

Featured Article

AI INTEGRATION IN CATHOLIC HIGHER EDUCATION: PHILOSOPHICAL REFLECTIONS IN THE LIGHT OF JOHN PAUL II'S *MEMORY AND IDENTITY*

Edward J. Alam
Notre Dame University-Louaize, Lebanon

*By reflecting on the history of the development of AI (from Aristotle till the famous workshop at Dartmouth College in 1956, when the term AI was first coined), this paper seeks, in the light of John Paul II's 2005 book, *Memory and Identity*, to reposition this intellectual achievement within the context of the long and venerable tradition of philosophical reflection on the intellectual and moral virtues. Cognizant and appreciative of those contemporary Catholic scholars who have sought to assess the present global educational shifts in pedagogy in the light of the Catholic intellectual tradition that has long appropriated and developed this venerable philosophical tradition of the virtues, this paper argues that deeper metaphysical and theological resources must be rediscovered and then applied to Catholic pedagogical movements in order to highlight the dignity and particularity of the human person. While not denouncing AI as a necessarily dehumanizing technology, the paper warns against its tendency to depersonalize human persons by deceptively blurring the distinction between reality and virtual reality.*

Keywords: Catholic education, logic, memory, mystery of iniquity, pedagogy, personalism, virtue

INTRODUCTION

At the dawn of the third millennium, a series of recorded historical and philosophical conversations took place in Rome among three prominent Polish philosophers, one of whom, at the time, was also the Bishop of Rome: Pope John Paul II. The other two, Krzysztof Michalski and Józef Tischner, founders of the Institute for Human Sciences in Vienna, had invited the Pope to critically reflect with them upon the two dictatorships that not only marked twentieth-century Polish history, but also shaped the world's history in that century: Communism and Nazism. About a decade later, the Pope returned to these transcribed conversations seeking to situate them in a

wider setting and in the context of the first few years of the new millennium. The result was his last book, *Memory and Identity* (John Paul 2005), published in 2005, the year he died. Although neither the general theme of Catholic Higher Education nor the specific one of AI integration in such education is explicitly addressed in this book, many of the philosophical presuppositions that inform its contents, some of which are explicitly drawn out and explained in various other contexts, are indeed relevant for the theme of AI in Education. The first two entries of Section One, *The Limit Imposed Upon Evil*, titled (1) “*Mysterium Iniquitatis*: The Coexistence of Good and Evil” and (2) “Ideologies of Evil,” are particularly pertinent to the theme in question.

Focusing on these two entries certainly does not unqualifiedly imply that AI is innately evil, but neither does it suggest that it is morally neutral. As with many sophisticated technological inventions (especially in modernity), the moral implications are complex and intricate. One position, not without value, but so obvious as to have become cliché, is that modern technology (or any technology for that matter) is neither good nor evil, but “is” or “will be” ethically judged only in terms of how we use it. Again, a position not without value, but one that is limited. It is well to recall here what Werner Heisenberg once wisely said: “[o]ne has to remember that every tool carries with it the spirit by which it has been created.”

For not only is it the case that technology is not morally neutral, but certain technologies are indeed inherently evil and demonstrably so even *before* being used. Consider, for instance, the very existence of nuclear weapons,¹ which is why the statements consistently found in Catholic Social Teaching regarding nuclear weapons are so categorical: “Nuclear weapons must be banned. . .[and] ultimately [abolished],” (Pope John XXIII 1963, para. 112, 113) despite the acknowledgment that such weapons do “indeed act as a deterrent” (Ibid., para. 11). This deterrent value, therefore, is only ‘good’ as a lesser evil and then only insofar as it is temporary. By extension, any AI technology that supports the build-up and technological refinement of such weapons is also inherently evil, whereas any AI technology that aids in eliminating these weapons is inherently good. Thus, certain weapons, certain algorithms, are inherently evil in and of themselves, whether they are used or not.

But as this is a philosophical reflection, it is necessary to back up and first examine the question of what ‘good’ and ‘evil’ are; and here, the first entry of Section One, “*Mysterium Iniquitatis*: The Coexistence of Good and Evil,” offers valuable insight. When asked about the meaning of the great evil that erupted in the twentieth century due to the totalitarian systems of Nazism and Communism, the Pope *re-minded* us that the same century which provided the setting for these great evils also furnished the context for their demise. His point was that despite the glaring failures resulting from the misguided ideologies of the twentieth century, rooted partially in the European Enlightenment, this same Enlightenment afforded the means and ideas to men of good will in this same century to overcome the evil, at least temporarily.

Reflecting on this historical fact, the Pope holds up the hermeneutical key that unlocks the door of that strange and enigmatic room called the *mysterium iniquitatis*: a cold, dark, entirely isolated and *empty* room. The very term *in-iquitatis* is a *negation*: the *absence of* equity, of justice, of fairness; no virtue dwells there. And yet, this utterly empty room is paradoxically *full* of horror, violence, and perversion. As the Holy Father states, echoing a long Augustinian and Thomistic tradition, evil “*is a privation*.”

..[but it is] never a total absence of good” (John Paul 2005, 4). He refers to Jesus’ parable of the good grain and the weeds to shed light on the *mystery* of this privation, noting that, “sometimes” evil originates “from the same soil” as the good (Ibid.).

Eschatologically extending the meaning of this parable to capture the dramatic journey of human history towards its final goal, the Pope writes: “The history of mankind is the ‘theater’ of the coexistence of good and evil. So even if evil exists alongside good, good perseveres beside evil and *grows*, so to speak, from the same soil, namely, human nature”(Ibid.). This is an important philosophical (Aristotelian) point; for although human beings are not born virtuous or vicious (virtues and vice grow through repeated action), human existence itself is still considered good by virtue of the fact that it *exists* (i.e., the transcendentals of being). Upon this view, even the most vicious person, in whom the intellectual and moral virtues are virtually absent, maintains some proximity to these virtues since each virtue or vice *grows* from the same human nature.

