Edukacja Filozoficzna 79/2025
ISSN 0860-3839, eISSN 2956-8269
DOI: 10.14394/edufil.2025.0007
ORCID: 0000-0003-0115-5787

Ethical Evaluation of Artificial Intelligence from the Perspective of the Catholic Church

Krzysztof Trębski
(Trnava University in Trnava, Slovakia)

Abstract: Artificial intelligence (AI) has emerged as a transformative force, profoundly reshaping many dimensions of human life. Its rapid growth, however, requires critical reflection on both benefits and risks. Ethical evaluation is not secondary but an opportunity to reconsider the meaning of human existence in a technology-driven world, while orienting progress with wisdom and foresight. The initial absence of clear frameworks has intensified debate on the urgent need for governance, legal safeguards, and moral principles to guide its invention, production, and use. This article analyzes the Catholic ethical evaluation of AI and the risks of unregulated development through documents of the Holy See, the teaching of recent popes, and their public pronouncements. It compares Catholic positions with existing governance instruments – such as the EU AI Act, UNESCO's *Recommendation on the Ethics of Artificial Intelligence*, and the *Rome Call for AI Ethics* with its *Hiroshima Addendum* – highlighting convergences and divergences, with particular attention to emerging ethical challenges. Based on the view that research and innovation are never morally neutral but always value-laden, the article underscores convergence between secular governance and Catholic teaching regarding the design, implementation, and responsible use of AI. At the same time, it highlights the Catholic emphasis on the centrality of the person – affirming that AI must serve humanity rather than replace or dominate it – on the inviolability of life (rejecting autonomous weapon systems), on human dignity (including principles such as non-discrimination, transparency, inclusion, accountability, reliability, safety, and privacy), on the dignity of work, social justice, and the universal call to fraternity. From this perspective, the Church supports a global ethical and regulatory framework, which it sees as essential not only to prevent harmful applications but also to promote virtuous practices and ensure continuous human oversight in the development and deployment of AI.

Key words: artificial intelligence, AI governance, ethical evaluation of AI in the Catholic Church

1. Introduction

We are witnessing the growing diffusion of artificial intelligence (AI), which elicits, on the one hand, uncritical enthusiasm and, on the other, excessive pes-

simism towards a tool that is at once "an exciting and fearsome tool,"[1] capable of generating immense benefits but also posing serious risks. This dual potential renders AI an inherently ambivalent system: it could become the most powerful multiplier of knowledge, bridging distances among people; yet it could equally evolve into a driver of injustice and social stratification. To prevent AI from becoming a multiplier of inequality – both between technologically advanced and developing nations, and between dominant and marginalized social groups – its development and implementation must be guided by robust political and ethical oversight.[2] Without such governance, AI risks undermining the "culture of solidarity and encounter," which is grounded in inclusion and dialogue,[3] and instead promoting a "culture of waste"[4] that fosters discrimination and marginalization.

2. Artificial Intelligence between Techno- and Human-Centrism

Technology, and particularly AI, with its capacity to shape material reality, mitigate risks, ease human labour, and enhance living conditions, embodies the objective dimension of human action. It must, however, be remembered that technology is not merely a human activity; rather, human nature itself constitutes a techno-human condition, insofar as the technical dimension is an intrinsic aspect of being human, an expression of existence as an individual, relational, and transcendent being.[5]

[1] The expression "an exciting and fearsome tool" was used by Pope Francis to emphasize that it is precisely the powerful technological progress that makes AI both a fascinating and a fearsome tool, calling for a level of reflection capable of meeting the challenge it presents. Cf. Francis, *Address of His Holiness Pope Francis*, Borgo Egnazia, 14.06.2024, URL: https://www.vatican.va/content/francesco/en/speeches/2024/june/documents/20240614-g7-intelligenza-artificiale.html.

[2] S. Quintarelli et al., *AI: profili etici. Una prospettiva etica sull'Intelligenza Artificiale. Princìpi, diritti e raccomandazioni*, "BioLaw Journal – Rivista di BioDiritto" 2019, Vol. 3, pp. 183–204.

[3] Cf. Francis, *Message of Pope Francis for the 48th World Communications Day: Communication at the Service of an Authentic Culture of Encounter*, 1.06.2014, URL: https://www.vatican.va/content/francesco/en/messages/communications/documents/papa-francesco_20140124_messaggio-comunicazioni-sociali.html; Francis, *Address of Holy Father Francis*, Cagliari, 22.09.2013, URL: https://www.vatican.va/content/francesco/en/speeches/2013/september/documents/papa-francesco_20130922_cultura-cagliari.html.

[4] Cf. Francis, *General Audience*, Saint Peter's Square, 5.06.2013, URL: https://www.vatican.va/content/francesco/en/audiences/2013/documents/papa-francesco_20130605_udienza-generale.html.

[5] Cf. P. Benanti, *Homo Faber: The Techno-Human Condition*, EDB, 2018, pp. 108, 110, 112.

This integral anthropological vision underscores the need for ongoing discernment to ensure that AI does not reduce the human being to a mere instrument of efficiency or productivity, but rather recognizes and safeguards the inalienable dignity of every person. Technology is born with a purpose and, through its interaction with human society, always represents a form of ordering social relations and a structure of power – empowering some to act while restricting others. This constitutive dimension of power inherently carries, whether explicitly or implicitly, the worldview of its creators and developers.[6]

Proponents of a techno-centric vision of development, who advocate for every form of technologization of the body and mind, envisage horizons in which the artificial becomes increasingly indistinguishable from the natural, intentionally erasing the difference between human and machine in a symbiotic fusion of humanity and technology, of organic and inorganic life. They promote the advancement of convergent technologies and robotics/AI, wherein the robot serves as the embodiment of AI, designed to replace and ultimately surpass the human being.[7] This is presented as the sole path towards overcoming the biological limitations of the body and the neurocognitive constraints of the mind, thereby moving towards a trans-human, post-human, or even super-human perfection. If the techno-centric worldview were to prevail, good would ultimately be reduced to what can be technologically achieved. In such a framework – where efficiency and utility become the sole criteria of judgement – authentic development is inevitably denied. True development, in fact, cannot be reduced merely to "doing." Its key lies in a mind capable of grasping the fully human meaning of action within a holistic vision of being.[8] Even when AI is employed, fundamental decisions remain human in nature and therefore require moral responsibility. There are strong anthropological, ontological, and ethical reasons to affirm that the non-reproducibility, non-substitutability, and uniqueness of human intelligence constitute a higher value.[9]

---

6 Cf. L. Winner, *Do Artifacts Have Politics?*, in: L. Winner, *The Whale and the Reactor: A Search for Limits in an Age of High Technology*, University of Chicago Press, Chicago 1988, p. 23.

7 Cf. E. Sadin, *Critica della ragione artificiale. Una difesa dell'umanità*, Luiss University Press, Milano 2019, pp. 10–33.

8 Cf. Francis, *Address Prepared by Pope Francis, Read by H.E. Archbishop Paglia, President of the Pontifical Academy for Life, Meeting with the Participants in the Plenary Assembly of the Pontifical Academy for Life*, Vatican City, 28.02.2020, https://www.vatican.va/content/francesco/en/speeches/2020/february/documents/papa-francesco_20200228_accademia-perlavita.html.

9 See L. Floridi, J.W. Sanders, *Artificial Evil and the Foundation of Computer Ethics*, "Ethics and Information Technology" 2001, Vol. 3, No. 1, pp. 55–66.

In the current scientific context, marked by the expanding presence of AI in vast domains of human activity, it becomes indispensable to develop a critical philosophical reflection on the human being – its meaning and value – in order to identify the potential limits of technology.[10] The challenge is not to exalt technology while disparaging the human person, nor to exalt the human while rejecting technology. Rather, the objective is to enable interventions upon the human condition without distorting its identity and without triggering irreversible transformations. In this sense, the task is not merely to acknowledge what remains human despite technology, but above all to discern what must remain human through technology.[11] If we understand the limits of what we can do with technology, we can make better choices about what we should do with it to make the world better for everyone.[12]

3. Core Ethical Principles in the Age of Artificial Intelligence

Given the vast scope of the phenomenon of AI and the significant progress achieved by such systems, many have sought to propose various initiatives aimed at defining the principles that should underlie AI, which must be viewed from a perspective that benefits humanity.

The four foundational principles of biomedical ethics – autonomy, beneficence, non-maleficence, and justice – developed by Tom L. Beauchamp and James F. Childress and first introduced in 1979,[13] embody fundamental moral values shared by individuals committed to ethical conduct and can therefore be regarded as central pillars in discussions on the ethical foundations that should guide the design, development, and use of AI.[14]

[10] Cf. T. Hagendorff, *The Ethics of AI Ethics: An Evaluation of Guidelines*, "Minds & Machines" 2020, Vol. 30, pp. 99–120.

