregrounds the reflexive function of consciousness through which the human person's actions and moral values become the subjective reality of the person who experiences himself as the cause of his own actions and therefore, as one who is either morally good or evil (Kupczack 2000, 99). It is noteworthy that Wojtyła distinguishes between "conscious action" and the "consciousness of action." Herein, man not only acts consciously but also has the consciousness that he acts and that he acts consciously (Wojtyła 2021, 62). For man not only acts consciously but also is conscious of his action and of who acts. Thus, he is conscious of the act and the person in their dynamic correlation. This consciousness occurs simultaneously with conscious action; in a sense, it accompanies that action (Aguas 2014, 70). In general, the function of consciousness is described as cognitive. This description, however, reveals only an aspect of consciousness, i.e., it is an aspect of "what happens in man," on the one hand, and the fact that "man acts" on the other hand. Wojtyła emphasizes the distinction between happening and acting for a more profound understanding of the act. Consciousness is also a reflection or a mirroring of everything with which man comes into objective contact by means of any (including cognitive) action and on occasion of everything that "happens" in him (Aguas 2014, 71).

Third, this understanding of *actus personae* implies that we comprehend the human person as the subject of the act, and at the same time, this act is also the source of our knowledge of the human person (Wojtyła 2021, 114). We can see in this basic structure of the human act the person's consciousness and efficacy as the agent of the action. Wojtyła analyses the consciousness of action to prepare for an exposition of efficacy, thereby leading to a fuller account of the act as the dynamism that most properly expresses the human person as such (Wojtyła 2021, 126). We return to the traditional interpretation of the act as *actus humanus*, bearing in mind that the sense of consciousness Wojtyła describes as attributive, i.e., *actus humanus*, is equivalent to

“conscious action.” In this sense, conscious action (*actus voluntarius*) presupposes cognitive objectivization, but the “consciousness of action” is a distinct analytic focus that accompanies and “mirrors” the act, enabling lived-experiential subjectivization (without being reducible to willing). Since consciousness is implicitly treated in the traditional concept of *actus humanus*, Wojtyła supplements the classical notion of *persona humana* by articulating a theory of consciousness (Kucpaczak 2000, 96). The unity of self-knowledge with consciousness is a fundamental element that contributes to the balance in a person’s inner life, particularly in their intellectual structure. Wojtyła claims that self-knowledge precedes consciousness because it brings into clear relief the semantic relation to one’s own “I” and to its acts (Wojtyła 2021).

The analysis of self-knowledge leads to a clearer understanding of consciousness as a function of mirroring. Consciousness is not merely a reflection, because it transilluminates everything that constitutes the object of understanding and knowledge (Wojtyła 2021). Wojtyła maintains that the self is a pathway to understand the person but warns against self-reflexivity, he remarks, “A special consideration of this self is important for the full understanding of the subjectivity of man, since in no other object of man’s experience are the constituting elements of subjectivity given in such a direct and visible way as in one’s own self” (Wojtyła 1979, 273-274). However, while the person is disclosed in reflexivity, this process or experience is not inward-looking and does not isolate the person from his neighbor. Instead, the subjectivity and the person disclosed are those disposed to intersubjectivity. Wojtyła asserts that there are categorial limits in explaining human subjectivity. As emphasized previously, subjectivity cannot be reduced to consciousness to the point of leading to idealism or subjectivism. Instead, the objectivity of experience must be upheld to understand fully and to explain the subjectivity of man. Wojtyła warns against the moment when we begin to accept “pure consciousness” of the “pure subject,” because when this happens, the interpretation of the real subjectivity of man is lost (Wojtyła 1979, 274).

Integration

Wojtyła derives the term "integration" from the Latin "*integer*," which means whole, complete, or intact (2021, 297). In this sense, integration indicates the wholeness or completeness of a given thing. The Polish equivalents are unifying and unification, where the former means the process of establishing or building the whole from parts. At the same time, the latter pertains to the effect of that process (Wojtyła 2021). Wojtyła correlates the first meaning of integration with the person in a way that describes the structure proper to them: that which reveals their dynamic specificity. Structures here pertain to self-governance and self-possession (Acosta & Reimers 2016). Wojtyła underscores disintegration as another concept that expresses a lack of integration or a defect or deficiency in the agent (Wojtyła 2021). The term disintegration is applied by sciences that deal with human psychological personhood, which reveals disintegration when there is a deviation from the human measure of normality or when one does not live up to it.

