From Machine Learning to Humane Learning.
Responsible AI adoption in Catholic Higher Education.

Ciro De Florio[1]

Paolo Gomarasca[2]

Abstract

This paper outlines a practical path for adopting AI responsibly in Catholic higher education. It starts from a simple claim: AI can support learning and research only if human judgment remains in charge. The paper reviews current policies (e.g., disclosure of AI use, human oversight, academic integrity) and aligns them with UNESCO guidance and Pope Francis's Global Compact on Education, which calls for an educational pact centred on human dignity, participation, social friendship, and care for our common home. Large language models are treated as cognitive artefacts that reshape, but must not replace, reasoning and authorship. The authors propose three complementary strategies: (0) Confinement – short-term limits to protect assessment and integrity; (1) Remodulation – redesign of goals, assessments, and literacies for an AI-rich environment; (2) Cooperation – responsible partnerships in teaching and research that preserve the right to publish and public scrutiny. Concrete actions include AI literacy, transparent disclosure, governance for sensitive data and dual-use risks, environmental accountability, integrity education, and support for at-risk scholars. Ultimately, Catholic higher education is the necessary workshop for this elaboration, the place where innovation is disciplined by human judgment, oriented to truth, and accountable to the common good envisioned by the Global Compact on Education.

Keywords: Responsible AI; Global Compact on Education; Large language models; Academic integrity; Common good.

[1] Università Cattolica del Sacro Cuore, Milan.
ciro.deflorio@unicatt.it
https://orcid.org/0000-0002-5503-7463
[2] Università Cattolica del Sacro Cuore, Milan.
paolo.gomarasca@unicatt.it
https://orcid.org/0000-0002-0539-9423

1. AI and Higher Education: Landscape, Opportunities, and Risks

Artificial intelligence (AI) is reshaping higher education across research, teaching, assessment, and administration. Systematic reviews show an acceleration in AI-in-HE publications since 2021-2022, with applications ranging from intelligent tutoring and automated feedback to analytics and student support, while also surfacing concerns about equity, privacy, and ethics, especially regarding the use of AI for predicting student academic performance (Crompton & Burke, 2023). Policy analysis of the first wave of university responses to generative AI (GenAI) emphasizes three early pillars (Dabis & Csáki, 2024):

(a) reaffirming that assessed work must evidence students' own learning (for example, as outlined in Oxford University's guidance for staff, "New Artificial Intelligence (AI) tools such as ChatGPT have the potential to change the way we teach and learn in many positive ways. However, the University made clear to students in the Student News today that the unauthorised use of AI tools in exams and other assessed work is a serious disciplinary offence." (University of Oxford, 2023);
(b) maintaining human responsibility and oversight vs. using AI detection platforms and tools (the crucial point is that we lack robust evidence that AI detection platforms effectively and consistently separate human-authored work from AI-generated material; in guidance issued to staff, the University of Cape Town, for example, cautions that AI detection tools may "disproportionately flag text written by non-first language speakers as AI-generated" (University of Cape Town, 2023, p. 8); and
(c) requiring transparent disclosure of AI use in syllabi and assignments. At Colorado State University (2023), faculty can choose among three syllabus statements: one that bans AI as plagiarism; one that permits AI only with explicit permission and under specified conditions; and one that assesses AI literacy, signalling that grades will also account for students' effective use of AI tools.

These trends suggest a shift from blanket bans toward principled enablement, with local academic freedom guiding course-level decisions, particularly around large language models (LLMs).
At the policy level, UNESCO's guidance offers a normative baseline that many universities have adopted. Adopted as the outcome document of UNESCO's International Conference on Artificial Intelligence and Education (Beijing, 16-18 May 2019), the Beijing Consensus on

Artificial Intelligence and Education sets out a human-centred, rights-forward agenda for system-level integration of AI in education (UNESCO, 2019). It explicitly affirms that the development of AI should be human-controlled and centred on people, and that deployment should be ethical, non-discriminatory, transparent and auditable. Operationally, the text couples ambitions – empowering rather than replacing teachers; leveraging data for adaptive and lifelong learning; and strengthening governance, monitoring and capacity building – with cautions regarding bias, privacy, and algorithmic accountability.

As a normative instrument, however, the Consensus also exhibits a characteristic tension of soft law: it urges large-scale, data-intensive personalization while simultaneously safeguarding equity and rights, an equilibrium that is programmatically clear but institutionally under-specified. Without sustained investment in teacher professional development, robust data-governance capacity, and independent evaluation frameworks, the human-centred principles remain aspirational rather than operational.

That said, UNESCO has repeatedly reaffirmed that "the use of AI should be at the service of the development of human capabilities for inclusive, just and sustainable futures. Such an approach must be guided by human rights principles, and by the need to protect human dignity and the cultural diversity that defines the knowledge commons. In terms of governance, a human-centred approach requires proper regulation that can ensure human agency, transparency and public accountability." (Miao, Holmes, Huang, & Zhang, 2023). UNESCO urges systems to align GenAI with inclusion, equity, data protection, age-appropriate use, and teacher/researcher capacity-building (Miao, Holmes, Huang, & Zhang, 2021). Complementary macro-evidence from the 2023 Global Education Monitoring (GEM) Report cautions that technology policy must be evidence-led, appropriately governed, and attentive to costs and inequalities, especially where connectivity gaps can deepen disadvantage (UNESCO, 2023). Together, these sources press higher education to balance innovation with guardrails and resourcing.

Academic integrity remains a prominent concern, but the emerging evidence base points beyond a simple "cheating crisis." Early syntheses and institutional scans highlight redesign of assessment (e.g., oral/iterative tasks, authentic projects, transparent use policies) over prohibition, coupled with guidance for ethical AI use (Cotton, Cotton, & Shipway, 2024). Besides, the policy of forbidding GenAI in HE was later condemned as unrealistic and detrimental to cultivating AI literacy. (Rudolph et al., 2023) Recent work in higher education research confirms that generative AI intensifies long-standing tensions around academic integrity, as institutions struggle to distinguish between tool-enabled performance and genuine student learning, and to redesign assessment accordingly (Bittle & El-Gayar, 2025). This emerging literature reinforces the need for approaches that move beyond mere prohibition or detection and toward a formation of habits and structures that support honest, reflective use of AI in academic work (Bittle & El-Gayar, 2025). Empirical studies of student attitudes similarly report a complex mix of enthusiasm and anxiety about AI: students perceive real benefits in terms of efficiency and support, but also fear deception, loss of creativity, and a weakening of academic values (Benke & Szőke, 2024). These findings suggest that policies and course designs cannot be limited to rule-setting, but must involve students in co-constructing a shared understanding of responsible AI use and in articulating the kind of academic community they wish to inhabit (Benke & Szőke, 2024).

A more inclusive approach in designing academic policies, as suggested by Luo, (2024, p. 662), is certainly more promising: "Rather than stressing originality from a surveillance angle, policies can place more emphasis on the available support to students in producing original work that is meaningful to their learning. It would also be important to acknowledge broader structural factors that influence the originality of students' work, such as cultural traditions on technology use and the accessibility of GenAI in the university. Open communication and collaboration with students and faculty members should be stressed over compliance in developing policies to foster a culture of trust and care. Policymakers could invite student feedback on their policy development to demonstrate partnership over policing. A stronger emphasis on transparency and two-way dialogue among higher education leaders, staff and students may lead to governance perceived as more equitable and learning-centred by all stakeholders."

Briefly, the most robust policy responses combine integrity education, clarity on permissible assistance, and constructive alignment of learning outcomes with assessment formats in which AI can be used as a scaffold rather than a substitute for cognition. (An, Li, & Zhang, 2025)

Ethical deployment also entails attention to the environmental profile of AI. The energy demands of training and deployment are non-trivial; scholarship has quantified significant compute-related costs and called for efficiency, measurement, and design choices that minimize carbon impacts (Strubell, Ganesh, & McCallum, 2019; Luccioni, Jernite, & Strubell, 2024). Institutional AI strategies increasingly confront a dual stewardship: preparing graduates to thrive with AI while managing AI's planetary footprint, an agenda well suited to universities' research and operational roles.

