

Izvirni znanstveni članek/Article (1.01)

Bogoslovni vestnik/Theological Quarterly 86 (2026) 2, 229—240

Besedilo prejeto/Received:11/2024; sprejeto/Accepted:08/2025

UDK/UDC: 004.89: 17.026.4

DOI: 10.34291/BV2026/02/Pabubung

© 2026 Pabubung, CC BY 4.0

*Michael Reskiantio Pabubung*¹

The Indefensibility of Algorithmic Bias: A Human Dignity Perspective on Artificial Intelligence

Neutemeljenost algoritemske pristranskosti: človeško dostojanstvo kot perspektiva na področju umetne inteligence

Abstract: AI-based (social) identification technology, which relies on machine learning performance supported by algorithmic complexity to achieve desired outcomes, has permeated humans' daily lives. On the one hand, this technology offers significant benefits, particularly in enhancing operational efficiency and facilitating the analysis of large volumes of data. On the other hand, however, it presents a severe ethical challenge, as AI technology operates through algorithms that have not been free from bias. This study critically examined the ethical issues of discrimination within the discourse on human dignity, particularly in the context of biased algorithmic applications from a (Catholic) moral theologian's perspective. This article is intended to reveal that (a) algorithmic bias has the potential to disproportionately affect vulnerable and minority groups, and (b) discrimination resulting from algorithmic bias is exceedingly difficult to eliminate, thereby necessitating a discourse grounded in the infinite dignity of every human being.

Keywords: AI debate, ethical AI, equal status, infinite dignity, original datum

Povzetek: Tehnologija (socialne) identifikacije, ki temelji na umetni inteligenci in se za doseganje zelenih rezultatov opira na učinkovitost strojnega učenja, podprto z algoritmično kompleksnostjo, je prodrla v vsakdanje življenje ljudi. Po eni strani ta tehnologija ponuja pomembne prednosti, zlasti pri izboljšanju operativne učinkovitosti in olajšanju analize velikih količin podatkov. Po drugi strani pa predstavlja resen etični izziv, saj tehnologija umetne inteligence deluje prek algoritmov, ki niso brez pristranskosti. Prispevek v okviru razprave o človekovem dostojanstvu kritično obravnava etična vprašanja diskriminacije, zlasti v kontekstu pristranskosti algoritmičnih aplikacij z vidika (katoliške) moralne teologije. Namen je razkriti, da (a) algoritmična pristranskost lahko nesorazmerno vpliva na ranljive in manjšinske skupine; (b) je diskriminacijo, ki izhaja

¹ M. Phil., Universitas Atma Jaya Yogyakarta, Indonesia, michael.pabubung@uajy.ac.id.

iz algoritmične pristranskosti, izjemno težko odpraviti, zato je potrebna razprava, utemeljena na neodtujljivem dostojanstvu vsakega človeka.

Ključne besede: razprava o umetni inteligenci, etična umetna inteligenca, enakopravni status, neodtujljivo dostojanstvo, izvirni podatki

1. Introduction

When the traces of an individual's life are captured as "data" and fragmented into components that become part of a database or big data system, a significant consequence emerges: the creation of correlations (Williams et al. 2018). Within these databases or systems, personal characteristics are unveiled, often coalescing to form discernible patterns. These patterns typically manifest in two ways: correlations within an individual's own data and correlations with the data of others. In the first instance, machines can identify recurring events in a person's life, enabling them to extrapolate and predict future occurrences based on historical trends. The second type of pattern-correlation with the data of others raises even more profound implications. Shared attributes such as names, skin color, purchasing habits, or watching preferences may lead to predictive assumptions. Big Data and databases that house extensive societal records are deeply entrenched in such correlations (2018). Machine learning algorithms thrive on these patterns, training themselves to derive conclusions from historical data to classify or assess new individuals. This process, however, carries significant ethical and practical risks (Ghasemaghaei and Kordzadeh 2024; Joseph 2025).