[11] Cf. L. Floridi, J.W. Sanders, *On the Morality of Artificial Agents*, "Minds & Machines" 2004, Vol. 14, No. 3, pp. 349–379.

[12] Cf. M. Broussard, *Artificial Unintelligence: How Computers Misunderstand the World*, The MIT Press, Cambridge, MA–London 2019, p. 12.

[13] T.L. Beauchamp, J.F. Childress, *Principles of Biomedical Ethics*, Oxford University Press, Oxford 1979.

[14] Beauchamp and Childress maintain that these norms have developed because the essential role of morality as a social institution is to support human flourishing by addressing the factors that diminish well-being and by preventing conditions such as indifference, conflict, suffering, hostility, scarcity, and misinformation. Historical evidence demonstrates that when such moral

The principle of autonomy recognizes the capacity of individuals to self-determine and to act according to their own moral values and convictions. It implies that every person must be able to exercise meaningful control over their choices, remaining free from external coercion and internal constraints that could compromise voluntariness and understanding.[15] As Beauchamp and Childress explain, autonomy is self-rule that is free from both controlling interference by others and from limitations, such as inadequate understanding, that prevent meaningful choice.[16] In this sense, the principle is expressed in the power to decide, including the power to choose whether and when to decide.[17] Such a capacity constitutes the core of moral self-determination and forms the foundation of all respect for human dignity.[18] In the context of AI ethics, the principle of autonomy acquires growing significance, as intelligent systems increasingly interact with human decision-making processes. Ethically sound AI design must therefore aim to preserve – and, where possible, enhance – human capacities for comprehension, deliberation, and informed decision-making. This entails ensuring that users understand how AI systems operate, what data they use, and how their outputs are generated, so that individuals can make genuinely voluntary and informed choices regarding their interaction with these systems.[19] Ultimately, respecting autonomy in the age of AI means promoting a balanced relationship between humans and machines – one in which AI serves as a tool for cognitive and decision-making empowerment, rather than as a replacement for human will or moral responsibility.

The principle of beneficence ("do good only")[20] mandates that AI be developed and applied with the primary objective of generating tangible benefits for individu-

norms are ignored, human life deteriorates into misery, violence, and distrust. Conversely, respecting and upholding these norms helps to reduce suffering and promote social harmony. Therefore, they are vital for improving human well-being and achieving the fundamental aims of morality. Cf. T.L. Beauchamp, *Standing on Principles: Collected Essays*, Oxford University Press, New York 2010, pp. 43–44.

15 Cf. T.L. Beauchamp, J.F. Childress, *Principles of Biomedical Ethics*, 8th ed., Oxford University Press, New York–Oxford 2019, pp. 99–111.

16 Cf. ibid., p. 101.

17 Cf. S. Hajkowicz, *Global Megatrends: Seven Patterns of Change Shaping Our Future*, CSIRO Publishing, Melbourne 2015, p. 91.

18 Cf. P. Lin, K. Abney, G. Bekey, *Robot Ethics: Mapping the Issues for a Mechanized World*, "Artificial Intelligence" 2011, Vol. 175, Nos. 5–6, pp. 942–949, https://doi.org/10.1016/j.artint.2010.11.026.

19 Cf. L. Floridi et al., *AI 4 People – An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*, "Minds & Machines" 2018, Vol. 28, No. 4, p. 698.

20 Cf. T.L. Beauchamp, J.F. Childress, *Principles of Biomedical Ethics*, 8th ed., op. cit., p. 217.

als and society as a whole.[21] This principle encompasses three fundamental dimensions: the promotion of well-being, the safeguarding of the intrinsic dignity of every person, and the sustainability of technological development, which includes the protection of the environment.[22] Specifically, the promotion of well-being entails that AI should contribute meaningfully to improving the quality of human life by enhancing cognitive, relational, and operational capacities, while simultaneously reducing social and economic inequalities. The protection of human dignity constitutes a second essential dimension of beneficence: every application of AI must respect and value the human being as an end in itself, avoiding any form of objectification, manipulation, or algorithmic discrimination. Ethically oriented AI must therefore be conceived as a tool of human empowerment – one that supports decision-making and action without replacing individual will or moral responsibility. Finally, beneficence requires a sustained commitment to sustainability, understood as a balance between technological progress and environmental responsibility. The development and deployment of AI systems should be designed to ensure efficient resource use, minimize ecological impact, and promote an innovation model that does not compromise the well-being of future generations.

Overall, the principle of beneficence, when applied to the domain of AI, calls for an ethical vision oriented towards the "digital common good," in which technology functions as an enabling force for the promotion of human welfare, the protection of dignity, and the preservation of the environment. Only within this framework can AI be regarded not merely as technologically advanced, but also as morally justifiable and socially sustainable.[23]

The principle of non-maleficence ("do no harm") requires the deliberate avoidance of actions that may cause harm to individuals or society.[24] It thus establishes

[21] Cf. A. Jobin, M. Ienca, E. Vayena, *The Global Landscape of AI Ethics Guidelines*, "Nature Machine Intelligence" 2019, Vol. 9, No. 1, pp. 389–399.

[22] Cf. M. Latonero, *Governing Artificial Intelligence: Upholding Human Rights & Dignity*, Data & Society, URL: https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf.

[23] Cf. L. Floridi et al., *AI 4 People*, op. cit. In particular, see point 4.1: "Beneficence: promoting well-being, preserving dignity, and sustaining the planet."

[24] Lorenzo D'Avack combined the principle of beneficence with that of non-maleficence, explaining that such systems, in addition to contributing to the improvement of human well-being, should also avoid causing harm to individuals and society. Cf. L. D'Avack, *La rivoluzione tecnologica e la nuova era digitale: problemi etici*, in: *Intelligenza Artificiale. Il diritto, i diritti, l'etica*, ed. U. Ruffolo, Giuffrè, Milano 2020, pp. 3–28.

a minimal threshold of ethically acceptable behaviour, below which practices become detrimental to the dignity or integrity of the person. As Beauchamp and Childress emphasize,[25] non-maleficence highlights the moral obligation not only to refrain from intentionally causing harm but also to anticipate and prevent potential risks that may arise from technological or procedural decisions. In the context of AI, this principle assumes particular significance in three key domains: privacy, security, and capability caution. First, with regard to privacy, AI systems must not violate the right to personal data protection or intrude upon individuals' private spheres, as such violations would constitute a direct harm to autonomy and human dignity.[26] Second, concerning security, AI systems must be designed to ensure robustness, reliability, and resistance to malicious use, errors, or unintended consequences that could result in physical, psychological, or social harm. Preventing malfunctions and ensuring safety therefore represent essential components of ethically responsible AI design. Finally, the concept of capability caution refers to the responsibility of avoiding the development or deployment of systems whose capacities could become dangerous if they were to exceed or escape human control. This includes both the containment of potentially harmful autonomous functions and the governance of AI systems whose operational scope may produce unforeseen or uncontrollable effects.[27]

The principle of justice concerns the promotion of prosperity and the preservation of solidarity within society. AI must function as an instrument to reduce, not exacerbate, social and economic inequalities, ensuring that its benefits are distributed fairly and that no one is left behind.[28] In this sense, justice in the domain of AI – understood as impartiality – is best described through the concept of "algorithmic fairness."[29]

According to Luciano Floridi and Josh Cowls,[30] the framework of the four principles derived from bioethics should be supplemented with a fifth principle,

---

25 Cf. T.L. Beauchamp, J.F. Childress, *Principles of Biomedical Ethics*, 8th ed., op. cit., pp. 133–136.

26 Cf. L. Floridi et al., *AI 4 People*, op. cit. In particular, see point 4.2: "Non-maleficence: privacy, security and 'capability caution.'"

27 Cf. A. Jobin, M. Ienca, E. Vayena, *The Global Landscape of AI Ethics Guidelines*, op. cit., p. 392.

28 Cf. L. Floridi et al., *AI 4 People*, op. cit. In particular, see point 4.4: "Justice: promoting prosperity and preserving solidarity."

29 Cf. J. Morley et al., *Ethics as a Service: A Pragmatic Operationalisation of AI Ethics*, "Minds & Machines" 2021, Vol. 31, No. 2, pp. 239–356, https://doi.org/10.1007/s11023-021-09563-w.

