Journal of Dharma 50, 1 (January–March 2025), 109–126

ECHOES OF POWER: BIAS, OBJECTIVITY AND CULPABILITY IN ALGORITHMIC AND STATISTICAL WORLDS

SHENGYING LI♦

Abstract: Algorithms, increasingly deployed in domains such as welfare, credit, hiring, policing, and healthcare, are often celebrated as neutral tools of efficiency and rationality. This study argues otherwise: algorithms do not merely process data but inscribe and reproduce the values of their designers, amplifying existing inequities and diminishing human agency. Drawing on Nietzsche's deconstruction of truth as a myth, Foucault's analysis of power relations, and Kierkegaard's concept of responsibility, the article contends that algorithmic outputs are not neutral projections of reality but socially sanctioned fictions—echoes of entrenched power that present themselves as objectivity. Philosophical critique is set in dialogue with technical accounts of decision-making, including Kochenderfer et al.'s *Algorithms for Decision Making* (2022) and Christian and Griffiths' *Algorithms to Live By* (2016). These works reveal how optimization under uncertainty translates complex moral and political questions into ostensibly technical procedures. In doing so, algorithms reshape normative life: reducing ethics to efficiency, justice to probability, and responsibility to statistical thresholds. The analysis challenges the persistent belief that bias can be eradicated through better data or technical refinement. Instead, it argues for a reconfiguration of objectivity—not as detached neutrality but as reflexive responsibility. By reframing culpability as distributed across coders, institutions, and societies, this study develops a theoretical grammar of socio-technical networks. Algorithms are not judge-like arbiters standing above human affairs but narrative actors embedded within them, shaping and reflecting humanity's moral order. They echo existing structures of power while amplifying new forms of asymmetry. To engage them critically

♦ **Shengying Li** is from Teachers College, Columbia University. E-mail: Lisringlee@gmail.com; lisringlee@outlook.com.

requires listening to these echoes, recognizing where they distort, and reclaiming responsibility for what is projected through their seemingly impartial voices.

Keywords: *Accountability, Algorithms, Bias, Ethics, Foucault, Legitimacy, Marginalization, Philosophy, Power, Responsibility, Statistics.*

1. Introduction

Algorithms have become indispensable to contemporary life, mediating decisions in welfare allocation, credit scoring, job recruitment, policing, healthcare and online commerce. Their appeal rests on the proposition that statistical models, unlike human judgment, can neutralize prejudice by grounding decisions in data rather than subjective bias. Proponents of algorithmic decision-making invoke the language of neutrality, insisting that "data speaks for itself." Within this frame, algorithms appear as impartial mediators producing outcomes through objective calculation, free from politics or morality. Yet this claim is deeply misleading: data is never separate from history, culture or interpretation. Far from neutral, algorithms inherit the assumptions and values of their creators and the contexts in which they are trained. They lack moral reasoning capacity and operate as what Veliz (2020) calls "moral zombies," acting on inputs without distinguishing right from wrong. In doing so, they risk amplifying structural inequalities while concealing them beneath the authority of statistical reasoning.

Algorithmic bias is not merely a technical flaw but a moral phenomenon exposing two persistent illusions: that algorithms can achieve objective neutrality and that responsibility for their harms rests solely with individual coders or corporations. Common approaches emphasizing fairness and transparency address symptoms without confronting deeper ethical questions. Central concerns involve the objectivity of algorithmic outputs and the distribution of responsibility for their consequences. Various philosophical thoughts illuminate these issues: Foucault's power/knowledge framework depicts algorithms as

instruments of normalization; Nietzsche reveals "truth" as socially constructed convention; and Kierkegaard emphasizes human responsibility and the ethical stakes in designing and deploying algorithms.

Technical accounts, including Christian and Griffiths' *Algorithms to Live By* and Kochenderfer et al.'s *Algorithms for Decision Making*, illustrate optimization under uncertainty, while highlighting that every goal encoded in an algorithm carries moral and political weight. Objectivity is better understood as reflexive responsibility, and culpability as distributed across coders, institutions, and society. Examining predictive policing, hiring, and credit scoring shows how neutrality claims perpetuate inequities, emphasizing the need to balance efficiency with justice and foster collective accountability.