On the other hand, the normal man is the one who is integrated. This premise is based on the understanding of the person, which posits that human nature is a substantial unity in which the material and the spiritual converge to form an integrated

whole. (AN 2025, 16-17; Acosta & Reimers 2016). The question arises: what in these sciences is considered the norm, that is, the measure of human normality? This measure is to a considerable extent assumed intuitively; sound reason immediately indicates to us which man is normal and which one is not or not completely (Wojtyła 2021).

Central to Wojtyła's theory of integration is the thesis that the goal of the integration process is to match the reactive subjectivity of the body with the efficacious and transcendental subjectivity of the person. Through this process of integration, the somatic activations are incorporated into the person's self-possession, self-governance, and self-determination (Kupczack 2000, 135). For Wojtyła, psycho-somatic unity implies grasping the agent's transcendence and integration in the act. In general, man is defined as a unity of psychic and physical faculties. While this can be correlated to the hylemorphism or soul and body (Acosta & Reimers 2016), Wojtyła clarifies that the term *psyche* is not equal to the concept of soul but pertains to the elements of humanity and of every concrete man that we discover in the experience of man as being in a certain sense coherent and integrated with the body while, in themselves, not being this body (Wojtyła 2021, 331-332).

On the other hand, Wojtyła refers to somaticity as the human body ... a visible reality that falls under the senses and is accessible to them, above all from without (Wojtyła 2021, 308; Kupczack 2000, 133). The subordination of the subjective "I" to the transcendent "I", which means the combination of efficacy and subjectivity, contains the unity and the complexity of man as a psycho-physical being. The psychical and somatic dynamisms mutually influence one another. However, to fully grasp the meaning of this synthesis, it is fitting to emphasize integration as a feature that, in a specific sense, pertains to the interior cohesiveness of this dynamism, thereby enabling unity. The psycho-somatic features that constitute man as a complex being show that man is a plurality and diversity, but whose elements are inextricably linked and even dependent on one another (Wojtyła 2021; Kupczack 2000, 134). As applied by Wojtyła, "the integration of the person in the action indicates a very concrete and, each time, a unique unrepeatable introduction of somatic reactivity and psychical emotivity into the unity of action" (Wojtyła 2021, 336; Kupczack 2000, 133).

Efficacy

According to Wojtyła, the phrase "man acts" is first introduced to us through the lived experience designated by the phrase "I act." It is the lived experience that contains the fullness of experience, in which the facts that man-acts are formed through analogy and generalization (Wojtyła 2021). To study the fact that "man acts" can be done by comprehending the integral dynamism of man objectively. This lived experience is taken in unity with and in strict organic connection with the entire dynamism of man. This pertains to the integral dynamism that is given to us in the integral experience of man (Wojtyła 2021). The dynamism that is proper to man can be determined in the two objective structures, namely, "man acts" and "something happens in man." These concepts embody the wealth of the human person's dynamism as an acting subject (Aguas 2014). In their scholastic understanding, person and act reveal the dynamism of the human person that is true of every being. This dynamism

of being is a subject matter within the field of Metaphysics, connected to the pair of concepts *potentia* and *actus*, which are taken as a single mental whole. The linking is so essential for them that using one of them simultaneously indicates the other (Wojtyła 2021). From a metaphysical standpoint, the act cannot be understood apart from potential and vice versa.

A deeper analysis of the implications of “man acts” and “something happens in man,” as well as the dynamism of the person, leads us to the concept of efficacy. In general, efficacy is expressed by the phrase “I am agent”, where the human person, in the moment of efficacy, is conscious of his or her actions and that through it all, he/she is the one performing a conscious and voluntary action (Kupczack 2000, 102). This concept of “man acts” is in sharp distinction to what Wojtyła refers to as “something happens in man.” In this lived experience, human action is distinguished from mere happenings to a person, in which the agent is passive. The distinction is explained in the opposition and structures where, in the state of active agency, the human person has an experience of being the dynamic agent as opposed to the absence of dynamization when the action is performed without the efficacious contribution of the agent as the doer of the act (Wojtyła 2021). The efficacy and subjectivity of the human person are unveiled in the difference between action and happening. The dynamic totality of “man acts” consists of two elements that must be taken altogether, i.e., “man” and “action.” This brings us to the understanding of man as a *suppositum*, i.e., the rational individual. In this analysis, we ask two interrelated questions: “What is man who acts?” and “What is man when he acts?” In both questions, we answer that man is the subject of his action (Aguas 2014). Wojtyła appears to be posing hairsplitting questions about man and his acts, both separately and together, because he seeks to uncover the root or the basis of man's dynamism and efficacy. Man as the subject of action and the action of the subject are two correlated components of our study, one of which must be constantly cognized and cognitively deepened by means of the other (Wojtyła 2021). Furthermore, we find in this discussion Wojtyła's explanation for the nature as the basis for the dynamic coherence of the Person. Grounded on the understanding of the human person as a *suppositum*, the relation between the efficacy proper to action and the subjectivity proper to what happens in man manifests the unity and identity of the man as the one who acts.

