Artificial Intelligence and Ethical Reflections from the Catholic Church

Teofilo Giovan S. Pugeda III ♦

Abstract: The article explores some views from the Catholic Church regarding artificial intelligence (AI). It covers the following: 1) a description of the development of AI and the ethical issues raised by it; 2) an overview of the ethical considerations of the Catholic Church through a survey of three events and two documents in relation to AI; 3) an interpretative commentary; and 4) an attempt to offer a Christological-ethical reflection on AI.

Keywords: Artificial Intelligence • Catholic Church • Ethics • Technology

Introduction

Artificial Intelligence (AI) often evokes *The Terminator* and *The Matrix* movie franchises—the more recent films, *Free Guy* and *Finch*, bring AI to another level. AI has a large and growing influence on global society's direction. It is one of the "signs of the times" because it is quite unthinkable today to be not affected by it—for better or for worse. With AI's ubiquity, members of the Catholic Church hierarchy, in conversation with other stakeholders, propose some ethical guidelines for everyone. This article presents these guidelines and further suggests developing an AI ethics inspired by the words and deeds of Jesus of Nazareth.

♦ Teofilo Giovan S. Pugeda III is a lecturer at the Theology Department of Ateneo de Manila University. He earned his Master of Arts in Theological Studies with a concentration in moral theology at Ateneo de Manila University and Loyola School of Theology. Email address: tpugeda@ateneo.edu.

MST Review 23/2 (2021): 52-93
Received, 03 April 2021; Accepted, 03 Oct. 2021; Published 16 Dec. 2021

The Development of AI

The notion of intelligent machines dates to the philosophers and mathematicians of the 17th century, with the likes of René Descartes, Gottfried Wilhelm Leibniz, and Blaise Pascal designing "calculating machines that mechanized arithmetic, which had hitherto been the province of learned men called 'calculators,' but they never made the claim that the devices could think."[1] Today, AI, being the "the science and engineering of making intelligent machines,"[2] takes that for granted.[3]

The modern understanding of AI comes from the 1950s with the pioneering work of John Von Neumann and Alan Turing.

> They made the transition from computers to 19th century decimal logic (which thus dealt with values from 0 to 9) and machines to binary logic (which rely on Boolean algebra, dealing with more or less important chains of 0 or 1). The two researchers...formalized the architecture of our contemporary computers and demonstrated that it was a universal machine, capable of executing what is programmed.[4]

---

[1] Bruce G Buchanan, "A (Very) Brief History of Artificial Intelligence," *AI Magazine* 26, no. 4 (December 2005): 53.

[2] John McCarthy, "What Is Artificial Intelligence?" What Is Artificial Intelligence? (Stanford, 2007), http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html (accessed 1 April 2021).

[3] AI is "the field of study devoted to making machines intelligent." Paul Scharre and Michael Horowitz, "Artificial Intelligence What Every Policymaker Needs to Know," Center for New American Security (Center for a New American Security, June 19, 2018), last modified June 19, 2018; https://www.cnas.org/publications/reports/artificial-intelligence-what-every-policymaker-needs-to-know (accessed 31 March 2021).

[4] "History of Artificial Intelligence," Artificial Intelligence (Council of Europe, n.d.); https://www.coe.int/web/artificial-intelligence/home (accessed 31 March 2021).

Turing wrote the article "Computing Machinery and Intelligence" in 1950, which proposed the intriguing question "Can machines think?" Although he considered the question ill-defined, he developed the "Turing Test" to examine the problem.[5] According to this thought experiment, an individual would use a computer to ask questions to two entities, one human and the other machine. Based on the answers received on the computer screen, the individual must determine within a specified period which entity is human and which is the machine. If the machine could successfully fool the individual into thinking that it was human, then it passed the Turing Test.[6] Anticipating the theological implications, Turing presented a theological objection (one of nine objections) against machine thinking in his article, though he admits that he is not "very impressed with theological arguments whatever they may be used to support."[7] He claims: "Thinking is a function of man's immortal soul. God has given an immortal soul to every man and woman, but not to any other animal or to machines. Hence no animal or machine can think."[8] But he suggests against the theological objection that it is not impossible, so he thinks, for God to confer souls into machines.[9] In

[5] Turing called it the "Imitation Game," which is also how some refer to it today.

[6] The Turing Test has spawned variations of tests depending on the AI aspect. Among them, Colin Allen, G. Varner, and J. Zinser proposed in 2000 the "comparative moral Turing Test," which determined the ethical judgement of an AI by putting it against the ethical judgement of a human in response to a moral dilemma. If it could be identified as the more ethical one, then it passed the test.

[7] Alan M. Turing, "Computing Machinery and Intelligence," *Mind*, 59/236, (October 1950): 433–460.

[8] Ibid, 443.

[9] Selmer Bringsjord offers a rebuttal to Turing's theological objection; see, Selmer Bringsjord, "God, Souls, and Turing: in Defense of the Theological Objection to the Turing Test," *Kybernetes* 39, no. 2 (May 4, 2010): 414-422.

1951, Turing designed the first computer program to learn a game, specifically chess.[10] A year later, Arthur Samuel designed a similar computer program but for checkers. In 1955, Allen Newell, Cliff Shaw, and Herbert Simon authored the Logic Theorist, a computer program designed to mimic human problem-solving skills.[11]

Following these developments, John McCarthy organized a two-month workshop conference at Dartmouth College in 1956,[12] where he coined the term "artificial intelligence." One of the attendees, Marvin Minsky, defined it as "the construction of computer programs that engage in tasks that are currently more satisfactorily performed by human beings because they require high-level mental processes such as perceptual learning, memory organization, and critical reasoning."[13] In 1958, McCarthy created LISP (list processing), the second oldest computer programming language used by AI today.[14] Minsky published in 1961 his paper entitled "Steps toward Artificial Intelligence," which would help

---

[10] In 1912, Torres Quevedo had developed a rudimentary electromagnetic device that could do end-game chess moves.

[11] Leo Gugerty, "Newell and Simon's Logic Theorist: Historical Background and Impact on Cognitive Modeling," Proceedings of the Human Factors and Ergonomics Society Annual Meeting 50/9 (October 1, 2006): 880-884.

[12] Although the 1956 Dartmouth workshop-conference was most noteworthy in AI history, Herbert Bruderer suggests that a 1951 Paris conference on computing and human thinking was significant as a predecessor.

[13] "History of Artificial Intelligence," *Artificial Intelligence* (Council of Europe, n.d.); https://www.coe.int/web/artificial-intelligence/home (accessed 31 March 2021).

[14] John McCarthy, "History of Lisp," Professor John McCarthy (Stanford University, February 12, 1979), last modified February 12, 1979; http://jmc.stanford.edu/articles/lisp.html (accessed 1 April 2021).

inspire researchers of the time to consider AI as a serious discipline.[15]

Over time, different types of AI developed. From Turing up to the 1980s, AI was understood mainly within a symbolic paradigm, which "is that intelligence results from the manipulation of abstract compositional representations whose elements stand for objects and relations."[16] In a sense, it treated the matter of cognition as essentially the same as logical computation. But one significant limitation of symbolic (or classical) AI was its reliance on a controlled environment. Symbolic AI computation only works if it handles data in a structured manner. By the 1980s, interest shifted toward understanding AI in relation to environmental adaptation, known as the Physical Grounding Hypothesis, which gave rise to embodied AI. This required programming AI into robots, which moved within unpredictable environments, real or simulated. Cognition was measured by how well the AI responded to different environmental stimuli. As Rodney Brooks remarked, "The new methodology bases its decomposition of intelligence into individual behavior generating modules, whose coexistence and co-operation let more complex behaviors emerge."[17] However, Riccardo Manzotti, citing Tom Ziemke, notes that,

> A problem with embodied AI, or in fact embodied cognitive science in general, is that it seems to be much

---

[15] Patrick Henry Winston, "Marvin L. Minsky (1927–2016)," *Nature*, 530/7590 (February 18, 2016): 282.

[16] Marta Garnelo and Murray Shanahan, "Reconciling Deep Learning with Symbolic Artificial Intelligence: Representing Objects and Relations," *Behavioral Sciences* 29 (October 2019): 17-23.

[17] Rodney Brooks, "Elephants Don't Play Chess," *Robotics and Autonomous Systems* 6, no. 1-2 (June 1990): 3-15; https://www.sciencedirect.com/science/article/abs/pii/S0921889005800259 (accessed 16 June 2021).