Reflecting, then, upon the challenge of AI integration in Catholic Higher Education within this framework of virtue-centered ethics (presupposed in *Memory and Identity*) promises to yield valuable insight. This is especially the case if the focus is not only, as is often the case in such AI ethical reflections, on how the intellectual virtues are related to the moral virtues, but first upon how the intellectual virtues are related to one another. To be sure, the goal is ultimately the integration of the intellectual with the moral, but a closer examination of the intellectual affords an insight into the *history* of AI, which in turn sheds light on how it might be properly integrated into Catholic Higher Education.

Most histories of AI trace its origins back to the now famous brainstorming workshop at Dartmouth College in the summer of 1956, which brought together leading international researchers² from the field of “machine intelligence (MI),” a branch of computer science. Since the term AI (replacing MI) was indeed coined by these scholars during the workshop, it is accurate to pinpoint 1956 as the beginning of AI, but the intellectual antecedents go back just a bit further—about twenty-four hundred years earlier to be precise—yes, to Aristotle himself. In what follows, I attempt to draw a line from Aristotle to the founders of AI (John McCarthy and company) via Gottfried Wilhelm Leibniz (1646-1716), George Boole (1815-1864), Gottlob Frege (1848-1925), Alfred North Whitehead (1861-1947), Bertrand Russell (1872-1970) and Alan Turing (1912-1954). My claim is that because the success of the 1956 Dartmouth workshop (i.e., a new branch of computer science) was the direct result of giants standing on the shoulders of giants, the first being Aristotle himself, then *unless* this new branch of computer science remains rooted in Aristotle, who never separated the *intellectual* from the *moral* virtues (nor the intellectual virtues from *one another*) it is destined to crumble and, given its unprecedented power, destroy much of what is good and beautiful on our planet. *Catholic* integration of AI into higher education, therefore, must be just that—*katholikos*, according-to-the-whole, a cosmic universal *integration* of all things with *everything* else.

To get a good idea of how solid and influential Aristotle’s logical system really was, it is enough to listen to no one less than Immanuel Kant (1795, 17), who claimed that, although there were great logicians after Aristotle, none of them were really able to “take a single step forward, and therefore [Logic as a discipline] seems ... to be

finished and complete.” As it turned out, “seems” was the key word here because a significant step forward was finally taken, but it came half a century after Kant’s death, when a largely self-taught philosophical genius, the son of a shoemaker, published a monograph titled *An Investigation of the Laws of Thought*; the year was 1854, the name of this genius was George Boole. Anticipating Boole’s aspirations some two centuries earlier, Leibniz dreamt of a universal concept language (formal mathematical language) that could *symbolically* capture the truth value in all claims to knowledge expressed through ordinary language. This dream became somewhat of a reality when he built a calculating machine to manipulate these symbols to enable a kind of *decision* regarding the truth value of all and any knowledge claims and statements.³ Though successful at one level, his machine never achieved what Leibniz ultimately wanted due primarily to shortcomings in his formal language—limitations that would be partially (but never fully) overcome by Boole, Frege, Whitehead, Russel, Turing, and then, finally, by McCarthy and company. But, first, back to Aristotle, though not yet to his *Organon* (logical works) but to his *De Anima* (his work on the soul), which critically included not only investigations into the *human* soul, but into the soul as such—the life principle in a (any) body. For him, research into the souls of plants and the non-human animals formed an indispensable part of the study of the human soul.

For Aristotle, the human soul has two distinct capacities, the rational and the non-rational, the latter being further distinguished between *vegetative* and *desiderative*. The vegetative (or nutritive) powers (fixed, instinctual, and almost automatic) are common in all plants and animals; the desiderative (or aspiring) powers, not present in plants per se, are indeed present in animals. But in human animals, unlike in non-human animals, these non-rational powers have access to, and somehow ‘communicate’ with, the rational (practical) powers of the soul. Somewhere in the midst of this interaction emerges something which makes all the difference in terms of that which makes the human animal soul different, not only in degree, but in kind from the non-human animal, namely, the crucial distinction between needs and wants. It is this distinction (and the conscious awareness of it) that makes *virtue* possible—something of which only human animals are capable. Many (but not all) of the needs and wants of the human animal are also found in non-human animals, though in the latter there is no such distinction: needs *are* wants and wants *are* needs. Now, in addition to the needs that human animals share with non-human animals, there are other needs that are exclusively human, and *chief* among them is the *need to be wise*. If this need is not met, it is impossible for the human animal to achieve its final *end* and ultimate purpose: happiness or what Aristotle calls *eudaimonia* (εὐδαιμονία).

Whereas the general ‘need’ *to know* is something we naturally ‘want’ or ‘desire,’ Aristotle starts his *Metaphysics* (1924) by asserting that everybody desires to know, just as the need to eat is also a natural want or desire, the ‘need’ *to be wise* must be discovered and cultivated before it becomes a real want or desire. Just as the non-rational powers of the human soul are hierarchically distinguished into vegetative and desiderative parts, with the latter having some indirect contact or communication with the rational (practical) powers, so too are the rational powers distinguished between the aptitude for pure thought, which, if fully realized, results in the soul’s possession of the theoretical virtue of wisdom (*Sophia*), and the soul’s capacity for practical

thought, which, if achieved, likewise results in the practical virtue of wisdom (*Phronesis*).