30 Cf. L. Floridi, J. Cowls, *Unified Framework of Five Principles for AI in Society*, "Harvard Data Science Review" 2019, Vol. 1, pp. 2–15.

explicability, specifically designed to address the unique ethical challenges posed by AI systems. This principle is crucial because it enables the effective implementation of all other ethical principles.[31] Given that AI systems are often characterized by significant technical and conceptual opacity, explicability encompasses two complementary dimensions: intelligibility, that is, the capacity to understand how a system functions ("How does it work?"), and accountability, understood as the ability to identify who is responsible for the system's functioning and its consequences ("Who is responsible for the way it works?").[32] There is broad consensus that accountability with respect to moral and legal norms, as well as the associated liability, represents an essential requirement for any AI technology. The central issue, however, particularly concerning autonomous systems and robots with independent decision-making capacities, is how such responsibility can be effectively ensured and how moral and legal accountability can be assigned in the event of unintended or harmful outcomes.[33]

In this context, the principle of explicability goes beyond promoting technical transparency; it constitutes a necessary condition for ensuring public trust, the traceability of algorithmic decisions, and the ethical and legal legitimacy of AI deployment in contemporary society.

Building on these principles, it becomes necessary to define how AI research should develop so as not to harm humanity: it must remain under human control, be designed transparently and intelligibly, and be developed and applied fairly, in such a way that it neither perpetuates nor exacerbates existing inequalities.[34] A central challenge lies in the difficulty of achieving full transparency in the decision-making processes of AI systems based on deep neural networks. For this reason, a balance must be sought between the efficiency of results and their interpretability. Through the systematic recording and ongoing analysis of AI actions, it is possible to verify their compliance with ethical and legal principles, to

[31] Cf. L. Floridi et al., *AI 4 People*, op. cit. In particular, see point 4.5: "Explicability: Enabling the other principles through intelligibility and accountability."

[32] Cf. J.A. Kroll et al., *Accountable Algorithms*, "University of Pennsylvania Law Review" 2017, Vol. 165, No. 3, p. 645.

[33] Cf. J. Morley et al., *From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices*, "Science and Engineering Ethics" 2020, Vol. 26, No. 4, pp. 2141–2168, https://doi.org/10.1007/s11948-019-00165-5.

[34] Cf. M. Taddeo, L. Floridi, *How AI Can Be a Force for Good: An Ethical Framework Will Help to Harness the Potential of AI while Keeping Humans in Control*, "Science" 2018, Vol. 361, No. 6404, pp. 751–752.

identify and correct potential biases or errors, and to strengthen user trust. This process not only improves AI models but also ensures that their development remains ethical and sustainable.[35] To this end, it is crucial to distinguish between AI "decisions," which can be traced back to computational activity, and human "choices."[36] The latter require profound ethical reflection, drawing upon history, culture, and a shared system of values, since every act of choosing is the product of judgement rather than mere calculation.[37] It is therefore indispensable that human beings establish the boundaries and rules necessary to guarantee a responsible use of this technology – one that should always serve the highest potential and aspirations of humankind,[38] while safeguarding those human functions that cannot and must not be replaced by machines: judgement, respect, understanding, caring, and love.[39]

4. Secular Models of Artificial Intelligence Governance

The accelerated evolution of AI technologies has given rise to profound ethical, social, and legal challenges, thereby necessitating the establishment of robust and coherent governance frameworks.[40] In this context, instruments such as UNESCO's

[35] Cf. L. Floridi, *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*, Oxford University Press, Oxford 2023, pp. 105–112.

[36] Cf. L. Floridi, F. Cabitza, *Intelligenza artificiale. L'uso delle nuove macchine*, Bompiani, Firenze–Milano 2021, p. 70.

[37] Cf. D.M. Berry, *The Limits of Computation: Joseph Weizenbaum and the ELIZA Chatbot*, "Weizenbaum Journal of the Digital Society" 2023, Vol. 3, No. 3, https://doi.org/10.34669/WI.WJDS/3.3.2.

[38] Francis, *Message of His Holiness Pope Francis for the 57th World Day of Peace: Artificial Intelligence and Peace*, 1.01.2024, URL: https://www.vatican.va/content/francesco/it/messages/peace/documents/20231208-messaggio-57giornatamondiale-pace2024.html.

[39] Cf. J. Weizenbaum, *Il potere del computer e la ragione umana. I limiti dell'intelligenza artificiale*, EGA-Edizioni Gruppo Abele, Torino 1987, p. 192.

[40] Floridi clarifies the term governance and emphasizes that digital governance, digital ethics (also known as computer, information, or data ethics), and digital regulation represent distinct normative approaches. Digital governance refers to the practice of defining and implementing policies, procedures, and standards for the proper development, use, and management of the infosphere. It may include guidelines and recommendations that overlap with digital regulation, without necessarily coinciding entirely with it. Digital regulation, on the other hand, refers to the system of laws developed and enforced by social or governmental institutions to regulate the behaviour of agents. Not every aspect of digital regulation pertains to digital governance, and not every aspect of digital governance falls under regulation. Floridi highlights the need for

Recommendation on the Ethics of Artificial Intelligence[41] and the European Union's AI Act[42] represent two pivotal regulatory models. While differing in scope and legal enforceability, both initiatives converge on a set of foundational ethical principles, thereby contributing to the broader debate on global AI governance. Their significance lies not only in establishing normative ethics standards for the responsible development and deployment of AI[43] but also in fostering international dialogue aimed at reconciling diverse ethical traditions and regulatory approaches in the pursuit of a shared, human-centred digital future.[44]

As a transformative force, AI gives rise to global ethical, social, and political questions. In this context, UNESCO's *Recommendation on the Ethics of Artificial Intelligence*, adopted unanimously in November 2021 by all 193 Member States, represents a significant attempt to establish a shared international framework. It identifies four foundational values: human dignity and human rights, social justice, inclusiveness, and environmental sustainability.[45] These values underpin the formulation of guiding principles and policy actions intended to ensure that AI development and deployment serve the common good while respecting fundamental rights. Human dignity occupies a central place in the *Recommendation*, understood as the intrinsic and equal worth of every individual, which cannot be compromised at any stage of the AI lifecycle. Technologies must therefore contribute to enhancing human well-being without objectifying, subordinating, or discriminating against individuals or communities, with particular attention to vulnerable groups.[46] Environmental protection constitutes another key principle,

ethical guidance in the governance of AI. Cf. L. Floridi, *Soft Ethics, the Governance of the Digital and the General Data Protection Regulation*, "Philosophical Transactions of the Royal Society A" 2018, Vol. 376, No. 2133, https://doi.org/10.1098/rsta.2018.0081.

41 UNESCO, *Recommendation on the Ethics of Artificial Intelligence*, Paris 2022, URL: https://unesdoc.unesco.org/in/documentViewer.xhtml?v=2.1.196&id=p::usmarcdef_0000381137&file=/in/rest/annotationSVC/DownloadWatermarkedAttachment/attach_import_75c9fb6b-92a6-4982-b772-79f540c9fc39%3F_%3D381137eng.pdf&locale=en&multi=true&ark=/ark:/48223/pf0000381137/PDF/381137eng.pdf#1517_21_EN_SHS_int.indd%3A.8946%3A.

42 European Union, *Regulation (EU) 2024/1689 of the European Parliament and of the Council*, 13.07.2024, URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689.

43 V.C. Müller, *Ethics of Artificial Intelligence and Robotics*, in: *Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), ed. E.N. Zalta, URL: https://plato.stanford.edu/archives/sum2021/entries/ethics-ai/.

44 M. Coeckelbergh, *AI Ethics*, The MIT Press, Cambridge, MA, 2020, p. 57.

45 UNESCO, *Recommendation on the Ethics of Artificial Intelligence*, op. cit. p. 8.

46 Cf. ibid., p. 10.

as ecosystems are considered indispensable for human welfare and for future generations. Stakeholders involved in AI development and use are required to minimize environmental impacts through sustainable practices and adherence to the precautionary principle.[47] The *Recommendation* further underscores the importance of inclusion and diversity, which must be safeguarded by avoiding social, digital, or cultural exclusion and by promoting the active participation of all groups, regardless of origin, gender, age, religion, disability, or socio-economic condition.[48] It also stresses the need to foster peaceful, just, and interconnected societies in which AI serves as a tool for solidarity, justice, and equity, without undermining human autonomy or fuelling social or environmental conflicts.[49] Among its operational principles, the document highlights proportionality and the imperative to "do no harm," restricting AI applications to legitimate and proportionate purposes, particularly in contexts directly affecting human life and death.[50] It also requires safety and security mechanisms to prevent risks and vulnerabilities, fair access to the benefits of AI, and continuous assessment of the social, economic, and environmental consequences of technology. Further principles include the protection of privacy and personal data through adequate regulatory frameworks, human oversight and accountability – ensuring that ultimate responsibility rests with natural or legal persons – together with transparency and explainability as essential conditions for trust, traceability, and avenues of redress.[51] The *Recommendation* emphasizes the importance of digital literacy and public awareness, enabling citizens and communities to understand the implications of AI and make informed choices. It calls for a multilevel, collaborative, and adaptive governance model engaging governments, civil society, the private sector, academia, and local communities, in full respect of cultural diversity and territorial specificities.[52] In addition, clear requirements for transparency and explainability must be complemented by measures to counteract bias and stereotypes in datasets. Diversity and inclusion in technological development and access should be actively promoted, while States are encouraged to contribute to the formulation of international standards ensuring safety, reliability, and respect for