Finally, ethics in AI-in-Education is not an "add-on" but touches pedagogy, assessment validity, the distribution of benefits/harms, and the purposes of education itself (Holmes, Porayska-Pomsta, & Nemorin, 2024). A unified response integrates policy (institutional rules, procurement), pedagogy (assessment redesign, AI literacy), and governance (participatory oversight, impact assessment), aligned with global guidance and local mission.

2. AI and Catholic Higher Education: Toward a Human-Centered, Integral Ecology of Learning

Catholic higher education brings distinctive theological and philosophical resources to AI adoption. *Ex corde Ecclesiae* defines the Catholic university by the unity of knowledge, the dialogue of faith and reason, and service to the common good (John Paul II, 1990). *Veritatis gaudium* renews ecclesiastical studies for today's "epochal change," calling for outreach to all disciplines and cultures (Francis, 2017). These foundations imply an AI posture that is neither technocratic optimism nor reactionary refusal, but discernment in service of the human person and the common good.

Pope Francis' *Laudato Si'* offers a critical lens on the "technocratic paradigm," urging that technology be governed by ethics, integral ecology, and care for the poor (Francis, 2015, §§102–114, 156-158). *Fratelli tutti* reframes social life around "fraternity and social friendship," inviting dialogue across differences and insisting on the inviolable dignity of every person (Francis, 2020, §§94, 99, 106). The Global Compact on Education explicitly asks educators to form persons open to encounter, inclusion, ecological responsibility, and solidarity, priorities that speak directly to AI's opportunities and asymmetries (Francis, 2020b). These touchstones provide criteria for AI use in Catholic universities: (1) the centrality of the person; (2) integral ecology; (3) participation and subsidiarity; and (4) solidarity and justice.

Recent magisterial and Vatican-adjacent initiatives engage AI explicitly. The *Rome Call for AI Ethics* (2019/2020) proposes "algor-ethics" rooted in transparency, inclusion, responsibility, impartiality, reliability, and security: "We are all aware – argues the Pope – of how artificial intelligence is increasingly present in every aspect of daily life, both personal and social. It affects the way we understand the world and ourselves. Innovation in this field means that these tools are increasingly decisive in human activity and even compelling in human decision-making. I encourage you, then, to continue in this endeavour. I am pleased to know that you also wish to involve the other great world religions and men and women of goodwill so that “algor-ethics” – ethical reflection on the use of algorithms – will be increasingly present not only in public debate, but also in the development of technical solutions. Indeed, every person must be able to enjoy a human and supportive development, without anyone being excluded. We must therefore be vigilant and work to ensure that the discriminatory use of these instruments does not take root at the expense of the most fragile and excluded. Let us always remember that the way we treat the last and least of our brothers and sisters speaks of the value we place upon all human life. We could take the example of asylum seekers: it is not acceptable that the decision about someone’s life and future be entrusted to an algorithm.” (Pontifical Academy for Life et al., 2020).

Francis’ 2024 *Message for the World Day of Peace* (“Artificial Intelligence and Peace”) and his address to the G7 on AI stress that AI is “first and foremost a tool,” whose benefits and harms depend on human governance oriented to peace, equity, and encounter (Francis, 2024a, 2024b). In 2025, Pope Leo XIV’s *Urbi et Orbi* and Angelus messages urged the world to “build bridges through dialogue and encounter, to be one people, to be at peace,” a hermeneutic that also guides Catholic institutions toward AI for reconciliation, inclusion, and human flourishing (Leo XIV, 2025a, 2025b).

Against this backdrop, a new scholarly conversation is emerging on Catholic higher education and AI. Ocampo and Gozum (2025) survey “diversions and intersections” between Catholic HE and AI, juxtaposing UNESCO’s human-centred recommendations with Catholic social teaching and urging institutions to craft policies that are at once ethically robust and pedagogically pragmatic.

Bringing the Catholic tradition into operational focus, a Catholic university AI strategy can be articulated along five commitments:

1. **Dignity and the centrality of the person.** AI should augment, not replace, formative teacher-student relationships and student agency; course policies should state what forms of AI assistance support learning versus displace it (Francis, 2020; Dabis & Csáki, 2024).
2. **Integral ecology.** Procurement and deployment should include energy-efficiency and carbon metrics, prefer efficient models, and consolidate workloads, linking campus operations with research on sustainable AI (Francis, 2015; Strubell et al., 2019/2020; Luccioni et al., 2024).
3. **Justice, inclusion, and accessibility.** Guided by UNESCO (2019/2021; 2023), ensure equitable access to AI-enabled learning, mitigate bias, protect privacy, and support

multilingual/cross-cultural contexts, especially for the most vulnerable (Miao et al., 2021; Miao & Holmes, 2023; UNESCO, 2023).

4. **Synodality, dialogue, and subsidiarity in governance.** Establish participatory oversight (faculty, students, IT, ethicists, campus ministry) and clear escalation pathways; adopt the *Rome Call* principles institution-wide; and encourage course-level discernment in assessment redesign (Pontifical Academy for Life et al., 2020; Cotton et al., 2024; Luo, 2024).
5. **Formation for citizenship and the common good.** Embed AI literacy (capabilities, limits, bias, data protection), interreligious and intercultural dialogue, and social-impact projects that use AI to serve community needs—enacting *Fratelli tutti*'s call to social friendship and the Global Compact's call to encounter (Francis, 2020; Francis, 2020b).

In short, Catholic higher education can be a laboratory for "integral" AI – pedagogically excellent, ethically grounded, environmentally responsible, and socially inclusive – where dialogue across disciplines and traditions turns technology toward communion and the common good. That arc accords with Francis' and Leo XIV's invitations to build bridges through encounter, and with UNESCO's insistence that GenAI be governed for human flourishing.

3. Focus on Large Language Models: Governance, Epistemology, and Pedagogy in a Catholic University Context

3.1 What Do We Talk About When We Talk About AI?

It is certainly not an exaggeration to claim that, over the last decade, we have been witnessing a genuine revolution in Artificial Intelligence. This revolution has been punctuated by a number of symbolic moments, whose cultural significance has arguably exceeded their actual technological impact. Yet it is well known that technology and science are inextricably intertwined with society, history, and their respective dynamics. Consider, for instance, the year 2016, when AlphaGo! defeated Lee Se-dol, the human world champion of Go!: the new software thus conquered the last stronghold of human excellence in board games (chess had already fallen two decades earlier, when Kasparov lost to the then highly advanced DeepBlue). Even more striking, however, was November 2022, when the first publicly accessible version of ChatGPT was released online. This chatbot displayed a capacity for text processing — comprehension, analysis, and production — markedly superior to that of earlier systems. Media storm, financial bubble, epochal revolution: it is exceedingly difficult, if not impossible, to maintain a disenchanted and objective perspective on a phenomenon of such breadth and complexity while it is still unfolding.

The aims of this work, however, are more modest, and we do not require an overall (if any) vision of AI. Before reflecting on the role that certain computational technologies may play in the integral formation of persons within the framework of a non-reductionist anthropology, it is necessary to establish a few preliminary points. First, we must clarify what we mean when we speak of AI. Second, we must consider the role that AI systems play in human cognition. Finally, we shall examine the impact of this technology on learning processes. These steps structure the following sections. In particular, readers who are already familiar with AI-related issues may proceed directly to Section 3.3.

It is well known in the literature that providing a definition of Artificial Intelligence that is both sufficiently informative and broadly shared is a problematic task. The reasons for this difficulty are manifold: one of the most relevant is undoubtedly the essentially interdisciplinary nature of AI research. As we shall see below, this feature has important consequences for the use of AI in education and in formative practices more generally. Let us return, however, to the issue of definition, drawing on the brilliant proposal of Minsky and McCarthy (Minsky, 1962):

(Def.) AI aims at the construction of machines capable of performing tasks which, if carried out by human beings, would require intelligence.