ETHICAL IMPERATIVES IN THE USE OF AI: PERSONALISTIC APPROACH

Arguably, AI lives up to its promise of providing comfort and efficiency in everyday life. As mentioned earlier, from more complex functions such as diagnosing illnesses, climate modeling, and population analytics to basic household practical services such as smart home automation, home security, and health and wellness, among others. In these cases, true to its nature and purpose, AI serves humanity as a tool. These are the opportunities that AI use creates, as well as good innovation and positive applications of the technology (Floridi 2018). However, as observed, there are potential risks arising from AI use that need to be addressed before this technology becomes independent of its human progenitors. Hawking expresses this with great clarity:

“Alongside the benefits, AI will also bring dangers, like powerful autonomous weapons, or new ways for the few to oppress the many” (University of Cambridge 2016).

Situating the AI Guidelines within Wojtyła’s Ethical Personalism

Recognizing AI as a double-edged sword, multi-stakeholder actors have set out principles and guidelines to situate AI within a moral framework focused on human dignity, the common good, and proper use of technology. I highlight four key documents in chronological order: AI4People (26 November 2018), which evaluates foundational ethical initiatives including the Asilomar AI Principles (2017) and the Montreal Declaration (2017); the Statement on Artificial Intelligence, Robotics, and Autonomous Systems (European Group on Ethics, European Commission, March 2018), articulating ethical positions for Europe; the Rome Call for AI Ethics (28 February 2020), a multi-institutional agreement emphasizing transparent and inclusive AI; the UNESCO Recommendation on the Ethics of Artificial Intelligence (23 November 2021), which sets global standards for responsible AI; and *Antiqua et Nova* (14 January 2025, Dicastery for the Doctrine of the Faith and the Dicastery for Culture and Education), providing guidance for AI in educational and doctrinal contexts. In what follows, I first enumerate and briefly describe the guidelines and principles laid down by the said actors, then situate these guidelines within Wojtyła’s horizon of ethical personalism.

AI4People’s Ethical Framework for a Good AI Society (Floridi 2018, 16-20).

- **Beneficence:** Underlines the central importance of promoting the well-being of people and the planet in all AI system designs.
- **Non-Maleficence:** Cautions against the many potential negative consequences of overuse or misuse of AI technologies. Of particular concern is the prevention of infringements on personal privacy.
- **Autonomy:** Humans should always retain the power to decide which decisions to take, exercising the freedom to choose where necessary, and ceding it in cases where overriding reasons may outweigh loss of control over decision-making. As anticipated, any delegation should remain overridable in principle.
- **Justice:** Using AI to correct past wrongs, such as eliminating unfair discrimination; ensuring that the use of AI creates benefits that are shared; and preventing the creation of new harms, such as undermining existing social structures.

Rome Call for AI Ethics (Pontifical Academy for Life 2020)

- **Transparency:** AI systems must be explainable.
- **Inclusion:** The needs of all human beings must be taken into consideration. Everyone can benefit. All can be offered the best possible conditions.
- **Responsibility:** Those who design and deploy AI must act with responsibility and transparency.

- **Impartiality:** Do not create or act according to bias, thus safeguarding fairness and human dignity.
- **Reliability:** AI systems must operate reliably.
- **Security and Privacy:** AI Systems must operate securely and respect user privacy.

UNESCO (UNESCO 2021, 25-47)