> more defined in terms of what it argues against than what it argues for [...]. Many embodied AI researchers reject the idea that intelligence and cognition can be explained in purely computational terms, but it is left unclear exactly what the alternative is.[18]

To resolve the ambiguity, researchers tried to understand intelligence in terms of pattern recognition. AI was trained to recognize various data and discern desired patterns from them. Eventually, the AI system, which is known as connectionist (or emergentist) AI, "learns" to produce outputs or take action on its own following the patterns, with greater precision achieved the more data is available. This learning process is known as Machine Learning, which Arthur Samuel defined in 1959 as the "field of study that gives computers the ability to learn without being explicitly programmed."[19] It does this by "systematically [applying] algorithms to synthesize the underlying relationships among data and information."[20] A subset of Machine Learning is Deep Learning, which "is a special kind of learning with deep artificial neural networks"[21] inspired by the human brain. Today, "it has become established as

---

[18] Riccardo Manzotti, "Embodied AI beyond Embodied Cognition and Enactivism" *Philosophies* 4/3 (16 July 2019): 39. https://doi.org/10.3390/philosophies4030039.

[19] Arthur L. Samuel, "Some Studies in Machine Learning Using the Game of Checkers," *IBM Journal of Research and Development* 44/1.2 (1959): 210–229.

[20] Mariette Awad and Rahul Khanna, "Machine Learning," in Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers (Berkeley, CA: Apress Open, 2015), pp. 1-268; https://link.springer.com/book/10.1007/978-1-4302-5990-9#about (accessed 16 June 2021).

[21] Sandro Skansi "Preface." Preface. In *Introduction to Deep Learning From Logical Calculus to Artificial Intelligence (*Springer, 2018), v.

one of the most impactful research areas within [AI]."[22] Some limitations of connectionist AI are that it does not provide a clear step-by-step layout of the learning process, which takes a long time, and it may produce poor generalizations if the data sets are deficient.

Technological innovation soon led to efforts to synthesize symbolic and connectionist AI. Nils Nilsson theorized that "AI systems that achieve human-level intelligence will involve a combination of symbolic and non-symbolic processing."[23] What came to be were hybrid systems loosely known as neuro-symbolic AI. But current neuro-symbolic AI research is "still a long way from a satisfying synthesis."[24] As Garnelo and Shanahan remarked,

> A truly satisfying synthesis of symbolic AI with deep learning would give us the best of both worlds. Its representations would be grounded, learned from data with minimal priors. It would be able to learn representations comprising variables and quantifiers as well as objects and relations. It would support arbitrarily long sequences of inference steps using all those elements, like formal logic. But it would not be constrained by the rules of formal logic, and would be able to learn forms of inference that transcend the strictures they imply. Given an architecture that combined all these features, the desired properties of data efficiency, powerful generalisation, and human interpretability would likely follow.[25]

---

[22] Garnelo and Shanahan, "Reconciling Deep Learning with Symbolic Artificial Intelligence," 17.

[23] Ben Goertzel, "Artificial General Intelligence: Concept, State of the Art, and Future Prospects," *Journal of Artificial General Intelligence* 5/1 (2014): pp. 1-46, https://doi.org/ 10.2478/jagi-2014-0001.

[24] Garnelo and Shanahan, "Reconciling Deep Learning with Symbolic Artificial Intelligence", 21.

[25] Ibid.

On top of the different AI types, there are two broad categories of AI. Currently, existing AI is considered as Artificial Narrow Intelligence (ANI) or "Weak" AI. This AI category can achieve very advanced processes, though only for specific (narrow) tasks. In other words, it is "domain-specific, excellent at specific tasks...however, [it] cannot transfer to another domain."[26] Still, with what specific tasks it can do, current ANI has proven to significantly beat human counterparts. On the other hand, Artificial General Intelligence (AGI) or "Strong" AI is the category of AI that deals more with the theoretical aspects of machine intelligence, such as consciousness, mind, and moral agency. Ben Goertzel advances the core hypothesis of AGI as,

> The creation and study of synthetic intelligences with sufficiently broad (e.g. human-level) scope and strong generalization capability, is at bottom qualitatively different from the creation and study of synthetic intelligences with significantly narrower scope and weaker generalization capability.[27]

John Searle suggests that "the appropriately programmed computer is a mind, in the sense that computers can be literally said to understand and have other cognitive states."[28] AGI is still hypothetical and eludes a single definition because, to date, no machine has fully demonstrated human-level capabilities. Searle expresses concern about whether Strong AI could even be

---

[26] Wim Naude and Nicola Dimitri, "The Race for an Artificial General Intelligence: Implications for Public Policy," *AI and Society* 35 (April 22, 2019): 367-379; https://link.springer.com/article/10.1007/s00146-019-00887-x (accessed 16 June 2021).

[27] Goertzel, "Artificial General Intelligence," 3.

[28] John. R. Searle, "Minds, brains, and programs." *Behavioral and Brain Sciences* 3 no. 3 (1980): 417-457.

achieved. In 1980, he proposed a thought experiment to illustrate his point called the Chinese Room argument.[29] Searle defended the idea that semantics (meaning) is irreducible to syntax, which alone does not constitute understanding. Since computers operate on syntax, "This leads to the conclusion that computers do not possess any intentional (meaningful) mental states. The core of Searle's argument against the strong hypothesis of artificial intelligence hinges upon the inability to derive semantics out of syntax."[30]

By the 21st century, AI had already extended beyond the academic realm. Its applications in online search engines, social media, video games, and voice recognition devices are common in consumerist commerce. But AI also offers opportunities for service-based sectors. The health sector is especially open to AI. According to Fidelma Fitzpatrick et al., "AI has the potential to detect transmission events during outbreaks or predict high-risk patients, enabling [the] development of tailored [infection prevention and controlled] interventions."[31] They use the example of Bluedot, an AI platform. According to them, "Bluedot first alerted on COVID-19 on

---

[29] Searle imagines himself, an English monoglot, to be inside a room. Outside the room were Chinese speakers who would slip questions in Chinese underneath the door. Searle, with no knowledge of Chinese, would use an English manual on Chinese grammar to decode the messages and form replies. He would then release his replies in linguistically correct Chinese underneath the door. The Chinese speakers outside would believe that the one who produced the responses understands Chinese. Yet, Searle does not understand Chinese.

[30] Wojciech P Grygiel, "Artificial Intelligence and Scholastic Epistemology," *Annales UMCS Informatica AI* 3 (2005): 93-103; http://ai/annales.umcs.pl (accessed 26 June 2021).

[31] Fidelma Fitzpatrick, Doherty A, Lacey G. Using Artificial Intelligence in Infection Prevention. Curr Treat Options Infect Dis. 2020 Mar 19:1-10. doi: 10.1007/s40506-020-00216-7. Epub ahead of print. PMID: 32218708; PMCID: PMC7095094.

December 31, 2019, almost a week ahead of national surveillance centers and WHO."[32] Anita Ho suggests that AI could assist in elderly care.[33] The global population of those aged at least 65 is projected to reach 1.6 billion by 2050, with no guarantee of a commensurate increase in healthcare professionals.[34] In this regard, "AI monitoring data regarding older adults' risk levels for adverse events can inform medical decision making and transform healthcare delivery. They can help healthcare providers to triage cases to ensure that the right patients have timely access to appropriate care."[35] Apart from these uses, AI can also be used for recreational or peculiar purposes. One example is Deep Nostalgia, an AI program of MyHeritage which animates old photographs to appear like a recording.[36] According to a media article,

> Users are invited to supply old photos of their loved ones, and the program uses deep learning to apply predetermined movements to their facial features. It also makes up for little moments that aren't in the original photo, like the reveal of teeth or the side of a head. Together it creates, if not an entirely natural effect, [then] a deeply arresting one.[37]

While Deep Nostalgia can provide comfort for the deceased's relatives by digitally editing photographs, the

---

[32] Ibid, 138.

[33] Anita Ho, "Are We Ready for Artificial Intelligence Health Monitoring in Elder Care?," *BMC Geriatrics* 358 (2020).

[34] Ibid.

[35] Ibid.

[36] AJ Willingham, "A New Program Can Animate Old Photos. But There's Nothing Human about Artificial Intelligence," *CNN Business* (CNN Worldwide, March 5, 2021), last modified March 5, 2021; https://edition.cnn.com/2021/03/02/business/deep-nostalgia-myhertiage-ai-learning-trnd/index.html (accessed 31 March 2021).

[37] Ibid.

developers warned the public not to submit photographs of living people to prevent abuse of the program.