47 Cf. ibid., p. 12.
48 Cf. ibid., p. 16.
49 Cf. ibid., pp. 22–25.
50 Cf. ibid., p. 20.
51 Cf. ibid., p. 8.
52 Cf. ibid., p. 21.

human dignity. With regard to data governance, quality, security, and protection are paramount, together with corrective feedback mechanisms. Privacy safeguards should be rooted in privacy by design, impact assessments, and legislation aligned with international law, ensuring that individuals retain full control over their personal data, including rights of access, erasure, and enhanced protection for sensitive categories such as biometric, genetic, and health information.[53]

The main criticisms of UNESCO's *Recommendation on the Ethics of Artificial Intelligence* focus on both theoretical and practical limitations. First, scholars emphasize its non-binding character: although it represents the first global attempt to establish a shared ethical framework, it lacks legal force and delegates the responsibility for implementation to Member States. This feature raises doubts about its operational effectiveness, particularly in political contexts where AI governance does not constitute a strategic priority.[54] A second critical point concerns the generality of the principles, which are often formulated in broad and indeterminate terms. While this vagueness facilitates international consensus, it risks undermining the translation of these principles into concrete guidelines and regulatory mechanisms.[55] Moreover, the *Recommendation* fails to adequately address emerging issues, such as the legal responsibility of autonomous systems, the impact of generative technologies, and the geopolitical challenges linked to data sovereignty.[56] For these reasons, the document is often regarded as a preliminary ethical framework: valuable as a general point of reference, yet insufficient to govern the complexity of the ongoing transformations.

The European Regulation on Artificial Intelligence (EU AI Act)[57] constitutes the first comprehensive attempt to regulate AI systems within the European Union, establishing a normative framework designed to reconcile technological innovation with the protection of fundamental rights. It is inspired by the principles developed by the European Commission's High-Level Expert Group on AI,[58]

53 L. Floridi, *The Ethics of Artificial Intelligence*, op. cit., pp. 112–115.

54 L. Floridi, J. Cowls, *A Unified Framework of Five Principles for AI in Society*, in: L. Floridi, ed., *Ethics, Governance, and Policies in Artificial Intelligence*, Springer Verlag, Cham, 2021, p. 15.

55 A. Jobin, M. Ienca, E. Vayena, *The Global Landscape of AI Ethics Guidelines*, op. cit., p. 392.

56 C. Cath, *Governing Artificial Intelligence: Ethical, Legal and Technical Opportunities and Challenges*, "Philosophical Transactions of the Royal Society A" 2018, Volume 376, No. 2133, 20180080, https://doi.org/10.1098/rsta.2018.0080.

57 European Union, *Regulation (EU) 2024/1689*, op. cit.

58 High-Level Expert Group on AI (AI HLEG), *Ethics Guidelines for Trustworthy AI*, 8.04.2019, URL: https://www.europarl.europa.eu/cmsdata/196377/AI%20HLEG_Ethics%20Guidelines%20for%20Trustworthy%20AI.pdf.

which identified four fundamental ethical principles regarded as the foundation of trustworthy AI: (1) respect for human autonomy; (2) prevention of harm; (3) fairness; and (4) explicability. However, in order to effectively achieve reliable AI, they outlined seven key prerequisites that, in their view, must be continuously monitored and managed throughout the entire lifecycle of AI systems: (1) human agency and oversight; (2) technical robustness and safety; (3) privacy and data governance; (4) transparency; (5) diversity, non-discrimination, and fairness; (6) societal and environmental well-being; and (7) accountability. Furthermore, the group emphasized the potential necessity of introducing new legal measures and control mechanisms capable of ensuring adequate protection against negative effects, while enabling effective human ethical oversight in the processes of design, development, and deployment of AI technologies.

Among its most significant aspects, the AI Act introduces a regime of explicit prohibitions targeting practices deemed incompatible with human dignity and collective security.[59] AI systems that may adversely affect safety or fundamental rights are classified as "high-risk" under the EU AI Act. This category encompasses, on the one hand, systems integrated into products already subject to EU product safety legislation, such as toys, aviation technologies, motor vehicles, medical devices, and lifts. On the other hand, it includes applications operating in sensitive domains, such as critical infrastructure management, education and vocational training, employment and labour relations, access to essential private and public services, law enforcement, migration and border control, as well as systems used in legal interpretation and application.[60] Similarly, the regulation bans the use of technologies exploiting vulnerabilities related to age, disability, or socio-economic conditions, where such exploitation results in behavioural distortion with damaging consequences.

A further prohibition concerns social scoring mechanisms, namely the classification of individuals based on behaviours or personal characteristics. This practice is considered harmful, as it may generate discriminatory or dispropor-

[59] For further developments on the topic, see the 2025 updates: European Commission, *Commission Guidelines on Prohibited Artificial Intelligence Practices Established by Regulation (EU) 2024/1689 (AI Act)*, C(2025) 5052 final, 29.07.2025, URL: https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act.

[60] European Commission, *Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)*, COM(2021)206final, 21.04.2021, URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206.

tionate treatment, particularly when applied in contexts different from those in which the data were originally collected.[61] Likewise, predictive systems that assign criminal risk to individuals solely on the basis of automated profiling are banned, with the exception of tools that support human evaluation grounded in objective and verifiable evidence.

The regulation also restricts the creation of biometric databases through indiscriminate scraping of facial images from the internet or surveillance systems, a practice deemed invasive of privacy and likely to foster mass surveillance. Similarly, it prohibits the use of systems intended to infer emotional states in professional or educational contexts, except in narrowly defined medical or security circumstances. Furthermore, biometric categorization aimed at deducing sensitive attributes – such as race, religious belief, sexual orientation, or political opinion – is forbidden, with exceptions limited to legitimate purposes, like dataset labelling for research or security activities.

Equally significant are the obligations imposed on generative and general-purpose models, which must provide adequate technical documentation, comply with copyright law, and disclose transparency regarding training data.[62] These provisions are designed to mitigate risks associated with violations of fundamental rights while reinforcing public trust through enhanced traceability of decision-making processes. Ultimately, the EU AI Act represents an innovative regulatory model capable of translating ethical principles into binding legal obligations, thereby consolidating a European approach centred on human dignity, fairness, and sustainability. It serves as a bridge between ethical reflection and political action, fostering a digital ecosystem where innovation is guided by the common good and respect for fundamental values.

5. Catholic Church's Vision of Artificial Intelligence

The development of AI in contemporary society represents one of the most profound ethical and anthropological challenges of our time. In this context, there

[61] A. Atabekov, A. Yastrebov, *Legal Status of Artificial Intelligence across Countries: Legislation on the Move*, "European Research Studies Journal" 2018, Vol. 21, No. 4, pp. 773–782.

[62] European Commission, *White Paper on Artificial Intelligence: A European Approach to Excellence and Trust*, 19.02.2020, URL: https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en.

emerges an urgent need to formulate an ethics of discernment and decision-making that, in the light of Catholic teaching, reaffirms the primacy of the spirit over matter[63] and ensures that technology remains at the service of the human person, rather than becoming its master. The defining risk of our age lies in the emergence of a "technocratic paradigm," a worldview that tends to subordinate the human person to the power of machines and the logic of efficiency, thereby obscuring the spiritual, moral, and relational dimensions of human existence.[64] It is therefore essential to understand these profound transformations and to orient them towards serving the human person, while safeguarding and promoting inherent human dignity. Given the complexity and unpredictability of such developments, this task calls for particularly deep ethical discernment.[65]

The Catholic Church's support for the ethical moderation of algorithms reflects an awareness that, given the complexity of today's technological landscape, a more sophisticated ethical framework is required to ensure that this commitment is genuinely effective.[66] It is therefore essential to maintain a robust ethical framework throughout the entire process of AI development – from design to deployment and use – in order to guide the values shaping this ongoing transformation for the common good.[67] From this necessity arises the proposal of

---

63 Cf. Benedict XVI, *Encyclical Letter Caritas in Veritate*, Rome, 29.06.2009, par. 69–70, URL: https://www.vatican.va/content/benedict-xvi/en/encyclicals/documents/hf_ben-xvi_enc_20090629_caritas-in-veritate.html.

64 Cf. Francis, *Apostolic Exhortation Laudate Deum*, Rome, 4.10.2023, par. 21, URL: https://www.vatican.va/content/francesco/en/apost_exhortations/documents/20231004-laudate-deum.html.