This two-fold rational virtue of wisdom, which *theoretically* contemplates *truth* and *practically* chooses the *good*, complements, and is complemented by two other theoretical capacities, namely, *nous*, which allows for an intuitive grasp of first principles, and *episteme*, the aptitude of demonstrative scientific knowledge, primarily (though not exclusively) involving deduction. The former, distinct (but not separate) from imagination and sense knowledge, is sometimes characterized as “the mind’s eye,” which is able to seize upon “why” things are *what they are* and what *they are ultimately for*; the latter, also distinct (but not separate) from imagination and sense knowledge, has the ability to know and deductively demonstrate how the different causes of things are properly ordered. We may call the two theoretical virtues associated with these theoretical rational powers *right intuition* and *proper logical demonstration*, respectively, bringing the number of the intellectual *theoretical* virtues to three. Finally, in addition to the practical virtue of *Phronesis*, Aristotle identifies the practical intellectual power of skill or knowhow, which potentially results in the practical virtue of art or what he calls *techne*; properly speaking, it is these two practical powers of the human soul with which the *desiderative* powers interact to cultivate and develop the moral virtues, but all five intellectual virtues cooperate together in the task of producing *the philosopher*—the one who knows the truth, chooses the good, and loves the beautiful—the *happy* and fulfilled man, the *lover of wisdom*.

For the purposes of this brief reflection, *episteme* and *techne* are particularly relevant since, beginning with Leibniz, it was precisely the hyper-development of these capacities (potentially finding their fulfillment in the intellectual virtues of *episteme* and *techne*) that led to the breakthroughs during the 1956 summer workshop at Dartmouth. But before chronologically examining some of the details of this historical development (from Leibniz to 1956) an anachronistic word about the role of AI in either the promotion or elimination of nuclear weapons, with which I began this reflection, is in order, because both the invention of electronic computers (i.e., MI/AI) and the splitting of the atom emerge after the world’s first world war, and at the dawn of a second. Thus, the critical questions arise: is it possible to separate either the origins and motivations of these achievements or their *subsequent* developments from the intrinsic evil of modern warfare, in particular, from the *evil* of World War II and its corresponding evil ideologies of Communism and Nazism? And if we cannot separate them, are we to conclude that the inventions themselves were evil?

Here, the Pope’s remarks regarding the coexistence of good and evil provide part of the answer because the disordered and bloody 20th century that produced these evils was, as he astutely reminds us, the same century that overcame them. My point is that although both the splitting of the atom and the invention of electronic algorithms emerged from, and then were driven in their rapid development by, the evil of the world’s most brutal and inhumane war in history, we must at least distinguish (if we cannot totally separate) these inventions from the evil that resulted. If this is the case for the culmination of these inventions, then it must also be the case for the historical antecedents from which they spring—as I attempt now to make clear by looking more closely at the historical development of Aristotelian logic from Leibniz up to and including the 1956 Dartmouth workshop. Attending to the patterns and details of this

development promises to shed light upon the paths we must take in the task of integrating AI into Catholic Higher Education.

I have already referred to Leibniz's dream of a universal concept language, a formal mathematical language that could symbolically capture the truth value in all claims to knowledge expressed through ordinary language. Surely, such a dream was only possible in the aftermath of Descartes' attempt at a universal mathematics⁴, which, as it turned out, paradoxically led to an unintentional but radical turn towards subjectivity (See my discussion of this, Alam 2002). Leibniz admired Descartes' work but was also aware of its limitations. Like Descartes, Leibniz too had mastered both Aristotle's logical system of categorical syllogisms⁵ and the work of medieval logicians who refined this system. But unlike Descartes, who had little use for this tradition, (Ibid.) Leibniz admired it. His appreciation of the long and venerable tradition of logical debate and discussion during the medieval period inspired him to extend this system by including not only the 256 possible combinations of categorical syllogisms generated from Aristotle's famous 'square of opposition,'⁶ but also propositional and mathematical statements. I have also already mentioned that Leibniz's dream became somewhat of a reality when he built a calculating machine to manipulate these symbols to enable a kind of *decision* regarding the truth value of knowledge claims and statements, and that though the machine was successful *at one level*, it never achieved what he ultimately wanted due primarily to the limitations of his formal symbolic language.

Part of Leibniz's success was that he made it possible for logicians after him to appreciate the power and precision of how mathematical symbols could represent the logical relationships between words inherent in ordinary language and potentially eliminate ambiguity when it came to claims regarding true knowledge. His insights into the role and reality of binary relations in language anticipated a breakthrough in logical theory, which later logicians, especially Boole, would utilize. While the logical specifics of these insights are fascinating and significant at various levels and for different reasons, not the least of which is the decisive role his work played in paving the way towards AI, what is more important for our purposes concerns how *integrated* and humane his work really was. Though not a Catholic, his personal appropriation of the Protestant Christian faith can rightly be described as a 'catholic' (universal) approach to knowledge; the centrality of his desire to create a 'universal language' is one indication of this and an argument could be made that a main impetus in his desire to 'create' or discover a universal language was to foster communication among all the peoples of the world in order to avoid the kinds of misunderstandings that lead to war. In addition to this, he was a theologian who never separated his scientific work from his theological efforts, and unlike many Protestant theologians who spurned any use of Greek philosophy in theological endeavors, he benefited greatly from this tradition while appreciating the work of Catholic theologians who employed Aristotle in their reflections.

For Leibniz, as for Aristotle, the intellectual and moral powers of the soul were meant for *virtue* and the true possession of one entailed the possession of all the others, and while it is true that he devoted incredible energy to cultivating the intellectual virtue of *episteme*, exposing him to the charge of an isolated overemphasis in this regard, or at least setting the stage for other less conscientious (and less virtuous)

thinkers to exploit this emphasis, he himself prevented his novel insights into the power of mathematical calculation from excessively spinning out of control. As responsible for providing some of the earliest and most momentous intellectual antecedents of AI⁷, Leibniz's work goes a long way in debunking the idea that AI is inherently evil. On the contrary, for someone whose theological work focused on reconciling evil with the omnipotence and benevolence of God, Leibniz remains a model for contemporary educators today tasked with integrating AI into Catholic Higher Education.