Ethical Issues Raised by AI

The ethical issues raised by AI are complex and involve theoretical and practical aspects. However, to avoid the wrong assumption that theoretical is necessarily abstract, the theoretical ethical issues discussed in this section are very much rooted in current AI developments. Moreover, there are a variety of practical ethical issues raised by AI, but only three of these will be considered.

Theoretical Ethical Issues

David Chalmers asks "What happens when machines become more intelligent than humans?"[38] He suggests that what could occur is an intelligence explosion leading to a singularity (the point of physical limit for intelligence.) In one scenario, humans would create super-intelligent machines capable of creating more super-intelligent machines, and so on until singularity is achieved. There are two possible non-exclusive routes in this regard. The first involves creating human-based AI, and the second involves creating non-human-based AI. According to Chalmers, "Under human-based AI, each system is either an extended human or an emulation of a human. The resulting systems are likely to have the same basic values as their human sources."[39] One factor worth carefully considering here is how far we would allow any individual to be extended by AI. Extended humans could escalate any conflict. Dangerous organizations could use

[38] David Chalmers, "The Singularity: A Philosophical Analysis," *Journal of Consciousness Studies* 17 (2010): pp. 7-65.

[39] Ibid, 25.

the technology for nefarious purposes. Even if such technology were used therapeutically, achieving global equity is difficult like basic healthcare today. Eventually, the human mind could be programmed into a computer forming a transhuman entity.[40] In contrast to Christian anthropology of human beings as a body-soul unity, Grzegorz Osinski points out that "Transhumanists consider the problem of mind transfer only in material terms, treating the mind only as a product of neural activity and therefore completely ignore the concept of the human soul."[41] The human body is viewed as a limitation to overcome, rather than our way of experiencing the world. This negative view may lead to the exploitation of the human body in pursuit of freeing the mind. Benedikt Paul Göcke warns that "the greatest danger of synthetic biology is that we may treat human beings as simply disposable and run the risk of violating their inalienable human dignity. A person who knows how to genetically modify particular features of a human being will normally also know how to erase them."[42] The quest to enhance the human condition may ultimately result in its degradation.

As for non-human-based AI, Chalmers asks "What sort of values should we aim to instil in a non-human-based AI [?]"[43] The worst case is "If at any point there is a powerful AI+ [greater than human intelligence] or AI++ [far greater than human level] with the wrong value system, we can expect disaster (relative to our values) to

[40] The idea is not as far-fetched when one remembers that current medical technology can already connect with the human brain.

[41] Grzegorz Osiński, "Theological and Ethical Aspects of Mind Transfer in Transhumanism," *Scientia et Fides* 9/1 (2021): 149-176.

[42] Benedikt Paul Göcke, "The Ideals of Humanity in light of Synthetic Biology and Artificial Intelligence," *Concilium* 3 (2019): 25-33.

[43] David Chalmers, "The Singularity," 26.

ensue."[44] With the many competing values within society today, it is important to clearly define what we would want to impart on AI, or it might mirror our worst qualities, such as unbridled belligerence or racism. As a Christian, Yong Sup Song argues "that some religious aspects should be considered as necessary foundations"[45] for AI+. He believes that "AI could be good and safe only when it remains a creature among humans, remembering God as the ultimate Creator."[46] Whatever the values imparted on AI, Michael Anderson and Susan Anderson state that,

> If we believe that machines could play a role in improving the lives of human beings—that this is a worthy goal of AI research—then, since it is likely that there will be ethical ramifications to their behavior, we must feel confident that these machines will act in a way that is ethically acceptable.[47]

Theoretically, AI is only under obligation to act ethically if it has moral agency, which is as yet undetermined if it can possess. As Roman Yampolskiy and Joshua Fox remarked, "Defining an ethical system for a superhuman and inhuman intelligence takes us to areas inadequately explored by philosophers to date."[48] If greater-than-human-intelligence AI is achieved, then do we ascribe moral agency to it? Agency, "at least as it is

---

[44] Ibid, 27.

[45] Yong Sup Song, "Religious AI as an Option to the Risks of Superintelligence: A Protestant Theological Perspective," *Theology and Science* 19/1 (October 6, 2020): 65-78.

[46] Ibid, 70.

[47] Michael Anderson and Susan Anderson, "Machine Ethics: Creating an Ethical Intelligent Agent," *AI Magazine* 28/4 (2007): 15-25.

[48] Roman Yampolskiy and Joshua Fox, "Safety Engineering for Artificial General Intelligence," *Topoi* 32 (August 24, 2012): pp. 217-226, https://doi.org/10.1007/s11245-012-9128-9.

typically characterized and understood, requires that there be some kind of animating 'intention' behind the observed action."[49] In turn, moral agency is the capacity to intentionally act following specified moral standards. Concerning AI, Goertzel subscribes to the instrumental theory, which views AI as instruments or tools similar to a hammer. For Goertzel, a human being is ultimately responsible for any consequences of machines. Mikko Siponen and Abbe Mowshowitz suggest that to ascribe moral agency to AI is to divert attention away from someone who should be held accountable.[50] Deborah Johnson and Keith Miller express the overall sentiment of instrumental theory as "it is dangerous to conceptualize computer systems as autonomous moral agents."[51] For instrumentalists, AI is not a moral agent. Yet, the instrumental theory of AI has been called anthropocentric. According to critics, not all machines are directly operated by and dependent on humans; some are autonomous and capable of greater intelligence. Goertzel admits that,

> Not too far in the future, however, things are going to be different. AI will possess true artificial general intelligence (AGI), not necessarily emulating human intelligence, but equaling and likely surpassing it. At this point, the morality or otherwise of AGI's will become a highly significant issue.[52]

David Gunkel states that "If these predictions are even partially correct and accurate, then what has been defined and largely limited to the status of a mere instrument will, at some point in the not too distant

---

[49] David J. Gunkel, *The Machine Question: Critical Perspectives on AI, Robots, and Ethics* (Cambridge, MA: MIT Press, 2017), 19.

[50] Ibid, 28.

[51] Ibid, 29.

[52] Ibid, 32.

future, no longer be just a tool or an extension of human capabilities."[53] It may be a moral agent. History has shown the risks of limiting one's perception of moral agency to specific categories of persons. Over the past two hundred years, people of color and women were recognized as having as much moral agency as Caucasian males. By the 1970s, animals with high levels of intelligence had interested philosophers, such as Peter Singer, enough to think that animals could be "persons" loosely defined.[54] With this in mind, machines could be considered in light of the same impulse towards an inclusive understanding of moral agents.

Yampolskiy and Fox suggest that "At an early stage, when AIs have near-human intelligence, and perhaps humanlike mind architectures and motivation systems, humanlike morality, regulated by law, trade, and other familiar constraints towards mutual cooperation, may be enough."[55] But such a suggestion becomes unwieldy the more intelligent machines become. They speculate that "There is no reason that a superintelligence [AI] would necessarily have goals favoring human welfare, which are a tiny part of the space of possible goals."[56] So they propose safety engineering AI to desire certain values, like propagating those values to succeeding generations of AI. However, the more intelligent the machines become, "the number of errors in the design increases proportionately or perhaps even exponentially."[57] Even if certain values are initially programmed, the risk of

---

[53] Ibid, 33.

[54] Singer's 1975 book entitled "Animal Liberation" provides a good description of his position regarding an expanded notion of personhood within the animal kingdom.

[55] Yampolskiy and Fox, "Safety Engineering for Artificial General Intelligence," 4.

[56] Ibid, 6.

[57] Ibid, 9.

diminishing returns thereafter is still possible. As things are, AI moral agency remains an open question.

Assuming AI is not a moral agent, then does it follow that it is not a moral patient? Moral patiency refers to the extent to which an entity "might constitute an *other* to which or to whom one would have appropriate moral duties and responsibilities."[58] It is generally accepted that moral agents are also moral patients, but not always vice-versa. As an example, while animals are considered moral patients, they are generally not considered moral agents. In other words, though they are the objects of moral action, they are not moral actors. While Descartes characterized animals and machines on the same level to the detriment of animals, developments in animal ethics gave cause to reconsider the validity of the Cartesian characterization to the benefit of machines. Is AI a moral patient? From an instrumentalist's perspective, AI is not a moral patient for the same reason that it is not a moral agent, i.e. because it is a tool to be used. Logically, if we assume that AI is a moral agent, then there is a presumption that it is also a moral patient. But, as noted earlier, AI moral agency remains an open question. Perhaps AI may be a moral patient without being a moral agent, like animals? David Calverley suggests that,

> As a result of modern science, animals have been shown to possess...characteristics that...make them something more than inanimate objects...but less than human...These characteristics...are similar to characteristics designers are seeking to instantiate in androids. If designers succeed with the task they have set for themselves, then logically androids...could assert claims to moral considerations in a manner similar to those claimed for animals.[59]

---

[58] David J. Gunkel, *The Machine Question*, 93.
[59] Ibid, 110.