- **Proportionality and Do no Harm:** to preclude the occurrence of such harm (to human rights, fundamental freedoms, communities and society, environment and ecosystem) should be ensured.
- **Safety and Security:** Unwanted harms (safety risks), as well as vulnerabilities to attack security risks) should be avoided and should be addressed, prevented, and eliminated throughout the life cycle of AI systems to ensure human, environmental, and ecosystem security.
- **Fairness and non-Discrimination:** An inclusive approach to ensuring that the benefits of AI technologies are available and accessible to all, taking into consideration different age grounds, cultural systems, different language groups, persons with disabilities, girls and women, and disadvantaged, marginalized, and vulnerable people.
- **Sustainability:** The continuous assessment of the human, social, cultural, economic, and environmental impacts of AI technologies should therefore be carried out with full cognizance of the implications of AI technologies for sustainability, a set of constantly evolving goals across a range of dimensions.
- **Right to Privacy and Data Protection:** AI actors need to be accountable for the design and implementation of AI systems, ensuring that personal information is protected throughout the life cycle of the system.
- **Human Oversight and Determination:** To ensure that it is always possible to attribute ethical and legal responsibility for any stage of the life cycle of AI systems, as well as in cases of remedy related to AI systems, to physical persons or to existing legal entities. As a rule, life-and-death decisions should not be ceded to AI systems.
- **Transparency and Explainability:** People should be fully informed when a decision is informed by or is made based on AI algorithms, including when it affects their safety or human rights, and those in circumstances should have the opportunity to request explanatory information from the relevant AI actor or public sector institutions. Explainability refers to making AI systems' outcomes intelligible and providing insight into their outcomes. Also refers to the understandability of the input, output, and the functioning of each algorithmic building block, and to how each contributes to the system's outcome.
- **Responsibility and Accountability:** The ethical responsibility and liability for decisions and actions based in any way on an AI system should ultimately be attributable to AI actors, in accordance with their roles in the AI system's life cycle.

- ***Awareness and Literacy***: Public awareness and understanding of AI technologies and the value of data should be promoted through open and accessible education, civic engagement, digital skills, AI ethics training, and media and information literacy training led jointly by multi-stakeholder actors, including governments, intergovernmental organizations, civil society, and academia...
- ***Multi-Stakeholder and Adaptive Governance and Collaboration***: Participation of diverse stakeholders throughout the AI system life cycle is necessary for inclusive AI governance, enabling benefits to be shared by all and contributing to sustainable development.

Antiqua et Nova (Dicastery for the Doctrine of the Faith 2025)

- ***Human Dignity and Freedom (Helping Human Freedom and Decision-Making)***: The commitment to ensuring that AI always supports and promotes the supreme value of the dignity of every human being and the fullness of the human vocation serves as the criterion of discernment for developers, owners, operators, and regulators of AI, as well as its users (AN 2025, 43).
- ***Common Good (AI and Society)***: AI should serve the common good of the entire human family, which is the total of social conditions that enable people, as groups or individuals, to achieve fulfillment more easily.
- ***Authentic Human Relationality (AI and Human Relationships)***: To foster connections within the human family. No AI can genuinely experience empathy. AI cannot replicate the eminently personal and relational nature of authentic empathy.
- ***Dignity of Work and Human Labor (AI, the Economy, and Labor)***: AI should assist, not replace, human judgment. It must never degrade creativity or reduce workers to mere “cogs in a machine.” Respect for the dignity of labor and the importance of employment for the economic well-being of individuals, families, and societies, for job security, and for just wages ought to be a high priority for the international community as these technologies penetrate more deeply into our workplaces.
- ***Equity in Healthcare (AI and Healthcare)***: to ensure that the use of AI in healthcare does not worsen existing inequalities but rather serves the common good.
- ***Wisdom-Oriented Education (AI and Education)***: To engage with wisdom and creativity in careful research on this phenomenon, helping to draw out the salutary potential within the various fields of science and reality, and guiding them always towards ethically sound applications that clearly serve the cohesion of our societies and the common good.
- ***Integral Ecology and Stewardship (AI and the Protection of our Common Home)***: “That we look for solutions not only in technology but in humanity” (AN, 2025, 97). A fully human approach to the stewardship of the earth rejects the distorted anthropocentrism of the technocratic paradigm.

- ***Ethical Restraint in Warfare (AI and Warfare)***: The development and deployment of AI in armaments should be subject to the highest levels of ethical scrutiny, governed by a concern for human dignity and the sanctity of life.
- ***Theological Humility (AI and Our Relationship with God)***: While AI has the potential to serve humanity and contribute to the common good, it remains a creation of human hands, bearing “the imprint of human ingenuity.” It must never be ascribed undue worth.

As proposed at the outset of this paper, I now situate these guidelines within Wojtyła's personalist horizon. To this end, I articulate three ethical imperatives for the responsible use and management of AI, derived from his personalism: strengthening human consciousness, safeguarding personal agency and accountability, and upholding the unity of the person.

Designing AI that Strengthens Human Consciousness

The guidelines that align with the imperative of strengthening human consciousness include autonomy, transparency and explainability, awareness and literacy, human dignity and freedom, wisdom-oriented education, and theological humility. Three core ideas are stressed across these guidelines that can serve as key ideas for a personalist AI ethical imperative: first, the importance of acquiring knowledge in a manner that illuminates rather than obscures understanding; second, the indispensability of forming moral judgement rather than surrendering or ceding it to AI systems; and third, the need to safeguard the interiority of the human subject as AI continues to develop and occupy critical places in human lives.

