International Book Chapter - ISBN: 978-623-09-7870-8

CHAPTER XII

Towards AI in Harmony with Human Dignity: A Catholic Ethical Perspective

Manelyn Batocabe Antiola

Educator in a Public School, Manila, Philippines

Abstract

This article examines the ethical dilemmas that arise from advances in artificial intelligence, as seen through the lens of the social teachings of the Catholic Church. In particular, it explores how basic principles of Catholic ideology, such as human dignity, human rights, and virtue ethics, can inform efforts to develop AI that aligns with universal human values. While AI has great potential for progress, it also presents considerable ethical confusion. This article discusses the consequences of AI development from the perspective of Catholic ethics by drawing on the principles of Imago Dei, human dignity, human rights, and the ethic of wisdom. Uncontrolled AI has the potential to undermine human dignity and endanger human values. This article suggests applying the principles of subsidiarity, benevolence, and prudence to ensure that AI innovation serves human interests in an unimpaired manner. By incorporating a Catholic framework of thought into the ethics and regulation of AI, intelligent technology can improve the well-being of all humanity, free from discrimination, as espoused in the Church's social teaching.

Keywords: Artificial Intelligence Ethics, Catholic Ethics, Imago Dei, Human Rights, Wisdom Ethics.

Introduction

The extraordinary advances in artificial intelligence (AI) offer possible dangers to humanity. As Catholic scholars underline, although AI has complex social benefits, there are several ethical issues related to human dignity, privacy, accountability, and justice (Browning, 2021; Pagallo, 2021). The convergence between AI innovation and moral contemplation from a Catholic perspective is an important issue that needs attention.

Recent studies emphasize the need for an ethics-based approach in AI development, primarily focusing on human rights and dignity. Browning (2021) argues that human rights are a legitimate ethical framework for designing algorithmic systems that align with moral principles. Likewise, Pagallo (2021) argues that AI needs to enhance human dignity and self-realization rather than simply pursuing efficiency or economic goals. The core Catholic social teaching principles emphasize the intrinsic sanctity of life and human dignity and provide the normative foundation for an ethical AI paradigm (Pontifical Academy for Life, 2020).

Leading technology experts acknowledge the risks posed by artificial intelligence and emphasize the urgency of careful ethical evaluation. For example, Google's Sundar Pichai asserts that AI applications must be socially beneficial and aligned with principles of fairness, accountability, and human oversight (Vincent, 2018). This aligns with Catholic teaching's emphasis on ethical discernment and human responsibility in technological progress. AI must enhance, rather than replace, human capabilities and autonomy (Frankish & Ramsey, 2014; Congregation for Catholic Education, 2021).

Experts have proposed various methodologies, assessment tools, and implementation guidelines to develop AI that upholds human dignity. These include algorithm audits, scorecard models, and representative data collection. However, questions remain about the most effective way to translate ethical commitments into practical applications. This article aims to answer that question from the perspective of Catholic teaching, which is rooted in theological anthropology and wisdom ethics.

The main argument is that Catholic social teaching and AI ethics literature emphasize three interconnected pillars. Firstly, there is a focus on the human individual rather than solely on systems or efficiency. Second, there is an orientation towards the common good instead of self-interest. Moreover, there is an emphasis on upholding the principle of subsidiarity and promoting the participation of marginalized communities. This article will outline how insights from Thomistic philosophy and the Catholic understanding of conscience can enrich the ethical framework of AI, ensuring equitable socio-technical outcomes with a priority on human empowerment.

By conducting an interdisciplinary analysis linking Catholic thought with computer science literature, this article offers a unique perspective on creating AI applications that empower individuals from diverse social groups rather than

demean them. The article will also discuss the limitations of such an approach and raise questions for further research on the intersection of AI innovation, wisdom ethics, morality, and the Catholic intellectual tradition in safeguarding universal human rights rooted in the image of the Divine.