Data plays a crucial role in both government and business sectors. Major corporations, particularly big-tech companies, are heavily dependent on data, which serves as the fuel that powers the mechanisms of these megatech companies. Governments, too, require data to advance the welfare goals of their citizens (OECD 2019). The data collected becomes a vital tool for decision-making at the state level. However, concerns arise when this data is intertwined with the interests of governments and businesses in ways that infringe upon individual rights (Tilmes 2022), since both governments and corporations have stakes in data (Martin 2022; Zhao 2024).

However, discussions around (massive) data raise significant questions about issues such as data breaches (Schlackl et al. 2022) and discrimination (Gillis and Spies 2018; Joseph 2024; Shams et al. 2025). However, data alone does not determine outcomes; the algorithms processing this data play a crucial role in shaping decisions. The ways in which algorithms categorise, interpret, and weigh data can amplify existing biases or introduce new forms of discrimination. Algorithmic bias, for instance, manifests in areas such as recruitment (Tilmes 2022), financial services, healthcare (Johnson 2022; Kumar et al. 2024), judicial administration, criminal cases, socio-political life or international affairs between Global South and North (Hasan 2023), and many other programs reliant on algorithmic perfor-

mance particularly related to gender and race (Yang 2025).

According to Girasa (2020), bias can be categorised into four types: sample bias, stereotype bias, systemic value distortion, and model insensitivity (Girasa 2020). In the first place, sample bias occurs when the sample statistics fail to accurately represent the target value or, at the very least, fail to capture the average value of the studied object. Secondly, stereotype bias arises when assumptions are made about particular characteristics, uniqueness, advantages, privileges, or qualities attributed to specific groups of people. In many cases, this kind of bias has its roots in social institutions, practices and attitudes which originate in subcultures, society at large, private or public institutions or organizations (Friedman and Nissenbaum 2022). Third, systemic value distortion occurs when a system limits choices to binary options – good or bad, right or wrong, or beautiful or ugly – without allowing for middle-ground possibilities. Lastly, model insensitivity refers to bias that occurs mathematically due to the inability to fairly partition data (Girasa 2020). Model insensitivity is particularly prevalent in the era of big data, where achieving “data fairness” is nearly impossible. Algorithms favour larger datasets and discriminate against those who do not contribute data or are underrepresented (Manzoor et al. 2024; OECD 2024).

One significant ethical dilemma posed by self-driving cars arises in emergency situations (Gilli et al. 2019). On the road, autonomous vehicles face immense challenges in making “human decisions” when confronted with life-and-death scenarios -such as whether to swerve left and it will hit a child or veer right, resulting in a 50% probability of hitting a married couple. But what if the dilemma involves a famous celebrity and an ordinary individual? Given the available data and algorithms, how would an autonomous vehicle’s machine learning system decide? The action would inevitably depend on the underlying algorithm and data, which might factor in gender, age, probability, the severity of potential fatalities, an individual’s societal contributions, and even their social status, all derived from the accessible data. Turner-Lee (2018) explains that AI can have varying levels of accuracy for different demographic groups. As a result, an algorithm can exhibit bias, make different decisions when applied to different populations (Heilweil 2020). In other words, AI algorithms can be highly discriminatory (Zowghi and Bano 2024), thus contradicting fundamental human values that uphold equality.

This article pursues two primary objectives. First, it seeks to demonstrate, on the basis of previous research, that algorithmic bias constitutes an inherent element in the development of artificial intelligence, which is exceedingly difficult to eradicate and therefore a significant gateway to discrimination and injustice. Second, it aims to delve further into the meaning of human dignity, both from a philosophical perspective and, more specifically, within the framework of Catholic theology, and to highlight its relevance amid the contemporary challenge of algorithmic bias in artificial intelligence. Accordingly, the argument unfolds in four major sections. The text begins with an introductory account of bias in artificial intelligence and explains why such bias demands urgent analysis in light of the concept of human dignity. The subsequent section advances the claim that human

dignity is not merely an axiomatic postulate to be taken for granted, but rather a crucial idea that must be rationally justified and yet remains central to the very meaning of humanity. In the age of AI, human dignity must serve as the guiding principle in addressing the anomalies of algorithmic systems, ensuring that every technological development remains oriented toward “greater justice, deeper fraternity, and a more humane structure of social relations” (AN 38).