While there were other logicians and philosophers after Leibniz who contributed to the rush of the stream (leading to AI) flowing from the spring that he unplugged, their work might be described as only small tributaries when compared to George Boole's discovery of another spring, flowing from the same water source deep beneath the surface, but one that provided enough new energy to transform this stream into a mighty river. Contrary to what is too often taught, Boole, like Leibniz before him, never challenged or undermined the basic principles of Aristotelian logic. In fact, concurring with Kant, he acknowledged his debt to Aristotle when Boole (1854) writes:

In its ancient and scholastic form, indeed, the subject of Logic stands almost exclusively associated with the great name of Aristotle. As it was presented to ancient Greece in the partly technical, partly metaphysical disquisitions of *The Organon*, such, with scarcely any essential change, it has continued [and for good reason] to the present day.

Apparently, neither Boole nor Kant recognized the achievements of Leibniz, or if they did, they did not consider them essential or groundbreaking. Whatever the case may be in this regard, there is no question that Boole's extension and expansion of Aristotelian logic by translating the syllogism into an algebraic calculus gave it a new mathematical foundation that allowed for novel and powerful applications—somewhat analogous to the way Descartes had extended Euclidian geometry by converting lines into numbers, numbers into lines, and diagrams into formulas, thus allowing the manipulation of symbols to go beyond mere spatial intuition (See my discussion on this, Alam 2002). Boole's principle of “wholistic reference,” in general, and his theories of concept and proposition formation, in particular, broke new ground in the extension of Aristotelian logic and paved the way for novel developments in mathematical logic and for the establishment of the new field of computer science. As with Leibniz's work, the nuanced mathematical details of this transformation are captivating, but what is most relevant for this reflection, in addition to the role his work played in moving us much closer to the advent of AI, for which these details obviously play a major role, was the way he integrated, again like Leibniz, his scientific work into his ideas about other branches of philosophy, particularly Ethics. Though not a theologian and thus lacking the metaphysical depth of Leibniz, his religious convictions informed his reflections on morality and then guided his speculations on how this morality was related to the laws of thought and rationality in general. Together with Leibniz, Boole underscores the necessity of unified knowledge and stirs

us away from the dangerous practice of excessively isolating particular mental aptitudes to the detriment of other intellectual and moral capacities of the human soul.

At the same time Boole was publishing his mature ideas, another philosophical genius was born in Germany by the name of Gottlob Frege, who would likewise play a major role in the extension of Aristotle's *Organon* and in the development of a new kind of mathematical logic that would play a decisive role in paving the way for the new field of computer science, eventually leading to the AI branch of that science. Without entering into the complex and somewhat controversial question regarding Boole's influence on Frege, since Frege was definitely aware of Boole's work, but did not explicitly acknowledge it, there is no doubt that his original genius certainly played a crucial role in the history of the AI story I am telling. Not only was he able to better formalize the very notion of proof through a new analysis of quantified statements, but his entire philosophy of language allowed for an original way of conceiving the relation of mathematics to logic in the context of determining truth values in propositional statements.

Unlike Leibniz and Boole, Frege's *Begriffsschrift* (1967), a work in logic, did not engage religion or metaphysics and largely ignored Aristotle's *Nicomachean Ethics* and *De Anima*. While others were quick to apply Frege's insights into verifiability and proof to ethical questions concerning objective moral truth, Frege himself showed little interest in ethics and focused almost exclusively on formal logic and the foundations of mathematics. His inattention to how his work in logic might be integrated into the other branches of philosophy, something so important for Leibniz and, to a lesser degree, perhaps for Boole, is highly relevant to our subject. For such inattention causes epistemic fissures that end up breaking the edifice of unified knowledge. To his credit, Frege was searching for and trying to construct a unified mathematics; but to his great discredit, he apparently failed to see that mathematics is simply one epistemic region among many, and certainly not the central one. By excessively isolating and amplifying the one mental capacity of logical deductive calculation, to the exclusion of the others, not only is the virtue of *episteme* undermined, but all the other vital intellectual and moral powers of the human soul risk being damaged as well. I am reminded here of how G. K. Chesterton famously defined insanity: "The madman is not the man who has lost his reason. The madman is the man who has lost everything except his reason" (1908, 10). In any case, any isolation of one intellectual capacity over another or any separation of intellectual virtue from moral virtue tends to strip the search for truth and knowledge of its most exciting and humane features, while jeopardizing the very meaning and goal of human life: *eudaimonia*.

The category of 'the mystical,' moreover, was still considered a genuine realm of knowledge (mystical theology) for Leibniz and, to some degree, for Boole, despite their strong preference for rationalism. For Frege, not only does this long tradition of mystical theology, so significant in Christian and Islamic philosophy (both of which entailed important and intricate logical investigations), have no place whatsoever in Frege's work, but it seems he did not even consider it a legitimate form of knowledge. Given the time-honored and revered tradition of the Jewish Mystical Tradition, and given that Frege was Jewish, this neglect is curious and raises several questions. One concerns his failure to consider the metaphysical (or mystical) implications of the

meaning and function of the term “is” in its *relation* to the symbol of the “equal sign” in mathematics; another, which naturally follows, concerns his inattention to the metaphysical reasons underlying how his insightful reflection on this *relation* led him to overcome his older content theory of one-layer meaning with his content theory of two-layer meaning, namely, his famous sense/reference distinction?