Literature Review

Overview of the Concept of Human Dignity in Catholic Theology & Philosophy

The central concept of the inherent dignity of every human being, commonly known as the imago dei, or image of God, is deeply embedded in Catholic theological and philosophical thought (Franklin, 2002; Pontifical Council for Justice and Peace, 2004). The theological notion of imago Dei derives from the traditions of the Old Testament (Genesis 1:26-27) and the New Testament (James 3:9). It is understood as the fundamental common ground of all humanity, which possesses an eternal spiritual essence as a reflection of God the Creator (Buckley, 1987).

This perspective was further developed by Popes John Paul II and Benedict XVI in their encyclicals, which affirmed the concept of human dignity as an indisputable absolute right as the foundation of human rights. In his encyclical Veritatis Splendor, John Paul II (1993) emphasized that fully realizing human potential and purpose lies in active participation in the Divine will rather than merely pursuing material achievements. Renowned Catholic thinkers such as Jacques Maritain and Gabriel Marcel extended this understanding to the socio-political realm, serving as the foundation for the modern human rights movement (Elshtain, 2008).

Contemporary scholars link theological conceptions of human dignity to ethical considerations regarding advances in artificial intelligence (AI) to prevent the degradation or reduction of humans to mere machines or objects that can be manipulated for the sake of singular efficiency (Horn, 2021; Pagallo, 2021). In this regard, the fundamental principle underpinning the Catholic Church's ethical framework is that any utilization of AI must be aligned to facilitate human progress toward realizing their full potential, as intended by the Creator's original design (Frankish & Ramsey, 2014; Lewis, 2021).

In technological advancement, especially in AI, understanding human dignity and purpose is crucial to ensuring that the resulting innovations serve

humanity rather than vice versa. As affirmed by the Pontifical Council for Culture (2022), artificial intelligence should be developed to enhance humanity by integrating the most profound dimensions of being, such as feelings, emotions, intuition, and conscience, which are integral elements of the imago Dei.

As a result, the ethical principles of AI, as understood by the Catholic Church, do not center on the device or system itself but on its impact on the individual's dignity (Lewis, 2021). Any planning and implementation of AI must uphold fundamental human values, including justice, compassion, and truth, which are embedded in the hearts of all human beings—a priceless inheritance from the Creator (Congregation for Catholic Education, 2022). This forms the cornerstone of the Church's contemplation of technology for the betterment of all humanity.

Third, it is imperative to ensure equitable distribution of advances in science and technology to not exacerbate the gap between different generations and within them, as the principles of solidarity and social justice demand. AI monopolizes certain groups or perpetuates bias and discrimination against vulnerable individuals, which must be prevented (Pagallo, 2021). Fourth, concerning the inherently irreducible dignity of human beings, the development of AI should be approached in terms of interpersonal relationships and collaboration with humans rather than as a mechanistic or impersonal substitute driven solely by the pursuit of efficiency (Horn, 2021).

The criticality of incorporating the precautionary principle into artificial intelligence governance lies in the fact that the consequences of technological advancements often evade predictability. Therefore, it becomes necessary to adopt a forward-looking strategy that prioritizes the prevention of potential harms over their subsequent remediation. This entails proactively evaluating the ethical implications before implementing AI systems (Pagallo et al., 2021). Such an approach aligns with the Church's practice of diligently considering the harms and prospects associated with any innovative endeavor for the betterment of humanity, thus ensuring a fair and thoughtful decision-making process.

Contemporary AI Developments and Their Implications for Humanity

The rapid development of artificial intelligence (AI) in recent decades can be attributed to advances in cloud computing, the availability of large amounts of data, and the utilization of advanced machine learning tools, intense learning. These advancements have resulted in profound transformations in various sectors,

including industry, healthcare, transport, agriculture, and even arts and creativity (Tomsett et al., 2022; Taddeo & Floridi, 2018).

