

# Generating Insights from Catholic Social Teaching: Ethical Guidelines for Artificial Intelligence in Health Care Ministries

Nicholas Kockler, PhD, MS, HEC-C

## INTRODUCTION

1. In the annals of human history, few innovations have been as transformative as the printing press and the machinery of mass production in the Industrial Revolution. Today, we stand on the precipice of another monumental shift: the rise of generative artificial intelligence (gAI). This technological marvel, with its potential to revolutionize sectors like Catholic health care, beckons us to “read the signs of the times and respond to them in light of the Gospel.”<sup>1</sup> The complexities of gAI are vast, and our aim is to offer some practical ethical guidance on this immense topic. Catholic moral and social teaching has insights that can guide the development, deployment, and evaluation of artificial intelligence in health care. This essay argues that Catholic social teaching (CST) in particular offers insights that shape axioms for generative AI in health care applications.

## WHAT IS GENERATIVE AI?

AI, or Artificial Intelligence, refers to the simulation of human intelligence in machines. It's a branch of computer science aiming to create systems able to perform tasks that usually require human cognition, such as decision-making, pattern recognition, understanding language, and problem-solving. When AI is qualified as “generative,” it means that it is a type of artificial intelligence that has the capacity to produce outputs that mimic human-created content (text, images, etc.).

At its core, AI is a tool. Just as a hammer is an extension of the hand for driving nails, AI can be seen as an extension of human cognition for processing data, recognizing patterns, and performing tasks. One may think of AI as a highly advanced cognitive-arithmetic-linguistic-algorithmic tool: AI has multifaceted capabilities from basic arithmetic to complex cognitive tasks like natural language understanding.

Another way to conceptualize artificial intelligence is through its three primary functions: automation, augmentation, and assistance. First, AI's automation capabilities transfer specific tasks from human oversight to algorithmic control, thereby reducing manual labor and increasing efficiency. Second, AI serves to augment human capabilities by enhancing cognitive processes and expanding the collective knowledge base, thereby enabling more informed decision-making and problem-solving. Lastly, AI assists in streamlining operations by providing real-time support and guidance, which in turn lightens the human workload and improves overall productivity.

However, the line can get blurred because of AI's ability to mimic certain human-like qualities, such as conversation, generating digital images, or playing games. Regardless of how advanced or "intelligent" an AI might seem, it does not have feelings, consciousness, or self-awareness. It operates based on the code and algorithms it's been designed with, making it a tool created and directed by humans.

There is a parallel between the rise of AI and the era of industrialization. The following five points suggest this: One, with AI, many anticipate the displacement of jobs. Just as industrialization led to the automation of many manual, labor-intensive tasks (making certain crafts obsolete), AI has the potential to automate many "thinking-intensive" jobs that involve data analysis, customer service, and even some aspects of decision-making.

Two, relatedly, AI could require a skill shift. Industrialization required workers to acquire new skills to operate machines. Similarly, the AI era requires a workforce that understands how

to work with, manage, and even program these new tools. At the time of this writing, a great many early adopters of AI have focused on the art of the prompt: how to get AI like ChatGPT to produce the intended results the user wants by making the prompt "perfect."

Three, many expect AI to increase productivity of certain work. Just as machines increased the scale and efficiency of production of goods, AI can increase efficiency in various sectors, from finance to healthcare, by handling large datasets and performing complex calculations at speeds unimaginable to humans.

Four, many see immense societal implications of AI. The industrial era brought about significant societal changes, from urbanization to changes in work-life balance. Similarly, AI has the potential to bring profound societal shifts, such as changes in how we view privacy, the nature of work, or even what tasks are deemed valuable.

Finally, there are ethical concerns with AI. Industrialization raised concerns about worker safety, fair wages, and working conditions. AI introduces its own set of ethical issues, from bias in algorithms to surveillance concerns.

Just as industrialization transformed societies, economies, and ways of life, AI promises (or threatens, depending on one's perspective) to bring about its own set of transformative changes. The challenge for societies is to harness the benefits while mitigating potential drawbacks and ensuring equitable outcomes. We may think of AI, and generative AI in particular, as marking the industrialization of thought.