This is not the place to speculate on the answers to these questions, even if such speculation helps to answer a few more general questions, which are more to the point for the purposes of this short reflection, such as, does his neglect of metaphysics and virtue-centered ethics detract from the solid edifice of the logical/epistemological system he constructed, and which played such a key role in the development of AI? Would his results have been different had it not neglected these other realms? Would the development of mathematical logic have taken a different turn, been delayed, or perhaps hastened? I cannot possibly address these questions here, but at the very least, I will say that his decision to disregard the metaphysical and ethical categories established a precedent in applied mathematics, whether in physics or in logic, or in any other field for that matter, whereby no distinction is made between what we are “able” to do and what we “should” do. In other words, without the virtue of both theoretical and practical wisdom, *Sophia* and *Phronesis*, human beings are destined for fragmentation and despair. I return in my conclusion to this theme (in the context of AI integration in Catholic Higher Education) when I consider the historical and scientific relationship between the splitting of the atom and the decisive breakthrough in machine intelligence.

Now less than a quarter of a century after Boole and Frege had made their invaluable contributions, the field of mathematical logic was in full swing and took a major step forward with the 1910-1913 publication of *Principia Mathematica* (1963) by Bertrand Russell and Alfred North Whitehead, both of whom were deeply influenced by both Boole’s algebraic logic and Frege’s formal language, including especially his sense/reference distinction.⁸ No one could have predicted then, not even Russell or Whitehead, how their ambitious effort to unify mathematics under a painstakingly exact framework of deductive logic was on the verge of erupting into that AI branch of computer science that would drastically change the modern world forever. The first few decades of the 20th century had witnessed the invention of mechanical devices and prototype computers that could calculate differential equations at amazing speed, but this was just the beginning. One such machine, an analog computer,⁹ was developed by a certain Vannaver Bush, Professor of Electrical Engineering at MIT, who would go on to play a key role in the Manhattan Project during World War II. But it was Bush’s student, Claude Shannon, an original member of the 1956 Dartmouth summer workshop, who made the decisive breakthrough.

As an undergraduate, Shannon was required to take several courses in general philosophy. In addition to being introduced to Plato and Aristotle, he was also exposed to the philosophy of George Boole, which caused a spark to go off in the young mind of Shannon that would set an entire field ablaze. He began to see the connection between his own field of electrical engineering and the deepest modern philosophical foundations of symbolic mathematical logic rooted in Aristotle’s logical works. Perhaps we could say that the modern field of computer science was partially conceived in this spark. The subsequent publication of his 1938 paper, “A Symbolic

Analysis of Relay and Switching Circuits,” turned out to be one of the most important academic papers of the century. Equipped with the insights and achievements of Boole and Frege, and encouraged by the systemic power of the *Principia Mathematica*, Shannon combined electrical engineering with the groundbreaking insights of the previous century and positioned himself to become one of the key participants in the 1956 Dartmouth seminar, during which AI was officially born.

At the same time Shannon was doing his work at MIT, a young Englishman in his early twenties by the name of Alan Turing was independently working on a paper titled, “On Computable Numbers, With an Application to the [Decision Problem] *Entscheidungsproblem*.” The origin of this problem went all the way back to Leibniz, but a modern and more advanced version of the problem had been presented in 1928 by David Hilbert and Wilhelm Ackermann, and Turing had set himself the task of solving it, which, in a way, he did—precisely by showing it could not be solved—at least not by any computational procedure.¹⁰ In a telling 2006 article, 50 years after the famous Dartmouth seminar, and five years before he died in 2011, Professor John McCarthy (who, as mentioned in note 2 above, was the main organizer of the Dartmouth seminar and the one who coined the term AI) recounts that since “Alan Turing might have been the first to really understand that programming computers was the main way to realize AI, he would have played an important role at the seminar; unfortunately, he had died just two years before.”¹¹

Based on what McCarthy wrote and often said up until his death, 55 years after the seminar, we can surmise that, for him, although the 1956 seminar was a monumental success, it had not achieved the kind of human-level AI breakthrough he had hoped for. But he was confident that this breakthrough would soon come (surely during his lifetime) and knew that it required a sustained collaboration with leading scientists from various fields; it was not the work of one person. He especially appreciated Turing’s work because it had redirected research away from trying to solve a problem that could not be solved towards other areas of research that would yield more fruit. That is to say, Turing demonstrated that no systematic computational procedure could solve the age-old *decision problem* that Leibniz had first identified. He did this by providing a convincing analysis of what such a reliable computational procedure actually looked like, namely, his Turing machine. This machine, which embodied a reliable systematic computational process, proved that there could not be a systematic computational procedure that could solve every mathematical question; this not only laid the groundwork for some of the progress made during the 1956 seminar, but also foreshadowed the invention of the digital computer. For McCarthy, what was most promising about Turing’s work was that it seemed to be the beginning of achieving human-level AI.¹¹ McCarthy (2006) writes:

My hope for a breakthrough towards human-level AI was not realized at Dartmouth, and while AI has advanced enormously in the last 50 years, I think new ideas are still required for the breakthrough ... [b]esides proposals for extending logic, there are many systems that restrict logic in order to make computation more efficient. I'd prefer to use full logic, but want systems that can reason about their own reasoning

methods in order to decide on efficient reasoning. After all these years, I still have not been able to make specific proposals.

McCarthy gives expression here to a tension in, and a debate about, the future of AI that is still very much alive and shows no signs of going away. On one hand, he acknowledges the enormous advancement on the engineering level, but then admits his disappointment at the failure, including his own, to come up with any *new ideas* on the cognitive/ontological level. Today, the latest cutting-edge research in this field still grapples with the same problems that McCarthy identified a few years ago before his death. While progress at the level of engineering continues to skyrocket, virtually no progress has been made on the human-level AI front. In other words, narrow AI, designed for strictly defined tasks such as image recognition or language translation, has exploded and continues to become more and more efficient, while general AI remains basically where it was at the time of the 1956 seminar.