2. Neutrality as Myth: Power, Values and Limits of Ethical AI

The claim that algorithms can ever be neutral is itself a myth. Objectivity must therefore be redefined not as detachment but as reflexivity grounded in accountability, while culpability must be recognized as distributed across coders, institutions, and society. Algorithms do not stand outside power; they encode values and operate within networks of influence. As Paul Hayes (2020) observes, meeting this challenge requires an understanding of transparency that moves beyond technical interpretability, toward accessibility that enables democratic oversight (849–874). Similarly, Zhu et al. (2021) warn that proposals for "ethically aware algorithms" often lead to overly complex systems that remain opaque to non-specialists, thereby excluding publics from meaningful engagement (1–19). If neutrality itself is illusory, then transparency must be reimagined less as a technical fix and more as a democratization of accountability.

To assess the philosophical stakes, it is necessary to consider technlogical counterclaims. Computer scientists often argue that bias can be measured and mitigated, for instance by equalizing error rates across demographic groups in fairness-aware machine learning (Pagano, 1–26). Transparency initiatives such as explainable AI seek to make algorithmic processes

interpretable, while value-sensitive design attempts to embed ethical considerations explicitly (Dexe et al., 3–15). Yet these very efforts reveal that designers' own values inevitably shape outcomes. Gritsenko and Wood (2022) note that such choices can create an illusion of objectivity, masking subjective ethical stances (45–62). Datta et al. (2022) further caution that formalizing values risks presenting contested moral questions as if they were merely technical. John Locke's reminder that "No Man's Knowledge here can go beyond his experience" (Buckingham et al., 130–133) resonates strongly: algorithms, too, can never transcend the data and value-laden experiences that ground them. If so, can fairness ever be designed—or would it instead be politically negotiated?

This tension is acute. On one side, algorithmic systems streamline data gathering, reduce human effort and promise fairness, transparency and neutrality through metrics (Atmaja, 620–627). On the other, they prioritize technical efficiency over ethical reflection, producing systems that may satisfy statistical measures of fairness while leaving intact broader social inequities and entrenched power imbalances (Osasona et al., 322–335). As Selbst et al. (2019) caution, reducing ethical dilemmas to optimization problems risks obscuring the very harms algorithms reproduce (59–68). Should fairness mean equal opportunity, equal outcomes or proportional representation? Should transparency serve experts or the public? Each formulation embeds contested moral values while presenting itself as pure mathematics.

3. Frameworks of Algorithmic Behaviour: The Dizziness of Freedom in Bias

Søren Kierkegaard's (1813–1855) existentialist philosophy—particularly his account of anxiety as the "dizziness of freedom" (Buckingham et al., 194)—offers a powerful view to illuminate the moral stakes of algorithmic bias and the responsibility embedded in statistical decision-making. Against Georg Hegel's (1770–1831) determinism, which cast reality as a historical process (180–185), Kierkegaard argued that human existence is shaped not by inevitability but by radical freedom: the choices through which individuals bear responsibility for their being (194). In The

Concept of Anxiety (1844), he characterizes freedom as producing both angst and possibility. To confront multiple courses of action is to experience a vertigo of choice, where the subject alone bears the burden of decision (194–195).

This existential struggle can be vividly seen in Shakespeare's *Hamlet* (Act III, Scene III), where Hamlet hesitates to kill Claudius, caught between avenging his father and the fear that killing him in prayer would absolve his soul (Shakespeare, 392). As Buckingham et al. (2011) note, Hamlet's hesitation reveals the unbearable responsibility of freedom (195). The analogy resonates with algorithmic decision-making: beneath the mask of neutrality lies a similar moral burden, one in which designers and institutions must reckon with the consequences of their choices.

Machine learning systems embody this tension. Decisions about what data to collect, how to weight outcomes and where to apply predictive models mirror the Kierkegaardian anxiety of freedom. Though often framed as impartial, algorithmic systems are artifacts of human judgment, carrying the imprint of ethical responsibility. Just as Kierkegaard's man on the cliff feels both the pull of preservation and the temptation of self-annihilation (194), so too do designers confront the double edge of technology: the capacity to reproduce inequity through bias or to foster fairness through conscientious design.

Kierkegaard's insight reframes algorithmic bias not as a mere technical flaw but as a moral phenomenon. Anxiety here is not paralysis but a productive force, compelling recognition of agency, freedom and accountability in creating technological systems. Every statistical representation thus carries culpability — not because it is deterministic, but because it expresses the choices of actors who, in exercising freedom, define the moral horizon of algorithmic life.