This term implies the mechanization and systematic production and execution of cognitive and creative tasks, similar to how industrialization referred to the mechanization of physical labor. For example:

**Standardization and Scalability:** Just as industrialization led to the standardized production of goods on a large scale, AI allows for the standardized processing of data and decision-making on scales previously unattainable. An AI model, once trained, can be deployed countless times across different devices and platforms, producing consistent results.

**Efficiency and Speed:** Industrial machines increased the speed of production. Similarly, AI can process and analyze vast amounts of data at speeds far surpassing human capabilities.

**Specialization:** With industrialization, machines were often designed for specific tasks, leading to specialized production lines. In AI, there are specialized models for various tasks, from image recognition to natural language processing.

**Transformation of Human Labor:** Just as machines reduced the need for manual labor, AI reduces the need for human cognitive labor in certain areas. Tasks like data analysis, which might take humans hours, can be completed in moments by AI.

**Depersonalization:** A criticism of industrialization was that it could lead to the depersonalization of work, turning craftsmen into mere cogs in a machine. Similarly, there's a concern that relying too heavily on AI, especially in areas like decision-making, might

strip away the human touch, intuition, or ethical considerations.

**Transformation of Skill Sets:** As industrialization changed the skills workers needed, AI's rise emphasizes the need for new skills in the modern workforce, such as data literacy and understanding AI ethics.

In many ways, the phrase industrialization of thought aptly captures the transformative impact of AI on cognitive tasks and broader societal functions. However, it is essential to recognize that while AI can simulate many aspects of human thought, it lacks consciousness, emotions, and the nuanced understanding that humans bring to tasks.

Generative artificial intelligence (gAI) represents a significant shift in the realm of computational capabilities. Unlike traditional AI systems that primarily focus on analysis and prediction, gAI is designed to create. This creation can range from generating coherent text to simulating intricate biological processes. The potential applications of gAI are vast, especially in sectors like health care. For instance, gAI can revolutionize diagnostics by analyzing extensive datasets to identify patterns that might be imperceptible to the human eye. This could lead to the early detection of ailments even before they manifest. Additionally, by understanding a patient's unique genetic makeup, lifestyle, and medical history, gAI can offer personalized care, optimizing treatment outcomes. Another promising application is in the realm of research acceleration. The drug discovery process, which traditionally spans several years, could be significantly condensed with gAI simulating molecular interactions, predicting

drug efficacy, and ensuring safety.<sup>2</sup>

Sara Vaezy explains four strategic domains of gAI applications in health care.<sup>3</sup> First, in the clinical domain, gAI has the potential to support clinical decision-making, automate mundane tasks, and assist providers with documentation. Second, from the patient's perspective, gAI could augment patient experience in a highly personalized, precise way based on their unique needs, motivations, preferences, and history. Third, the administrative domain contains numerous gAI opportunities for various tasks such as predictive scheduling, billing applications, etc. Finally, Vaezy points to several gAI applications to back-office functions such as applications that intercept and redirect inquiries to the best channel or outlet to support the specific needs of the patient or consumer.

Generative AI technologies are rapidly maturing and finding applications in various domains, including software engineering. For instance, gAI can be used in software engineering use cases such as translating natural language to code, code translation, and code auto-completion.<sup>4</sup> However, the introduction of gAI into various sectors also brings forth a plethora of ethical considerations. Concerns range from potential infringements on copyrights due to the replication and production of content by gAI, the risk of job losses due to automation, to challenges in discerning truth from fiction given the ability of AI to create realistic content.<sup>5</sup>

The ethical implications surrounding generative AI are profound. While the technology offers promising advancements, it also underscores the multifaceted ethical landscape that demands

careful consideration and proactive measures as it continues to evolve.<sup>6</sup>

As the above suggests, the issues are vast and many:

1. **Theft of Intellectual Property:** Generative AI can replicate and produce content, leading to potential infringements on copyrights and the devaluation of original creations.
2. **Displacement of Workers:** As AI automates tasks, there's a risk of job losses, especially in sectors reliant on repetitive tasks, potentially leading to economic and social disruptions.
3. **Loss of Autonomy:** Over-reliance on AI recommendations can diminish human decision-making, making individuals overly dependent on algorithms for choices.
4. **Erosion of Human Dignity and Dignity of Work:** Beyond automating tasks, AI can reduce the perceived intrinsic worth of human contributions, undermining the unique value and experiences individuals bring.
5. **Data Privacy and Confidentiality:** AI models, especially those that generate content based on vast datasets, can inadvertently reveal private information or patterns, posing risks to individual privacy.
6. **Bias and Discrimination:** AI models can reflect and amplify societal biases present in their training data, leading to unfair or discriminatory outputs.
7. **Authenticity and Truth:** The ability of AI to create realistic content, like deepfakes, challenges our ability to discern truth from fiction, potentially enabling misinformation.
8. **Economic Inequality:** The concentration of AI capabilities among a few entities can exacerbate economic disparities,

with wealth and power becoming more centralized.

9. **Safety and Reliability:** Advanced AI models can produce unpredictable results, posing risks when deployed in critical sectors.
10. **Depersonalization:** An over-dependence on AI for personal tasks can diminish human-to-human interactions, leading to impersonal and detached experiences.
11. **Transparency and Accountability:** The "black box" nature of some AI models can obscure decision-making processes, challenging accountability and understanding.
12. **Environmental Concerns:** The computational demands of training AI models can lead to significant energy consumption, raising environmental and sustainability concerns.
13. **Over-reliance and Loss of Skills:** Excessive dependence on AI can result in the atrophy of essential human skills, as machines take over tasks previously done by humans.
14. **Anthropological Implications:** AI's ability to create art or philosophical content raises questions about human uniqueness, creativity, and consciousness.
15. **Regulatory and Legal Challenges:** The rapid advancement of AI can outpace legal and regulatory frameworks, leading to potential conflicts and ambiguities.

Each of these points underscores the multifaceted ethical landscape of generative AI, emphasizing the need for careful consideration and proactive measures as the technology evolves. The ethical implications are profound, and the Catholic Social Teaching (CST) offers a beacon.

## PILLARS OF CATHOLIC SOCIAL TEACHING

CST, rooted in millennia of theological reflection, provides a moral compass. Drawing from many documents of modern CST, we find that a number of key values, ends, and mechanisms to effectuate change. Each of these, in turn, point to general ethical principles guiding Catholic health care. I will organize our reflections of CST around three pillars – axiological, eschatological, and sociological – to shed light on the key principles these bring to bear on generative AI questions.

*Axiological Pillar:* The axiological pillar describes essential values at the center of human personhood, communal living, and relationship with God. These are human dignity, the common good, and stewardship.

To begin, human dignity is the inherent dignity rooted in being created in the image and likeness of the divine: "Human persons are willed by God; they are imprinted with God's image. Their dignity does not come from the work they do, but from the persons they are."<sup>7</sup> As an essential value, human dignity corresponds to two general principles: respect human dignity and respect human life from conception to death. These are interrelated, of course, but distinct principles guiding behavior.

Next, the common good refers to the context and capacity for human flourishing in community. These words describe the common good at Vatican II:

...the sum of those conditions of social life which allow social groups and their individual members relatively thorough and ready access to their own fulfillment,

today takes on an increasingly universal complexion and consequently involves rights and duties with respect to the whole human race. Every social group must take account of the needs and legitimate aspirations of other groups, and even of the general welfare of the entire human family.<sup>8</sup>

Corresponding to the value of the common good we have the general principles of promoting the common good and enabling participation in the common good itself.