Much could be said about this definition. We wish to emphasize three points that, in our view, are indispensable in any reflection on the nature, limits, and applications of AI. First, this is a definition with a *counterfactual* component. Second, and relatedly, it neither claims that machines are intelligent nor denies that they are; it thus remains *agnostic* with respect to the ascription of intelligence. Third, what it does imply is a possible separation between *intelligent action* and *intelligence* itself. Translating a text from one language into another, composing a symphony, organizing a series of meetings, or playing chess are all activities that, when performed by human beings, require some form of intelligence. Now, if we succeed in building machines capable of carrying out such activities (and indeed we have built such machines), these machines act intelligently without thereby, according to our definition, being declared intelligent. Obviously, for many authors, machines are intelligent.

The point, however, is that this does not necessarily follow from the definition, and this is in fact one of its merits: it does not commit us to any specific theoretical stance.

The definition proposed by Minsky and McCarthy refers, at least implicitly, to a gradualist conception of intelligence. This is because the domain of possible human actions displays, at least at an intuitive level, varying degrees of *intelligence requirement*. Consider, for instance, a thermostat: it regulates the temperature of an environment by adapting to general conditions and to the preset value. Can we say that the thermostat is intelligent? Well, it obviously depends. The crucial point, however, is that in AI research it is impossible to speak of intelligence without reference to the environment and to the machine's capacity for adaptation to it. Now, such adaptation can occur in a very simple, linear fashion, so to speak, or it can take the form of a highly flexible adjustment to a complex environment rich in parameters. The environment of a thermostat is a single value along a one-dimensional scale; the environment of a self-driving car may be a busy street in Rio de Janeiro. As is easy to see, these represent two extremes that entail a continuum of intermediate positions.

At the core of the concept of intelligence shared within AI research lies the capacity of systems to adapt extensively to their environment in order to pursue a given class of goals. (Russell & Norvig, 2020) Adaptation to the environment may require — and in many cases in fact does require — the capacity to obtain reliable information from the environment itself. It is fairly evident that this line of reflection suggests that continuist and gradual conception of intelligence we mentioned before; in a nutshell, there are systems that adapt poorly and inadequately, and systems that adapt with great flexibility. It is, however, important to note that this conception of intelligence does not in itself preclude consideration of other, more classical, notions of intelligence — those related to the search for meaning, the interpretation of reality, and thus to an essentially human dimension. In other words, what we wish to emphasize here is that AI research does not, *ipso facto*, commit us to a particular anthropology, and that the two conceptions of intelligence just outlined can very well coexist — provided, of course, that we are clear and precise in our use of terms. In hindsight, perhaps, the choice of the expression *"Artificial Intelligence"*, made at the now-legendary Dartmouth Conference of 1956, has ultimately proved to be a source of ambiguity regarding the very nature of intelligence.

3.2 A Geography (and the Fortune of a Region)

There exist many approaches to AI; this means that there are many different ways of constructing machines capable of carrying out tasks that require intelligence. For our purposes, it is important to mention a possible taxonomy of AI technologies, since, as we shall shortly see, the debate has in fact concentrated on a specific type of approach and, more precisely still, on a particular set of software applications with a specific use. What has occurred, sometimes unconsciously, is a kind of *pars pro toto* operation: AI has come to be treated as synonymous with the chatbot, or, more technically, with chatbots based on Large Language Models. This is not merely a socio-technological contingency but has significant implications for the way in which we interrogate these technologies, the way in which we attempt to locate them within the world, and the way in which we conceive of the relations between ourselves and these artefacts.

Research conducted under the heading of AI is as old as computer science itself; indeed, one might say that, at a certain level of abstraction, it coincides with the very aims of computing. From the very beginning, software was developed to perform cognitive tasks: from *Turochamp*, Alan Turing's chess-playing program, to *Logic Theorist*, capable of proving certain theorems of propositional and predicate logic, to *Eliza*, which simulated simple conversations with a psychotherapist.

In brief, there are two main strategies for teaching a machine to perform a complex task. The first is to write a program, that is, a series of instructions the machine follows in order to accomplish the task at hand. Such a program must be sufficiently flexible to enable the machine to cope with a wide range of situations. In the case of a chess-playing program, for example, we must anticipate the machine's responses to the possible moves of its opponent. Within such programs, a great deal of knowledge must be encoded, in particular the knowledge relevant to the specific domain of the task in question. Returning to the example of chess, a chess-playing software system must contain, in a form accessible to the machine, a considerable amount of chess-related knowledge.

In other words, this knowledge must be *represented*. A program designed to help us catalogue cooking recipes will, according to this approach, contain a representation of the culinary world in terms of categories and subcategories, together with the relations between these items. For this reason, one speaks of *computational ontologies*: just as in metaphysics, the goal is to *carve reality at its joints*, that is, to identify an effective partitioning of reality that can be represented through concepts on the basis of which the machine can build classifications and inferences.

The other path is to build a machine that learns autonomously how to perform the task. The program (to put it in very simplified terms) will not consist of instructions for playing chess directly, but rather of instructions for learning how to play chess. It will therefore be a different kind of program, whose aim is to implement choices (mostly random), to evaluate the responses of the environment, and to modify the general strategy accordingly. This is the essence of *machine learning*, the approach through which software systems are trained.
There exist many approaches to machine learning; at present, the most widespread systems are those based on so-called deep neural networks. These are, again speaking in highly simplified terms, AI architectures composed of a series of layers made up of neurons. The neurons — whose inspiration from the biology of the brain is evident — are connected to one another by informational links and are equipped with specific activation thresholds. The first layer of the network is the input layer, where an initial set of neurons is activated; the last layer is the output layer. For instance, the input layer may consist of the values corresponding to the pixels of a digitalized image, while the output layer produces the name of the object represented in that image (within a certain confidence interval). Between the two lies the entire functioning of the network, which consists in nothing other than learning and improving its performance by adjusting the activation values and thresholds of its neurons.
Neural networks, in other words, are capable of adapting remarkably well to complex stimuli and prove to be extraordinarily effective in complex environments. For example, as already mentioned, in image recognition and, even more famously, in language. Large Language Models (LLMs) are computational technologies based on neural networks trained on massive amounts of linguistic data, enabling them to understand, analyze, and generate natural language. As noted at the outset, the commercial success and media explosion of these systems have led to a situation in which an entire field of research (AI) has come to be regarded as coextensive with a particular subfield of it (namely, subsymbolic models based on machine learning through neural networks trained on vast datasets).
We may therefore speak of a kind of computational synecdoche: LLMs become the very archetype of AI. This is, after all, not an entirely new phenomenon in the history of technology (when we think of a robot, we almost immediately imagine a humanoid robot, although such machines represent only an infinitesimal and, in practice, marginal fraction compared to the vast number of robots employed in industry). Yet it is important to take this into account, since when we speak of education and pedagogy in the AI era, we are often — albeit implicitly — assuming that AI is reducible to LLMs. And, as we shall see, this is by no means a neutral assumption.

3.3 AI technologies as cognitive artefacts

The synthetic overview presented in the previous section has attempted to map the macro-region of AI by providing a general subdivision of the approaches and technologies involved. What, however, eludes this kind of consideration are answers to more fundamental questions: what kinds of entities are AIs? What is their place in the world? What is their proper function? These are, as is evident, questions posed at a different level from that of any inter-theoretical classification.

A seemingly straightforward answer is to regard AIs — correctly — as particular kinds of software. But what is software? And what, more generally, is a computer? The debate on these questions is vast, and many proposals have been advanced. In the present context, however, it appears dialectically useful and strategically fruitful to shift perspective and to consider AI systems as instruments constructed by human beings for the execution of specific tasks. We shall call this view the *artefactual conception of AI*. More specifically, we advance the following thesis, namely that

(1) AI systems constitute a particular class of cognitive artefacts.

In order to evaluate (1), it is necessary to examine, albeit briefly, the conceptual category of cognitive artefacts. Cognitive artefacts (Heersmink, 2013) exhibit two fundamental characteristics: (a) they are physical and artificial objects; (b) they serve to accomplish a cognitive task.