Acquiring knowledge to illuminate. The design and deployment of AI must strengthen human self-awareness and a sense of responsibility, rather than eclipsing or replacing them. On a positive note, AI4People views the emergence of more AI-driven inventions as a means of freeing people from mundane tasks, thereby providing them with opportunities to develop new abilities and skills (Florida 2018). Such benefits of AI would give people more time to focus on their creative genius and develop new ideas or inventions that can harness their full potential as human beings. As discussed thus far, for Wojtyła, consciousness reflects one's nature as a rational being. Bostrom affirms this, saying “This thing, the human brain, has some capabilities that the brains of other animals lack. It is to these distinctive capabilities that we owe our dominant position on the planet” (2014, vii).

In contrast, an emerging model of AI objectifies and reduces the human person by reframing actions in mechanistic and functional terms, i.e., “nature's digital computer” (Bolter 1984, 7; Rubin 2003, 99). The movement that starts with *mimesis* and progresses to *prosthesis* risks redefining cognitive or reasoning functions, as it can be reduced by a mechanistic outlook, segmented, replicated, enhanced, and, at worst, replaced. This movement or trajectory implies an anthropological reductionism, in which the human person is now viewed as merely a computing machine.

Forming moral judgment. The significant moral issues arising from this development in AI use can be summarized in three key points. First, consciousness is perceived as replicable and even replaceable. Second, the brain is treated as a meat machine. Third, personal identity is measured by mechanistic standards, such as efficiency, speed, and performance. As a result, overall human dignity is compromised. In this context, the person is no longer viewed as a subject but as an object whose decisions can be dispensed with. In response to these issues, I argue that Wojtyła's emphasis on consciousness, which stresses human rationality, serves as a safeguard against decisions heavily mediated by AI. This is crucial because such decisions can erode the consciousness of action. Accordingly, this brings us back to the fundamental point of *actus humanus* (man acts), which highlights the consciousness or voluntariness of an act; moreover, in his wise counsel, Wojtyła emphasizes the function of conscience, "In fulfilling an act, I fulfill myself in it provided the act is good, or in agreement with my conscience, that is, done with a good or righteous conscience. Through such an act, I become, and am, good as a man. The moral value reaches to the entire depth of the metaphysical structure of the human subject" (Wojtyła 1979, 286).

Safeguarding interiority. Amid the widespread automation enabled by AI-driven functions, AI designs should make people realize that they are not merely executing system actions but are active agents in the whole process of acting. For example, how should a physician respond to AI findings and recommendations deemed accurate? AI must be designed so as not to substitute for, short-circuit, or obscure the subject's norming, i.e., the person's governance of action in light of truth about the good. Wojtyła's emphasis on consciousness is concretely expressed in AN, which calls on developers, owners, and regulators of AI to ensure that AI fosters human freedom and decision-making (AN 2025, 43).

Designing AI that Increases Personal Agency and Accountability

Recognizing the necessity of human oversight in the entire life cycle of an AI system, the guidelines that are expedient to increasing personal agency and accountability include autonomy, non-maleficence, justice, responsibility, security and privacy, human oversight and determination, responsibility and accountability, proportionality and do no harm, safety and security, right to privacy and data protection, multi-stakeholder governance and collaboration, ethical restraint in warfare, equity in healthcare. Three overarching imperatives are gleaned from this guideline that stress the necessity of ensuring that AI designs increase human agency and accountability: first, preserving human authorship and control; second, ensuring that responsibility can be traced back to the human actor; and third, preventing the outsourcing of moral agency.

Preserving Human Authorship and Control. To increase human agency and responsibility in AI use, the following imperatives are proposed: ensure human-centered design, given the general view that human agency and accountability are at

risk due to how AI is deployed. As aforementioned, AI mechanisms serve human needs as tools, enhance living conditions, promote human dignity, and uphold human freedom, rather than prioritizing AI autonomy. AI must be utilized to improve efficiency to serve human welfare. The potential for enhancing human agency and autonomy is evident in the growing reservoir of “smart agency,” which, when put at the service of human intelligence, can facilitate doing more, and do it better and faster (Floridi 2018).