4. On the Reality of Truth

Nietzsche (1873) dismantles the assumption that human intellect provides access to an objective reality. For him, intellect is not divine reason but a fragile evolutionary trick, serving survival through deception and illusion (79–80). What humans call "truth"

is not knowledge of the real but a practical mechanism for preserving life. Language exposes this mechanism most clearly. Words are not mirrors of reality but metaphors, socially negotiated signs detached from what they claim to represent (81–83). When such constructions are mistaken for truth, illusion emerges into so-called reality. Moreover, can universality survive once it is recognized that linguistic and conceptual systems are resting on arbitrary simplifications? Clearly not.

Even concepts themselves embody falsehood. They erase differences, impose categories and stabilize the flux of reality by fabricating rigid systems (83–86). Science is often celebrated as the path to truth, but it merely codifies these illusions. No "correct" or universal world exists, since perception itself varies across species; what humans take as truth is only a construct conditioned by time, space and number (86–88). Belief in truth, Nietzsche insists, is simply forgetfulness—that is, the erasure of its status as invention. Science obscures its dependence on abstraction, whereas art openly acknowledges its illusory foundations. On the one hand, scientific discourse always disguises its claims as objective revelation. On the other hand, art embraces its deceptions as generative, producing creativity, beauty and finding meaning (88–91).

Again, for Nietzsche, truth is never absolute but entangled with illusion, belief and utility. Science, language and morality do not rest on reality but on fictions. What sustains life is not truth itself but illusion—whether in the rational system or the creative play of art (91–97). Nietzsche argued that what we call "truth" is, in actuality, a product of made-up images, comparisons, and human-constructed ideas. If that's the case, then today's claims that algorithms are "objective" are just another form of human power disguised as truth. Knowledge, in his view, grows out of human needs and desires, not out of some pure, neutral reality. When we treat statistics or algorithms as if they were brute facts, we overlook the reality that they are shaped by human choices, assumptions, and limitations. Algorithms, like language, compress the complexity of human life into rigid categories. In the process, they risk erasing differences while presenting outcomes

as if they were natural laws. What appears to be neutral mathematics is, in truth, a survival narrative disguised as universality and necessity.

From this perspective, algorithmic bias is not a technological glitch but a deeper problem: the illusion of neutrality conceals older social prejudices now reframed through the language of data and fairness. Nietzsche reminds us of the fragility of 'truth' and the illusory nature of 'objectivity.' Applied to algorithms, this critique suggests that bias is not an anomaly to be eliminated but an inherent feature of statistical representation. Accordingly, those who design and enforce these systems should be held morally accountable for the values and power structures their technologies perpetuate.

5. Space, Knowledge and Power in Algorithmic Bias

Michel Foucault's critical reflections on how 'space,' 'knowledge,' and 'power' affect each other cannot be ignored in the questioning of algorithmic systems. In an interview with Paul Rabinow (1991), Foucault shows how rationalities of government, disciplinary processes and spatial social order intertwine. When applied to algorithm bias in detection and surveillance, do we need to understand how the objectivity in these representations conceals practices of governance, standardization, and exclusion?

5.1. Governmentality and Algorithmic Rationality: Can algorithms ever be treated as neutral? Foucault would argue otherwise. Algorithmic rationality is best understood through his concept of governmentality, where rationalities and techniques are mobilized to shape conduct within prescribed systems of rules. Just as the eighteenth-century state reimagined territory as a city to be policed, contemporary institutions imagine populations as datasets to be governed (Rabinow, 239). Algorithms, then, are not impartial instruments of prediction but modern technologies of governmentality. Decisions about what to measure, how to classify, and whose optimality to prioritize are deeply political. Their supposed objectivity is grounded in political rationalities, embedding power directly into statistical

representation. Algorithmic bias, therefore, is not merely a technical feature but a manifestation of power itself.