Let us begin with (b). Clearly, a screwdriver is not a cognitive artefact because, although it is a physical object and is constructed by human beings, it does not assist us in performing a cognitive task: screwing or unscrewing a screw does not pertain to our cognitive functions. It clearly does not follow from this that some basic cognitive functions are not required in order to drive a screw. By contrast, a map is a cognitive artefact because it enables us to orient ourselves, and orientation is (widely considered) a cognitive function. In the same sense, abaci, calculators, thermometers, dictionaries, and languages all qualify as cognitive artefacts. One may in fact question whether being physical objects is a necessary condition; after all, one might doubt the concreteness of entities such as languages, software, or even mathematics itself. These metaphysical reflections, however, lie beyond the scope of the present work and would lead us too far afield. (Liggins 2024) In the present context, we shall assume that at least the instances or implementations that perform the function of cognitive artefacts must have a spatio-temporal location. Thus, one may regard software as an abstract (or quasi-abstract) object, but its exemplification within a hardware system constitutes a complex functional state that is concretely realized; that is, it has a spatio-temporal location and is embedded within a chain of causal relations.

How are cognitive artefacts to be classified? A fairly widespread taxonomy in the literature divides them into two broad macro-categories: *representational* artefacts and *non-representational* artefacts. The latter will not concern us here (also because they are, all things considered, relatively uncommon). The former — that is, representational cognitive artefacts — convey information in three different ways, corresponding to three subcategories: *iconic*, *indexical*, and *symbolic*.

Iconic cognitive artefacts are those that assist us in performing a cognitive task by conveying information through their representational and visual nature. The classic example is the map, which essentially resembles the territory. Indexical artefacts, by contrast, are those that represent specific information about a system by means of an index, as in the case of a thermometer. Finally, symbolic cognitive artefacts are those in which information is transmitted through a conventional structure. The paradigmatic example of this latter type is language.

This taxonomy has been criticized for involving both cases of overdetermination and cases of underdetermination. In the present context, we shall not engage in a discussion of the theoretical merits and shortcomings of this classification. What we shall retain from these preliminary reflections are two essential points: namely, that cognitive artefacts must in some way convey information, and that the user's intention plays an important role in their functioning [1]. A promising approach is to consider cognitive artefacts not in themselves, but in relation both to the cognitive task under consideration and to the human user (and, as we shall see, this strategy will prove fruitful in the discussion of AI systems in educational contexts). With respect to a given cognitive task, a cognitive artefact may be: *complementary*, *innovative*, or *substitutive*. Let us examine these three categories in order.

A cognitive artefact is *complementary* when it enables the accomplishment of a cognitive task that, at least in principle, could be carried out by an epistemic agent without the artefact in question. Among the most familiar examples are those artefacts that enhance our mnemonic capacities: remembering is a paradigmatic cognitive activity that can be complemented through the use of various kinds of external support. Another example is the electronic calculator, in its most basic form. It is certainly possible to compute mentally the result of 45 + 98, but this operation can be complemented by the use of a calculator. Note that even writing and the algorithms of the four basic operations may be classified as complementary cognitive artefacts.

The point, however, is that they are not only that. There are tasks that necessarily require the use of a cognitive artefact in order to be carried out. Consider, for instance, the multiplication of rather large integers: 98,334 × 1,998,344. It does not seem possible to perform this calculation mentally; the intervention of a cognitive artefact is required. More significantly, the use of such technologies transforms the way in which we perform the relevant operations. For this reason, we speak of *innovative cognitive artefacts*: the technology alters the manner in which we carry out these activities. The most illuminating example in this respect is probably that of writing, which has profoundly reshaped our cognitive life. More generally, it is difficult to identify cognitive tasks that are not, in practice, mediated by some cognitive technology: much information acquisition, for example, occurs through reading, which in turn requires written language.

The final category of artefacts is that of *substitutive artefacts*, namely technologies that allow for the complete replacement of the epistemic agent in the execution of the task. Given the relevance of this category for reflections on AI, it is worth devoting a few additional words to it.

Let us begin with an example somewhat more sophisticated than those previously discussed. Suppose we need to purchase a sheet to cover the surface of our garden. Our garden has a trapezoidal shape: the measurements are 24 meters for the larger base, 21 meters for the smaller base, and 18 meters for the height. How much material do we need to buy? To find out, we must calculate the area of the plot and perhaps add an extra 10% as a margin of safety.

For the sake of discussion, let us assume that we do not quite recall the formula for calculating the area of a trapezoid. We may vaguely remember that a trapezoid is equivalent to a triangle, but little more. The best course of action would likely be to resort to a cognitive artefact that complements our memory. Since this is a thought experiment, let us imagine opening an old geometry book and finding the formula for calculating the area of a trapezoid. At this point, we must carry out the calculation and, again, let us assume that we are somewhat out of practice in computing the following operation:

(2) $a = ((21\text{m}+24\text{m}) \times 18\text{m}):2$

We take paper and pen and obtain the result: 405 m². We therefore need 405 m² of covering material. But wait, we had agreed to purchase a little more as a safety margin. How do we calculate 10%? Percentages are by now a faded memory, and we cannot turn to the geometry book for help. What should we do? We take a calculator and enter:

(3) 405 + 10% =

The result is 445.5; at last, we can go to the store to purchase our covering sheet. It is quite evident that in the final step — namely, the cognitive task of calculating the percentage — we relied on a *substitutive* technology: the calculator performs the computation even if we have no idea what a percentage actually is. All that is required of us is to provide the input data.

It is now important to focus on one final point concerning the category of *substitutive artefacts*. The very notion of a cognitive task is multi-layered — as we shall see more clearly below. In the previous example, calculating the addition within (2) is itself a task, but calculating (2) as a whole is also a task. Obviously, in order to perform the latter, one must first carry out the former. Yet it is also a cognitive task to recall the formula for calculating the area of a trapezoid, and it is a cognitive task to perform the following means/ends inference:

(p1) I want to cover the surface of the garden. (p2) I want to know how much material I need to procure. (c) It is necessary to calculate the area of the garden.

In fact, cognitive tasks can be decomposed in many different ways, many of which are likely to be equivalent. Moreover, the granularity of our analysis is contextual. The consequence of this is that cognitive artefacts may appear complementary at one level of analysis, innovative at another, and substitutive at yet another. The use of the calculator replaces the task of calculating percentages; yet, in turn, it is complementary to the higher-level, more general task of calculating the area of the trapezoid. In other words, the calculator assists us in determining the area of the trapezoid, which is itself a precondition for knowing how much covering material to purchase.

To summarize, then: cognitive artefacts are physical objects that convey information and contribute to the performance of a cognitive task. Depending on how we identify the task in question, cognitive artefacts can supplement our "natural" cognitive capacities in ways that are more or less dispensable, up to cases in which they simply replace us in carrying out the task. So far, so good. Now, the conceptual challenge we face is the following: where should we place recent information technologies based on AI approaches that make use of machine learning algorithms?

In general, “classical” information technologies occupy, respectively, all the categories under consideration. Take, for instance, a spreadsheet: it certainly complements our ability to organize and store large amounts of data. Yet it would be more accurate to say that it innovates the way in which we process information: without it, it is simply impossible to derive new information from the database at our disposal. Finally, as is well known, in a spreadsheet it is possible to insert macros that autonomously supervise certain sub-tasks, thereby automating and accelerating a series of operations. But what happens when we employ an advanced AI software system? And, above all, a generalist AI software, that is, one capable of handling text, images, and sounds for an extremely wide range of functions?

We submitted the issue in question to ChatGPT-5 with the following prompt:

I need to buy some tarpaulin to cover a plot of land in the shape of a trapezoid whose bases measure 21 and 24 meters. The height is 18 meters, and I would like 10% more tarpaulin just to be safe. How much tarpaulin do I need to buy?