A third essential value in CST I will name as stewardship, which pertains to the dignity of work: humankind's participation in God's plan as created co-creators. St. John Paul II had this to say:

Even though it bears the mark of a bonum arduum, in the terminology of Saint Thomas, this does not take away the fact that, as such, it is a good thing for man. It is not only good in the sense that it is useful or something to enjoy; it is also good as being something worthy, that is to say, something that corresponds to man's dignity, that expresses this dignity and increases it. If one wishes to define more clearly the ethical meaning of work, it is this truth that one must particularly keep in mind. Work is a good thing for man—a good thing for his humanity—because through work man not only transforms nature, adapting it to his own needs, but he also achieves fulfilment as a human being and indeed, in a sense, becomes 'more a human being'.<sup>9</sup>

More recently, Pope Francis had this to say in

his encyclical on caring for the Earth:

We were created with a vocation to work. The goal should not be that technological progress increasingly replace human work, for this would be detrimental to humanity. Work is a necessity, part of the meaning of life on this earth, a path to growth, human development and personal fulfilment.<sup>10</sup>

In sum, stewardship calls upon us to abide by several general principles: (1) honor the spirituality of work, (2) respect the dignity of work itself and the workers (cf. above), (3) prioritize the worker over utility and efficiency, (4) exercise just use and allocation of resources corresponding to the universal destiny of goods (versus private property), and (5) act to maximize sustainability of resources.

*Eschatological Pillar:* The eschatological pillar orients humankind to the ends of God's invitation: a transcendent horizon fulfilled by our love for God and for neighbor. This is a 'now and not yet' pillar. The horizon includes ends such as responsibility and religious liberty, social justice, integral human development, and integral ecology.

The eschatological horizon in our tradition calls for responsibility and religious liberty. In the words of *Dignitas Humanae*:

In all his activity a man is bound to follow his conscience in order that he may come to God, the end and purpose of life. It follows that he is not to be forced to act in a manner contrary to his conscience. Nor, on the other hand, is he to be restrained from acting in accordance with his conscience, especially in matters religious.



The reason is that the exercise of religion, of its very nature, consists before all else in those internal, voluntary and free acts whereby man sets the course of his life directly toward God. No merely human power can either command or prohibit acts of this kind. The social nature of man, however, itself requires that he should give external expression to his internal acts of religion: that he should share with others in matters religious; that he should profess his religion in community. Injury therefore is done to the human person and to the very order established by God for human life, if the free exercise of religion is denied in society, provided just public order is observed.<sup>11</sup>

Thus, in terms of general principles related to responsibility and religious liberty, we have the following. One, persons and corporations should act responsibly and be held accountable. Two, respect for personal and corporate conscience should be established in law within the parameters of the public order. Finally, respect for diversity of views should be a hallmark of communal living.

Social justice is another key component of our eschatological horizon. Lisa Cahill defines social justice as “inclusive participation in the common good.”<sup>12</sup> The Compendium of the Social Doctrine of the Church states, “Ever greater importance has been given to social justice, which represents a real development in general justice, the justice that regulates social relationships according to the criterion of observance of the law.”<sup>13</sup>

The general principles as they relate to social justice include (a) promoting participation in

society, (b) establishing commutative fairness between parties, (c) encouraging contributive fairness of individuals and groups, (d) ensuring proper distribution of benefits and burdens, and (e) exhibiting charity in the absence of justice.

Next, the eschatological component includes integral human development in our horizon. Benedict XVI states,

The truth of development consists in its completeness: if it does not involve the whole man and every man, it is not true development. This is the central message of *Populorum Progressio*, valid for today and for all time. Integral human development on the natural plane, as a response to a vocation from God the Creator, demands self-fulfilment in a 'transcendent humanism which gives [to man] his greatest possible perfection: this is the highest goal of personal development.' The Christian vocation to this development therefore applies to both the natural plane and the supernatural plane; which is why, 'when God is eclipsed, our ability to recognize the natural order, purpose and the 'good' begins to wane.'"<sup>14</sup>

The general principles of integral human development include but are not limited to the following. One, design, development, and deployment of technology should be in service to the person, not vice versa. Two, persons should be afforded the opportunity to develop competencies and talents. Three, institutions should cultivate an appreciation of the human person in totality.