If McCarthy had really comprehended and then appropriated the timeless insights into the human soul as expressed by Aristotle 2,400 years ago, I do not think he would have spent much of his life chasing a pipedream, especially when the very pursuit of that pipedream potentially spawns new nightmares of unimaginable destruction. Human beings are capable of incredible technology, and AI is certainly a dazzling and impressive achievement, which potentially has the ability to right many of the technological wrongs of the 21st century, but if those controlling this technology have exploited the mental aptitudes which are ordered towards their completion in the intellectual virtues of *episteme* and *techne*, then the chance of possessing the even higher intellectual virtues of theoretical and practical wisdom is virtually impossible, to say nothing of the moral virtues. If, in fact, the deepest and oldest antecedents of AI are anchored in an extension of Aristotelian logic, and if this logical system is part and parcel of his metaphysics and ethics, then if present-day AI technology is to cultivate the good, the true, and the beautiful, then a return to Aristotle is imperative.

Without the virtue of wisdom, the other intellectual powers run amuck, making it impossible to see that, sometimes, what “can” be done (splitting the atom, for instance) might not always be what “should” be done. I am not claiming that Ernest Rutherford and all those before him who anticipated his work of splitting the atom, nor that Turing or McCarthy and all those before them that anticipated AI, were vicious and totally lacking any intellectual or moral virtue. On the contrary, those responsible for these two technological breakthroughs (historically and scientifically related and complementary) were, for the most part, pursuing scientific progress for noble purposes. But not only were these breakthroughs almost immediately usurped by vicious people for ignoble purposes, but the seeds of this usurpation seem to have been sown unwittingly by people like Frege, who, in myopically pursuing progress in one isolated area, set the stage for a dangerous imbalance in another. Both Turing and McCarthy seem totally unaware of what was perhaps the most important implication of the Turing Machine, namely, that the category of certainty cannot be restricted to those things we can mathematically demonstrate. We can be totally certain of things we cannot prove or see—something that John Henry Newman proposed in his *Essay in Aid of a Grammar of Assent* (1906)—and something that made its way into the official *Catechism of the Catholic Church* (1993) under the title of the “Certainty of

Faith,” paragraph 157. The many-layered and rich approach to the broad category of certainty, which began with Aristotle and was then refined by those medieval scientific/philosophical men of Jewish, Christian, and Muslim faith, is completely lost on Turing and McCarthy. Not once, it seems, do they reflect on the fact that, following in their footsteps and standing on their shoulders, most of the philosopher/scientists in the modern world who contributed to the long and arduous process of the extension of Aristotelian logic, which they inherit and without which they could have never made the progress that they did, were motivated by that classical definition of faith proposed by St. Paul: “The assurance of things hoped for, the conviction of things not yet seen” (Hebrews 11: 1).

I have already mentioned the participation of Professor Bush of MIT in the shameful Manhattan Project—a project that will go down in infamy as the first project in history designed to drop an atomic weapon on a civilian population. The perversity of such an act of terror was intensified when the decision was made to drop a second bomb even after the devastating consequences of the first were well-known. More unfathomable, perhaps, and even more obscene, were the names of these bombs, “Little Boy” and “Fat Man,” while the overall project was dubbed “Trinity.” In the face of such facts, the “mystery of iniquity” takes on new proportions. The point is that those responsible now for integrating AI into Catholic Higher Education must not be naïve; they must know the history of AI and teach this history in a dynamic and critical way. To date, there are scholars who have written fairly extensively about the role of AI in education and specifically in Catholic education. To their credit, some have focused on the ethical implications and potential benefits of AI while insisting upon fidelity to the core values of Catholic education.¹³ But much more spiritual and philosophical (metaphysical and ethical) work needs to be done.

CONCLUSION

Catholic pedagogy must be stimulated by the fire and power of the Holy Spirit so as to inspire young people to reflect upon the philosophical and moral questions of AI in a spiritual and even mystical way. Young scientists ought to be grounded early on in that school of philosophy which, for all intents and purposes, has become the ‘official’ Catholic philosophical school for our times, namely, Philosophical Personalism. And *Scientists Saints*, such as St. Maximilian Kolbe,¹⁴ whom Pope John Paul II declared to be the “Patron Saint of Our Difficult [20th] Century,” need to be held up as models and heroes in our present 21st century. As we approach the end of the first quarter of the 21st century, I think the goal of integrating AI into Catholic Higher Education must be guided along these lines; and it is indeed a hopeful sign that the new Pontiff almost immediately indicated that part of the reason he chose the name of Leo XIV was motivated by the need to address the “new things” of our times, AI topping the list, just as his predecessor, Leo XIII, addressed the “new things” *Rerum Novarum* (1891) of his times—the industrial revolution—a revolution that contained the seeds of AI. The new Pope’s *memory* of Leo XIII is now playing an essential role in his own *identity* as the Successor of Peter—an identity that for the 267th time in history, stretching back to the Fisherman himself (St. Peter), promises to guide the

World out of the ‘mechanistic’ horror of ecocide and genocide into the fields of life and light.

NOTES

1. Of course, a distinction must be made between technology per se and the result of technology. In this case, there is nothing inherently evil about the technology used to split the atom (nuclear fission). However, if the intention of splitting the atom is to create weapons of mass destruction, then the technology itself could be considered inherently evil. Historically, one impetus (though not the original one) for pursuing the technology necessary for nuclear fission was to create such weapons.

2. John McCarthy (main organizer and the one who coined the term AI), Marvin Minsky, Claude Shannon, Nathaniel Rochester, Ray Solomonoff, Oliver Selfridge, Julian Bigelow, W. Ross Ashby, W.S. McCulloch, Abraham Robinson, Tom Etter, John Nash, David Sayre, Arthur Samuel, Kenneth R. Shoulders, Alex Bernstein, Herbert Simon, Allen Newell, Trenchard More, Peter Milner, R. Culver, and Bill Shutz. Gloria Minsky, Bernard Widrow, W.A. Clark, and B.G. Farley all had visited the seminar/workshop, but did not play a major role in the discussions. In February of 2008, I had the pleasure of briefly meeting and listening to one of the key Dartmouth workshop participants, Mr. Raymond Solomonoff (present during the entire eight-week period of the workshop), when he gave the keynote address at a Conference titled “Current Trends in the Theory and Application of Computer Science” at my university (NDU) in Lebanon. Mr. Solomonoff graciously stayed on for a few weeks to give a short course on his own theory of algorithmic information. On the last day of the course, many students and teachers participated in conversing and taking pictures with Mr. Solomonoff. He died one year later.