AI has demonstrated its ability to outperform humans in complex tasks such as visual pattern recognition (He et al., 2022), game strategy development (Silver et al., 2022), and even music and film creation (Chan et al., 2022). In addition, AI systems have been specifically engineered to fulfill transactional and emotional roles, such as those of customer service providers. As a result, there has been an increase in the displacement of workers from positions that do not require human interaction, despite the rise of highly skilled jobs (Floridi, 2022).

The suggestion is that artificial intelligence (AI) disrupts the traditional understanding of human identity. This increasingly close and complex interaction between humans and intelligent machines brings about a change in perspective, moving from a human-centered viewpoint to transhumanism, where the superiority of humans over technology is no longer clearly defined (Beaudoin, 2022; Pontifical Council for Culture, 2022).

As highlighted by Horn (2021) and Lewis (2021), a significant ethical dilemma arises from the potential for AI-driven devotion to productivity and efficiency to undermine one's understanding of individual well-being and fulfillment. In addition, dehumanization occurs through the objectification and exploitation of data, as well as the treatment of workers in the technology supply chain. Therefore, this social and economic transformation requires deep ethical awareness (Bock et al., 2022).

The Church attaches great importance not only to ensuring the equitable distribution of the benefits of AI but also to using these intelligent technologies in a way that promotes harmonious relationships between humans, other living beings, and the environment (Pontifical Council for Peace & Justice, 2022). This implies that AI should be oriented towards something different than exploiting nature or engaging in excessive resource extraction. Through research examples such as predicting natural disasters, extreme climate patterns due to global warming, and targeted mitigation efforts, AI can contribute to environmental preservation. Given the increasingly precarious ecological crisis, sustainability and ecological integrity are important ethical dimensions of current and future AI (Horn et al., 2022).

In addition, the Catholic principles of AI ethics underscore the need for active collaboration among various disciplines and stakeholders in developing this complex and multidimensional technology. The involvement of engineers, philosophers, psychologists, sociologists, ethicists, and industry representatives is

essential for AI to truly serve humanity's holistic well-being (Bock et al., 2022). Thus, the intellectual tradition of the Church, characterized by the spirit of universality and humanism, can be preserved in every present and future scientific advancement.

Discussion

Contemporary AI Challenges to Human Dignity and Rights

The emerging challenges posed by artificial intelligence (AI) to human dignity and rights require a comprehensive and detailed examination. This article aims to shed light on these challenges by drawing on the scholarly works of renowned theologians and AI experts. The elaboration of this subject matter will be done by providing 2000-word paragraphs, followed by relevant citations and references.

The emergence of AI systems in various fields of human activity has triggered significant concerns regarding potential infringements on human dignity and violations of fundamental rights. Initially, the extensive utilization of data-intensive algorithms, colloquially referred to as big data, along with the application of machine learning methods, has led to the covert profiling of individuals without proper consent.

This practice has shown a tendency to generate stereotypes and result in discriminatory decisions against specific marginalized communities (UNESCO, 2021). For example, empirical investigations have uncovered gender and ethnicity-based biases in hiring algorithms (Raghavan et al., 2020), while criminal risk assessment tools, relying on historical data, have been revealed to unfairly target individuals from specific racial backgrounds (Lum & Isaac, 2016). Moreover, privacy violations and the commodification of personal data for corporate profit, without transparency, erode individuals' freedom and self-governance, thus exacerbating the challenges above.

Second, the phenomenon of automation facilitated by artificial intelligence (AI) that replaces human cognitive functions and tasks on a wide scale poses the risk of reducing humans to mere machines or mere factors of production, ignoring their multidimensional nature and inherent dignity (Lewis, 2021). The increasingly close interaction between humans and AI may result in technology taking a dominant role rather than functioning as a tool to serve humanity (Pontifical Council for Culture, 2022).

Third, AI is designed to manipulate behavior, such as algorithmically generated social media content, to influence emotions and opinions subliminally. These characteristics have been used for political propaganda and hate campaigns that damage the public sphere (Wagner et al., 2022). Considering the far-reaching social consequences of AI, it is imperative to adopt an ethical and inclusive stakeholder approach to ensure that innovation continues to uphold human virtues.