5.2. The Police and the Logic of Data Regulation: For Foucault, "the police" is not merely a body of officers enforcing law, but a broader apparatus of regulation that reduces life to calculable functions—health, labour, habits and everyday transactions (241–242). When we transpose this framework to algorithmic responsibility, the parallel becomes clear: predictive policing, credit scoring, and automated hiring are not neutral tools but data-driven systems of regulation, constructing self-governing zones of control. The more we describe algorithms as "neutral," the more we conceal their role in extending a deeper logic of surveillance and regulation. Once life is translated into data and models, the promise emerges that order will sustain itself without intervention. Yet here lies the moral danger: algorithms embody precisely the fantasy Foucault warned against—that society can be governed by pure reason and calculation. At the same time, they normalize the illusion that no one is accountable, hiding human responsibility behind the facade of objectivity. Algorithms, therefore, do not merely organize data; they reproduce an old dream of absolute control while erasing the traces of those who design, deploy, and benefit from them.

5.3. Discipline and Statistical Objectivity: Is algorithm bias just a technical error, or is it a disciplinary tool? Foucault's account of discipline as a micro-technique for shaping and directing behaviour is directly relevant here. Classification, categorization, and normalization are not neutral statistical operations but instruments of social regulation (247–252). As Calhoun et al. (2022) note, Jeremy Bentham's Panopticon—an architectural model where a central tower enables constant, unseen surveillance—is a key metaphor in Foucault's analysis (211). Its power lies in generating self-regulating subjects: the prisoner becomes "an object of information," perpetually visible, isolated, and deprived of solidarity (212). Aware of the possibility of being watched, individuals internalize surveillance, acting as both

observer and observed, and ultimately disciplining themselves (Mills, 46). Power, in this sense, is "visible and unverifiable," embedded in perception rather than force. Algorithms operate in a similar fashion. When they learn to associate certain names with lower employability or particular postal codes with higher criminality, they are not just reproducing bias—they are rehearsing disciplinary norms. The danger lies in mistaking this process for objectivity. Bias here is not deviation but disciplinary truth: algorithms do not merely mirror society; they actively constitute it by enforcing what counts as "normal." Statistical neutrality, then, is less a safeguard against prejudice than a subtle technique of power.

5.4. *Heterotopias* of Data and the Space of Exceptions: What occurs when algorithms carve out hidden zones of exclusion? Foucault's notion of *heterotopias*—real and virtual "other spaces" that mirror society while simultaneously contesting it—offers a platform to understand these algorithmic dynamics. A credit score threshold, for instance, produces an invisible *heterotopia*: a space of exclusion where individuals denied financial access live in the shadow of absence, defined by what the system refuses to recognize (254). Similarly, predictive models generate zones of perpetual suspicion, marking particular neighbourhoods as "high risk" and effectively transforming them into digital *ghettos*. These are not physical spaces but algorithmic ones, yet they embody the same principle: power materialized in spatial form, where the map of society is redrawn through the logic of data.

5.5. Techne, Responsibility and the Morality of Bias: Foucault's account of *techne*—the practical rationalities for governing life (255–256)—pushes us to rethink responsibility in algorithmic systems. Algorithms are not neutral scientific instruments but technologies of governance. To treat them as "objective" is to mistake them for natural science rather than *techne*, infused from the outset with moral and political choices. Culpability, therefore, cannot be displaced onto "the algorithm." It resides in how these systems are designed and deployed—what they render visible,

whom they represent, and how they distribute risk and opportunity. The morality of algorithmic bias lies not in fixing representational errors but in interrogating the governing rationalities that generate those errors. Foucault's insights reveal that bias is no accident: it is the continuation of governmental rationalities disguised as neutrality. By embedding power into statistical representation, algorithms normalize inequality while concealing responsibility. Moral accountability, then, requires not technical correction but critical engagement with the political rationalities underpinning algorithmic governance.

6. Living by Algorithms: Rationality, Bias and Moral Responsibility

Brian Christian and Tom Griffiths's *Algorithms to Live By* (2016) presents a provocative thesis: algorithms are not merely computational procedures but philosophical artifacts that shape human life. Their framework, which draws from computer science, philosophy and cognitive psychology, demonstrates how algorithms can guide ordinary decisions—whether in relationships, housing or scheduling—with a rigour that blends efficiency with insight. Yet, as Tosin Adeoti (2021) observes, algorithms still evoke the mystique of something only a "computer nerd" could understand. Christian and Griffiths challenge this perception by showing that algorithms are not esoteric codes but systematic steps that can inform everyday reasoning (Adeoti, Medium). Similarly, Amir Ghobadi (2024) notes that the book highlights how principles such as optimal stopping, game theory, or the explore/exploit dilemma can illuminate daily choices, enabling people to become more decisive and innovative (Medium).