Unlike the Cartesian counterpart, Calverley's analogy between animals and machines seeks to expand rather than limit the notion of moral patiency. However, the analogy has its limitations. For one thing, animals are considered moral patients because they suffer similarly to human beings. The early advocates of animal ethics capitalized on the consequentialist philosophy of Jeremy Bentham to assert that animal suffering is deplorable, like human suffering. But, as far as it is known, machines with AI do not suffer similarly to animals. It is as yet uncertain whether machines can suffer at all. Instrumentalists logically point out that the push against anthropocentrism in the context of AI leaves us with no definitive frame of reference of what constitutes suffering apart from humanity's understanding of it. Even then, suffering remains a questionable standard for moral patiency because it is not a static term. Suffering is more than just pain, which is a biological response to strong stimuli. Suffering is firstly a psychological state and presumes the existence of a mind. Yet, harkening back to Searle's Chinese Room argument, it is contentious whether AI has a mind understood in human terms. Thus, as with AI moral agency, AI moral patiency remains an open question.

Practical Ethical Issues

One practical ethical issue which AI raises is the spreading of fake news. This issue is apparent in Deepfake, which is "a video or sound recording that replaces someone's face or voice with that of someone else, in a way that appears real."[60] Deepfake is often used to advance a controversial agenda by using someone's

[60] "DeepFake," in Cambridge Dictionary (Cambridge University Press); https://dictionary.cambridge.org/us/dictionary/english/deepfake (accessed 23 June 2021).

identity without his or her consent. It can also manipulate someone's image to appear nonsensical. In December 2020, a Deepfake video of Queen Elizabeth II stirred the public when she appeared to address several pressing issues in controversial manners. After which, she appeared to dance on her desk. The video prompted at least two hundred complaints addressed to the UK Office of Communications.[61] What was ironic was that it was made and released by a British public-service television network to raise awareness of the dangers of Deepfake. The misinformation caused by Deepfake is divisive and contrary to the common good.

A second practical ethical issue is the infringement of data privacy. As stated by a media article, "According to the AIGS (Artificial Intelligence Global Surveillance) index, at least 75 out of 176 countries globally are actively using AI-based surveillance technologies."[62] While they help prevent security threats, there is concern about their possible invasive nature for regular people. This concern also holds for the private sphere. One example is a US company which "created an algorithm that figured out if people were pregnant based on purchase patterns, and the company then sent coupons to...those customers. That kind of predictive action was problematic, especially...when a young woman hadn't yet told her father she was pregnant, but mailed coupons informed

---

[61] BBC, "Deepfake Queen Prompts 200-plus Complaints to Ofcom," *BBC News* (BBC, December 29, 2020), last modified December 29, 2020; https://www.bbc.com/news/technology-55478579 (accessed 31 March 2021).

[62] Steve Nouri, "Council Post: How AI Is Making An Impact On The Surveillance World," Forbes (Forbes Magazine, December 3, 2020), last modified December 3, 2020; https://www.forbes.com/sites/forbestechcouncil/2020/12/04/how-ai-is-making-an-impact-on-the-surveillance-world/?sh=22a93632265e (accessed 31 March 2021).

him instead."[63] At the core of this issue is the right to privacy. Today, we speak of data privacy. Someone's digital identity can be summed up into a set of online personal data, which ideally remains private. In reality, however, the data can be exploited. Recalling that machine learning precision is greater the more data is available, many AI systems contain hundreds of thousands of people's data, some in the millions.[64] It would be precarious for everyone if the AI systems are breached.[65]

A third practical ethical issue is the changing employment landscape. Fortunately, AI can do routine tasks to allow people to pursue more meaningful work. It can also generate in the long term many jobs which do not currently exist. But in the short term, the job displacement it causes is economically troubling for the affected. One of the most affected in this regard is the call center industry. Companies are expanding the use of AI-powered robots to provide telecommunication services. For a North American company, clients had communicated with robots around 10% of the time before the COVID-19 pandemic, but it rose to almost 25% during the

[63] David A Teich, "Artificial Intelligence And Data Privacy – Turning A Risk Into A Benefit," Forbes (Forbes, August 10, 2020), last modified August 10, 2020; https://www.forbes.com/sites/davidteich/2020/08/10/artificial-intelligence-and-data-privacy--turning-a-risk-into-a-benefit/?sh=1c79e43c6a95z (accessed 31 March 2021).

[64] Another ethical issue is that AI might discriminate against certain groups based on the data. Companies using AI to filter employment candidates might unknowingly give advantages to some groups over others. Similarly, input data that is drawn from a limited portion of the population may produce output that reflects the limitation.

[65] For a good reference on the Catholic response to data-sharing in the health sector, see: Jean Baric-Parker and Emily E. Anderson, "Patient Data-Sharing for AI: Ethical Challenges, Catholic Solutions," *The Linacre Quarterly* 87/4 (2020): 471-481.

pandemic.[66] The Asian Development Bank projected that AI could displace at least 283,000 call center workers in the Philippines by 2030.[67]

Catholic Church's Ethical Considerations

The Catholic Church, in the form of the Holy See, takes interest in providing a moral presence in the ambit of AI. Within four years, it organized a symposium, a conference, and a workshop, which were all held in Rome. From 30 November to 1 December 2016, the Pontifical Academy of Science organized the symposium entitled "Powers and Limits of Artificial Intelligence." Three years later, from 26 to 28 September 2019, the Pontifical Council for Culture and the Dicastery for Promoting Integral Human Development jointly organized the conference entitled "The Common Good in the Digital Age." Five months later, from 26 to 28 February 2020, the Pontifical Academy for Life organized the workshop entitled "The 'good' algorithm? Artificial Intelligence, Ethics, Law, Health."[68] Additionally, the Commission of the Bishops' Conferences of the European Union (COMECE), which is the association of the Catholic Church in the EU, issued a document in January 2019 entitled "Robotisation of Life – Ethics in view of new

---

[66] Bruce Einhorn, Siegfrid Alegado, and Ditas B Lopez, "Empathetic Robots Are Killing Off the World's Call-Center Industry," Bloomberg.com (Bloomberg, March 17, 2021), last modified March 17, 2021; https://www.bloomberg.com/news/articles/2021-03-16/artificial-intelligence-chatbots-threaten-call-center-industry-human-operators (accessed 31 March 2021).

[67] Ibid.

[68] The Pontifical Academy of Life held another workshop from 25 to 26 February 2019 entitled Roboethics: Humans, Machines and Health. For brevity, this article did not touch upon it. For more information on this workshop, see: http://www.academyforlife.va/content/pav/en/events/general-assembly-2019.html.

challenges" as a complement to the EU's creation of a High-Level Expert Group on AI. Later on, COMECE would follow up by publishing on 14 June 2020 an annex to the EU's public consultation with a White Paper on Artificial Intelligence.[69]

Each event and document will be chronologically surveyed with some interpretative commentary for further consideration. While there were dozens of speakers from various backgrounds in the three events in Rome, attention will be given only to the views of some Church officials to distill what magisterium-informed digital ethics looks like. Similarly, while many informative paragraphs constituted the documents, only excerpts, summaries, or highlights will be presented below.

Powers and Limits of Artificial Intelligence

In two days, twenty participants from seven countries gathered in the Casina Pio IV to discuss artificial intelligence in connection to an array of topics which included mathematics, evolution, children, and consciousness. Of these participants, the most globally known of them was the late Professor Stephen W. Hawking, who offered a commentary on the ethics of

[69] The European Commission initiated a public consultation from 19 February to 14 June 2020 in the form of a White Paper which proposed several policies and regulations to make Europe the most attractive, secure and dynamic data-agile economy in the world. The White Paper came with an online survey to gauge the public's reaction to the document. The final report on the White Paper revealed that 1,215 responses were received from citizens, business operators, civil representatives, academics, and non-EU persons, with mixed responses to the proposals. Generally, there was an attitude of openness towards more AI integration in the EU. Draft regulations were published in 2021.

artificial intelligence. By its end, the symposium issued a two-page statement.[70]

The statement's more salient points include [1] "While progress is impressive, no evidence suggests the imminent emergence of a runaway intelligence with a will of its own. Artificial intelligence remains far from human and lacks an overarching mathematical framework,"[71] [2] "Used as a toolkit, AI has the potential to advance every area of science and society,"[72] [3] "Unless channeled for public benefit, AI will soon raise important concerns for the economy and the stability of society. We are living in a drastic transition period where millions of jobs are being lost to computerized devices, with a resulting increase in income disparity and knowledge gaps,"[73] and [4] "The effort to develop intelligent machines must remain continuously directed to the greater good, reducing the poverty gap and addressing general needs for health, education, happiness and sustainability."[74]

It should be noted that, because the symposium was not organized with the Catholic perspective as the only criterion, evident by the participation of Professor Hawking, a non-believer but a then-member of the Pontifical Academy of Science, there is nothing exclusively Catholic with the statement issued. What is remarkable about the symposium is that the Holy See made it clear that AI is within the purview of Catholic interest.