Here the result:

Ok, let’s calculate step by step:

1. **Trapezoid data**

- Larger base B=24 mB = 24\ \text{m}B=24 m

- Smaller base b=21 mb = 21\ \text{m}b=21 m

- Height h=18 mh = 18\ \text{m}h=18 m

2. **Area of the trapezoid**

A=(B+b)×h2A = \frac{(B + b) \times h}{2}A=2(B+b)×h A=(24+21)×182=45×182A = \frac{(24 + 21) \times 18}{2} = \frac{45 \times 18}{2}A=2(24+21)×18=245×18 A=8102=405 m2A = \frac{810}{2} = 405\ \text{m}^2A=2810=405 m2

3. **Adding the 10% safety margin**

Atot=405×1.10=445.5 m2A_{\text{tot}} = 405 \times 1.10 = 445.5\ \text{m}^2Atot =405×1.10=445.5 m2

👉 You need to buy **about 446 square meters** of tarpaulin (rounding up, it would be safer to get at least 450 m²).

As can be seen, the software autonomously carried out both the main task and the subsidiary ones, even providing a rounding derived from the context specified in the prompt. There is no doubt that we are facing a *discontinuity* with respect to the cognitive technologies we have considered so far. An LLM is capable of grasping the problem in its generality and of performing a series of operations entirely on its own, without the intervention of the epistemic agent. We may disregard the formula for calculating the area of a trapezoid, just as we may disregard the operational procedure for computing a percentage. But, more profoundly still, we may disregard the very concepts of area and trapezoid. Every cognitive function involved in the task has, in fact, been assumed by the software: we have consulted the oracle, and we have obtained an answer [2].

LLMs thus appear to be cognitive artefacts belonging to yet another category: their capacity to replace human epistemic agents in the resolution of complex tasks is incomparably more advanced than that of existing information technologies. The fact that the prompting of these systems can take place in natural language removes the barrier of programming, or at least the need to regiment language into a series of instructions. This has many consequences, as one might expect. The one that concerns us most directly, however, is precisely that it is possible to ignore (or only partially grasp) the conceptual structure embedded in the problem and nonetheless obtain an answer from the software. The dimension of cognitive substitution is complete: AI systems no longer merely supplement human beings in cognitive tasks but perform cognitive tasks on behalf of human beings.

3.4. A conceptual framework for learning

At this point, we have all the premises in place to undertake a reflection on the impact of AI technologies on certain aspects of education. To this end, we shall outline a Learning Conceptual Framework (LCF). This is, of course, a model, and as such it inevitably involves certain simplifications.

The LCF comprises three components: aim, task, and steps. Let us consider them briefly. The *aim* concerns the ultimate purpose of the learning process. Aims may include, for example: [learning to write a few short meaningful sentences]; [handling the concepts of area and perimeter]; [composing a descriptive text from a particular perspective]; and so forth. Aims pertain to the general objectives of education, and they obviously vary depending on the learners' age, their prior training, their chosen specializations, and even their cultural backgrounds. What the aims are—namely, what the goals of education should be—is a matter of debate within pedagogical communities, and even a cursory presentation of the positions at stake would clearly fall beyond the scope of this article [3]. It is important to emphasize that aims possess an internal structure. Consider, for instance, the aim of being able to compose a descriptive text. This can be situated within a hierarchy of inclusion, insofar as it constitutes a specification of the more general aim of being able to compose a text. In turn, there may be sub-aims, such as the ability to compose a descriptive text in a journalistic style.

In order to achieve the aims, specific tasks are designed—that is, the actual assignments to be carried out. Our students are primarily exposed to tasks rather than to aims (which sometimes remain at the level of the teachers' intentions). A task connected to the aim discussed above might be: *describe your room* or *describe a typical day of your holidays*.

The formulation of a task may be more or less specific, depending on which aim is intended to be achieved. Similarly, the question of which tasks should be assigned in order to accomplish a given aim, or a class of aims, remains entirely open. Indeed, one might say that a significant part of instructional design precisely consists in structuring a series of tasks suited to educational needs and to particular contexts. Finally, it may happen that the conceptual distance between aim and task is greatly reduced, to the point of almost disappearing. For example, if the aim is to be able to read short sentences with correct intonation and proper observance of punctuation, the task may consist precisely in reading exercises.

In order to solve a task, one typically proceeds by decomposing it into sub-tasks, which in turn can be carried out through a series of steps, that is, elementary cognitive operations. Here again, we do not commit ourselves to identifying such ultimate atoms of learning, beyond which further decomposition is impossible; nor do we wish to defend the thesis that there exists a single or well-defined number of correct ways of executing a task. What matters for our purposes is a more basic consideration: namely, that the task itself has structure, and that this structure unfolds in a series of steps and sub-steps [4].

Let us return to our earlier example: the task was to determine how much fabric would be required. In order to do this, we had to devise a non-trivial procedure. First, to understand that the area of the garden was needed; then, to recall the formula for calculating the area; subsequently, to carry out the calculations; to add 10%, and so forth. It must also be noted that certain cognitive sub-tasks serve as prerequisites for others. For instance, the ability to move from an instruction expressed in natural language (in this case, English) to a mathematical formalization underlies many of the mathematical competences required in compulsory education.

We have thus completed the presentation of our Learning Conceptual Framework. It is now time to ask where cognitive artefacts come into play and whether AI systems constitute a significant discontinuity.

3.4.1. Where Does Technology Intervene?

On the basis of what has been said above, it is easy to recognize that the presence of cognitive artefacts permeates much of our mental life. There is evidence that the use of such technologies has shaped our very cognitive processes—and that this continues to occur. With what outcomes, however, remains a matter of debate. Let us recall that our definition of cognitive artifact is highly flexible and, as noted, *cognitive-centered*, that is, attentive to the dimension of the epistemic subject. It does not directly concern the intrinsic complexity of the artifact in question. A table of logarithms is a cognitive artifact, just as the software Wolfram Mathematica is. There is, moreover, no doubt that certain cognitive artefacts are readily employed in addressing the various steps into which a task is structured. Consider, for instance, numerical operations, or looking up a word in a dictionary while translating. Such uses are not merely tolerated in traditional educational contexts but, for the most part, encouraged; and they can themselves become the object of specific training (for example, when one is taught how to use a dictionary).

With the exception of particular cases – namely, the early stages of learning, where it is necessary to build basic competences – the general objectives are not grounded in the execution of these steps. Put differently, using a calculator does not significantly alter a student's mathematical abilities; nor does the possibility of consulting a dictionary improve their writing skills. This is because, once we move to a slightly higher level of competence, what is required is not so much the autonomous execution of individual steps as, crucially, the transition from a task to its solution structure. The ultimate goal is the capacity to structure solutions in relation to the individual tasks that are posed.

"Understanding mathematics" precisely means being able to solve the problem—that is, to devise a procedure that can lead us to the indicated result of the task on the basis of the data at our disposal (together with background knowledge). As is well known, this *creative* step has a substantial inventive component. Indeed, the more refined and sophisticated the task, the more this characteristic is emphasized: the student is required to make a productive conceptual effort rather than merely reproduce manipulative techniques mechanically.

The more general the task, the greater the need for adequate operationalization in order to arrive at the sequence of steps necessary for its completion. This will be a complex undertaking requiring a broad spectrum of cognitive abilities and, consequently, a diversified use of cognitive artefacts.

It might therefore be thought that AI systems, in particular those based on Machine Learning and Deep Neural Networks, belong within this continuum between cognitive technologies and the epistemic competences of subjects. This is true, but only in part. What appears to mark a discontinuity with respect to the cognitive artefacts with which we have interacted for thousands of years is that the plasticity of use of these models allows them to replace the human agent even in the very choice of the solution procedure. That creative and inventive component, which is the true *upshot* of a learning process, can be (almost) entirely outsourced to an LLM, for example. We have already seen this with the simple problem of the fabric. And this bypasses altogether the *conceptual competences* required to determine the steps for solving the task. What is at stake is not merely carrying out a calculation, but setting it up – that is, understanding what type of operation is required in the given context. Certainly, the performance of LLMs is the object of ongoing evaluation, and there is no doubt that their reliability remains a matter of debate – for, after all, they are machines trained to provide plausible rather than true answers. Yet it is equally reasonable to expect that technological progress will enhance their performance in this respect. In any case, this issue does not in itself affect the educational problem they raise. In what follows, we shall consider some possible responses to the advent of these formidable cognitive machines in the classroom.