Julian Humphrey (2016) emphasizes that Christian and Griffiths show how principles from computer science can offer practical, well-tested guidance for making decisions in everyday life, including professional areas like coaching (122–124). However, Suraj Deshmukh (2020) offers a sobering counterpoint: while the analogies are compelling, life's complexity often resists simplification into algorithmic form. Real-world decision-making

is entangled in constraints—emotional, cultural and historical—that no formula can fully capture. The tension between promise and limitation reveals something crucial: algorithms are never simply neutral aids. Each line of code reflects hidden commitments about rationality, objectivity and responsibility. If algorithms embody human priorities, then they must be understood as moral instruments as much as technical ones.

Christian and Griffiths begin by noting that human decision-making is inevitably constrained by time, limited knowledge and irretrievable choices (p. 1). Griffiths, speaking at Talks at Google (2016), argued that everyday dilemmas—organizing one's house, scheduling time—are structurally similar to the computational problems solved by machines. This reframing, he suggests, can provide new insights, even elegant solutions, for human life. Among their most famous illustrations is the "optimal stopping problem." The so-called 37% rule recommends surveying the first third of available options before committing, maximizing the chance of success (2–3). Applied to apartment hunting, dating or even parking, this strategy seems to offer rational clarity. Yet, as Christian (2016) himself admits, such efficiency smuggles in implicit standards of prudence and responsibility.

The crucial insight here is that algorithms not only help us decide but also establish the very standards by which we assess decisions. They offer the allure of objectivity, seemingly free from wavering emotions or second-guessing. But objectivity itself is hardly innocent. To label a choice "optimal" presupposes what is worth maximizing. The 37% rule elevates probabilistic efficiency but remains silent about regret, loss or meaning. Should flourishing be reduced to statistical prescriptions? This question opens the door to the ethics of algorithmic life. At this juncture, the problem of bias emerges. Algorithms are not empty containers but vessels of values. To use one is to tacitly accept its embedded priorities. When predictions fail, responsibility becomes contested: was it the individual's fault for trusting the system, or the system's fault for constraining their options? Bias, then, is not an error but the shadow of rationalization itself. Christian and

Griffiths remind us that every systematic attempt to decide entails exclusions, trade-offs and judgments that cannot be dissolved into numbers.

Peter Norvig, speaking at the same Google event, noted that what matters most is not coding *per se* but learning to model the world differently—to think algorithmically about choices. Algorithms shape not only micro-decisions such as dating or renting but also macrostructures of governance. Systems of sorting, caching, scheduling and attention allocation—once confined to machines—now regulate social life itself (6–7). But every act of optimization hides a trade-off: should efficiency outweigh fairness? Should accuracy eclipse inclusivity? Can speed justify the sacrifice of reflection? To optimize is always to prioritize, and thus to decide whose values matter most. This recognition shifts the moral discussion of algorithmic bias. It is not primarily about correcting statistical errors but about confronting the philosophy of representation itself. No algorithm is bias-free, because no algorithm can avoid the burden of choice. Bias is not a glitch; it is the condition of algorithmic rationality. Algorithms, therefore, should be read not only as technical tools but as moral artifacts—crystallizations of objectivity, culpability and value within mathematical form.

The task before us is not merely to execute algorithms, but to interpret and judge them. Every time we press "ENTER," we are not simply running a computation; we are endorsing a philosophy of what counts as rational, responsible and just. If algorithms carry values—and they do—then the ethical imperative is to continually ask: whose values are we living by?

7. Algorithmic Decision-Making Under Uncertainty

Let us now examine the decision-making framework under uncertainty as described by Mykel J. Kochenderfer, Tim A. Wheeler, and Kyle H. Wray in *Algorithms for Decision Making* (2022). The authors argue that both human and machine agents must often act under conditions of incomplete, ambiguous, or unavailable information. They identify four key types of uncertainty:

1. **Outcome Uncertainty**: The effects of actions are uncertain.
2. **Model Uncertainty**: The model of the problem is itself uncertain.
3. **State Uncertainty**: The actual state of the environment is unknown.
4. **Interaction Uncertainty**: The behavior of other agents within the environment is unpredictable (1-2).