Finally, an integral ecology is a component of

the eschatological pillar of CST. In *Laudato Si'*, Pope Francis writes,

Since everything is closely interrelated, and today's problems call for a vision capable of taking into account every aspect of the global crisis, I suggest that we now consider some elements of an integral ecology, one which clearly respects its human and social dimensions. [...] When we speak of the 'environment', what we really mean is a relationship existing between nature and the society which lives in it. Nature cannot be regarded as something separate from ourselves or as a mere setting in which we live. We are part of nature, included in it and thus in constant interaction with it.<sup>15</sup>

Ensuring that technology is not “severed” from ethics, at least two general principle(s) apply: (1) understand the interconnectedness of all things, and (2) exercise sustainable development and use of technology.

**Sociological Pillar:** The sociological pillar provides specific mechanisms to be exercised in social contexts in pursuit of the ends and values mentioned above. First, solidarity, based on the connection and relationship of humankind, is “a firm and persevering determination to commit oneself to the common good. That is to say to the good of all and of each individual, because we are all really responsible for all.”<sup>16</sup> Thus, the general principles of solidarity include (1) embracing a culture of encounter, (2) exercising empathy, (3) build unity with diversity, and (4) engage inclusive practices.

Second, subsidiarity ensures that decision-making should be localized, ensuring

community relevance. Pius XI writes,

Just as it is gravely wrong to take from individuals what they can accomplish by their own initiative and industry and give it to the community, so also it is an injustice and at the same time a grave evil and disturbance of right order to assign to a greater and higher association what lesser and subordinate organizations can do. For every social activity ought of its very nature to furnish help to the members of the body social, and never destroy and absorb them.<sup>17</sup>

General principles of subsidiarity include (1) shift power to those more proximate to the issues, (2) democratize technology and access to it, (3) disclose information appropriately to exercise due transparency with stakeholders, and (4) obtain consent from appropriate parties as possible.

Third, the preferential option for the poor or marginalized entails concrete actions that are always just and partial to those in need. From the *Church's Compendium* we learn,

The principle of the universal destination of goods requires that the poor, the marginalized and in all cases those whose living conditions interfere with their proper growth should be the focus of particular concern. To this end, the preferential option for the poor should be reaffirmed in all its force. “This is an option, or a special form of primacy in the exercise of Christian charity, to which the whole tradition of the Church bears witness. It affects the life of each Christian inasmuch as he or she seeks to



imitate the life of Christ, but it applies equally to our social responsibilities and hence to our manner of living, and to the logical decisions to be made concerning the ownership and use of goods. Today, furthermore, given the worldwide dimension which the social question has assumed, this love of preference for the poor, and the decisions which it inspires in us, cannot but embrace the immense multitudes of the hungry, the needy, the homeless, those without health care and, above all, those without hope of a better future.”<sup>18</sup>

For the preferential option of the poor, we see the following general principles: (1) promote health equity and equal opportunity, (2) invite those marginalized to participate in design and decision-making procedures.

Fourth, the sociological pillar includes corporal works of mercy as a call to help those in need. Again, the Compendium states,

The Church's love for the poor is inspired by the Gospel of the Beatitudes, by the poverty of Jesus and by his attention to the poor. This love concerns material poverty and also the numerous forms of cultural and religious poverty. The Church, “since her origin and in spite of the failing of many of her members, has not ceased to work for their relief, defence and liberation through numerous works of charity which remain indispensable always and everywhere.” ... [T]he Church teaches that one should assist one's fellow man in his various needs and fills the human community with countless works of corporal and spiritual mercy. ... [E]

ven if the practice of charity is not limited to alms-giving but implies addressing the social and political dimensions of the problem of poverty. In her teaching the Church constantly returns to this relationship between charity and justice: “When we attend to the needs of those in want, we give them what is theirs, not ours. More than performing works of mercy, we are paying a debt of justice.”<sup>19</sup>

To perform corporal works of mercy, these general principles apply: (1) monitor job displacement caused by internal and external factors; (2) provide reasonable access to necessary education and training; (3) measure impact on beneficiaries and on workers, not merely intention alone; and (3) mitigate biases and eliminate all forms of unjust discrimination.