2. Beyond the Axiom: Human Dignity as an Original Datum

The concept of human dignity has long been regarded as an axiomatic idea, often left unexamined and rarely brought to the surface (Kateb 2011; Ilesanmi 2023; Pabubung 2023). It was only after the end of World War II that the notion of human dignity began to resurface. Under the banner of human dignity, various charters addressing human rights were promulgated. Since then, nations worldwide have claimed their rights in the name of human dignity, which is fundamentally understood as equality free from discrimination and oppression (Becchi 2019). It is even challenging to conceive any aspect of human existence or significant event in recent history that does not, in some way, implicate human dignity, from public discourse – addressing issues such as genetic experimentation, freedom of expression – to legal debates concerning forced labour, torture, privacy, oppression, and religious freedom (Ilesanmi 2023).

The notion of human dignity can be traced back to Ancient Rome (Silva-Tapia 2023). Within the Roman intellectual tradition, the term *dignitas* acquired a dual philosophical significance that endures to this day. On the one hand, *dignitas* signifies humanity’s unique position within the cosmos. On the other hand, it denotes an individual’s distinctive status in relation to others within the public sphere. These dual interpretations reveal two critical dimensions of *dignitas*. First, it underscores humanity’s distinction from other elements of nature, rooted in the fact that human is rational being (*animal rationale*). Second, it highlights the differentiation of individuals from one another in the context of public life. In the first sense, human dignity is concerned with the “hierarchy of nature”, whereas in the second, it pertains to the “social hierarchy” (Becchi 2019).

Undoubtedly, the concept of the hierarchy of nature finds fertile ground within Christianity. God created humans in His image. Humanity possesses dignity as a natural endowment, derived directly from God, the Creator. However, the concept of being created in the image of God does not imply a social hierarchy. On the contrary, Christianity provides a foundation to affirm the universal dimension of human dignity. In other words, it advocates for equality among all humans. Every human being, without exception, is created in the image of God. Every person, without exception, possesses equal dignity before God. The emphasis lies on the likeness between humans and God, which explains their unique position among all prior creations. Scripture even declares that humanity was created “very good”

(Gen 1:31). God created humans in His image and bestowed upon them transcendent dignity. In other words, humanity's privileged status among creation is rooted in its resemblance to the Creator.

Unfortunately, secularisation has posed a complex and previously unimaginable challenge to this religious concept. The Italian Humanist manifesto, *De Hominis Dignitate* (1486) by Giovanni Pico della Mirandola, asserts that humans are the architects of their own purpose in life. This leads to the notion that humanity, in its essence, is incomplete. Therefore, "...as free beings, individuals must shape themselves according to their own desires" (Mirandola 1965).

This new humanistic vision imbues human dignity with the dual meanings of duty and conquest: the conquest of new territories and the achievement of knowledge novelties. The human vocation now becomes one of active engagement: an existence centred on labour and productivity. This marks the inception of the glory of *homo faber*, the human as a creator who now harnesses the power of intellect. Humanity has nearly dispensed with the need for God in the process of self-understanding. The individual, increasingly proud of their accomplishments, perceives their dignity to be enhanced by their achievements. The higher the achievement, the greater the dignity.

This perspective on social hierarchy flourished within the ideas of Thomas Hobbes. Hobbes contended that human dignity resides in one's standing within the public sphere. According to him, dignity is no longer tied to natural foundations (human nature) or transcendence. For Hobbes, dignity becomes something that is achieved through reciprocal relations, specifically in terms of "mutual recognition" between individuals. A person's worth is determined by their "market value". The value of each individual, which constitutes their dignity, lies in the recognition granted by others. Hobbes emphasised: "The worth of a person is their market price, which depends on the needs and judgments of others." (Hobbes 1997)

Nonetheless, a different understanding emerged from Samuel Pufendorf. He believed that freedom is the distinctive hallmark of humanity. This freedom is the precondition for the existence of a moral order, which stands apart from the natural order (Pufendorf 2009). According to Pufendorf, human dignity does not arise from the ability to conquer nature, but from the fact that humans are moral agents. For Pufendorf, humans stand on two distinct foundations: *entia physica* and *entia moralia* (Pufendorf 2009). This distinction relates to Kant's differentiation between the realm of nature and the realm of ends: human dignity is not conferred upon humans due to their position within the universe (the realm of nature) but because of their possession of purpose (the realm of ends). Humans possess intrinsic, absolute value as subjects of the categorical imperative, which is only realised through moral imperatives toward others (Kant 1991). It is within the ideas of Immanuel Kant that we find the notion that respect for others is grounded in a personal moral value, understood as an end in itself.