3.5. What Is to Be Done? Possible Strategies

The technological evolution of AI systems is a fact. As such, it is non-negotiable. What we must do, rather, is to understand it, frame it within an explanatory network, and, if necessary, adopt behaviors that take this fact into account. The worst strategy would be to deny reality or to trivialize the phenomenon. Although the future of the relationship between society and AI is marked by deep uncertainty, it seems implausible that an increase of such magnitude in the performance of a technology should remain without consequence. Let us therefore consider three strategies that are intended to be neither exhaustive nor mutually exclusive. They are three proposals that call for further elaboration and, even more importantly, possible empirical verification in practice.

3.5.1. Strategy 0: Confinement

In the short (or very short) term, one may decide to confine the use of AI systems within educational contexts. This is clearly a defensive and strongly conservative strategy. Yet in the absence of concrete proposals, it is evident that the liberalization of such technologies in the educational settings to which we are accustomed is simply disruptive. We therefore do not feel that limited containment measures, restricted in time, should be ruled out a priori—even though, as noted at the outset, the presence of AI systems in HE, as well as in CHE, is already so developed and structured as to render an enduring containment strategy both very costly and probably meaningless.

3.5.2. Strategy 1: Remodulation

Cognitive technology induces a remodulation of learning aims. This seems invariably to be the case; let us think, for a moment, of the introduction of low-cost electronic calculators. It is clear that the significance of students' computational abilities was redefined in the light of such technologies. To be sure, *knowing how to calculate* was not eliminated from curricula. But two things became evident: first, that these skills of symbolic manipulation are meaningful primarily as a preparation for higher-level abilities; and second, that logical-mathematical education does not ultimately rest upon these competences.

The advent of AI systems therefore requires a remodulation of teaching practices. But how? We do not yet have an answer – except to note that it will certainly not be univocal but will vary according to disciplines and contexts.

The point we wish to emphasize here is that learning aims must increasingly focus on the process - that is, on performance - rather than on the expected outcome. In a post-AI educational landscape, the structural framing of the problem becomes the true upshot. While this shift is, overall, achievable without profound changes in the assessment phase, what genuinely requires substantial remodulation is the training phase. The very concept of homework must be thoroughly reconsidered, as there is virtually no task that cannot be replicated by current technologies.

The rethinking of the learning system therefore calls, almost inevitably, for a form of pedagogical alliance. And this leads us directly to the next strategy: cooperation.

3.5.3. Strategy 2: Cooperation

The real challenge for educational institutions is cooperation: these systems must become allies in the learning process. One possible path – still to be explored both conceptually and in empirical implementation – is to exploit the generative capacities of these systems. We may call this the *sparring partner* sub-strategy. LLMs can very effectively replicate artificial tasks and can assess the steps produced by students. They can construct exercises, compare solutions, and suggest translational renderings. In this way, the product of a training session, perhaps carried out at home, can subsequently become the object of a joint meta-analysis in the classroom. The alliance here is between learners, the teaching body, and AI systems. This is a strongly innovative didactic paradigm, which requires considerable maturity on the part of all those involved, as well as at least an architectural understanding of these machines.

4. Discussion: AI and Integral Education

In what follows, we shall sketch some ideas concerning the relationship between AI and Integral Education. Artificial intelligence now functions as a pervasive cognitive artifact, a tool that reorganizes how attention, memory, and judgment are distributed across persons and technologies.

AI does not only accelerate pre-existing tasks; it re-architects cognition in three interlocking ways. First, there is a functional redistribution of cognitive work: working memory and retrieval shift toward prompting and selective evaluation; drafting shifts toward curation and revision; and explanation shifts toward metacognitive oversight (asking for alternatives, counter-examples, or uncertainty bounds). Second, AI introduces new forms of epistemic dependence on systems whose performance is powerful but opaque, with failure modes (e.g., hallucination, training-set blind spots) that users do not easily anticipate. Third, as Schneider (2025) argues, AI assistance reduces users' metacognitive monitoring during decision-making, fostering overconfidence in AI capabilities and impairing metacognitive oversight in subsequent decisions that do not depend on AI – a phenomenon that can be labelled "metacognitive outsourcing".

The relevance of these technologies for education, and in particular for Catholic Higher Education, is profound. In this paper we have explored a possible interpretative key, one that decisively points toward an integral approach, aimed at the formation of an *academic conscience*, as has recently been advocated, for instance, by Duncan Pritchard. Within the IAU-UNESCO Lecture Series on "The Academic Conscience & Artificial Intelligence" (Lecture: How Does AI Shape Human Cognition and Pedagogy?, 11 September 2025), Pritchard points out that universities must decide not merely whether to use AI but how to govern its integration so that it strengthens, rather than supplants, the learner's intellectual agency and responsibility. The decisive claim, consistent with Pritchard's broader work in virtue-theoretic epistemology, is that the relevant unit of analysis for knowledge is increasingly the person-plus-artifact system, yet such extended processes yield knowledge only when they are appropriately integrated and guided by the knower's intellectual character. The educational task is therefore to calibrate trust and reliance: to know when to lean on such tools, how to check and corroborate their outputs, and how to attribute intellectual labour responsibly. Consequently, critical thinking is no longer optional in universities because our students are graduating into socio-technical systems, especially AI-mediated ones, where hidden assumptions, incentive structures, and power asymmetries shape real human outcomes; rigorous inquiry must therefore move beyond critique to *counter-design*, the disciplined practice of articulating and prototyping better alternatives (including justified non-deployment) that make values, trade-offs, and accountability explicit. In this sense, counter-design is the practical horizon of critical thought: if we can name a harm or a bias, we must also experiment with interfaces, policies, and governance forms that reduce it and can be publicly evaluated. Catholic universities add a distinctive voice here by grounding counter-design in Catholic social teaching: human dignity and the sanctity of the person set red lines for what must not be optimized away; the common good and solidarity demand designs that distribute benefits and burdens fairly; subsidiarity favours empowering local communities rather than centralizing opaque control; the preferential option for the poor centers those at the margins in requirement-gathering and testing; and "integral ecology" (e.g., *Laudato Si'*) widens impact assessment to creation itself. (Gozum & Eballo, 2025) Together these principles push counter-design toward humane, truth-seeking, and often restraint-valuing alternatives.

Counter-design skills and attitudes finally ripen into academic conscience: The "academic conscience", again according to Pritchard – names the cluster of institutional and pedagogical responsibilities that follow. If AI is here to stay as an ingredient of routine scholarly practice, then the formation of the user becomes as important as the specification of the tool. Put sharply: without virtue-guided integration, AI-assisted performance risks becoming merely performance. With such integration, extended processes can be knowledge-productive and character-forming. This distinction is not semantic. It shapes curricular aims (what we are trying to form), assessment design (what we measure), and governance (what we permit or proscribe in tool adoption). It also suggests a criterion for institutional prudence: technologies should be welcomed insofar as they make human judgment more visible, accountable, and truth-oriented, not less.

When our prototypes and policies are judged not only by technical merit or social utility but by an unwavering preference for the human person, the common good, and creation itself, so that innovation is answerable to truth and charity, not merely to feasibility or scale. This makes the Catholic orientation constitutive rather than cosmetic: it names the criteria of judgment (human dignity, solidarity, subsidiarity, care for creation) that govern design choices, reframes success beyond optimization metrics, and treats the refusal to deploy as a morally serious outcome when dignity or the common good would be compromised; in short, counter-design becomes an exercise in truthful, responsible love within the university's vocation to form conscience.