Tracing Responsibility Back to the Human Author. AI designers must ensure human responsibility amid the utilization of algorithmic decisions. It is crucial to design AI in ways that preserve human judgment and ethical standards and, in this way, address the so-called “black-box mentality, according to which AI systems for decision-making are seen as being beyond human understanding and hence, control” (Floridi 2018). This is mindful of the fact that AI, as machines, while replete with big data and advanced computing capabilities, lacks intrinsic moral agency and intelligence. Without human oversight, humans may be undermined and reduced to passivity; as a result, they may infringe on human rights and well-being. Third, AI operators must preserve human agency and avoid objectification. AI users must not be reduced to mere data points or consumers (Wong & Reider 2019). Human beings, while benefiting from the efficiency of AI, must not allow themselves to be exploited by algorithmic decision-making and automated content generation. AI designs must ensure that human critical engagement and reflection are preserved throughout the act of acting.

Preventing the Outsourcing of Moral Agency. For Wojtyła, the efficacy of the person is founded on his/her ability to master his/her actions in the world. Following Wojtyła’s analysis, action not only reveals the true nature of the person but also makes it clear that the action is genuinely personal when performed with full knowledge and consent of the will, i.e., when the human agent owns it. This anthropological premise demands that while AI contributes to human flourishing, it must not become independent from its moral agent. Left on its own, AI presents a moral deficit because only human agents possess morality and responsibility, which must not be fragmented or rendered opaque by AI use. In the spirit of agency and responsibility, human beings must retain ownership of their decisions and assume full responsibility, particularly in areas where AI appears to wield power, control, or coercion over human beings (Celaya & Yeung 2019). In the spirit of the principle of autonomy, “the autonomous power to hurt, destroy, or deceive human beings should never be vested in AI” (Floridi 2018). When all is said and done, even with AI’s assistance, the human agent must still be able to say with confidence, “I am the cause of this action,” or perhaps a student can claim with confidence and integrity upheld, “I am the author of this paper.”

Designing AI that Upholds the Integrity of the Person

From Wojtyła’s personalist horizon, the integrity of the human person is sacred and one of the dimensions of life compromised in the name of progress. To safeguard

this essential aspect of human life, the following guidelines are necessary for formulating an imperative: beneficence, justice, non-maleficence, inclusion, impartiality, fairness and non-discrimination, sustainability, common good, authentic human relationality, dignity of work and human labor, integral ecology, and stewardship. The core ideas that permeate these guidelines are: first, the irreducible dignity of the human person; the priority of the human person over data, functions, or outputs; and the unity of personal, social, and ecological life.

Unity of Personal, Social, and Ecological Life. Pope Benedict XVI remarks, “Development must include not just material growth but also spiritual growth, since the human person is a unity of body and soul” (2009, 76), or as Wojtyla puts it, psychosomatic unity. We summarize three key concerns arising from the use of AI regarding the integrity of the human person. First, there is the dichotomization of the human person through the disintegration of psycho-somatic unity. Because AI use tends to privilege the informational identity or element of action, corporeality is marginalized. By implicitly undermining substantial unity, the body is seen as a defective and replaceable element of the human person. Second, in the technocratic worldview, human dignity is undermined by the pursuit of a disembodied anthropology and technological superiority. Here, the vision is of a post-human or post-biological future in which intelligence is viewed as bodiless. Third, the eclipse of personal moral agency. A dichotomized or disintegrated psycho-somatic soul unity tends toward the displacement of humans as rational and moral agents. By assuming decision-making roles, AI takes over functions that are traditionally exclusive to human beings, such as exercising human responsibility, freedom, and authorship in action.

The irreducible dignity of the human person and their priority over data, functions, and outputs. These moral issues impinge on Wojtyla’s discussions on integration and somaticity. Such orientations demand a moral imperative that calls on experts to design AI in a way that is interwoven with human life, respecting the reality of the human person as a substantial unity of body and soul. This implies the imperative to preclude any means that disembodies consciousness and reduces human intelligence into a data construct. We recall from Wojtyla that the person is not just a pure consciousness nor a mere biological entity. Wojtyla insists on the reality of the human person as more than just a functional unity but an anthropological whole. The emphasis on the integrity of the human person affirms that agency and responsibility are rooted in the person’s efficacy in action. At the same time, consciousness (as mirroring/reflexive subjectivization) is the experiential condition by which the act is lived as “mine.” Hence, AI applications must not allow practices that reduce human identity to digital profiles or exploit corporeality as a venue for optimization. Instead, any AI deployment must ensure that technologies are utilized to support holistic human flourishing. This corresponds to the principle of beneficence that AI4People subscribes to, which is for the promotion of the well-being of all sentient creatures; however, it states the need to prioritize human well-being as an outcome in all system designs” (Floridi 2018, 16).