The central idea of human dignity is that, on this earth, humanity is the greatest type of being and that each of its members deserves to be treated in accor-

dance with the intrinsic worth of the human species. The dignity of each individual is equal to that of every other individual, meaning that all people share the same status (Kateb 2011). According to Kateb, the principle of equality in human status is assured when the state refrains from undermining the dignity of any individual and treats all people equally (2002). Likewise, each individual is obligated to recognise and respect the equal human status of others. This equality is founded on the conviction that the rights of every individual are absolute (Kateb 2015). As such, any form of discrimination stands in direct opposition to human dignity.

Discussions of human dignity would be insufficient without addressing the teachings of the Church, which have laid the groundwork for the global recognition of human dignity, irrespective of religious boundaries. The encyclicals “*Rerum Novarum*” (1889) by Pope Leo XIII and “*Quadragesimo Anno*” (1931) by Pope Pius XI represent two pivotal social doctrines that emerged from a deep concern over the widespread injustice and warfare that severely violated human dignity, necessitating immediate intervention. The principles of charity, solidarity, and equality configure the basis of these teachings, advocating against all forms of oppression and discrimination, regardless of race, ethnicity, class, nationality, or religion. As a “humanity expert”, the Church’s teachings have become a crucial reference in the study of human dignity (Glendon 2010).

The Catholic Church, through the Dicastery for the Doctrine of the Faith, affirms that “Every human person possesses an infinite dignity, inalienably grounded in his or her very being, which prevails in and beyond every circumstance, state, or situation the person may ever encounter. This principle, which is fully recognizable by reason alone, underlies the primacy of the human person and the protection of human rights.” (DI 1) Every individual, without exception, possesses equal and immeasurable dignity. There is no distinction between people based on the colour of their skin, whether white or black, race A or B, high or low education, wealth or poverty, age, or even whether one is born or unborn—all share the same boundless dignity.

Human dignity is an original datum that must be consciously recognised in faith and accepted with deep gratitude. “Based on this recognition and acceptance of human dignity, a new coexistence among people can be established that develops social relationships in the context of authentic fraternity. Indeed, only by acknowledging the dignity of each human person can we contribute to the rebirth of the universal aspiration to fraternity.” (DI 6) The Church views human dignity not as an individualistic concept but as the foundation for fostering genuine fraternity. Only through the acknowledgement of human dignity can each individual respect the existence of others with equal dignity and build a shared life in fraternity and harmony.

The Catholic Church opposes misguided views of human dignity that are hierarchical and discriminatory:

“To clarify the concept of dignity even further, it is essential to point out that dignity is not something granted to the person by others based on their gifts or qualities, such that it could be withdrawn. Were it so bestowed, it would

be given in a conditional and alienable way, and then the very meaning of dignity (however worthy of great respect) would remain exposed to the risk of being abolished. Instead, dignity is intrinsic to the person: it is not conferred subsequently (*aposteriori*), it is prior to any recognition, and it cannot be lost. All human beings possess this same intrinsic dignity, regardless of whether or not they can express it in a suitable manner.” (DI 15)

There is no class distinction among human beings based on human dignity, and it cannot be removed by anyone. Therefore, all forms of discrimination, in any form, including in the digital world and within the realm of artificial intelligence, are detrimental to human dignity. In our age, we should reclaim the centrality of human dignity for “greater justice, deeper fraternity, and a more human structure of social relations”.