---

[70] Pontifical Academy of Science, "Statement on Artificial Intelligence" 1 December 2016, http://www.pas.va/content/dam/accademia/pdf/scripta_varia_132.pdf.

[71] Ibid.

[72] Ibid.

[73] Ibid.

[74] Ibid.

Robotization of Life - Ethics in view of new challenges

On the European regional level, COMECE had noted the rapid advances of AI on the continent, and the discussions and measures formed by the EU. It contributed to the European Parliament's 2017 consultation on Robotics and Artificial Intelligence, when it "expressed perplexity on the possible creation of a new dedicated EU Agency, as robotics is an extremely sectorial domain that can be covered in broader contexts (e.g. innovation and technology)."[75] But the EU ended up creating a High-Level Expert Group on AI.[76] In complement to it, COMECE established an *ad hoc* committee whose members specialized in theology, philosophy, law, and engineering. Professor Antonio Autiero of the University of Munster led the committee, which aimed to produce a document on AI vis-à-vis Catholic ethics.

The resulting document begins with a cautious overview of robotisation.[77] While praising the advantages

---

[75] Commission des Episcopats de la Communauté Européenne, "Annex to the public consultation on the White Paper on Artificial Intelligence – A European Approach," June 2020, http://www.comece.eu/dl/kmsNJKJKkKoOJqx4KJK/COMECE-contribution-and-annex_paper_for_the_public_consultation_on_the_White_Paper_on_AI-final.pdf.

[76] The European Commission established a High- Level Expert Group on Artificial Intelligence (HLEG AI) in June 2018 composed of 52 members from the academe, civil society, and several industries. The European AI alliance, currently composed of thousands of interested parties, was also formed as a forum to discuss the outputs of the HLEG AI.

[77] Commission des Episcopats de la Communauté Européenne, "Robotisation of Life - Ethics in view of new challenges," January 2019, http://www.comece.eu/comece-publishes-reflection-on-robotisation-of-life.

proffered by robotisation such as the minimization of production and labor costs, the reduction of risks and dangers to humans, the augmentation of human skills, especially for the disabled, and others, it warns that robotisation of the everyday occurs in an atmosphere that has little regard for the limitations of the human person. The world today is driven by the maximization of the human experience sometimes at the cost of respecting the human condition. In this regard, robotisation, if not properly utilized, can become a tool in this incessant quest for maximization. The document not only advocates for ethical analyses of robotisation but also is one. The eight-page document touches upon two ethical issues.

The first ethical issue is the primacy of the person and recognition of human dignity. It poses a paradox, "the more human power over the environment increases thanks to machines, the more human beings are deprived of agency and control."[78] In other words, the more we rely on machines, the less self-reliance we have. Consider the following example: the average smartphone empowers us to acquire knowledge faster than any other point in human history. However, if an individual loses his or her smartphone, he or she experiences a sense of helplessness, insecurity, and "nakedness." The document calls for a positive relationship between humans and machines. To this end, the document introduces the idea of "creaturality" in two points. First, just as humans are created by God, so, too, machines are created by humans. In each case, there is a dynamic relationship between creator and creation at work. The only difference is that, unlike humans, machines are not moral agents because they lack a will of their own (instrumentalist approach.) As Sinibaldi et al. state, "In Christian Ethics, moral agency is grounded [in] free actions and deliberate

[78] Ibid.

decisions taken in conscience."[79] Whereas God provides room for humans to have self-responsibility, humans are responsible for machines. Second, the governing criterion of the relationship between humans and machines is the primacy and dignity of the human person. Drawing from biblical wisdom, the document explains that humans are tasked with cultivating, developing, increasing, and preserving creation. According to the document, "[t]his dynamic sense of humanity's role in creation supports not a conservative ethics, but rather a future-oriented one which is open to and responsible for creation as it grows and develops."[80] Underlying this dynamic sense is the freedom of the human person, which implies that there must be a reasonable non-dependence on machines. The document did not speak of this, but vehicular accidents caused by phone use while driving are signs of the need to re-evaluate our relationship with machines. Ironically, this number may be reduced in the future with the use of self-driving cars, which use AI to maneuver. This possibility brings us to the second ethical issue found in the document.

COMECE also discussed the rights of robots. This discussion was prompted by the European Parliament's 2017 proposal to grant the most advanced machines a form of electronic personhood. COMECE is skeptical of using the word "persons" and its variants when referring to robots, even if used in the legal sense. It argues that the notion that robots being accorded the same personhood level as humans is in conflict with Article 6 of the Universal Declaration of Human Rights, which states that "everyone has the right to recognition

---

[79] Edoardo Sinibaldi et al., "Contributions from the Catholic Church to Ethical Reflections in the Digital Era," *Nature Machine Intelligence* 2 (May 11, 2020): 242-244.

[80] Commission des Episcopats de la Communauté Européenne, "Robotisation of Life - Ethics in view of new challenges."

everywhere as a person before the law."[81] For COMECE, authentic legal personhood is rooted in human rights and duties. In turn, human rights and duties entail responsibility, but responsibility derives from freedom. As noted above, machines lack their own will, understood in human terms, so if an advanced automated machine is programmed to do something legally and morally questionable, can it be truly responsible? If not, then its "personhood" is misleading. The document acknowledged the existence of legal personhood for entities, but this presupposes a natural person or persons behind it. Neither is COMECE convinced by the suggestion that rules for liability for robots be akin to those for animals. The document states that "This would represent a perilous shift towards the recognition of robots as [belonging] to the world of the living."[82] Assuming the European Parliament had not intended what COMECE critiques, the document issued by COMECE highlights the need for prudence in language. Whatever the European Parliament's intention was when it proposed a legal framework for AI, COMECE argues against using the word "persons" and its variants because it further complicates our understanding of personhood.[83]

The Common Good in the Digital Age

The conference **"**started with the participation of

---

[81] The United Nations, Universal Declaration of Human Rights, 10 December 1948, https://www.un.org/en/about-us/universal-declaration-of-human-rights.

[82] Commission des Episcopats de la Communauté Européenne, "Robotisation of Life - Ethics in view of new challenges".

[83] This position of COMECE was shared by more than 150 professors, researchers, and professionals in an open letter to the European Commission, though there was no collaboration with COMECE; https://www.politico.eu/wp-content/uploads/2018/04/RoboticsOpenLetter.pdf.

about 100 participants from all over the world. Experts in new technologies, [ethicists], academics and representatives of companies of digital assets gather to discuss and reflect on how the digital technologies can be for the good of human beings."[84] Some of the companies represented included Facebook, Mozilla, and LinkedIn. Pope Francis, in his speech to the participants, stated,

> If technological advancement became the cause of increasingly evident inequalities, it would not be true and real progress. If mankind's so-called technological progress were to become an enemy of the common good, this would lead to an unfortunate regression to a form of barbarism dictated by the law of the strongest.[85]

In his opening address, Cardinal Turkson refers to the "explosion of worldwide interdependence"[86] within the previous decade. He rightly calls this phenomenon "globalization." Globalization has expanded our understanding of the common good. In the past, one usually spoke of the common good in terms of community, city, and nation-state at most. Today, more often than

---

[84] "The Common Good in the Digital Age (26-28 September 2019)," Dicastery for Promoting Integral Human Development; https://www.humandevelopment.va/en/eventi/2019/the-common-good-in-the-digital-age-25-27-settembre-2019.html (accessed 1 April 2021).

[85] 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 and the Pontifical Council of Culture" 27 September 2019, Vatican Archives, https://www.vatican.va/content/francesco/en/speeches/2019/september/documents/papa-francesco_20190927_eradigitale.html.