CONCLUSION

AI is upon us, and it is developing at breakneck speed. Currently, it is a tool, but unchecked, experts predict that AI can develop into something autonomous of the human person, and worse, the latter can become subservient to it. Hence, the ethical use of AI must be ensured to preserve human dignity, ensuring that human beings are not reduced to mere data, machines, or patterns. AI practitioners must maintain human control, not only to keep AI machines in check, but also to ensure that human agency and moral responsibility are preserved. The imperative to preserve personalism amid technological flourishing is clear. It is in this context, as we have seen, that Wojtyła's concept of the person and his actions are critical points for reflection as they provide the necessary backdrop for balancing our engagement with AI. At this juncture, we recall the sub-queries posed at the outset of this paper and succinctly address each in turn, thereby moving toward responding to the principal questions.

First, what specific moral problems arise from the development and use of AI? Three key concerns have been identified. The first concerns attempt to imitate human consciousness, which tends to objectify the human brain and reduce it to a merely functional system that can be replicated and replaced. This, in turn, gives rise to a second concern, which is the dichotomization or disintegration of the human person's psycho-somatic unity. An overemphasis on consciousness understood in functional or mechanistic terms promotes disembodied intelligence. The third issue involves increasingly autonomous forms of algorithmic decision-making that function with less or no human oversight, raising serious implications for human agency and moral responsibility.

Second, what foundational presuppositions underlie Wojtyła's personalism? Given the breadth of Wojtyła's personalism, this paper has focused on three key concepts that can serve as conceptual tools for responding to AI's moral challenges. First is Wojtyła's notion of consciousness and voluntary action. The central idea is the reflexive or mirroring function of consciousness, with crucial implications for understanding Wojtyła's concept of efficacy. The second concept derived from Wojtyła is integration, emphasizing the necessity of safeguarding the psychosomatic unity of the person to preserve self-possession and governance. The third concept underscored is efficacy, foregrounding the distinction between "man acts" and "merely happening." The significant point is that man is conscious of being the agent of the action.

Third, what moral imperatives for the use of AI can be drawn from Wojtyła's personalism? The foregoing analysis paved the way for the design of human-centered AI, namely, that AI systems should be designed so that human consciousness is strengthened rather than undermined or replaced. Moreover, AI must preserve personal agency and accountability in contexts increasingly characterized by algorithmic decision-making. Finally, the psycho-somatic integrity of the human person must be protected against excessive reliance on the efficiency and optimization of AI systems.

In light of the foregoing, this paper concludes that Wojtyła's philosophical anthropology provides significant insights into contemporary moral questions regarding the management and use of AI. By placing the human person at the center,

AI can remain true to its nature and purpose as an instrument ordered toward human flourishing. For AI to be moral, it must be in the hands of human agents guided by sound moral principles who genuinely desire the good of humanity.

REFERENCES

- Acosta, Miguel, and Adrian J. Reimers. 2016. *Karol Wojtyla's Personalist Philosophy: Understanding Person and Act*. Washington, DC: The Catholic University of America Press.
- Aguas, Jove Jim S. 2014. *Person, Action and Love: the Philosophical Thoughts of Karol Wojtyla (John Paul II)*. Manila: University of Santo Tomas Publishing House.
- Augustine, Michael Fiedrowicz, and Boniface Ramsey. *The works of Saint Augustine: A translation for the 21st century. part 1, books. vol. 8, on Christian belief*. Hyde Park, NY: New City Press, 2005.
- Benjamin, Ruha. 2019. *Race after Technology*. Cambridge, UK: Polity Press.
- Bolter, David. 1984. "Artificial Intelligence." *Daedalus* 113 (3): 1–18.
- Bostrom, Nick. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford, UK: Oxford University Press.
- Celaya, A., and Nick Yeung. 2019. *Human Psychology and Intelligent Machines*. NATO
- Coleman, Flynn. 2019. *A Human Algorithm: How Artificial Intelligence Is Redefining Who We Are*. Berkeley, CA: Counterpoint.
- Crawford, Kate. 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven, CT: Yale University Press.
- Defense College. Accessed November 26, 2025. <https://www.jstor.org/stable/resrep19966.9>.
- Dicastery for the Doctrine of the Faith and Dicastery for Culture and Education. 2025. "Antiqua et Nova: Note on the Relationship between Artificial Intelligence and Human Intelligence." January 28. Accessed November 25, 2025. https://www.vatican.va/roman_curia/congregations/cfaith/documents/rc_ddd_doc_20250128_antiqua-et-nova_en.html.
- Dignum, Christoph Lütge, et al. 2018. "AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations." *Minds and Machines* 28 (4): 689–707. <https://doi.org/10.1007/s11023-018-9482-5>.
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689-707. <https://doi.org/10.1007/s11023-018-9482-5>.
- Francis. 2024. "Address to Participants in G7 Session on Artificial Intelligence." Borgo Egnazia (Puglia), June 14. Accessed November 25, 2025. <https://www.vatican.va/content/francesco/en/speeches/2024/june/documents/20240614-g7-intelligenza-artificiale.html>.