3. Reclaiming the Centrality of Human Dignity Amid Algorithmic Bias

Hao (2019) argues that bias can emerge in three stages: (1) Framing the problem -the type of scientists designing the algorithm and their underlying objectives; (2) Collecting the data -the inadvertent inclusion of biased data, which often excludes individuals with darker skin tones, as input datasets tend to be predominantly composed of data from white individuals or specific racial and gender groups; (3) Preparing the data – the selection of particular attributes by scientists or researchers that they intend the algorithm to recognize. In essence, bias can occur both intentionally and unintentionally. Intentional bias arises when algorithm developers deliberately seek to discriminate against certain racial or ethnic groups. Unintentional bias occurs when data inputs disproportionately represent a specific racial group due to a lack of awareness about potential biases (Hao 2022; Suveren 2022). Algorithmic bias, in this context, refers to the inherent limitations of an algorithm in adapting to and learning from diverse datasets.

Various forms of bias exist within AI-based technologies, fundamentally rooted in algorithms. Bias manifests across numerous domains, including recruitment processes, financial services, healthcare, judicial administration, criminal justice, and many other programs that rely on algorithmic performance. As previously mentioned, bias can be categorised into four distinct types: sample bias, prejudice bias, systemic value distortion, and model insensitivity (Girasa 2020). This form of bias is particularly pronounced in the era of big data, where achieving “data fairness” is nearly impossible. Algorithms inherently favour larger datasets and, consequently, marginalise groups that contribute less data. In a Twitter post, AI expert Yann LeCun asserted that machine learning systems will inevitably exhibit bias if the data they process is biased (LeCun 2019). Obtaining unbiased data is highly challenging, as data is inherently shaped by the biases of those who produce it (Obermeyer et al. 2021; Agan et al. 2023). While technical measures can

mitigate bias to some extent, it is crucial to recognise that the primary source of algorithmic bias stems from data itself – data that is, by nature, never entirely free from bias (Flowerman 2023; Mullainathan 2019; Kleinberg et al. 2020).

Data generated from social media platforms, when shaped by specific interests, becomes inherently biased, as it is processed through pre-programmed algorithms. This dynamic underscores the fundamental mechanics of machine learning. Conceptually, machine learning can be understood as a form of training. A computer is exposed to an immense and heterogeneous dataset, from which it learns to generate assessments and predictions by identifying recurring patterns. However, AI systems do not exhibit uniform accuracy across different demographic groups. Consequently, an algorithm may yield divergent decisions when applied to distinct populations, potentially leading to systematic disparities in outcomes (Heilweil 2020; Wellman 2024).

Eloquently, the advancements across various domains of artificial intelligence (AI) have driven extraordinary economic and social growth. Fields such as communication, healthcare, education, agriculture, transportation, space exploration, science, and entertainment have all reaped substantial benefits from AI breakthroughs. However, inherent risks accompany the way these emerging technologies are researched, designed, developed, and deployed. It is crucial to move beyond a purely benefits-driven perspective and critically examine the broader implications. Discussions and debates on AI ethics must remain open to addressing indirect consequences. One of the most immediate concerns is the emergence of new forms of discrimination due to algorithmic bias, particularly in social identification technology. Such biases directly contradict one of the fundamental principles of human dignity: equality.

The dignity of every individual is equal to that of others, meaning that every person holds the same moral and existential status as their fellow human beings. Therefore, the principle of human equality must be upheld in all human interactions, without exception (Kateb 2015). One can only affirm his/her own dignity by recognising and treating others as possessing the same inherent dignity. Moreover, dignity cannot be claimed exclusively for one's own group, race, ethnicity, or social class; it must be extended to all individuals without exception. "Each person must claim for all, and all for each." (2011) In the same tone, "Antiqua et Nova" stated that individuals and communities are called to discern how AI can be employed to advance the good of humanity while resisting uses that may undermine human dignity. Within this framework, responsibility ought to be understood as a responsibility of care for others (NA 47).

At the heart of the concept of human dignity lies the fundamental recognition of a shared humanity within each individual as an essential aspect of personal identity. Human dignity is rooted in the belief that while each person is unique, this uniqueness is universally inherent to all human beings, without exception (DI 1). Therefore, in an existential sense, no one is inherently superior to another. We are all part of the same human species, equally endowed with the same intrinsic

and infinite dignity.