[86] Peter A. Turkson, "Welcome Greetings," Dicastery for Promoting Integral Human Development, 26 September 2019, https://www.humandevelopment.va/content/dam/sviluppoumano/eventi/digitalage19/26-09-2019%20Common%20good%20Digital%20age%20welcome.pdf.

not, the common good is used to refer to the global community. Humanity is more interdependent than ever. It is becoming less feasible to be concerned with and responsible for the development only of our immediate locales. Globalization challenges us to broaden our vision for a just society. Our pursuit of the common good has become more rigorous with this paradigm shift. In this context, the conference highlights our collective responsibility for each other. Turkson characterizes digital technology as our tool for the common good. Digital technology not only allows us to encounter each other across vast distances, but it also allows us to foster meaningful relationships, which must underlie the common good at any level. Based on Turkson's words, integral human development is the responsibility of those developing and using digital technologies with AI.

In his introductory address, Msgr. Bruno-Marie Duffé suggests that the digital age is characterized by the dynamic transmission of knowledge.[87] Unlike in the past, the distinction between intelligence and instruments has blurred in AI. For most of history, human beings used technology to arrive at more knowledge. But with the invention of AI, humans are now also led by technology to more knowledge which was not even contemplated at the beginning. In a way, AI use has a revelatory experience to it. Issues that were not even considered initially were analyzed in the process. Yet, the notion of the common good is something that digital technology on its own can never fully capture. As Duffé notes, "The incessant movement of informative knowledge may exhaust both word and silence, both of which are

[87] Bruno-Marie Duffé, "Introductory Reflections: "'Digital Age": the Paradoxical Culture of Process and the Instantaneous,'" Dicastery for Promoting Integral Human Development https://www.humandevelopment.va/content/dam/sviluppoumano/eventi/digitalage19/MsgrBMDuffe_DigitalAge19_eng.pdf.

inherent to encounter and to ethical discernment." To avoid this scenario, the digital age must be imbued primarily by a culture of encounter rather than mediation. A genuine encounter requires patience (not to be confused with moral patiency), which only humans can exhibit. Patience is not something that can be programmed. Delay in computation can, but not patience. This patience exists "against a backdrop of digital impatience." Thus, "'the culture of encounter' allows for the reconciliation of digital impatience with the patience necessary in the process of building the common good through patient listening, comprehension and, ultimately, love for one another." In the context of Duffé's words, AI may analyze a person's data, but it cannot comprehend that person's intentionality. A robot may provide answers to someone's problems, but it cannot be compassionate with that person's struggles.

The 'good' algorithm? Artificial Intelligence, Ethics, Law, Health

The workshop gathered together around eighty participants. The term "algorethics" was developed to describe the newly emerging discipline of ethical values to be applied to the development of technology.[88] In his address to the participants of the General Assembly of the Pontifical Academy for Life, Pope Francis acknowledged that,

> The digital age is changing our perception of space, of time and of the body. It is instilling a sense of unlimited possibilities, even as standardization is becoming more and more the main criterion of aggregation. It has become increasingly difficult to recognize and appreciate differences. On the socio-economic level,

[88] The term "Roboethics" is used similarly.

> users are often reduced to "consumers", prey to private interests concentrated in the hands of a few. From digital traces scattered on the internet, algorithms now extract data that enable mental and relational habits to be controlled, for commercial or political ends, frequently without our knowledge. This asymmetry, by which a select few know everything about us while we know nothing about them, dulls critical thought and the conscious exercise of freedom. Inequalities expand enormously; knowledge and wealth accumulate in a few hands with grave risks for democratic societies. Yet these dangers must not detract from the immense potential that new technologies offer. We find ourselves before a gift from God, a resource that can bear good fruits.[89]

The most significant output of the workshop was the jointly signed document known as the "Rome Call for AI Ethics." According to its website:

> The Pontifical Academy for Life, Microsoft, IBM, FAO, the Italian Ministry of Innovation (part of the Italian Government), signed as first the "Call for an AI Ethics", a document developed to support an ethical approach to Artificial Intelligence and promote a sense of responsibility among organizations, governments, institutions and the private sector with the aim to create a future in which digital innovation and technological progress serve human genius and creativity and not their gradual replacement.[90]

---

[89] Francis, "Address Prepared by Pope Francis Read by HE Archbishop Paglia President of the Pontifical Academy of Life," 28 February 2020, Vatican Archives, http://www.vatican.va/content/francesco/en/speeches/2020/february/documents/papa-francesco_20200228_accademia-perlavita.html.

[90] "Rome Call For AI Ethics – A Human-Centric Artificial Intelligence," Rome Call For AI Ethics; https://www.romecall.org/ (accessed 31 March 2021).

The document discusses three impact areas of ethics, education, and rights. Each of these areas essentially advocates for a human-centric approach to AI. Six principles are enumerated to condition this approach. They are Transparency, Inclusion, Responsibility, Impartiality, Reliability, and Security and Privacy. While the text fills only four pages, it builds upon the two-page document of "Powers and Limits of Artificial Intelligence" issued four years previously by being a collaborative document between the Church and secular institutions, both governmental and private. As described in its website, "The idea behind [the document] is to promote a sense of shared responsibility among international organizations, governments, institutions and the private sector in an effort to create a future in which digital innovation and technological progress grant mankind its centrality."[91] In the one year from the original signing on 28 February 2020, more signatories have since joined, with invitations sent by the Pontifical Academy of Life to leaders of other world religions. It could be that inter-religious dialogue would be the next dimension of AI and the Church.

Annex to the public consultation on the White Paper on Artificial Intelligence

The seven-page annex begins by reiterating some of what the Church has said thus far on AI, e.g., that it must be human-centric and serve the common good.[92] The annex expresses disappointment on the absence of churches in the White Paper's discussion on stakeholder participation, but also expresses readiness to participate

---

[91] Ibid.

[92] Commission des Episcopats de la Communauté Européenne, "Annex to the public consultation on the White Paper on Artificial Intelligence – A European Approach".

should the EU extend a formal offer. It also reinforces COMECE's call to distinguish AI from human beings. It notes with satisfaction that the move to adopt a form of legal personality for robots was discarded, although it still cautions against applying the words "autonomy" and "behaviour" for AI, as they connote human beings. Regarding "autonomy," Sinibaldi et al. claim that it "can only be attributed to human beings, due to the underlying cognitive processes (entailing self-awareness, self-consciousness and self-authorship according to reasons and values) closely and pre-eminently identified with the dignity of human persons and human agency."[93] The annex also acknowledges that the most vulnerable group in terms of AI is children. Companies must consider carefully the effects that humanoid robots may have on them. Notably, the annex goes beyond the contents of the White Paper by commenting on AI and military systems, specifically "to ban completely autonomous armed systems without human supervision for their critical functions, and to work towards the start of international negotiations on a legally binding instrument prohibiting lethal autonomous weapon systems."[94] The annex includes a copy of COMECE's submission to the White Paper's online survey.

Developing a Christological AI Ethics

The three events and two documents presented thus far provide strong foundations for AI ethics in the Church, but an overview of the events and a reading of the documents reveal an inadequate reference to Jesus,

[93] Edoardo Sinibaldi et al., "Contributions from the Catholic Church to Ethical Reflections in the Digital Era."

[94] Commission des Episcopats de la Communauté Européenne, "Annex to the public consultation on the White Paper on Artificial Intelligence – A European Approach."

whose figure must animate any ethical discourse of the Church. It could be said that such is the case because non-Catholics were involved. Granted that was the case, how, then, would the Church consider AI in light of the narrative of Jesus? What can he teach us about AI?

The following points are suggested: With Jesus as our model, our humanity is characterized by our capacity to compassionately care for one another and not only by our personal consciousness, logical thinking, or ability to solve puzzles, which are the main gauges by which AI researchers seek to replicate the human experience in machines. All are essential for authentic human progress but to characterize humanity solely in these terms echoes the quest of humanity to "be like God, knowing good and evil" (Genesis 3:5) and the Tower of Babel. AI can stretch our knowledge capabilities, but being human is also about sharing our vulnerabilities with one another.[95] AI may perhaps replicate human thought but it cannot do so for compassion. Some AI research has argued that emotions are key components of intelligence.[96] But compassion has a suprarational dimension because it can move oneself to self-giving for another even unto death, as Jesus did on the cross.[97] While AI can undertake logical analyses, solve problems beyond human abilities, and may one day achieve a semblance of consciousness, it cannot experience compassion, which is authentically

[95] Cf. 1 Corinthians 13:2, And if I have the gift of prophecy and comprehend all mysteries and all knowledge; if I have all faith so as to move mountains but do not have love, I am nothing (NAV); see also the parable of the Good Samaritan in the Gospel of Luke (10: 30-37).