65 Cf. Francis, *Letter of His Holiness Pope Francis to the President of the Pontifical Academy for Life for the 25th Anniversary of the Establishment of the Academy: Humana Communitas*, Vatican City, 6.01.2019, par. 12, URL: https://www.vatican.va/content/francesco/en/letters/2019/documents/papa-francesco_20190106_lettera-accademia-vita.html.

66 Cf. Francis, *Address of His Holiness Pope Francis to the Participants in the Congress on "Child Dignity in the Digital World"*, Vatican City, 6.10.2017, URL: https://www.vatican.va/content/francesco/en/speeches/2017/october/documents/papa-francesco_20171006_congresso-child-dignity-digitalworld.html.

67 Cf. Pontifical Academy for Life, *Rome Call for AI Ethics*, Rome, 28.02.2020, URL: https://www.vatican.va/roman_curia/pontifical_academies/acdlife/documents/rc_pont-acd_life_doc_20202228_rome-call-for-ai-ethics_en.pdf. The *Rome Call for AI Ethics* is a document promoting a shared ethical approach to AI. It aims to ensure that digital innovation and technological progress serve humanity by putting the human person at the centre. The signatories advocate for a new "algor-ethics" to guide the development of AI that respects human dignity, benefits everyone, and does not focus solely on profit or the replacement of workers.

algor-ethics,[68] a fully human and responsible approach to AI, as promoted by the Catholic Church.

Algor-ethics, understood as applied ethics in the field of AI, requires an assessment not only of the ways in which AI models are designed, developed, and used by human beings, but also of the social and environmental impacts that these systems may exert on society and the natural environment through their operation and behaviour. It thus assumes a dual nature. On the one hand, it seeks to identify the principles that human beings must observe to ensure that AI systems are developed exclusively to promote sustainable social well-being, adopting not merely a technical approach but a multidisciplinary one that integrates perspectives from computer science, engineering, psychology, anthropology, philosophy, religion, and political science. On the other hand, algor-ethics also represents an attempt to encode within AI systems a set of behavioural rules that enable machines to act in ways that respect the human person.

In this context, the global initiative *Rome Call for AI Ethics*[69] – launched by the Pontifical Academy for Life (Holy See, Vatican) with the support of the RenAIssance Foundation,[70] established by Pope Francis on 12 April 2021 to promote an ethical approach to the development and use of artificial intelligence worldwide – assumes particular significance. The aim of this initiative was to propose, with broad international and interfaith consensus, that AI development should adopt, from the very beginning of algorithm design, an "algor-ethical" approach – ethics integrated into the design itself, or "ethics by design." This effort seeks to promote algor-ethics, ensuring that AI is used in an ethical manner. To this end, the *Rome Call for AI Ethics* proposes six ethical evaluation criteria for AI: transparency, inclusion, responsibility, impartiality, reliability, and respect for security and privacy, so that AI benefits all individuals and safeguards human dignity.[71] In light of the *Rome Call*, which articulates the Catholic Church's position

[68] See P. Benanti, *Oracoli. Tra algoretica e algocrazia*, Luca Sossella Editore, Roma 2018.

[69] Pontifical Academy for Life, *Rome Call for AI Ethics*, op. cit.

[70] Cf. RenAIssance Foundation, URL: https://www.romecall.org/renaissance-foundation/.

[71] The contribution of Floridi and Cowls influenced the six AI governance principles proposed in the *Rome Call*. The explainability principle proposed by Floridi clearly shaped the content of the document, as he was directly involved in its development. However, it is important to emphasize that the five original principles elaborated by Floridi and Cowls do not fully coincide with the six principles set forth in the *Rome Call*.

on AI,[72] shared ethical principles acquire crucial importance in addressing contemporary challenges.[73] Foremost among these is the need for transparency, ensuring that all machine-generated content is immediately recognizable. This principle is closely linked to accountability, which requires establishing standards to trace the origin and authenticity of digital content, thereby countering the spread of disinformation and fake news.

Moreover, the development of AI systems must prioritize inclusivity, respecting the diversity of cultures, traditions, and languages that define humanity. This entails a strong commitment to fairness, ensuring that generative AI does not perpetuate or amplify existing biases. Given their far-reaching societal impact, the reliability and robustness of such systems are of primary importance.[74] Finally, safeguarding user security and privacy remains imperative, particularly in view of the significant power these technologies exert.[75]

Another significant initiative promoted by the Holy See is the *Hiroshima AI Process Addendum on Generative AI*, a key document emerging from the Hiroshima AI Process launched by the G7 leaders and officially adopted on 30 October 2023.[76] Although not a legally binding text, the document – also signed by the Vatican – serves as a foundational reference for global AI governance. The *Addendum* emphasizes the need for ethical oversight of generative AI, reiterating the core principles advanced by the *Rome Call* and underscoring the imperative of developing AI that is inclusive, fair, and – given its profound social impact – reliable, safe, and privacy-preserving, so that its potential may be harnessed for the good of humanity.

---

72 Cf. Francis, *Address of His Holiness Pope Francis to Participants in the "Minerva Dialogues"*, Vatican City, 27.03.2023, URL: https://www.vatican.va/content/francesco/en/speeches/2023/march/documents/20230327-minerva-dialogues.html.

73 Francis, *Address Prepared by Pope Francis*, op. cit.

74 Generative AI systems can create coherent texts, but this does not ensure reliability. They may "hallucinate," producing statements that seem plausible but are false or biased. This is particularly dangerous in disinformation campaigns that undermine trust in the media. Privacy, data ownership, and intellectual property are also at risk. Misuse of these technologies can further lead to discrimination, electoral manipulation, mass surveillance, digital exclusion, and rising individualism detached from society. Cf. Francis, *Message of His Holiness Pope Francis for the 57th World Day of Peace: Artificial Intelligence and Peace*, op. cit., par. 4.

75 Cf. A. Adam, *Delegating and Distributing Morality: Can We Inscribe Privacy Protection in a Machine?*, "Ethics and Information Technology" 2005, Vol. 7, No. 4, pp. 233–242.

76 *Hiroshima Addendum*, URL: https://www.romecall.org/wp-content/uploads/2024/07/Hiroshima-Addendum-2.pdf.

In the search for an ethical framework for AI, the social doctrine of the Catholic Church reminds us that technologies must be studied and developed according to criteria that ensure their genuine service to the entire human family,[77] proposing an ethics of technological development grounded in the principles of human dignity, justice, subsidiarity, and solidarity.

The technology is not merely a tool but a complex force that requires careful ethical evaluation to ensure that it serves human dignity and the common good.[78] This common good is something towards which all people naturally aspire, and no ethical framework worthy of the name can fail to acknowledge it as a fundamental guiding principle.[79] It must therefore respond to the biblical mandate to "till and keep the earth" (Gen 2:15), strengthening the covenant between humanity and creation in accordance with God's creative love.[80] In the Catholic understanding, the human person possesses an irreducible spiritual transcendence that no machine or algorithm can replicate or replace. Only the human being, created "in the image and likeness of God" (Gen 1:27), has a spiritual and immortal soul, capable of moral discernment and free self-determination.

Technological development can contribute significantly to the progress of humanity, but it can also foster the illusion of human self-sufficiency when people focus solely on how to act, neglecting the deeper why that gives moral and spiritual meaning to their actions. However, such progress cannot truly benefit humanity unless it is accompanied by genuine moral and spiritual maturity: technological advancement, while representing a potentially great benefit for hu-

---

77 Francis, *Message of His Holiness Pope Francis to the Executive Chairman of the "World Economic Forum" on the Occasion of the Annual Gathering in Davos-Klosters*, 23–26.01.2018, URL: https://www.vatican.va/content/francesco/en/messages/pont-messages/2018/documents/papa-francesco_20180112_messaggio-davos2018.html.

78 For further discussion, see S.P. Chalmers, *Papal Teaching on the Ethical Challenges of Artificial Intelligence*, in: *New Trends in Disruptive Technologies, Tech Ethics and Artificial Intelligence*, eds. D.H. de la Iglesia, J.F. de Paz Santana, A.J. López Rivero, Springer, Cham 2023, pp. 167–177, https://doi.org/10.1007/978-3-031-14859-0_15.

79 Cf. Francis, *Address of His Holiness Pope Francis to the Participants in the Seminar "The Common Good in the Digital Age," Organized by the Dicastery for Promoting Integral Human Development (DPIHD) and the Pontifical Council for Culture (PCC)*, Vatican City, 27.09.2019, URL: https://www.vatican.va/content/francesco/en/speeches/2019/september/documents/papa-francesco_20190927_eradigitale.html.