5. Conclusion

This paper began with a very concrete question: how can higher education, and Catholic higher education in particular, adopt AI in ways that genuinely support learning and research without displacing human judgment? Section 1 mapped the current landscape: AI is already deeply embedded in university life, from assessment and tutoring to analytics and administration, and institutions are moving from prohibition to more nuanced forms of enablement. Policies on disclosure, integrity, and human oversight are emerging, often guided by UNESCO and related international frameworks, and they share a common intuition: AI is neither a neutral tool nor an irresistible fate, but a socio-technical choice that must be governed. In that sense, AI in higher education is a test case for our understanding of what education is for.

Section 2 then argued that Catholic higher education brings distinctive resources to this discernment. *Ex corde Ecclesiae* and *Veritatis Gaudium* frame the university as a community ordered to the unity of knowledge, the dialogue of faith and reason, and service to the common good. *Laudato Si'*, *Fratelli tutti*, the *Global Compact on Education*, and the *Rome Call for AI Ethics* provide a thick set of criteria - centrality of the person, integral ecology, participation and subsidiarity, solidarity and justice - for judging technologies and their institutional uptake. These sources resist both technocratic optimism and reactionary refusal: they call instead for a pedagogy and governance of AI that protects the vulnerable, widens participation, and orients innovation toward communion and care for our common home.

The subsequent sections translated these broad orientations into a more specific analysis of AI as a family of cognitive artefacts, with particular attention to large language models. By situating LLMs within the continuum of tools that extend human cognition, we highlighted both the continuity with earlier technologies (from calculators to search engines) and the discontinuity introduced by systems that can autonomously select and execute solution procedures. Through the Learning Conceptual Framework (aim–task–steps), we showed that

AI can now intervene at precisely the level where educational value is concentrated: the transition from a task to its appropriate solution structure, the inventive moment in which a learner must mobilize concepts, judgment, and creativity.

From this perspective, the three sketched strategies - confinement, remodulation, and cooperation - can be read as different ways of protecting and reconfiguring this inventive core. Confinement is a short-term protective response, aiming to preserve assessment validity while institutions learn to cope with disruption. Remodulation acknowledges that aims and assessments must change in an AI-rich environment, so that we do not confuse mechanical performance with understanding. Cooperation, finally, explores how AI might become a "sparring partner" in learning, generating prompts, counter-arguments, and alternative pathways that can be critically appropriated rather than passively consumed.

If we now bring this analysis back to the specifically Catholic vision outlined in Section 2, we can state the central claim of the paper more forcefully. The challenge is not simply to *AI's presence in higher education with an integral ecology of learning - one that holds together intellectual* regulate AI, but to align AI's presence in higher education with an integral ecology of learning - one that holds together intellectual, spiritual, social, and ecological dimensions. Catholic universities are called to be laboratories where this alignment is worked out in practice. This requires that the ethical and theological criteria advanced in Section 2 be allowed to shape, in a visible and verifiable way, the policies, pedagogies, and governance mechanisms sketched in Sections 3 and 4.

We may therefore propose, by way of conclusion, an expanded test for discerning educationally and mission-appropriate uses of AI in Catholic higher education. An institutional use of AI passes this test when it satisfies at least four interrelated conditions:

a) Truth-seeking and understanding (Section 1). AI-supported practices must demonstrably promote understanding rather than merely optimize outputs. This means that assessment design, course policies, and research practices should make it possible to distinguish between performance generated by an artefact and insight acquired by a person. Where AI assists with drafting, problem-solving, or data analysis, students and staff should be able to articulate what they have learned, which judgments they have made, and how they would justify those judgments independently of the tool.

b) Formation of intellectual virtues (Sections 1 and 3–4). AI integration should foster, not erode, the intellectual character of learners and teachers. Epistemic virtues such as honesty, intellectual courage, humility, and attentiveness are tested in AI-mediated environments: the temptation to outsource effort, conceal assistance, or suspend critical scrutiny is real. Educationally appropriate uses of AI will therefore include structured opportunities for metacognitive reflection, transparent disclosure of AI use, and practices that reward intellectual initiative and responsible authorship. Here, the notion of "academic conscience" becomes pivotal: universities must cultivate habits of self-scrutiny and institutional reflexivity about when and how AI is used. (Dicastery for the Doctrine of the Faith & Dicastery for Culture and Education, 2025)

c) Respect for human dignity, solidarity, and integral ecology (Section 2). AI adoption must be compatible with the inviolable dignity of every person and with the pursuit of the common good, including care for creation. Concretely, this entails resisting uses of AI that treat students or staff as data points to be optimized, that exacerbate exclusion or bias, or that externalize environmental costs. It also entails a preferential attention to those at the margins - students with limited access to technology, communities disproportionately affected by surveillance or automation, and the non-human environment that bears the burden of AI's energy demands.

d) Participation, subsidiarity, and synodal governance (Section 2). Decisions about AI in teaching, research, and administration should not be made solely "from above" or outsourced to vendors. Instead, they should emerge from processes that involve faculty, students, professional staff, and, where appropriate, community partners. Such participatory governance expresses the Catholic commitment to subsidiarity and synodality: it recognizes that those closest to the educational process have privileged insight into how AI helps or harms learning, and it creates channels for ongoing discernment and course correction.

When these four conditions are taken together, the earlier "concise test" of AI use - supporting truth-seeking and understanding, forming intellectual virtues, and respecting human dignity and the common good - acquires a more concrete institutional profile. It links the micro-level of classroom design and supervision, the meso-level of faculties and academic support services, and the macro-level of university-wide strategy, procurement, and public witness. It also makes clear that some uses of AI, however efficient or innovative they may appear, will be judged incompatible with the university's mission and therefore not adopted or actively discontinued. The non-deployment of certain tools is thus recognized as a positive, conscientious decision rather than as a failure to innovate.

In sum, moving "from machine learning to humane learning" is not a slogan but a demanding orientation. It asks Catholic universities to receive AI as a sign of the times that calls forth new forms of critical literacy, institutional courage, and creative governance. It also invites them to contribute, alongside other higher education institutions and international bodies, to the global conversation on responsible AI. If Catholic universities can show, in concrete practices, that high-quality research and teaching with AI can coexist with - and even strengthen - human judgment, solidarity, and care for our common home, they will not only safeguard their own identity. (AI Research Group of the Centre for Digital Culture, 2023) They will offer a credible and hopeful model of how technological power can be ordered to truth and to the flourishing of all. (Corpuz, 2025)

In this sense, the recent volume *Artificial Intelligence and Care of Our Common Home: A Focus on Industries, Finance, Education and Communication* offers a convergent horizon for Catholic higher education and for the concerns articulated in this paper. In his introduction, Cardinal José Tolentino de Mendonça insists that discernment about AI cannot be reduced to technical feasibility alone: "Faced with artificial intelligence, we cannot limit ourselves to asking what is technically possible. We must also ask, is it right to do so?" (Tolentino de Mendonça, 2025, p. 13). This shift from the question Can we? to the question Should we? resonates directly with the ethical and theological criteria recalled in Section 2, where Catholic social teaching and the language of care for our common home set boundaries and priorities that are not captured by optimization metrics. The same volume brings this horizon down to the level of educational practice in the Education section. There, Danielle Morin argues that "Higher learning institutions who prepare their students to be ethical users of AI have a very positive impact on the leaders of tomorrow. We can seriously ask ourselves what the role of universities and professors will be when generative AI can offer everything" (Morin, 2025, p. 277). Her contribution reinforces two central claims of this paper: first, that universities are being judged not only by how they regulate AI, but by how they form the practical judgment and responsibility of future professionals who will inhabit AI-mediated environments; and second, that Catholic universities in particular must link this formation to an "academic conscience" oriented toward the common good, integral ecology, and the protection of the most vulnerable. Read alongside our analysis of AI as a cognitive artefact and of cooperation as the most promising long-term strategy, Tolentino de Mendonça's and Morin's perspectives underscore that genuinely humane learning with AI will require institutions to cultivate ethical competence and critical literacy as core learning outcomes, so that powerful technical systems are consistently integrated into patterns of discernment, solidarity, and care for our common home rather than into logics of extraction or indifference.