[96] Cf. Matthias Scheutz, "Artificial Emotions and Machine Consciousness," in Keith Frankish and William M. Ramsey, eds., *The Cambridge Handbook of Artificial Intelligence*, 247-266 (Cambridge: Cambridge University Press, 2014).

[97] Cf. John 15:13, No one has greater love than this, to lay down one's life for one's friends (NAV).

human as demonstrated by Jesus himself.[98] With AI research questioning what it means to be human, an AI ethics inspired by Jesus could help us understand our identity in God.

Practically speaking, AI may not replicate compassion but it can help us be compassionate like Jesus. To be like him includes upholding the truth, respecting the dignity of another, and empowering the poor.[99] As AI provides more (processing) power than ever before, Jesus, as the servant-leader, teaches us to use this power for service.[100] AI must not be used to propagate falsehoods, infringe privacy, or unjustly prejudice people because these are contrary to the teachings of Jesus.

Let us highlight an issue. Can people bring AI to extend compassion and care to people who do not matter to accounting ledgers? AI is often appreciated only for its commercial value of deriving maximum profit at the expense of people. What comes to mind is the notion of the "technocratic paradigm," which was discussed by Pope Francis in his encyclical *Laudato Si'*. According to Francis, the technocratic paradigm is "the way humanity has taken up technology and its development according to an undifferentiated and one-dimensional paradigm. This paradigm exalts the concept of a subject who, using logical and rational procedures, progressively approaches and gains control over an external object."[101] AI today is

---

[98] Cf. Matthew 15:32, Jesus summoned his disciples and said, "My heart is moved with pity for the crowd, for they have been with me now for three days and have nothing to eat. I do not want to send them away hungry, for fear they may collapse on the way" (NAV).

[99] Cf. Luke 6:21, And raising his eyes toward his disciples he said: "Blessed are you who are poor, for the kingdom of God is yours (NAV),

[100] Cf. Matthew 20:16, Thus, the last will be first, and the first will be last (NAV).

[101] Francis, "On care for our common home," *Laudato SI'* no. 106, 24 May 2015, Vatican Archives, https://www.vatican.va/

susceptible to the technocratic paradigm. Francis joins the tradition of critique against the instrumentalization of the world advanced by philosophers and theologians, such as Heidegger, Marcuse, Habermas, Boff, and Gutiérrez. The Gospel narratives are a strong foundation for the Church to contribute in this regard by showing the centrality of compassion and care in the ministry of Jesus to the poor as opposed to the tendency of prioritizing rational calculation, tradition-keeping, and legal positivism over the care of human beings. An AI ethics inspired by Jesus may help us "recover the values and the great goals swept away by our unrestrained delusions of grandeur."[102]

It might be said that Jesus is incompatible with discussions on AI ethics and that speaking about him might disenfranchise the Church from AI fora. But such concerns limit discussions. The Church must make it clear that Jesus is the model *par excellence* for the numerous Christians immersed in the digital world. AI allows humanity to reach possibilities previously unknown, but to what end? Based on Christian anthropology, humanity is created in the image and likeness of God, which means that we are to share in God's work of transforming the world according to God's plan, while also reflecting the relationality of the Godself by exercising solidarity with each other. As the *eschaton Adâm*, Jesus inspires us, especially Christians, to rethink AI to that end.[103] *Gaudium et Spes* lends basis to this notion when it states,

---

content/francesco/en/encyclicals/documents/papa-francesco_20150524_enciclica-laudato-si.html#_ftn53.

[102] Ibid, 114.

[103] Cf. Paolo Benanti, "Artificial Intelligence, Robots, Bio-engineering, and Cyborgs: New Challenges for Theology?" *Concilium* 3 (2019): 34-48.

> Let [the faithful] blend new sciences and theories and the understanding of the most recent discoveries with Christian morality and the teaching of Christian doctrine, so that their religious culture and morality may keep pace with scientific knowledge and with the constantly progressing technology. Thus they will be able to interpret and evaluate all things in a truly Christian spirit.[104]

The Second Vatican Council occurred within a decade from the early developments of AI, as described in the first part of this article. AI was not prolific enough then to warrant a comment from the Church. Yet over fifty years later, with the near-ubiquity of AI, the Council's words remain relevant. Just as the events and documents presented are efforts by the Church to "keep pace with scientific knowledge and with the constantly progressing technology," so also developing an AI ethics inspired by Jesus may be a legitimate attempt "to interpret and evaluate all things in a truly Christian spirit." These points are an invitation to include Jesus in future discussions on AI ethics.

Conclusion

As suggested in the introduction, one could think of AI as one of the contemporary "signs of the times." It offers many benefits and problems for humanity, which all require ethical considerations. The Church, in the spirit of Vatican II, acted with that in mind. From what has been presented in this article, the Church has demonstrated an open but cautious attitude towards AI. The caution of the Church towards AI should be seen as

---

[104] *Gaudium et Spes: Pastoral Constitution on the Church in the Modern World*, no. 62, Vatican Archives, https://www.vatican.va/archive/hist_councils/ii_vatican_council/documents/vat-ii_const_19651207_gaudium-et-spes_en.html.

a necessary counterbalance for a rapidly digitizing world. Aware of the science involved, the Church sought to initiate and engage in conversations on the topic. These conversations were with various secular institutions and experts from different fields, not all of whom subscribe to the Church's beliefs. Without ethical parameters, only the privileged would be secure from the problems of AI and the poor would bear the brunt of them. The Church has been active in developing AI ethics, but it must include Jesus in future discussions. By doing so, AI can be better imbued with a pastoral dimension that keeps it at the service of the poor. All things considered, there is reason to think that AI and the Church can enrich each other for the benefit of humanity.[105]

[105] The Church may benefit from creatively integrating AI in its evangelical work. A precedent of this was when the Catholic media organization *Aleteia* had developed in 2017 a Chatbot named ePaul in honor of the apostle to the gentiles for the celebration of World Communications Day. A person could post an inquiry or sentiment to the Chatbot, and in return, the Chatbot would request further information from the person. And based on the information received, the Chatbot would scour its archives and provide articles that best respond to the person's original message. Although the Chatbot has been deactivated, it was a primarily faith-orientated use of AI, albeit on a limited scale. Just as the Church integrated new technologies in its evangelical work in previous periods, so, too, it can integrate AI in the contemporary period.

Bibliography

Anderson, Michael, and Susan Anderson. "Machine Ethics: Creating an Ethical Intelligent Agent," *AI Magazine* Vol. 28, no. 4, (2007): 15-25.

Awad, Mariette, and Rahul Khanna. "Machine Learning." Essay. In Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers, 1–268. Berkeley, CA: Apress Open, 2015; https://link.springer.com/book/10.1007/978-1-4302-5990-9#about (accessed 16 June 2021.).

Commission des Episcopats de la Communauté Européenne, "Robotisation of Life - Ethics in view of new challenges, " January 2019, http://www.comece.eu/comece-publishes-reflection-on-robotisation-of-life.

Commission des Episcopats de la Communauté Européenne, "Annex to the public consultation on the White Paper on Artificial Intelligence – A European Approach," June 2020, http://www.comece.eu/dl/kmsNJKJKkKoOJqx4KJK/COMECE-contribution-and-annex_paper_for_the_public_consultation_on_the_White_Paper_on_AI-final.pdf.

Chalmers, David. "The Singularity: A Philosophical Analysis," *Journal of Consciousness Studies* 17/9-10 (2010): 7-65.

BBC. "Deepfake Queen Prompts 200-plus Complaints to Ofcom." *BBC News*. BBC, December 29, 2020. Last modified December 29, 2020; https://www.bbc.com/news/technology-55478579 (accessed 31 March 2021).

Benanti, Paolo. "Artificial Intelligence, Robots, Bio-engineering, and Cyborgs: New Challenges for Theology?" *Concilium* (2019/3): 34-48.

Brooks, Rodney. "Elephants Don't Play Chess." Robotics and Autonomous Systems 6, no. 1-2 (June 1990): 3–15; https://www.sciencedirect.com/science/article/abs/pii/S0921889005800259 (accessed 16 June 2021).

Buchanan, Bruce G. "A (Very) Brief History of Artificial Intelligence." *AI Magazine* 26, no. 4 (December 2005): 53–60.

Einhorn, Bruce, Siegfrid Alegado, and Ditas B Lopez. "Empathetic Robots Are Killing Off the World's Call-Center Industry." *Bloomberg*. Bloomberg, March 17, 2021. Last modified March 17, 2021; https://www.bloomberg.com/news/articles/2021-03-16/artificial-intelligence-chatbots-threaten-call-center-industry-human-operators (accessed 31 March 2021).