Consequently, in the age of artificial intelligence, utilitarianism should be reviewed thoroughly. Prioritising the majority (great number) over the minority (small number) should be re-examined in constructing AI algorithms, as this inevitably leads to the marginalisation and discrimination of smaller groups, since, in the name of human dignity, the majority does not have a higher status than the minority. The existence of a single human being is sufficient reason to uphold the principle of equality among all members of the human species without exception. Furthermore, in the age of AI, philosophers and technocrats must reexamine “[neo] virtue ethics”, especially when it comes to the scoring system of governance (Lainez and Gardner 2023), since we must reject any discrimination based on virtue. It is highly susceptible to being used by those in power to assign credit scores based on AI algorithms. Individuals considered virtuous or with higher credit scores do not possess a higher status than those deemed less virtuous. Additionally, machines merely execute technical selections among various possibilities, guided by predetermined criteria or statistical inferences. Human beings, by contrast, not only make choices but also possess the inner capability to decide at the level of the heart (NA 46).

There are individuals who, due to their inability to perform tasks effectively or to contribute in a tangible manner – whether as a result of age, health conditions, cognitive limitations, or other constraints- perceive themselves as insignificant. Does the concept of human dignity encompass them as well? The answer is an unequivocal “yes” in the most absolute sense. Their inherent human status remains unaltered, and they must be afforded the same recognition and respect as those deemed productive and or virtuous. They are to be treated as fully human, not as subhuman, nor as belonging to a category inferior to humanity. This elucidates the fundamental premise that the concept of human dignity, grounded in the uniqueness of the individual, ultimately affirms the centrality of the human being.

“Antiqua et Nova” reminds us that human worth is not defined by particular abilities, cognitive or technological accomplishments, nor by personal achievements, but rather by an intrinsic dignity rooted in being created in the image of God (AN 34). In this regard, the very appearance and impression of a human being suffice as a measure of dignity. In a certain sense, the awareness of human equality necessitates activating the “*aesthetic antenna*” of our intellect, enabling us to perceive the inherent beauty in every human and find the “infinite dignity”.

4. Conclusion

Algorithmic bias in artificial intelligence, manifesting through existing human bias or technical errors, fundamentally disregards the imperative of placing humanity at the centre. To place humanity at the centre means to uphold human dignity

in its holistic sense, without distinction based on race, gender, age, status, or social group. A human-centred approach also ensures that human dignity remains the ultimate objective of every human action, thereby precluding any action that compromises the dignity of others. In other words, all forms of discrimination -whether intentional or unintentional- are indefensible. Algorithmic bias has led to racial, gender, age-based, status-related, and social discrimination. As such, algorithmic bias is indefensible, for it becomes a mechanism of discrimination that undermines the fundamental principle of equal human dignity. Therefore, those who employ AI must take care not to grow excessively reliant on it for decision-making, as this tendency risks deepening the already significant dependence of contemporary society on technology. The use of AI must be guided by an ethic rooted in the vision of equality, responsibility, and fraternity that promotes the integral development of persons in relation to others.

Data Availability Statement

Because this study is based on a theoretical approach, no new research data were generated or analyzed.

Abbreviations

AN – Dicastery for the Doctrine of the Faith 2025 [Antiqua et Nova].

DI – Dicastery for the Doctrine of the Faith 2024 [Dignitas Infinita].