3. Leibniz identified this as the *Entscheidungsproblem* [Decision Problem]—the same problem that Alan Turing wrestled with many years later in his famous 1936 paper, “On computable numbers.” I address this important paper below.

4. See Rule IV of Descartes’ *Rules for the Direction of the Mind*, originally published posthumously in 1701. *Great Works of the Western World*. Volume 31. Edited by Robert Manyard Hutchins. Encyclopedia Britannica, 1952.

5. See Leibniz’s *New Essays on Human Understanding*, Book IV, Ch. XVII, where he praised the Aristotelian syllogism, calling it “the greatest instrument of human reason.” Translated by Alfred Langley. London: The Macmillan Company, 1896, 557.

6. Aristotle worked only with 4 categorical statements: All p are q, No p are q, Some p are q, and Some p are Not q. The medieval logicians famously referred to these as A, I, E, and O statements, which were derived from the Latin words **A**ffirmo (I affirm) and **nEgO** (I deny). Of the 256 possible combinations of categorical syllogisms, Aristotle showed why only 15 were valid. Subsequent logicians worked on showing which of these 15 were *unconditionally* valid, as distinct from being *conditionally* valid, but always on the lookout (unlike the majority of modern logicians who inherited this same logical edifice) for the metaphysical implications of this distinction. Leibniz, to his credit, was not among these.

7. Leibniz's pioneering mathematical work, in particular his independent invention of the calculus and of binary arithmetic, which became the mathematical foundation of all modern digital computer hardware and AI processing systems. Leibniz also invented the other number system commonly used in computing, namely, base 16 or hexadecimal in modern parlance. Lloyd Strickland and Harry Lewis. *Leibniz on Binary: The Invention of Computer Arithmetic*. Cambridge, Massachusetts, London, England: The MIT Press, 2022, xi.

8. Other notable influences upon Russel and Whitehead, which I have not dealt with in this brief reflection, include the work of Georg Cantor's Set Theory, David Hilbert's Formalism, and L.E.J. Brouwer's Intuitionism.

9. Analog computers eventually became obsolete with the invention of digital computers; these latter machines applied the use of symbolic language in a way that allowed for much greater complexity, efficiency, and a much broader general range.

10. At a talk given in Calgary on 24 January 2012 by Richard Zach celebrating the Alan Turing Centenary, Zach stated that Turing "showed that no Turing machine, and hence no computational procedure at all, could solve the *Entscheidungsproblem*. Many scientific questions are considered solved to the best possible degree when we have a method for computing a solution. This is especially true in mathematics and those areas of science in which phenomena can be described mathematically: one only has to think of the methods of symbolic algebra in order to solve equations, or laws of physics, which allow one to calculate unknown quantities from known measurements. The crowning achievement of mathematics would thus be a systematic way to compute the solution to any mathematical problem. The hope that this was possible was perhaps first articulated by the 18th-century mathematician-philosopher G. W. Leibniz. Advances in the foundations of mathematics in the early 20th century made it possible in the 1920s to first formulate the question of whether there is such a systematic way to find a solution to every mathematical problem. This became known as the *decision problem*, and it was considered a major open problem in the 1920s and 1930s. Alan Turing solved it in his first, groundbreaking paper, "On computable numbers" (1936). In order to show that there cannot be a systematic computational procedure that solves every mathematical question, Turing had to provide a convincing analysis of what a computational procedure is. His abstract, mathematical model of computability is that of a Turing Machine. He showed that no Turing machine, and hence no computational procedure at all, could solve the *Entscheidungsproblem*."

11. See McCarthy's *The Dartmouth Workshop--as planned and as it happened*. John Vincent Atanasoff, too, played a key role in the invention of the digital computer, but has not really been given the credit he deserves.

12. "The Dartmouth Workshop—As Planned and as it Happened" McCarthy importantly stated: "I remember well only the events at Dartmouth that intersected with my own scientific interests, so this is not a comprehensive account of what went on. Good work that I am ignoring here includes Raymond Solomonoff's work on algorithmic information and E. F. Moore's further development of his ideas on automata. What came out of Dartmouth? I think the main thing was the concept of artificial intelligence as a branch of science. Just this inspired many people to pursue AI goals in their own ways. My hope for a breakthrough towards human-level AI was not realized at Dartmouth, and while AI has advanced enormously in the last 50 years,