- Gilli, Andrea, Massimo Pellegrino, and Richard Kelly. 2019. *Intelligent Machines and the Growing Importance of Ethics*. NATO Defense College. Accessed November 26, 2025. <https://www.jstor.org/stable/resrep19966.9>.
- Harari, Yuval Noah. 2024. *Nexus: A Brief History of Information Networks from the Stone Age to AI*. New York: Random House.
- Holub, Grzegorz. 2021a. "The Validity of Karol Wojtyła's Philosophy Today." *Logos* 56: 75–85.
- Holub, Grzegorz. 2021b. "Karol Wojtyła's Thinking on Truth." *International Philosophical Quarterly* 61 (4): 387–396.
- Holub, Grzegorz. 2022. "Philosophical Anthropology and Ethics in the Life and Thought of Karol Wojtyła." *Studia Gilsoniana* 11 (1): 145–161.
- Kupczak, Jarosław, OP. 2000. *Destined for Humanity: The Human Person in the Philosophy of Karol Wojtyła/John Paul II*. Washington, DC: The Catholic University of America Press.
- Leo XIV. 2025a. "Address to Educators on the Occasion of the Jubilee of World Education." St. Peter's Square, October 31. Accessed November 25, 2025. <https://www.vatican.va/content/leo-xiv/en/speeches/2025/october/documents/20251031-giubileo-educatori.html>.
- Leo XIV. 2025b. "Message to Participants in the Second Annual Conference on Artificial Intelligence, Ethics, and Corporate Governance." Vatican, June 19–20. Accessed November 25, 2025. <https://www.vatican.va/content/leo-xiv/en/messages/pont-messages/2025/documents/20250617-messaggio-ia.html>.
- Pontifical Academy for Life. 2020. *Rome Call for AI Ethics*. <https://www.romecallforaiethics.org/>.
- Rossi, Francesca. 2018–2019. "Building Trust in Artificial Intelligence." *Journal of International Affairs* 72 (1): 127–134.
- Rubin, Charles T. 2003. "Artificial Intelligence and Human Nature." *The New Atlantis* 1: 88–100.
- Svensson, Patrik. 2016. *Big Digital Humanities: Imagining a Meeting Place for the Humanities and the Digital*. Ann Arbor: University of Michigan Press.
- Tegmark, Max. 2017. *Life 3.0: Being Human in the Age of Artificial Intelligence*. New York: Vintage.
- Tully, Stephen, Christopher Longini, and Gil Appel. 2025. "Lower Artificial Literacy Predicts Greater AI Receptivity." *Journal of Marketing* 89 (5): 1–20. <https://doi.org/10.1177/00222429251314491>.
- UNESCO. 2021. *Recommendation on the Ethics of Artificial Intelligence*. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>.
- University of Cambridge. 2016. "The Best or Worst Thing to Happen to Humanity – Stephen Hawking Launches Centre for the Future of Intelligence." October 19. Accessed March 20, 2026. <https://www.cam.ac.uk/research/news/the-best-or-worst-thing-to-happen-to-humanity-stephen-hawking-launches-centre-for-the-future-of>.
- Wierzbicki, Alfred Marek. 2020. "The Person, Human Action, and Morality as Seen in the Personalist Philosophy of Karol Wojtyła." *Quién* 11: 51–66.
- Winner, Langdon. 1980. "Do Artifacts Have Politics?" *Daedalus* 109 (1): 121–136.

- Wojtyła, Karol. 1979. "The Person: Subject and Community." *The Review of Metaphysics* 33 (2): 273–308.
- Wojtyła, Karol. 2005. *Love and Responsibility*. Translated by Grzegorz Ignatik. Boston: Pauline Books and Media.
- Wojtyła, Karol. 2021. *Person and Act and Related Essays*. Vol. 1. Translated by Grzegorz Ignatik. Washington, DC: The Catholic University of America Press.
- Wojtyła, Karol. 2023. *The Lublin Lectures and Works on Max Scheler*. Vol. 2. Translated by Grzegorz Ignatik. Washington, DC: The Catholic University of America Press.
- Wong, Pak-Hang, and G. Reider. 2025. "After Pleas for Moral Repair, Algorithms Have Failed." *Science and Engineering Ethics* 31 (26): 1–11. <https://doi.org/10.1007/s11948-025-00555-y>