References

- Agan, A. Y., D. Davenport, J. Ludwig, and S. Mullainathan. 2023. *Automating Automaticity: How the Context of Human Choice Affects the Extent of Algorithmic Bias*. Cambridge, MA: National Bureau of Economic Research.
- Becchi, P., and Klaus Mathis, eds. 2019. *Handbook of Human Dignity in Europe*. Cham: Springer.
- Dicastery for the Doctrine of the Faith. 2024. *Dignitas Infinita*. Declaration. Vatican City: Libreria Editrice Vaticana.
- . 2025. *Antiqua et Nova*. Vatican City: Libreria Editrice Vaticana.
- Flowerman, C. H. 2023. (Some) Algorithmic Bias as Institutional Bias. *Ethics and Information Technology* 25, no. 2:24.
- Friedman, B., and H. Nissenbaum. 2022. Excerpt from Bias in Computer Systems. In: *Ethics of Data and Analytics*, 20–26. Boca Raton, FL: Auerbach Publications.
- Gengler, E., I. Hagerer, and A. Gales. 2024. Diversity Bias in Artificial Intelligence. In: *The Digital and AI Coaches' Handbook*, 229–240. London: Routledge.
- Ghasemaghaei, M., and N. Kordzadeh. 2024. Understanding How Algorithmic Injustice Leads to Making Discriminatory Decisions: An Obedience to Authority Perspective. *Information & Management* 61, no. 2:103921.
- Gilli, A., M. Pellegrino, and R. Kelly. 2019. Intelligent Machines and the Growing Importance of Ethics. In: *The Brain and the Processor: Unpacking the Challenges of Human-Machine Interaction*, 45–54. Rome: Nato Defense College.
- Gillis, T. B., and J. L. Spiess. 2019. Big Data and Discrimination. *The University of Chicago Law Review* 86, no. 2:459–488.
- Girasa, Rosario. 2020. *Artificial Intelligence as a Disruptive Technology*. Cham: Palgrave Macmillan.
- Glendon, Mary A. 2010. The Influence of Catholic Social Doctrine on Human Rights. In: Roland Minnerath, ed. *Catholic Social Doctrine and Human Rights*. Vatican City: Pontificia Academia Scientiarum

Socialum.