Fifth, liberation through structures of grace, as opposed to structures of sin, forms another sociological pillar from CST. *The Compendium* states,

The moral dimension of the economy shows that economic efficiency and the promotion of human development in solidarity are not two separate or alternative aims but one indivisible goal. Morality, which is a necessary part of economic life, is neither opposed to it nor neutral: if it is inspired by justice and solidarity, it represents a factor of social efficiency within the economy itself. The production of goods is a duty to be undertaken in an efficient manner, otherwise resources are wasted. On the other hand, it would not be acceptable to achieve economic growth at the expense of human beings, entire populations

or social groups, condemning them to indigence. The growth of wealth, seen in the availability of goods and services, and the moral demands of an equitable distribution of these must inspire man and society as a whole to practise the essential virtue of solidarity, in order to combat, in a spirit of justice and charity, those “structures of sin” where ever they may be found and which generate and perpetuate poverty, underdevelopment and degradation. These structures are built and strengthened by numerous concrete acts of human selfishness.<sup>20</sup>

We might think that the antidote or prophylaxis to structural sin is liberation through structures of grace. Thus, Benedict XVI writes in *Caritas in Veritate*:

The development of peoples is intimately linked to the development of individuals. The human person by nature is actively involved in his own development. The development in question is not simply the result of natural mechanisms, since as everybody knows, we are all capable of making free and responsible choices. Nor is it merely at the mercy of our caprice, since we all know that we are a gift, not something self-generated. Our freedom is profoundly shaped by our being, and by its limits. No one shapes his own conscience arbitrarily, but we all build our own “I” on the basis of a “self” which is given to us. Not only are other persons outside our control, but each one of us is outside his or her own control. A person's development is compromised, if he claims to be solely responsible for producing what he becomes. By analogy,

the development of peoples goes awry if humanity thinks it can re-create itself through the “wonders” of technology, just as economic development is exposed as a destructive sham if it relies on the “wonders” of finance in order to sustain unnatural and consumerist growth. In the face of such Promethean presumption, we must fortify our love for a freedom that is not merely arbitrary, but is rendered truly human by acknowledgment of the good that underlies it. To this end, man needs to look inside himself in order to recognize the fundamental norms of the natural moral law which God has written on our hearts.”<sup>21</sup>

General principles for the liberation of humankind through structures of grace include the following: (1) engage in inclusive, human-centered design, (2) apply structural competency to mitigate the social determinants of disease, and (3) cooperate appropriately with others to promote the common good.

### BRIDGING CST AND GAI: FORMULATION OF ETHICAL GUIDELINES

Now, I will attempt to synthesize these insights, particularly the general principles, as ethical guidelines for the design, development, and use of gAI in Catholic health care. While these guidelines’ specificity will be somewhere between principles and concrete moral norms that guide specific behaviors or choices, they should provide practical influence on gAI in Catholic health care. Overtime, additional guidelines derived from CST insights may be warranted; alternatively, the guidelines below may require further specification or elaboration in given circumstances. I have included some

guiding questions to prompt further reflection on these themes, too.

To begin, Catholic health care should engage human-centered design and inclusivity of gAI. Algorithms, data sets, and machine learning applications should reflect our diverse human tapestry, championing inclusivity and ensuring marginalized communities are not sidelined. This could include development of a sense of humanism and a spirituality of gAI and the related work. For example, questions to ask in the design, development, and deployment of gAI could include (1) How can the design process actively involve stakeholders from marginalized communities? (2) What measures are in place to ensure the AI system does not perpetuate existing biases? (3) Does the gAI reflect and enrich integrative human development as a whole (or does it compartmentalize and deconstruct in a way that adversely affects the human experience)?

Next, Catholic health care should aim to empower its workforce and enable continuous learning opportunities. As gAI reshapes work functions and workplaces, continuous training should ensure the workforce remains relevant and the connections among workers strengthened. Minimizing the adverse effects of disruptive technology is also key to the adoption and use of gAI applications. This should include translational skills-building as well as an emphasis on the humanities in AI. What training programs are available for employees to adapt to new AI technologies? How does the organization plan to maintain the relevance of human skills in an AI-driven environment?