80 Francis, *Laudato si'*: *Encyclical Letter on the Care for Our Common Home*, 24.05.2015, par. 109, URL: https://www.vatican.va/content/francesco/en/encyclicals/documents/papa-francesco_20150524_enciclica-laudato-si.html.

mankind, must always be guided by an ethical conscience capable of discernment and responsibility.[81]

Given the contemporary context, there is an urgent need to ground the design, development, and use of AI in a robust ethical, anthropological, and wisdom-based foundation. It is necessary to overturn the assumption that everything technically possible is therefore legitimate, and instead ask how we can ensure that what is truly just becomes possible.[82] From this standpoint, the central challenge identified by the Catholic Church lies in orienting AI towards fostering a network of authentic communication – one rooted in communion that unites, in truth that sets free, and in love that confers ultimate meaning to human action.[83]

AI must always remain a tool at the service of humanity and must never replace human conscience or ethical discernment. Its orientation must consistently aim at the integral development of both the human person and society as a whole.[84] One of the critical concerns highlighted is the growing tendency towards the anthropomorphization of AI, which risks displacing authentic human relationships, particularly among younger generations. For this reason, the Church strongly emphasizes the necessity of education in critical thinking and discernment in the use of data and content generated by intelligent systems.[85]

Recently, Pope Leo XIV reaffirmed the Church's position on the development of AI, stressing that this epochal transformation requires careful reflection and ethical guidance to ensure its orientation towards humanity and the common good.[86] As AI systems acquire the capacity to make autonomous, technically driven decisions, it becomes imperative to examine their ethical and anthropo-

---

81 Cf. Benedict XVI, *Encyclical Letter Caritas in Veritate*, op. cit., par. 68–70.

82 Cf. Francis, *Address Prepared by Pope Francis*, op. cit.

83 Cf. Dicastero per la Communicazione, *La Chiesa di fronte all'attuale fenomeno dell'"intelligenza artificiale"*, 22.05.2024, URL: https://www.comunicazione.va/it/notizie/notizie_2024/la-chiesa-di-fronte-all-attuale-fenomeno-dell-intelligenza-artif.html.

84 Cf. Dicastery for the Doctrine of the Faith, Dicastery for Culture and Education, *Antiqua et nova: Note on the Relationship between Artificial Intelligence and Human Intelligence*, 28.01.2025, par. 6, URL: https://www.vatican.va/roman_curia/congregations/cfaith/documents/rc_ddf_doc_20250128_antiqua-et-nova_en.htmlhttps://www.vatican.va/roman_curia/congregations/cfaith/documents/rc_ddf_doc_20250128_antiqua-et-nova_en.html.

85 Cf. ibid., par. 21.

86 Cf. Leo XIV, *Message of the Holy Father, Signed by the Cardinal Secretary of State Pietro Parolin, on the Occasion of the AI for Good Summit 2025*, Geneva, 10.07.2025, URL: https://www.vatican.va/content/leo-xiv/en/messages/pont-messages/2025/documents/20250708-messaggio-aiforgood-ginevra.html.

logical implications. While AI may simulate human reasoning and perform tasks with remarkable efficiency, it remains incapable of exercising moral judgement or fostering authentic human relationships. For this reason, technological advancement must be accompanied by a strong commitment to human values, moral conscience, and a deepened sense of responsibility.[87] This unprecedented stage of innovation thus calls for renewed reflection on the meaning of human existence itself. Ultimately, AI requires ethical guidelines and regulatory frameworks grounded in the primacy of human dignity, rather than being governed solely by criteria of utility or efficiency.

The Church's moral and social teachings offer valuable guidance to ensure that AI is employed in ways that respect and preserve human agency. Reflections on justice, for instance, should also encompass the promotion of equitable social structures, the safeguarding of global security, and the advancement of peace. By exercising prudence, both individuals and communities can discern responsible ways to harness AI for the benefit of humanity, while avoiding applications that might compromise human dignity or cause harm to the environment.[88]

In contemporary debates on AI governance, the Catholic Church underscores the necessity of meaningful human oversight as an essential condition for orienting technological innovation towards the service of the human person and the common good.[89] This perspective does not remain at the level of abstract principles but identifies operational criteria capable of translating the values of human dignity, responsibility, and social justice into concrete regulatory practices.[90]

In light of the personalist principle and the categorical rejection of delegating life-or-death decisions to machines, meaningful human oversight in the military domain requires: (1) human-in-command structures with clearly identifiable legal responsibility across the entire chain of command; (2) *ex ante* limits on the autonomous functions of weapon systems,[91] excluding target selection and

[87] Cf. Francis, *Address of His Holiness Pope Francis to the Participants in the Seminar "The Common Good in the Digital Age"*, op. cit.

[88] Cf. Dicastery for the Doctrine of the Faith, Dicastery for Culture and Education, *Antiqua et nova*, op. cit., par. 47.

[89] Cf. *Catechism of the Catholic Church*, Libreria Editrice Vaticana, Vatican City 1992, par. 1905–1912, URL: https://www.vatican.va/archive/ENG0015/_INDEX.HTM.

[90] Cf. Francis, *Message for the 57th World Day of Peace: Artificial Intelligence and Peace*, op. cit., par. 2–10.

[91] In discussions on lethal autonomous weapon systems, Pope Francis made a pivotal statement at the 2024 G7 summit: "No machine should ever choose to take the life of a human being," affirming that decisions affecting life and death must remain under human authority. Cf. Francis, *Address*

engagement without effective human control; (3) compliance testing with international humanitarian law and human rights standards, including mandatory red-teaming and kill-switch mechanisms; and (4) full traceability through independent auditing and periodic review of rules of engagement.[92] To align AI with the principle of person-centred care and equity in access, the Catholic Church advocates for: (1) clinical governance of AI with ultimate medical responsibility remaining with the physician; (2) ethical-clinical impact assessments and post-market surveillance of devices and algorithms; (3) clinically useful explainability for both doctors and patients; (4) specific informed consent procedures for AI use, with safeguards for vulnerable groups; (5) systematic audits of bias and performance across diverse populations; and (6) robust data protection measures (minimization, quality, security), combined with human override mechanisms for inappropriate algorithmic recommendations.[93]

In accordance with the Catholic principles of the dignity of work and social justice,[94] meaningful human oversight in the field of employment must include: (1) human-in-the-loop mechanisms for adverse decisions (hiring, promotion, dismissal), guaranteeing the right to explanation and appeal; (2) impact assessments on non-discrimination and inclusion, with periodic audits and corrective measures; (3) participation of workers' representatives in the design and deployment of AI systems; (4) prohibition of black-box models for high-impact uses, accompanied by decision logs for accountability; and (5) continuous training on the critical use of algorithmic tools.

The Catholic perspective offers a coherent and universally applicable ethical framework that translates core values into concrete operational guidelines: prioritizing the human person, ensuring balance and prudence, advancing justice and inclusion, guaranteeing traceable accountability, and protecting the most vulnerable. These guidelines provide a foundational reference for both public policy and private regulatory practices, fostering a digital ecosystem genuinely oriented towards the common good.

---

of His Holiness Pope Francis, op. cit.; cf. A. Sharkey, *Autonomous Weapons Systems, Killer Robots and Human Dignity*, "Ethics and Information Technology" 2019, Vol. 21, No. 2, pp. 75–87.

92 Cf. Pontifical Academy for Life, *Rome Call for AI Ethics*, op. cit.

93 Cf. Francis, *Laudato si'*, op. cit., par. 109–110.

94 Cf. Pontifical Council for Justice and Peace, *Compendium of the Social Doctrine of the Church*, Libreria Editrice Vaticana, Vatican City 2004, par. 270–275, URL: https://www.vatican.va/roman_curia/pontifical_councils/justpeace/documents/rc_pc_justpeace_doc_20060526_compendio-dott-soc_en.html.

6. Conclusion

The profound transformations shaping contemporary society through the widespread use of AI inevitably raise significant ethical questions. Within this context, the Catholic contribution offers a coherent and universalizable framework of principles that does not remain purely theoretical but provides operational criteria: the centrality of the person, proportionality and precaution, justice and inclusion, traceable responsibility, and the protection of the most vulnerable.[95] Translated into practical requirements – such as effective human oversight, auditability, context-appropriate explainability, data protection, impact assessments, and redress mechanisms – these criteria are capable of informing both public and private standards and regulations, fostering a digital ecosystem genuinely oriented towards the common good.[96]

The Catholic Church's ethical evaluation of AI does not constitute a rejection of technological progress, but rather an appeal to orient innovation according to principles that safeguard human dignity and the common good.[97] The overarching goal is to ensure that AI remains at the service of humanity, promotes justice, and contributes to the construction of a more equitable, peaceful, and fraternal society.[98]

In this vision, human beings, endowed with their distinctive "wisdom of the heart," possess the capacity to discern the interconnectedness of realities, to recognize the positive dimensions of existence, and to uncover its deeper meaning.[99] This wisdom is neither reducible to abstract theory nor to mere technical expertise; rather, it is expressed concretely in relationships, commitment, and care.[100]

---

95 Cf. P. Benanti, *L'uomo è un algoritmo? Il senso dell'umano e l'intelligenza artificiale*, Castelvecchi, Roma 2025, pp. 45–48.