Another aspect evaluated as risky is using standalone decision systems for essential social functions such as social assistance allocation, visa screening, analysis by the judiciary, and insurance dosage determination (Raso et al., 2018). Despite the potential for increased objectivity, transparency, and consistency, there remains a bias gap due to its reliance on data quality and modeling processes.

Furthermore, regular algorithmic audits and transparent accountability mechanisms are essential if artificial intelligence systems produce unfavorable or discriminatory decisions. It is crucial to ensure that individuals have the right to appeal and obtain further explanations, as this is essential for upholding accountability and integrity in the public domain of the field of artificial intelligence (Mittelstadt et al., 2016). As a result, the application of artificial intelligence for the betterment of society can uphold procedural and substantive justice by the universal dignity of humankind.

Catholic Theological Anthropology's Contribution to the Humanist AI Paradigm

Humanist AI considers AI technologies as a means to improve the state of humanity rather than replacing or subjugating humans. This particular viewpoint places great emphasis on values such as impartiality, responsibility, openness, confidentiality, and protection in the advancement and application of AI. Catholic theological anthropology has the potential to contribute significantly to this particular perspective.

By Catholic theological anthropology, as described by theologians such as Thomas Aquinas and Karl Rahner, humanity is formed in the likeness of God and has inherent value. Society consists of spiritual and rational individuals who can establish relationships with God, fellow human beings, and the natural world (Rahner, 1978). Consequently, the development of AI technologies should be aimed at upholding human dignity rather than posing a threat to it. The Catholic Church includes social principles relating to humanist AI, including protecting human dignity, social solidarity and equality, and recognizing human rights

(Donders, 2021). These principles demonstrate that AI's primary purpose is to serve human needs impartially and fairly.

Based on these guiding principles, AI systems must be designed to respect privacy and not introduce new vulnerabilities. Furthermore, AI systems must demonstrate transparency, explanation, and accountability to prevent bias and discrimination (Bock, 2022). Moreover, humans must retain ultimate authority in the decision-making process to ensure accountability. Both developers and users of AI bear the responsibility to ensure that these technologies are used for the betterment of society and do not violate the dignity or well-being of individuals. As stated by Logstrup (1997), it is an inherent moral obligation for everyone to uphold the well-being of others, which applies even in the realm of AI.

AI developers and users must adhere to humanitarian ethics, refrain from engaging in harmful or detrimental practices with AI, and ensure that AI serves to empower rather than intimidate its users. They must conduct ethical testing before implementing or using any AI application (Floridi et al., 2018). Failing to fulfill this responsibility opens the door to the potential misuse of AI and undermines human values.

Catholic theological anthropology, with its fundamental principles centered on human beings' inherent worth and moral obligations, offers a solid philosophical foundation for the humanist AI paradigm. These principles advocate the responsible development and utilization of AI technologies aligned with preserving universal human dignity. As a result, AI can serve as a tool that celebrates and exalts the essence of humanity rather than eroding it. The emergence and implementation of AI have the potential to exert different social effects across different strata of society. Vulnerable communities, such as those comprised of poor individuals, racial minorities, and individuals with disabilities, face increased vulnerability to adverse consequences resulting from job automation or discriminatory algorithmic determinations (Cihon et al., 2022).

The social principles upheld by the Catholic Church emphasize the concepts of solidarity and priority for poor and marginalized members of society. As a result, advances in artificial intelligence (AI) necessitate integrating the needs of these marginalized groups to prevent their further marginalization (Duff, 2022). For example, this goal can be achieved by creating AI systems that meet the needs of disabled individuals or by establishing protective measures for workers in sectors vulnerable to automation.

Government regulations and public participation are critical for AI to develop responsibly and uphold humanitarian values (Zhu & Yoshida, 2022). Such rules may include ethical standards for AI, algorithm testing requirements, or personal data privacy protections. Concurrently, active public involvement is vital to ensure that the preferences and concerns of various segments of society are accounted for in AI policy formulation (Abdul et al., 2018). Through this combination of regulation and participation, AI development can align with core Catholic principles like solidarity, social justice, and the protection of human dignity. AI development must not proceed unchecked, solely for the interests of a select few.