Beyond its workforce, Catholic health care

should establish collaborative partnerships and practice community engagement. Collaboration is key. By forging partnerships and engaging communities, we ensure gAI is grounded in real-world needs. Thus, it is important to ask, what partnerships can be formed to ensure the AI system meets holistic, real-world needs and promotes the common good? Moreover, its design, development, and use should not be siloed; rather, it should be done in a truly participatory, synodal way. How is community engagement and feedback integrated upstream and downstream in the development and deployment of the AI application?

To ensure gAI is continuously improved and maintains ethical integrity, Catholic health care should enact ethical deployment protocols and transparent governance structures and processes. Robust governance structures should oversee gAI, ensuring ethical considerations are integral. In addition, advocacy efforts should be aimed at defending the human person and common good. What governance structures are in place to oversee the ethical considerations of AI deployments? How is transparency maintained in the AI system's decision-making processes? Is it clear who trains the AI and how and on which data sets?

In addition, Catholic health care should inclusively develop and collaboratively use choice architecture and enhance a gAI-stakeholder's autonomy. In an AI-augmented world, human agency remains paramount. This entails proper disclosure to appropriate parties of gAI practices, opt-in versus opt-out protocols, and informed consent procedures. Leadership should be able to answer, how are stakeholders involved in the decision-making

process related to AI's choice architecture? And, what mechanisms are in place to ensure that an AI system enhances rather than diminishes user (or beneficiary) autonomy?

In the rapidly evolving realm of artificial intelligence, Catholic health care should clarify proper authenticity and veracity of AI-generated output. To aid this, the delineation between authenticity and truth becomes paramount. Authenticity, in this context, refers to the genuine origin or source of data, ensuring that the foundational elements of AI models are rooted in proper attribution exhibiting coherence and cogency (e.g., hallucinations are identified, studied to understand errors in the AI, and mitigated). Truth, on the other hand, pertains to the accuracy and fidelity of AI outputs. As AI systems increasingly influence decision-making in healthcare, it is crucial to address and actively mitigate biases that might skew these outputs. This not only ensures that the results reflect genuine realities but also guards against the inadvertent perpetuation of existing disparities. Furthermore, a transparent disclosure of data sources, emphasizing their authenticity and representativeness, becomes an ethical imperative, fostering trust and credibility in AI-driven processes. What protocols are in place to verify the authenticity of data used and generated by the AI application? How does the application ensure that AI-generated output is accurate and truthful?

Moreover, the reliability of AI systems transcends their initial accuracy; it encompasses their consistent performance over time. Therefore, as these systems become integral to healthcare, Catholic health care should ensure continuous monitoring and validation to

maintain gAI reliability. Establishing feedback mechanisms, where users, patients, and other stakeholders can report inconsistencies or anomalies, enhances the system's adaptability and resilience. This iterative process of validation and recalibration not only ensures the system's ongoing reliability but also fortifies trustworthiness. Trust, in this domain, is not merely about technical robustness; it's about building and nurturing a relationship of dependability with communities of concern, ensuring that they can confidently rely on AI outputs for critical health care decisions. What are the key performance indicators for assessing the reliability of the AI application? How do these intersect with existing health care related indicators? When and at what cadence should the AI application be audited for performance and compliance with key legal and ethical norms?

Lastly, as AI delves into realms of creativity and innovation, the boundaries of intellectual property and creative rights come to the fore. Catholic health care should exhibit proper attribution of AI output as well as choose open-source versus proprietary models in ways that promote the common good and defend social justice. Especially in cases where AI models generate content or make decisions based on pre-existing works, it becomes ethically and legally imperative to provide clear attribution to the original sources. Respecting the creative rights of individuals and entities ensures that AI does not inadvertently infringe upon or dilute the value of original creations. Moreover, the ethical landscape of AI is further nuanced by the dichotomy between open-source and proprietary models. While open-source models champion transparency and collaborative betterment, proprietary models underscore the

sanctity of intellectual property. Navigating this landscape requires a delicate balance, ensuring that the benefits of AI are harnessed without compromising the rights and contributions of original creators. What guidelines are in place for attributing authorship or artistic credit to the output generated by the AI system? How does the choice between open-source and proprietary models align with the organization's commitment to social justice and the common good? When would a proprietary model be justified for the fiscal security of the organization and under what conditions would this be effectuated?