96 Cf. Francis, *Message for the 57th World Day of Peace: Artificial Intelligence and Peace*, op. cit., par. 2–6.

97 Cf. Francis, *Laudato si'*, op. cit., par. 102–114.

98 Cf. Francis, *Fratelli tutti: Encyclical Letter on Fraternity and Social Friendship*, 3.10.2020, par. 114–121, URL: https://www.vatican.va/content/francesco/en/encyclicals/documents/papa-francesco_20201003_enciclica-fratelli-tutti.html.

99 Cf. Francis, *Message of His Holiness Pope Francis for the 58th World Day of Social Communications: Artificial Intelligence and the Wisdom of the Heart. Towards a Fully Human Communication*, 24.01.2024, URL: https://www.vatican.va/content/francesco/en/messages/communications/documents/20240124-messaggio-comunicazioni-sociali.html.

100 Cf. V. Corrado, S. Pasta, eds., *Intelligenza artificiale e sapienza del cuore. Commento al Messaggio di Papa Francesco per la 58ma Giornata mondiale delle Comunicazioni Sociali*, Scholé, Brescia 2024, p. 102.

It enables the perception of realities that data alone cannot reveal, while recalling that at the foundation of all things lies the relational bond among persons – a dimension that digital technologies can neither replace nor diminish.

Acknowledgement
The author used generative AI tools exclusively for linguistic and editorial support, and their use did not have any significant impact on the scientific content of the article.

Bibliography

Adam A., *Delegating and Distributing Morality: Can We Inscribe Privacy Protection in a Machine?*, "Ethics and Information Technology" 2005, Vol. 7, No. 4, pp. 233–242.

Atabekov A., Yastrebov A., *Legal Status of Artificial Intelligence across Countries: Legislation on the Move*, "European Research Studies Journal" 2018, Vol. 21, No. 4, pp. 773–782.

Beauchamp T.L., *Standing on Principles: Collected Essays*, Oxford University Press, New York 2010.

Beauchamp T.L., Childress J.F., *Principles of Biomedical Ethics*, Oxford University Press, Oxford 1979.

Beauchamp T.L., Childress J.F., *Principles of Biomedical Ethics*, 8th ed., Oxford University Press, New York–Oxford 2019.

Benanti P., *Homo Faber: The Techno-Human Condition*, EDB, 2018.

Benanti P., *L'uomo è un algoritmo? Il senso dell'umano e l'intelligenza artificiale*, Castelvecchi, Roma 2025.

Benanti P., *Oracoli. Tra algoretica e algocrazia*, Luca Sossella Editore, Roma 2018.

Benedict XVI, *Encyclical Letter Caritas in Veritate*, Rome, 29.06.2009, URL: https://www.vatican.va/content/benedict-xvi/en/encyclicals/documents/hf_ben-xvi_enc_20090629_caritas-in-veritate.html.

Berry D.M., *The Limits of Computation: Joseph Weizenbaum and the ELIZA Chatbot*, "Weizenbaum Journal of the Digital Society" 2023, Vol. 3, No. 3, https://doi.org/10.34669/WI.WJDS/3.3.2.

Broussard M., *Artificial Unintelligence: How Computers Misunderstand the World*, The MIT Press, Cambridge, MA–London 2019.

Catechism of the Catholic Church, Libreria Editrice Vaticana, Vatican City 1992, URL: https://www.vatican.va/archive/ENG0015/_INDEX.HTM.

Cath C., *Governing Artificial Intelligence: Ethical, Legal and Technical Opportunities and Challenges*, "Philosophical Transactions of the Royal Society A" 2018, Volume 376, No. 2133, 20180080, https://doi.org/10.1098/rsta.2018.0080.

Chalmers S.P., *Papal Teaching on the Ethical Challenges of Artificial Intelligence*, in: *New Trends in Disruptive Technologies, Tech Ethics and Artificial Intelligence*, eds. D.H. de la Iglesia, J.F. de Paz Santana, A.J. López Rivero, Spinger, Cham 2023, pp. 167–177, https://doi.org/10.1007/978-3-031-14859-0_15.

Coeckelbergh M., *AI Ethics*, The MIT Press, Cambridge, MA, 2020.

Corrado V., Pasta S., eds., *Intelligenza artificiale e sapienza del cuore. Commento al Messaggio di Papa Francesco per la 58ma Giornata mondiale delle Comunicazioni Sociali*, Scholé, Brescia 2024.

Cucci G., *Emozioni e ragione: due mondi antitetici?*, "La Civiltà Cattolica" 2015, Vol. 3, pp. 139–150.

D'Avack, L., *La rivoluzione tecnologica e la nuova era digitale: problemi etici*, in: *Intelligenza Artificiale. Il diritto, i diritti, l'etica*, ed. U. Ruffolo, Giuffrè, Milano 2020, pp. 3–28.

Dicastero per la Communicazione, *La Chiesa di fronte all'attuale fenomeno dell'"intelligenza artificiale"*, 22.05.2024, URL: https://www.comunicazione.va/it/notizie/notizie_2024/la-chiesa-di-fronte-all-attuale-fenomeno-dell-intelligenza-artif.html.

Dicastery for the Doctrine of the Faith, Dicastery for Culture and Education, *Antiqua et nova: Note on the Relationship between Artificial Intelligence and Human Intelligence*, 28.01.2025, URL: https://www.vatican.va/roman_curia/congregations/cfaith/documents/rc_ddf_doc_20250128_antiqua-et-nova_en.html.

European Commission, *Commission Guidelines on Prohibited Artificial Intelligence Practices Established by Regulation (EU) 2024/1689 (AI Act)*, C(2025) 5052 final, 29.07.2025, URL: https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act.

European Commission, *Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)*, COM(2021)206final, 21.04.2021, URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206.

European Commission, *White Paper on Artificial Intelligence: A European Approach to Excellence and Trust*, 19.02.2020, URL: https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en.

European Union, *Regulation (EU) 2024/1689 of the European Parliament and of the Council*, 13.07.2024, URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689.

Floridi L., *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*, Oxford University Press, Oxford 2023.

Floridi L., *Soft Ethics, the Governance of the Digital and the General Data Protection Regulation*, "Philosophical Transactions of the Royal Society A" 2018, Vol. 376, No. 2133, https://doi.org/10.1098/rsta.2018.0081.

Floridi L., ed., *The Cambridge Handbook of Information and Computer Ethics*, Cambridge University Press, Cambridge 2010.

Floridi L., ed., *Ethics, Governance, and Policies in Artificial Intelligence*, Springer Verlag, Cham 2021.

Floridi L., Cabitza F., *Intelligenza artificiale. L'uso delle nuove macchine*, Bompiani, Firenze–Milano 2021.

Floridi L., Cowls J., *Unified Framework of Five Principles for AI in Society*, "Harvard Data Science Review" 2019, Vol. 1, pp. 2–15.

Floridi L., Cowls J., *A Unified Framework of Five Principles for AI in Society*, in: L. Floridi, ed., *Ethics, Governance, and Policies in Artificial Intelligence*, Springer Verlag, Cham 2021, pp. 5–17.

Floridi L., Sanders J.W., *Artificial Evil and the Foundation of Computer Ethics*, "Ethics and Information Technology" 2001, Vol. 3, No. 1, pp. 55–66.

Floridi L., Sanders J.W., *On the Morality of Artificial Agents*, "Minds & Machines" 2004, Vol. 14, No. 3, pp. 349–379.

Floridi L., et al., *AI 4 People – An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*, "Minds & Machines" 2018, Vol. 28, No. 4, pp. 689–707.

Francis, *Address of His Holiness Pope Francis*, Borgo Egnazia, 14.06.2024, URL: https://www.vatican.va/content/francesco/en/speeches/2024/june/documents/20240614-g7-intelligenza-artificiale.html.

Francis, *Address of His Holiness Pope Francis to Participants in the "Minerva Dialogues"*, Vatican City, 27.03.2023, URL: https://www.vatican.va/content/francesco/en/speeches/2023/march/documents/20230327-minerva-dialogues.html.

Francis, *Address of His Holiness Pope Francis to the Participants in the Congress on "Child Dignity in the Digital World"*, Vatican City, 6.10.2017, URL: https://www.vatican.va/content/francesco/en/speeches/2017/october/documents/papa-francesco_20171006_congresso-childdignity-digitalworld.html.