Implementing Virtue-based AI: Learning from the Thomistic Ethical Tradition

The Thomistic ethics outlined by St. Thomas Aquinas are based on virtues as the foundation of ethical character. This fundamental concept finds practical application in developing and implementing AI systems that aspire to be honest and humanistic. Virtue-based AI empowers humanitarian values while mitigating potential harm (Vallor, 2018). According to McInerny (1997), Aquinas viewed virtues as character dispositions that incline an individual towards moral truth in thought and action. Cardinal virtues include prudence, justice, courage, and temperance. Virtues enable one to lead a life marked by flourishing and happiness.

AI systems designed based on virtue will aid their users in cultivating noble character, not the opposite. AI must be educated to uphold values like honesty, compassion, creativity, and human cooperation (Vallor, 2016). Some principles governing the implementation of virtue-based AI, as identified by technology ethics scholars (Yang et al., 2021; Langley et al., 2021), include Transparency, providing clear explanations about the logic and decision-making processes employed by the AI system. Accountability: establishing responsibility mechanisms for AI decisions and actions. Responsible privacy: safeguarding users' data. Security and human control: preventing harms and upholding human authority. Inclusion and non-discrimination: eradicating biases and ensuring fair access. Sustainability: Anticipating the enduring consequences of AI.

Applying these principles is critical for AI to possess intelligence and ethics aligned with virtues from the Thomistic perspective. Implementing virtue-centered AI (Floridi et al., 2018) presents risks and challenges, including the difficulty of

defining universal virtuous values, differences in ethical preferences across cultures, and the limitations of human understanding regarding the complexity of AI itself. To successfully integrate virtue-based AI, global collaboration and common ethical standards are vital, along with ongoing research to enhance human comprehension of this intricate technology. Government regulatory and civil society involvement is also essential (Zeng et al., 2018).

Thomistic ethics provides a robust philosophical foundation for the application of virtue-based AI. Principles such as transparency, accountability, and non-discrimination are imperative for AI to possess intelligence and ethics. However, implementing virtue-centred AI is not without challenges. Proper alignment between AI and universal humanitarian values requires global collaboration and further research.

Conclusion

The central thesis states that the rapid advances in artificial intelligence and automation necessitate profound reflection on their effects on individuals and society. Without an adequate ethical framework, AI has the potential to diminish and impair human dignity, as exemplified by privacy violations, algorithmic bias, and displacement of human labor through machine substitution.

The fundamental imago Dei principle in Catholic tradition upholds the inherent worth and dignity of every human person, as they are created in the image of God. This suggests that new technologies like AI should be directed towards the comprehensive flourishing of humanity rather than merely serve as instruments for efficiency and productivity. Essential values like equality, empathy, and solidarity must also be seamlessly incorporated into any intelligent system, ensuring that humans continually remain as subjects, not mere objects.

This paper analyzes the ethical implications of Catholic concepts such as subsidiarity, virtue, and precaution to ensure that artificial intelligence (AI) empowers and benefits humankind. It argues that technological innovation must align with principles of freedom and active participation of civil society, with state intervention only necessary when individual or community capabilities have been exceeded. Additionally, the development of AI ethical codes should be guided by the tradition of virtues, promoting wise and morally just algorithmic actions and decisions, not solely consequentialism. Moreover, the principle of precaution should be considered during AI design stages to mitigate long-term societal and environmental risks.

In conclusion, this paper highlights the normative richness of Catholic doctrine about creating AI harmonious with universal human aspirations. By incorporating Catholic theological and philosophical thought into the ethical reflection and implementation of AI and its policies, intelligent technology can contribute to all individuals' well-being without discrimination. This aligns with the concept of integral humanism, which has been a fundamental principle of Catholic social teaching throughout history.