## CONCLUSION

As we stand at the precipice of another technological advancement and ethical discernment, the teachings of the Catholic Church offer a beacon of light, guiding our path. The rise of generative AI, with its transformative potential to ignite an 'industrialization of thought,' calls us to navigate this new frontier with a moral compass rooted in centuries of wisdom. By grounding our approach in the pillars of Catholic Social Teaching, we are better equipped to ensure that AI serves not just as a tool but as an extension of our commitment to human dignity, the dignity of work, the common good, and the overall betterment of society. In embracing these ethical guidelines, we affirm our responsibility to harness the power of AI in ways that uplift humanity, honor our shared values, and pave the way for a future where technology and ethics walk hand in hand. ✚

### NICHOLAS KOCKLER, PH.D, MS, HEC-C

Vice President, System Ethics Services  
Providence St. Joseph Health  
Renton, Washington

[nicholas.kockler@providence.org](mailto:nicholas.kockler@providence.org)

## ENDNOTES

1. Second Vatican Council, *Gaudium et spes* (GS), Vatican.va, December 7, 1965, accessed August 30, 2023, no. 4.
2. Jiao Sun, et al., "Investigating Explainability of Generative AI for Code through Scenario-based Design," IUI '22, March 22-25, 2022, DOI: [10.1145/3490099.3511119](https://doi.org/10.1145/3490099.3511119), accessed August 31, 2023.
3. Sara Vaezy, "Generative AI: Technology Worthy of Leadership's Attention," *Boardroom Press*, governanceinstitute.com, August 2023, accessed October 18, 2023.
4. Hazem Zohny, et al., "Ethics of generative AI," *Journal of Medical Ethics*, [jme.bmj.com](https://jme.bmj.com), February 2023, 49 (2): 79, accessed August 31, 2023.
5. Supra note 2.
6. Katja Thoring, et al., "The Augmented Designer: A Research Agenda for Generative AI-Enabled Design," *International Conference on Engineering Design*, July 24-28, 2023, Cambridge.org, accessed August 31, 2023.
7. John Paul II, *Centesimus annus*, Vatican.va, May 1, 1991, no. 11, accessed August 31, 2023.
8. GS, no. 26.
9. John Paul II, *Laborem Exercens*, Vatican.va, September 14, 1981, no. 9, accessed August 31, 2023.
10. Francis, *Laudato Si'* (LS), Vatican.va, May 24, 2015, no. 128, accessed August 31, 2023.
11. Second Vatican Council, *Dignitatis Humanae*, Vatican.va, December 7, 1965, no. 3, accessed August 31, 2023.
12. Lisa Sowle Cahill, *Theological Bioethics: Participation, Justice, Change* (Washington, DC: Georgetown University Press, 2005), 7.
13. Pontifical Council for Justice and Peace, *Compendium of the Social Doctrine of the Church*, Vatican.va, accessed August 30, 2023.
14. Benedict XVI, *Caritas in Veritate* (CV), Vatican.va, June 29, 2009, Chapter 4, Paragraph 18, accessed August 31, 2023.
15. LS, Chapter 4, Paragraphs 137-139. In paragraph 136, he writes, "In the same way, when technology disregards

the great ethical principles, it ends up considering any practice whatsoever as licit. As we have seen in this chapter, a technology severed from ethics will not easily be able to limit its own power."

16. John Paul II, *Sollicitudo rei Socialis*, Vatican.va, December 30, 1987, no. 38, accessed August 31, 2023.
17. Pius XI, *Quadragesimo Anno*, Vatican.va, May 15, 1931, no. 79, accessed August 30, 2023.
18. Pontifical Council for Justice and Peace, *Compendium of the Social Doctrine of the Church (Compendium)*, Vatican.va, May 26, 2006, no. 182, accessed August 31, 2023.
19. *Compendium*, no. 184.
20. *Compendium*, no. 332.
21. CV, Chapter 6, no. 68.