Francis, *Address of His Holiness Pope Francis to the Participants in the Seminar "The Common Good in the Digital Age," Organized by the Dicastery for Promoting Integral Human Development (DPIHD) and the Pontifical Council for Culture (PCC)*, Vatican City, 27.09.2019, URL: https://www.vatican.va/content/francesco/en/speeches/2019/september/documents/papa-francesco_20190927_eradigitale.html.

Francis, *Address of Holy Father Francis*, Cagliari, 22.09.2013, URL: https://www.vatican.va/content/francesco/en/speeches/2013/september/documents/papa-francesco_20130922_cultura-cagliari.html.

Francis, *Address Prepared by Pope Francis, Read by H.E. Archbishop Paglia, President of the Pontifical Academy for Life, Meeting with the Participants in the Plenary Assembly of the Pontifical Academy for Life*, Vatican City, 28.02.2020, URL: https://www.vatican.va/content/francesco/en/speeches/2020/february/documents/papa-francesco_20200228_accademia-perlavita.html.

Francis, *Apostolic Exhortation Laudate Deum*, Rome, 4.10.2023, URL: https://www.vatican.va/content/francesco/en/apost_exhortations/documents/20231004-laudate-deum.html.

Francis, *Fratelli tutti: Encyclical Letter on Fraternity and Social Friendship*, 3.10.2020, URL: https://www.vatican.va/content/francesco/en/encyclicals/documents/papa-francesco_20201003_enciclica-fratelli-tutti.html.

Francis, *General Audience*, Saint Peter's Square, 5.06.2013, URL: https://www.vatican.va/content/francesco/en/audiences/2013/documents/papa-francesco_20130605_udienza-generale.html.

Francis, *Laudato si'*: *Encyclical Letter on the Care for Our Common Home*, 24.05.2015, URL: https://www.vatican.va/content/francesco/en/encyclicals/documents/papa-francesco_20150524_enciclica-laudato-si.html.

Francis, *Letter of His Holiness Pope Francis to the President of the Pontifical Academy for Life for the 25th Anniversary of the Establishment of the Academy: Humana Communitas*, Vatican City, 6.01.2019, URL: https://www.vatican.va/content/francesco/en/letters/2019/documents/papa-francesco_20190106_lettera-accademia-vita.html.

Francis, *Message of His Holiness Pope Francis for the 57th World Day of Peace: Artificial Intelligence and Peace*, 1.01.2024, URL: https://www.vatican.va/content/francesco/en/messages/peace/documents/20231208-messaggio-57giornatamondiale-pace2024.html.

Francis, *Message of His Holiness Pope Francis for the 58th World Day of Social Communications: Artificial Intelligence and the Wisdom of the Heart. Towards a Fully Human Communication*, 24.01.2024, URL: https://www.vatican.va/content/francesco/en/messages/communications/documents/20240124-messaggio-comunicazioni-sociali.html.

Francis, *Message of His Holiness Pope Francis to the Executive Chairman of the "World Economic Forum" on the Occasion of the Annual Gathering in Davos-Klosters*, 23–26.01.2018, URL: https://www.vatican.va/content/francesco/en/messages/pont-messages/2018/documents/papa-francesco_20180112_messaggio-davos2018.html.

Francis, *Message of Pope Francis for the 48th World Communications Day: Communication at the Service of an Authentic Culture of Encounter*, 1.06.2014, URL: https://www.vatican.va/content/francesco/en/messages/communications/documents/papa-francesco_20140124_messaggio-comunicazioni-sociali.html.

Hagendorff T., *The Ethics of AI Ethics: An Evaluation of Guidelines*, "Minds & Machines" 2020, Vol. 30, pp. 99–120.

Hajkowicz S., *Global Megatrends: Seven Patterns of Change Shaping Our Future*, CSIRO Publishing, Melbourne 2015.

High-Level Expert Group on AI (AI HLEG), *Ethics Guidelines for Trustworthy AI*, 8.04.2019, URL: https://www.europarl.europa.eu/cmsdata/196377/AI%20HLEG_Ethics%20Guidelines%20for%20Trustworthy%20AI.pdf.

Hiroshima Addendum, URL: https://www.romecall.org/wp-content/uploads/2024/07/Hiroshima-Addendum-2.pdf.

Jobin A., Ienca M., Vayena E., *The Global Landscape of AI Ethics Guidelines*, "Nature Machine Intelligence" 2019, Vol. 9, No. 1, pp. 389–399.

Kroll J.A., et al., *Accountable Algorithms*, "University of Pennsylvania Law Review" 2017, Vol. 165, No. 3, pp. 633–705.

Latonero M., *Governing Artificial Intelligence: Upholding Human Rights & Dignity*, Data & Society, URL: https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf

Leo XIV, *Message of the Holy Father, Signed by the Cardinal Secretary of State Pietro Parolin, on the Occasion of the AI for Good Summit 2025*, Geneva, 10.07.2025, URL: https://www.vatican.va/content/leo-xiv/en/messages/pont-messages/2025/documents/20250708-messaggio-aiforgood-ginevra.html.

Lin P., Abney K., Bekey G., *Robot Ethics: Mapping the Issues for a Mechanized World*, "Artificial Intelligence" 2011, Vol. 175, Nos. 5–6, pp. 942–949, https://doi.org/10.1016/j.artint.2010.11.026.

Morley J., Floridi L., Kinsey L., Elhalal A., *From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices*, "Science and Engineering Ethics" 2020, Vol. 26, pp. 2141–2168, https://doi.org/10.1007/s11948-019-00165-5.

Morley J., et al., *Ethics as a Service: A Pragmatic Operationalisation of AI Ethics*, "Minds & Machines" 2021, Vol. 31, No. 2, pp. 239–356, https://doi.org/10.1007/s11023-021-09563-w.

Müller V.C., *Ethics of Artificial Intelligence and Robotics*, in: *Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), ed. E.N. Zalta, URL: https://plato.stanford.edu/archives/sum2021/entries/ethics-ai/.

Pontifical Academy for Life, *Rome Call for AI Ethics*, Rome, 28.02.2020, URL: https://www.vatican.va/roman_curia/pontifical_academies/acdlife/documents/rc_pont-acd_life_doc_20202228_rome-call-for-ai-ethics_en.pdf

Pontifical Council for Justice and Peace, *Compendium of the Social Doctrine of the Church*, Libreria Editrice Vaticana, Vatican City 2004, URL: https://www.vatican.va/roman_curia/pontifical_councils/justpeace/documents/rc_pc_justpeace_doc_20060526_compendio-dott-soc_en.html.

Quintarelli S., Corea F., Fossa F., Loreggia A., Sapienza S., *AI: profili etici. Una prospettiva etica sull'Intelligenza Artificiale. Princìpi, diritti e raccomandazioni*, "BioLaw Journal – Rivista di BioDiritto" 2019, Vol. 3, pp. 183–204.

RenAIssance Foundation, *Call for AI Ethics*, 28.02.2020, URL: https://www.romecall.org/.

Sadin E., *Critica della ragione artificiale. Una difesa dell'umanità*, Luiss University Press, Milano 2019.

Sharkey A., *Autonomous Weapons Systems, Killer Robots and Human Dignity*, "Ethics and Information Technology" 2019, Vol. 21, No. 2, pp. 75–87.

Smart Dubai, *AI Ethics Principles & Guidelines*, 2018, URL: https://www.digitaldubai.ae/docs/default-source/ai-principles-resources/ai-ethics.pdf.

Taddeo M., Floridi L., *How AI Can Be a Force for Good: An Ethical Framework Will Help to Harness the Potential of AI while Keeping Humans in Control*, "Science" 2018, Vol. 361, No. 6404, pp. 751–752.

UNESCO, *Recommendation on the Ethics of Artificial Intelligence*, Paris 2022, URL: https://unesdoc.unesco.org/in/documentViewer.xhtml?v=2.1.196&id=p::usmarcdef_0000381137&file=/in/rest/annotationSVC/DownloadWatermarkedAttachment/attach_import_75c9fb6b-92a6-4982-b772-79f540c9fc39%3F_%3D381137eng.pdf&locale=en&multi=true&ark=/ark:/48223/pf0000381137/PDF/381137eng.pdf#1517_21_EN_SHS_int.indd%3A.8946%3A.

United Nations, *Universal Declaration of Human Rights*, URL: https://www.un.org/en/about-us/universal-declaration-of-human-rights.

Weizenbaum J., *Il potere del computer e la ragione umana. I limiti dell'intelligenza artificiale*, EGA-Edizioni Gruppo Abele, Torino 1987.

Winner L., *Do Artifacts Have Politics?*, in: L. Winner, *The Whale and the Reactor: A Search for Limits in an Age of High Technology*, University of Chicago Press, Chicago 1988, pp. 19–39.