I think new ideas are still required for the breakthrough. What has happened since 1956? AI research split, perhaps even before 1956, into approaches based on imitating the nervous system and the engineering approach of looking at what problems the world presents to humans, animals, and machines, attempting to achieve goals including survival. Neither has achieved human-level AI. Proposals that one approach should be abandoned and all resources put into the other are silly, as well as being unlikely to happen. I'll confine myself to engineering approaches. Within the engineering approach, the greatest success has been accomplished in making computer programs for particular tasks, e.g., playing chess and driving an off-road vehicle. None of these purports to have achieved general common-sense knowledge. Thus, the chess programs do not know that they are chess programs. Their ontology consists mainly of particular positions. The logical AI approach, starting with [#!McC59!#], which was entitled *Programs with common sense*, is in principle more ambitious. It requires representing facts about the world in the language of mathematical logic and solving problems by logical reasoning. It faces many difficulties, some of which have been overcome, and there are proposals for overcoming others. Nevertheless, there is still no well-accepted plausible plan for reaching human-level AI. For some years, I have thought that mathematical logic needs to be extended in order to represent common-sense knowledge and reasoning. That extensions are possible may seem paradoxical in the light of Gödel's 1929 completeness theorem for first-order logic. (Don't confuse this with his 1931 incompleteness theorem for formalized arithmetic.) The 1929 theorem tells us that any sentence true in all models of some premises has a proof from these premises. Therefore, any genuine extension of logic must allow inferring some sentences that are untrue in some models of the premises. The various systems of formalized nonmonotonic reasoning do precisely that. They allow inferring sentences true in *preferred* models of the premises. Human commonsense reasoning is often nonmonotonic, and human-level logical AI requires nonmonotonic reasoning, but how to do this in a sufficiently general way is still undiscovered. The need for nonmonotonic reasoning is well accepted in AI, although for specific domains, the human designer often decides what interpretations are preferred and relegates only monotonic reasoning to the computer. This is at the cost of generality. Besides nonmonotonic reasoning, I propose other extensions to logic to be able to do common-sense reasoning. These include systems with concepts as objects, systems with contexts as objects, and admitting entities that cannot be characterized by if-and-only-if definitions. I'm sure there's lots more needed before logic fully covers common sense." McCarthy continues: "Besides proposals for extending logic, there are many systems that restrict logic in order to make computation more efficient. I'd prefer to use full logic, but want systems that can reason about their own reasoning methods in order to decide on efficient reasoning. After all these years, I still have not been able to make specific proposals."

13. See the work, for instance, of scholars like Zyra Lentija, "The Protagonist and the Artifact: Reclaiming the Moral Narrative of the Person Against Algorithmic Reductionism," *Philosophia: International Journal of Philosophy* Volume 27, 2: 2026; Allan Basas, "Karol Wojtyła's Personalism and the Question of Human Agency in Artificial Intelligence," *Philosophia: International Journal of Philosophy* Volume 27, 2: 2026; Jeff Clyde Guillermo Corpuz, "Faith and Artificial Intelligence (AI) in

Catholic Education: A Theological Virtue Ethics Perspective,” *Religions* 2025, 16(8), 1083 are particularly helpful.

14. Kolbe was proficient in mathematics and physics, and he incorporated modern technology into his work of building cities dedicated to the love of his life, Mary the Immaculata.

REFERENCES

- Alam, Edward. 2002. “Descartes’ Discourse on Method: More Discourse?” *Budhi, A Journal of Ideas and Culture*; Vol. 6, No. 2 & 3: pp. 105-122.
- Aristotle. *Metaphysics*. 1924. Trans. by W. D. Ross. Oxford: Clarendon Press.
- Basas, Allan. 2026. “Karol Wojtyła’s Personalism and the Question of Human Agency in Artificial Intelligence.” *Philosophia: International Journal of Philosophy* Volume 27, 2: 284-306. doi.org/10.46992/pijp.27.2.a.3.
- Catechism of the Catholic Church*. 1993. Libreria Editrice Vaticana, Vatican. https://www.vatican.va/archive/ENG0015/_INDEX.HTM.
- Chesterton, Gilbert K. 1908. *Orthodoxy*. New York: Dodd, Mead & Co.
- Corpuz, Jeff Clyde Guillermo. 2025. “Faith and Artificial Intelligence (AI) in Catholic Education: A Theological Virtue Ethics Perspective.” *Religions* 16(8), 1083; <https://doi.org/10.3390/>.
- Descartes, René. 1952. *Rules for the Direction of the Mind*. In *Great Works of the Western World*. Volume 31. Edited by Robert Manyard Hutchins. Encyclopedia Britannica.
- Frege, Gottlob. *Begriffsschrift in Heijenoort, Jean Van*. 1967. *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*. Harvard University Press.
- George Boole, 1854. *Laws of Thought*. Cambridge: MacMillan & Co. The original publication has been digitized and is available online at: <http://www.ccapitalia.net/descarga/docs/1847-boole-laws-of-thought.pdf>
- John Paul II, 2005. *Memory and Identity: Conversations at the Dawn of a Millennium*. New York: Rizzoli.
- John XXIII. 1963. *Pacem In Terris*, Libreria Editrice Vaticana, Vatican. https://www.vatican.va/content/john-xxiii/en/encyclicals/documents/hf_j-xxiii_enc_11041963_pacem.html
- Kant, Immanuel. 1927. *Preface to the 2nd edition of his Critique of Pure Reason*. Trans. by Kemp Smith. London: Macmillan.
- Leibniz, Gottfried Wilhelm. 1896. *New Essays on Human Understanding*. Translated by Alfred Langley. London: The Macmillan Company.
- Lentija, Zyra Lentija. 2026. “The Protagonist and the Artifact: Reclaiming the Moral Narrative of the Person Against Algorithmic Reductionism.” *Philosophia: International Journal of Philosophy* Volume 27, 2: 244-264. doi.org/10.46992/pijp.27.2.a.1
- Leo XIII. 1891. *Rerum Novarum*. Libreria Editrice Vaticana, Vatican. https://www.vatican.va/content/leo-xiii/en/encyclicals/documents/hf_l-xiii_enc_15051891_rerum-novarum.html

- McCarthy, John. 2006. The Dartmouth Workshop--as planned and as it happened. <https://www-formal.stanford.edu/jmc/slides/dartmouth/dartmouth/node1.html>
- Newman, Henry Newman. 1906. *Essay in Aid of a Grammar of Assent*. London and New York: Longmans, Green, and Co.
- Russell, Bertrand, and Alfred North Whitehead. 1963. *Principia Mathematica*. Cambridge: Cambridge University Press.
- Strickland, Lloyd, and Harry Lewis. 2022. *Leibniz on Binary: The Invention of Computer Arithmetic*. Cambridge, Massachusetts, London, England: The MIT Press.