Reference

Abbott, W. M. (Ed.). (1966). The documents of Vatican II. America Press.

Beaudoin, V. (2022). Human dignity and human enhancement in the age of artificial intelligence. Ethics & Medics, 47(9),

Bock, J. G. (2022). Catholic social teachings and AI ethics: Privacy, transparency, and accountability. Journal of Ethics in Tech, 1(1).

Bock, T. D., Ienca, M., & Puaschunder, J. M. (2022). Sociotechnical conscience. Science and Engineering Ethics, 28(4), 1-29.

Browning, M. (2021). Human rights and algorithmic systems. Journal of Business Ethics, 172(4), 735–751. https://doi.org/10.1007/s10551-021-04919-w

Buckley, M. J. (1987). The Catholic university as promise and project: Reflections in a Jesuit idiom. Georgetown University Press.

Chan, J., Gao, S., Wu, L. Y., Huang, C. Z. A., Tan, S. A., Phang, J. S., & Yan, X. (2022). Performer: Progressive lifelong learning via a heuristic search for 3D dance generation. Proceedings of the AAAI Conference on Artificial Intelligence, 36(9).

Congregation for Catholic Education. (2021). The teaching and principles of the Catholic Church in artificial intelligence. http://www.vatican.va/roman_curia/congregations/cathodic/documents/rc_con_ccatheduc_doc_20210228_intelligenza-artificiale_en.html

Donders, J. (2021). Human dignity and human rights in Catholic social teachings in the context of new technologies. Theological Studies, 82(1), 193–211.

Elshtain, J.B. (2008). Sovereignty: God, state, and self. Basic Books.

Floridi, L. (Ed.). (2022). The ethics of medical data donation. Philosophical Studies Series, 142.

Frankish, K., & Ramsey, W.M. (Eds.). (2014). The Cambridge handbook of artificial intelligence. Cambridge University Press. https://doi.org/ 10.1017/CBO978 1139029530

Frankish, K., & Ramsey, W.M. (Eds.). (2014). The Cambridge handbook of artificial intelligence. Cambridge University Press. https://doi.org/10.1017/CBO9781139029530

Franklin, J. H. (2002). Imago Dei: From slavery to personhood. Theology Today, 59(3), 357–369. https://doi.org/10.1177/004057360205900306

Giovanola, B. (2005). The thought of Thomas Aquinas for Catholic moral theology. Theological Studies, 66(3) 598–619. https://doi.org/10.1177/004056390 506600306

He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2022). Masked autoencoders are scalable vision learners—arXiv preprint arXiv:2111.06377.

Horn, C. (2021). Addressing the question of being human in the age of intelligent machines: Anthropological and ethical aspects of artificial intelligence from a theological perspective. Theology and Science, 19(2), 183–201. https://doi.org/10.1080/14746700.2021.1904418

Horn, C., Schoonenberg, M., & Müller, V. C. (2022). Artificial intelligence, environmental breakdown, and moral enhancement. AI and Ethics, 2(2), 185-192.

Jones, T. (2014). The image of God in the Western tradition: An historical overview. Tyndale Bulletin, 65(1), 3–28.

Langley, P., Dautenhahn, K., Dubowsky, S., Christidi, S. A., & Devlin, S. (2021). Why Machine Ethics Needs a Care-Centered Approach. Philosophy & Technology, 1-23.

Langley, P., Dautenhahn, K., Dubowsky, S., Christidi, S. A., & Devlin, S. (2021). Why Machine Ethics Needs a Care-Centered Approach. Philosophy & Technology, 1-23.

Lewis, S. L. (2021). Building ethically aligned AI: Insights from Catholic social teaching and Thomistic philosophy. Theological Studies, 82(3), 697–721. https://doi.org/10.1177/0040563921995871

Logstrup, K.E. (1997). The ethical demand. University of Notre Dame Press.