- Hao, K.** 2019. This Is How AI Bias Really Happens – and Why It’s So Hard to Fix. *MIT Technology Review*. <https://www.technologyreview.com/2019/02/04/137602/this-is-how-ai-bias-really-happens-and-why-its-so-hard-to-fix/> (accessed 25. 2. 2025).
- Hao, K.** 2022. When Algorithms Mess Up, the Nearest Human Gets the Blame. In: *Ethics of Data and Analytics*, 418–419. Boca Raton, FL: Auerbach Publications.
- Hassan, Y.** 2023. Governing Algorithms from the South: A Case Study of AI Development in Africa. *AI & Society* 38:1429–1442. <https://doi.org/10.1007/s00146-022-01527-7>.
- Heilweil, R.** 2020. Why Algorithms Can Be Racist and Sexist. *Vox*, 18. 2. <https://www.vox.com/recode/2020/2/18/21121286/algorithms-bias-discrimination-facial-recognition-transparency> (accessed 18. 9. 2020).
- Hill, R. K.** 2016. What an Algorithm Is. *Philosophy & Technology* 29:35–59.
- Hobbes, T.** 1997 [1651]. *Leviathan*. Edited by Michael Oakeshott. New York: Touchstone.
- Ilesanmi, S. O.** 2023. Religious Ethics and the Human Dignity Revolution. *Journal of Religious Ethics* 51, no. 4:652–672. <https://doi.org/10.1111/jore.12465>.
- Johnson, C. Y.** 2022. Racial Bias in a Medical Algorithm Favors White Patients over Sicker Black Patients. In: *Ethics of Data and Analytics*, 10–12. Boca Raton, FL: Auerbach Publications.
- Joseph, J.** 2024. Predicting Crime or Perpetuating Bias? The AI Dilemma. *AI & Society*. <https://doi.org/10.1007/s00146-024-02032-9>.
- Kateb, G.** 2006. *Patriotism and Other Mistakes*. New Haven, CT: Yale University Press.
- . 2011. *Human Dignity*. Cambridge, MA: The Belknap Press of Harvard University Press.
- . 2015. *Lincoln’s Political Thought*. Cambridge, MA: Harvard University Press.
- Kumar, A., V. Aelgani, R. Vohra, S. K. Gupta, M. Bhagwati, S. Paul, and J. S. Suri.** 2024. Artificial Intelligence Bias in Medical System Designs: A Systematic Review. *Multimedia Tools and Applications* 83, no. 6:18005–18057.
- Lainez, N., and J. Gardner.** 2023. Algorithmic Credit Scoring in Vietnam: A Legal Proposal for Maximizing Benefits and Minimizing Risks. *Asian Journal of Law and Society* 10, no. 3:401–432.
- LeCun, Y.** 2019. People Are Biased. Twitter, 7. 12. <https://twitter.com/ylecun/status/1203211859366576128?lang=id> (accessed 4. 10. 2025).
- Manzoor, R., W. Hussain, and M. L. Anjum.** 2024. Out of Dataset, Out of Algorithm, Out of Mind: A Critical Evaluation of AI Bias against Disabled People. *AI & Society* 40:3941–3951. <https://doi.org/10.1007/s00146-024-02168-8>.
- Martin, K.** 2022. Algorithmic Bias and Corporate Responsibility: How Companies Hide behind the False Veil of the Technological Imperative. In: *Ethics of Data and Analytics*, 36–50. Boca Raton, FL: Auerbach Publications.
- Mullainathan, S.** 2019. Biased Algorithms Are Easier to Fix than Biased People. *The New York Times*, 6.
- Obermeyer, Z., R. Nissan, M. Stern, S. Eaneff, E. J. Bembeneck, and S. Mullainathan.** 2021. *Algorithmic Bias Playbook*. Chicago, IL: Center for Applied AI at Chicago Booth.
- OECD.** 2019. *Artificial Intelligence in Society*. Paris: OECD Publishing. <https://doi.org/10.1787/eedfee77-en>.
- . 2024. *Algorithm and Eve: How AI Will Impact Women at Work*. Paris: OECD Publishing. <https://doi.org/10.1787/a1603510-en>.
- Pabubung, M. R.** 2023. An Inquiry into Human Dignity According to George Kateb. *Jurnal Filsafat* 33, no. 2:290–332.
- Pico della Mirandola, Giovanni.** 1965. *On the Dignity of Man*. Translated by Charles Glenn Wallis, Paul J. W. Miller, and Douglas Carmichael. Indianapolis, IN: Hackett Publishing.
- Pufendorf, S.** 2009 [1672]. *Two Books of the Elements of Universal Jurisprudence*. Edited by Thomas Behme. Indianapolis, IN: Liberty Fund.
- Schlackl, F., N. Link, and H. Hoehle.** 2022. Antecedents and Consequences of Data Breaches: A Systematic Review. *Information & Management* 59, no. 4:103638.
- Shams, R. A., D. Zowghi, and M. Bano.** 2025. AI and the Quest for Diversity and Inclusion: A Systematic Literature Review. *AI Ethics* 5:411–438. <https://doi.org/10.1007/s43681-023-00362-w>.
- Silva-Tapia, A., R. Morales-Olivares, and P. Cárcamo-Petridis.** 2023. The Concept of Dignity. In: *Global Handbook of Inequality*, 1–29. Cham: Springer International Publishing.
- Suveren, Y.** 2022. Unconscious Bias: Definition and Significance. *Psikiyatrike Guncel Yaklasimler* 14, no. 3:414–426.
- Tilmes, N.** 2022. Disability, Fairness, and Algorithmic Bias in AI Recruitment. *Ethics and Information Technology* 24, no. 2:21.
- Thomsen, F. K.** 2023. Three Lessons for and from Algorithmic Discrimination. *Res Publica* 29, no. 2:213–235.
- Turner-Lee, N.** 2018. Detecting Racial Bias in Algorithms and Machine Learning. *Journal of Information, Communication and Ethics in Society* 16, no. 3:252–260. <https://doi.org/10.1108/JICES-06-2018-0056>.
- Wellman, M. L.** 2024. “A Friend Who Knows What They’re Talking About”: Extending Source Credibility Theory to Analyze the Wellness Influencer Industry on Instagram. *New Media & Society* 26, no. 12:7020–7036.

Williams, B. A., C. F. Brooks, and Y. Shmargad. 2018. How Algorithms Discriminate Based on Data They Lack: Challenges, Solutions, and Policy Implications. *Journal of Information Policy* 8.

Yang, Y. 2025. Racial Bias in AI-Generated Images. *AI & Society*. <https://doi.org/10.1007/s00146-025-02282-1>.

Zhao, R., Z. Dong, and H. Wu. 2024. Harnessing Big Data for Strategic Innovation: Insights into Corporate Entrepreneurship and Technological Advancements. *Journal of the Knowledge Economy*. <https://doi.org/10.1007/s13132-024-02037-1>.

Zowghi, D., and M. Bano. 2024. AI for All: Diversity and Inclusion in AI. *AI Ethics* 4:873–876. <https://doi.org/10.1007/s43681-024-00485-8>.