Endnotes

[1] The intention of the user (as well as of the maker) is fundamental in the philosophical reflection on artefacts in general. In the case of cognitive artefacts, it has been emphasized that there may be instances which escape the above-mentioned distinction. Consider, for example, the case of an empty milk bottle placed in a certain spot in the kitchen in order to remind us that we need to buy milk. The bottle is a physical object that has a function, but in this case it helps us to perform a cognitive task, namely remembering. Is this sufficient to regard it as a cognitive artefact?

[2] The reliability of LLMs in mathematical tasks is a matter of ongoing discussion and debate. Essentially, these systems were not designed for this purpose, although their performance has improved over time. For a critique of ChatGPT's mathematical capacities, see, for example, Wolfram 2023. This point, of course, is not directly relevant to the argument we develop in the present work.

[3] An epistemologically attentive reader may notice that the distinction between know-that and know-how is at play here. This issue has been widely discussed within the epistemology of education.

[4] It is not difficult to discern here a conceptual similarity with the notion of an algorithm, understood as a finite set of deterministic instructions that transform input information into output information. After all, in his seminal 1936 paper, Turing himself took as an example precisely the algorithms for performing elementary arithmetical operations.

References

AI Research Group of the Centre for Digital Culture. (2023). *Encountering artificial intelligence: Ethical and anthropological investigations*. *Journal of Moral Theology, 1*(Theological Investigations of AI), i–262. https://doi.org/10.55476/001c.91230

An, Y., Li, Y., & Zhang, S. (2025). Investigating the higher education institutions' guidelines and practices regarding generative AI: A mixed methods study. *International Journal of Educational Technology in Higher Education, 22*(1). https://doi.org/10.1186/s41239-025-00507-3

Bittle, K., & El-Gayar, O. (2025). Generative AI and academic integrity in higher education: A systematic review and research agenda. *Information, 16*(4), 296. https://doi.org/10.3390/info16040296

Colorado State University. (2023). *What should a syllabus statement on AI look like?* https://tilt.colostate.edu/what-should-a-syllabus-statement-on-ai-look-like/

Corpuz, J. C. G. (2025). Faith and Artificial Intelligence (AI) in Catholic Education: A Theological Virtue Ethics Perspective. *Religions, 16*(8), 1083. https://doi.org/10.3390/rel16081083

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 61*(2), 228–239.

Crompton, H., & Burke, D. (2023). *Artificial intelligence in higher education: The state of the field*. *International Journal of Educational Technology in Higher Education, 20*(22). https://doi.org/10.1186/s41239-023-00392-8

Dabis, A., & Csáki, C. (2024). AI and ethics: Investigating the first policy responses of higher education institutions to the challenge of generative AI. *Humanities and Social Sciences Communications, 11*, 1006. https://doi.org/10.1057/s41599-024-03526-z

Dicastery for the Doctrine of the Faith, & Dicastery for Culture and Education. (2025, January 28). *Antiqua et nova: Note on the relationship between artificial intelligence and human intelligence*. Holy See Press Office. https://www.vatican.va/roman_curia/congregations/cfaith/documents/rc_ddf_doc_20250128_antiqua-et-nova_en.html

Francis. (2015). *Laudato Si': On care for our common home*. Libreria Editrice Vaticana. https://www.vatican.va/content/francesco/en/encyclicals/documents/papa-francesco_20150524_enciclica-laudato-si_en.pdf

Francis. (2017). *Veritatis gaudium (Apostolic Constitution on ecclesiastical universities and faculties).* Libreria Editrice Vaticana.

Francis. (2020). *Fratelli tutti: On fraternity and social friendship.* Libreria Editrice Vaticana.

Francis. (2020b, October 15). *A new Global Compact on Education: Video message*. Vatican. https://www.vatican.va/content/francesco/en/messages/pontmessages/2020/documents/papa-francesco_20201015_videomessaggio-pattoeducativo.html

Francis. (2024a, January 8). *Message for the 57th World Day of Peace: Artificial intelligence and peace*. Vatican.

Francis. (2024b, June 14). *Address to G7 Leaders on artificial intelligence (Borgo Egnazia, Italy)*. Vatican.

Gozum, I. E., & Eballo, A. Artificial intelligence, integral ecology, and the planetary polycrisis: Insights from *Laudato Si'* and *Laudate Deum* for sustainable and just AI governance. *AI and Society*, 1–13. https://doi.org/10.1007/s00146-025-02684-1

Heersmink, R. (2013). A taxonomy of cognitive artefacts: Function, information, and categories. *Review of Philosophy and Psychology, 4*(3), 465–481.

Holmes, W., & Porayska-Pomsta, K. (Eds.). (2022). *The ethics of artificial intelligence in education: Practices, challenges, and debates* (1st ed.). Routledge. https://doi.org/10.4324/9780429329067

John Paul II. (1990). *Ex corde Ecclesiae (Apostolic Constitution on Catholic universities).* Vatican.

Leo XIV. (2025a, May 8). *Urbi et Orbi blessing and message*. Vatican. https://www.vatican.va/content/leo-xiv/en/messages/urbi-et-orbi/2025/documents/urbi-et-orbi-08-may-2025.html

Leo XIV. (2025b, June 1). *Angelus*. Vatican. https://www.vatican.va/content/leo-xiv/en/angelus/2025/documents/angelus-01-june-2025.html

Liggins, D. (2024). *Abstract objects*. Cambridge University Press.

Luccioni, S., Jernite, Y., & Strubell, E. (2024). Power hungry processing: Watts driving the cost of AI deployment? *Proceedings of ACM FAccT 2024*, 85–99.

Luo, J. (Jess). (2024). A critical review of GenAI policies in higher education assessment: A call to reconsider the "originality" of students' work. *Assessment & Evaluation in Higher Education, 49*(5), 651–664. https://doi.org/10.1080/02602938.2024.2309963

Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research.* UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000386693

Miao, F., Holmes, W., Huang, R., & Zhang, H. (2021). *AI and education: Guidance for policy-makers.* UNESCO.

Minsky, M. L. (1962). Problems of formulation for artificial intelligence. In R. E. Bellman (Ed.), *Mathematical problems in the biological sciences* (Proceedings of Symposia in Applied Mathematics, Vol. 14, pp. 35–46). American Mathematical Society.

Ocampo, L.-M. A. R., & Gozum, I. E. A. (2025). Catholic higher education interface with AI: Diversions and intersections. *International Studies in Catholic Education,* 1-15. https://doi.org/10.1080/19422539.2025.2464734

Pontifical Academy for Life, Santa Clara University, Microsoft, IBM, FAO, & Italian Government. (2020). *Rome Call for AI Ethics.* https://www.romecall.org/the-rome-call

Rapaport, W. J. (2023). *Philosophy of computer science: An introduction to the issues and the literature.* John Wiley & Sons.

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching.*

Russell, S. J., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

Schneider, L. (2025). Human metacognition in AI-assisted decision-making. *ECIS 2025 Proceedings*, 11. https://aisel.aisnet.org/ecis2025/human_ai/human_ai/11

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *Proceedings of ACL 2019 / AAAI (updated).* https://ojs.aaai.org/index.php/AAAI/article/view/7123/6977

Tarantola, A. M. (Ed.). (2025). *Artificial intelligence and care of our common home: A focus on industries, finance, education and communication*. Vita e Pensiero.

UNESCO. (2019). *Beijing Consensus on Artificial Intelligence and Education.* https://unesdoc.unesco.org/ark:/48223/pf0000368303

UNESCO. (2023). *Global Education Monitoring Report 2023: Technology in education—A tool on whose terms?* UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000385723

University of Oxford. (2023b). *Unauthorised use of AI in exams and assessment.* https://academic.admin.ox.ac.uk/article/unauthorised-use-of-ai-in-exams-and-assessment

Wolfram, S. (2023). *What is ChatGPT doing … and why does it work?* (1st ed.). Wolfram Media.