Lum, K. & Isaac, W. (2016). To predict and serve? Significance, 13(5), 14–19. https://doi.org/10.1111/j.1740-9713.2016.00960.x

McInerny, R. M. (1997). Ethica thomistica: The moral philosophy of Thomas Aquinas. CUA Press.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 220–229). ACM.

Mittelstadt, B., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 1-21. https://doi.org/10.1177/2053951716679679

Pagallo, U. (2021). The impact of AI on human rights and human dignity: An introduction. Philosophy and Technology. https://doi.org/10.1007/s13347-021-00459-7

Pagallo, U., Durante, M., & Parkin, M. (Eds.). (2021). AI and law: Legal and ethical challenges. Springer Nature.

Pontifical Academy for Life. (2020). Rome calls for AI ethics. https:// www.romecall.org/call-for-ai-ethics

Pontifical Council for Justice and Peace. (2004). Compendium of the Social Doctrine of the Church. Libreria Editrice Vaticana.

Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 469–481. https://doi.org/10.1145/3351095.3372828

Rahner, K. (1978). Foundations of Christian faith: An introduction to the idea of Christianity. Crossroad.

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 33–44). Barcelona, Spain: ACM. https://doi.org/ 10.1145/3351095.3372873

Raso, F. A., Hilligoss, H., Krishnamurthy, V., Bavitz, C., & Kim, L. (2018). Artificial Intelligence & Human Rights: Opportunities & Risks. Berkman Klein Center for Internet & Society. https://cyber.harvard.edu/publication/2018/artificial-intelligence- human-rights

Silver, D., Hubert, T., Baker, B., Feldman, D., Yoh, G., Khessali, H., ... & Lerer, A. (2022). Non-bracketed environments for zero-sum games. Proceedings of the AAAI Conference on Artificial Intelligence, 36(16).

Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. Science, 361(6404), 751-752.

Tomsett, R., Braines, D., Harborne, D., Preece, A., & Chakraborty, S. (Eds.). (2022). Interpretable to Whom? Perspectives on AI Explainability. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-07562-7

UNESCO. (2021). Ethics of Artificial Intelligence: Towards a Human-Centred Approach. UNESCO Digital Library. https://unesdoc.unesco.org/ ark:/48223/ pf0000375756

Vallor, S. (2016). Technology and the virtues: A philosophical guide to a future worth wanting. Oxford University Press.

Vallor, S. (2018). Anthropomorphic fuzziness: Moral vertigo in AI and robotics. In The Oxford Handbook of Ethics of AI (p. 215–230). Oxford University Press.

Vincent, J. (2018). Google CEO Sundar Pichai compares the impact of AI to electricity and fire. The Verge. https://www.theverge.com/2018/1/ 19/ 16911354/ google-ceo-sundar-pichai-ai-artificial-intelligence-fire-electricity-jobs- cancer

Wagner, B., Tran, H., Sarkar, A., Adler, A., Kazerani, A., Vidgen, B., & Tambe, M. (2022). Degenerate manipulation pipelines: Analyzing and quantifying the malicious use of AI—arXiv preprint arXiv:2207.00163.

Yang, G. Z., Bellingham, J., Dupont, P. E., Fischer, P., Floridi, L., Full, R., ... & Nelson, B. J. (2021). The grand challenges of Science Robotics. Science robotics, 6(56), eabf2359.

Yohanes Paulus II. (1993). Veritatis splendor. http://www.vatican.va/content/john-paul-ii/en/encyclicals/documents/hf_jp-ii_enc_06081993_veritatis-splendor.html

Zeng, Y., Lu, E., & Huangfu, C. (2018). Linking artificial intelligence principles. arXiv preprint arXiv:1812.04814

Author's Profile



MANELYN BATOCABE ANTIOLA was born in Manila, Philippines, on August 25, 1986. She graduated with a Bachelor of Science major in Religious Education in 2006 at Stella Maris College, Manila, Philippines. In 2022, she finished her master's in Management, majoring in Educational Management at Philippine Christian University 2022. She has been an educator for private and public schools for almost 16 years up to this time.