Duffé, Bruno-Marie, "Introductory Reflections: "'Digital Age": The Paradoxical Culture of Process and the Instantaneous,'" Dicastery for Promoting Integral Human Development https://www.humandevelopment.va/content/dam/sviluppoumano/eventi/digitalage19/MsgrBMDuffe_DigitalAge19_eng.pdf.

Fitzpatrick F, Doherty A, Lacey G. Using Artificial Intelligence in Infection Prevention. Curr Treat Options Infect Dis. 2020 Mar 19:1-10. doi: 10.1007/s40506-020-00216-7. Epub ahead of print. PMID: 32218708; PMCID: PMC7095094.

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 and the Pontifical Council of Culture" 27 September 2019, Vatican Archives, https://www.vatican.va/content/francesco/en/speeches/2019/september/documents/papa-francesco_20190927_eradigitale.html.

Francis, "Address Prepared by Pope Francis Read by HE Archbishop Paglia President of the Pontifical Academy of Life," 28 February 2020, Vatican Archives, http://www.vatican.va/content/francesco/en/speeches/2020/february/documents/papa-francesco_20200228_accademia-perlavita.html.

Francis, "On care for our common home," Laudato SI' no. 106, 24 May 2015, Vatican Archives, https://www.vatican.va/content/francesco/en/encyclicals/documents/papa-francesco_20150524_enciclica-laudato-si.html#_ftn53.

Garnelo, Marta, and Murray Shanahan. "Reconciling Deep Learning with Symbolic Artificial Intelligence: Representing Objects and Relations." Current Opinion in Behavioral Sciences 29 (October 2019): 17–23. https://doi.org/10.1016/j.cobeha.2018.12.010.

Gaudium et Spes. Pastoral Constitution on the Church in the Modern World. Vatican Archives, https://www.vatican.va/archive/hist_councils/ii_vatican_council/documents/vat-ii_const_19651207_gaudium-et-spes_en.html.

Göcke, Benedikt Paul. "The Ideals of Humanity in light of Synthetic Biology and Artificial Intelligence," *Concilium*, (2019/3): 25-33.

Goertzel, Ben. "Artificial General Intelligence: Concept, State of the Art, and Future Prospects." Journal of Artificial General Intelligence 5, no. 1 (2014): 1–46. https://doi.org/ 10.2478/jagi-2014-0001.

Grygiel, Wojciech P. "Artificial Intelligence and Scholastic Epistemology." *Annales UMCS Informatica AI* 3 (2005): 93–103; http://ai/annales.umcs.pl (accessed 26 June 2021).

Gugerty, Leo. "Newell and Simon's Logic Theorist: Historical Background and Impact on Cognitive Modeling." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, no. 9 (October 1, 2006): 880–884.

Gunkel, David J. *The Machine Question: Critical Perspectives on AI, Robots, and Ethics*. Camb ridge, MA: MIT Press, 2017.

Ho, Anita. "Are We Ready for Artificial Intelligence Health Monitoring in Elder Care?" *BMC Geriatrics* 20/358 (2020): 1-7.

Manzotti, Riccardo, "Embodied AI beyond Embodied Cognition and Enactivism" *Philosophies* 4/3 (16 July 2019): 1-15. https://doi.org/10.3390/philosophies4030039.

McCarthy, John. "History of Lisp." Professor John McCarthy. Stanford University, February 12, 1979. Last modified February 12, 1979; http://jmc.stanford.edu/articles/lisp.html (accessed 1 April 2021).

McCarthy, John "What Is Artificial Intelligence?" What Is Artificial Intelligence? (Stanford, 2007). http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html (accessed 1 April 2021).

Naude, Wim, and Nicola Dimitri. "The Race for an Artificial General Intelligence: Implications for Public Policy." *AI and Society* 35 (April 22, 2019): 367–379; https://link.springer.com/article/10.1007/s00146-019-00887-x (accessed 16 June 2021).

Nouri, Steve. "Council Post: How AI Is Making An Impact On The Surveillance World." *Forbes*. Forbes Magazine, December 3, 2020. Last modified December 3, 2020; https://www.forbes.com/sites/forbestechcouncil/2020/12/04/how-ai-is-making-an-impact-on-the-surveillance-world/?sh=22a93632265e (accessed 31 March 2021).

Osiński, Grzegorz. "Theological and Ethical Aspects of Mind Transfer in Transhumanism." *Scientia et Fides* 9/1 (2021): 149–176.

Pontifical Academy of Science, "Statement on Artificial Intelligence" 1 December 2016, http://www.pas.va/content/dam/accademia/pdf/scripta_varia_132.pdf.

Samuel, Arthur L. "Some Studies in Machine Learning Using the Game of Checkers," *IBM Journal of Research and Development* 44/1.2 (1959): 210–229.

Searle, John. R. "Minds, brains, and programs." *Behavioral and Brain Sciences* 3/3 (1980): 417-457.

Scharre, Paul, and Michael Horowitz. "Artificial Intelligence What Every Policymaker Needs to Know." Center for New American Security. Center for a New American Security, June 19, 2018. Last modified June 19, 2018;

https://www.cnas.org/publications/reports/artificial-intelligence-what-every-policymaker-needs-to-know (accessed 31 March 2021).

Scheutz, Matthias. "Artificial Emotions and Machine Consciousness." In Keith Frankish and William M. Ramsey, eds. *The Cambridge Handbook of Artificial Intelligence*, 247-266. Cambridge: Cambridge University Press, 2014.

Sinibaldi, Edoardo, Chris Gastmans, Miguel Yáñez, Richard M Lerner, László Kovács, Carlo Casalone, Renzo Pegoraro, and Vincenzo Paglia. "Contributions from the Catholic Church to Ethical Reflections in the Digital Era." *Nature Machine Intelligence* 2/5 (May 11, 2020): 242–244.

Skansi, Sandro. "Preface." Preface. In Introduction to Deep Learning from Logical Calculus to Artificial Intelligence, i-xiii. Springer, 2018.

Song, Yong Sup. "Religious AI as an Option to the Risks of Superintelligence: A Protestant Theological Perspective," *Theology and Science* 19/1 (October 6, 2020): 65-78.

Teich, David A. "Artificial Intelligence and Data Privacy – Turning A Risk Into A Benefit." *Forbes*. Forbes, August 10, 2020. Last modified August 10, 2020; https://www.forbes.com/sites/davidteich/2020/08/10/artificial-intelligence-and-data-privacy--turning-a-risk-into-a-benefit/?sh=1c79e43c6a95 (accessed 31 March 2021).

The United Nations, Universal Declaration of Human Rights, 10 December 1948, https://www.un.org/en/about-us/universal-declaration-of-human-rights.

Turing, Alan M. "Computing Machinery and Intelligence." *Mind* 59/236 (October 1950): 433–460.

Turkson, Peter A.,"Welcome Greetings," Dicastery for Promoting Integral Human Development, 26 September 2019, https://www.humandevelopment.va/content/dam/sviluppoumano/eventi/digitalage19/26-09-2019%20Common%20good%20Digital%20age%20welcome.pdf.

Willingham, AJ. "A New Program Can Animate Old Photos. But There's Nothing Human about Artificial Intelligence." *CNN Business*. CNN Worldwide, March 5, 2021. Last modified March 5, 2021; https://edition.cnn.com/2021/03/02/business/deep-nostalgia-myhertiage-ai-learning-trnd/index.html (accessed 31 March 2021).

Winston, Patrick Henry. "Marvin L. Minsky (1927–2016)." *Nature* 530/7590 (February 18, 2016): 282.

Yampolskiy, Roman, and Joshua Fox. "Safety Engineering for Artificial General Intelligence." *Topoi* 32/2 (August 24, 2012): 217–26. https://doi.org/10.1007/s11245-012-9128-9.

"DeepFake." In the Cambridge Dictionary. Cambridge University Press; https://dictionary.cambridge.org/us/dictionary/english/deepfake (accessed 23 June 2021).

"History of Artificial Intelligence." Artificial Intelligence. Council of Europe, n.d.; https://www.coe.int/web/artificial-intelligence/home (accessed 31 March 2021).

"Rome Call For AI Ethics – A Human-Centric Artificial Intelligence" Rome Call For AI Ethics' https://www.romecall.org/ (accessed 31 March 2021).

"The Common Good in the Digital Age (26-28 September 2019)." Dicastery for Promoting Integral Human Development; https://www.humandevelopment.va/en/eventi/2019/the-common-good-in-the-digital-age-25-27-settembre-2019.html (accessed 1 April 2021).