Within this framework, algorithms cannot be regarded as neutral processors of information. They operate within the limits of incomplete knowledge, relying on statistical representations that are inherently partial or noisy (5). This raises a crucial philosophical question: how can algorithms claim objectivity when operating in unpredictable and complex environments? Kochenderfer et al. highlight that algorithms capable of decision-making under uncertainty can provide substantial social benefits, for example, in areas such as environmental sustainability, healthcare, urban planning and public infrastructure (12–13). However, these benefits are counterbalanced by serious challenges, including algorithmic bias, unintended consequences, adversarial manipulation and ethical and legal dilemmas.

The authors emphasize that rational decision-making is central when navigating uncertainty. They illustrate this by starting with simple, single-step choices and progressively moving to more complex scenarios involving unknown models, partial visibility and interactions among multiple agents. The framework employs probabilistic reasoning, utility theory, decision networks and sequential decision-making methods such as Markov Decision Processes (14). Both exact and approximate solution techniques are considered, depending on the complexity of the problem. As situations become more complex, distinctions between types of uncertainty become critical. Model uncertainty arises, for instance, when agents learn from experience, often via reinforcement learning (15). State uncertainty occurs when observations are probabilistically linked to the underlying reality, a scenario commonly modeled as a Partially Observable Markov Decision Process (POMDP) (Ibid.). Multi-agent interactions introduce additional layers of complexity, handled using models

such as Markov Games, partially observable Markov Games and decentralized POMDPs, with corresponding solution strategies (16).

Algorithmic bias, thus, creates a philosophical tension between objectivity and moral responsibility. Statistical modeling alone cannot guarantee objectivity; it is shaped by normative choices regarding which uncertainties to model and which outcomes or risks to prioritize. Similarly, culpability cannot be assigned solely to the machine. Responsibility extends to human decisions in algorithm design, data selection and computational strategy. Techniques such as supervision, optimization, and reinforcement learning further mediate this distribution of accountability between humans and algorithms. Again, we come to the conclusion that algorithmic bias is more than a technical flaw; it is a moral phenomenon arising from the interplay between epistemic constraints and human values. As Kochenderfer et al. note, "algorithms can amplify the impact of their users, regardless of their intention" (13). Consequently, a pressing question arises: how can ethical considerations be integrated into computational frameworks to mitigate bias while preserving the efficacy of algorithmic decision-making? This challenge makes us to think about the need for balancing responsibility and objectivity within a coherent philosophical and ethical paradigm.

8. Conclusion

Algorithmic bias is not merely a technical error but a moral and philosophical concern. Algorithms are not neutral; they encode human assumptions, priorities and cultural values into data-driven decisions. Nietzsche's critique of "truth" exposes how supposedly objective outputs are socially constructed illusions, flattening the richness of human experience into static, value-laden categories. Kierkegaard's "dizziness of freedom" highlights the ethical weight borne by designers, policymakers and institutions in shaping data, models, and deployment decisions. Foucault's insights on power and discipline show how algorithms function as instruments of governance, reinforcing disparities, creating exclusionary spaces, and masking human accountability.

Technical perspectives on uncertainty and optimization further reveal moral stakes: algorithms operate under incomplete knowledge, probabilistic reasoning and goal-oriented constraints, privileging some outcomes while marginalizing others. Statistical precision cannot substitute ethical judgment. Responsibility is not solely the coder's; it extends to institutions and society, demanding reflexive accountability. Ethical fairness requires more than technical fixes—it requires acknowledging the moral implications embedded in every algorithmic choice.

Religious and philosophical traditions reinforce this moral imperative: The *Bible* emphasizes stewardship of one's talents, the *Bhagavad Gita* urges action without attachment to outcomes, the *Qur'an* affirms accountability within capacity, and Zen Buddhism links freedom inseparably with mindful responsibility. Together, they create a world where freedom is inseparable from ethical engagement. Algorithms must therefore, be understood as moral and social artifacts, not neutral arbiters. They reflect, shape, and critique the societies they serve. Redefining objectivity as reflexive responsibility and distributing culpability across designers, institutions and society is essential. The myth of neutral technology must be dismantled to realize algorithmic governance that is both effective and just. Future research must continue integrating technical sophistication with ethical rigour, ensuring that algorithmic systems are capable not just of precision but of principled responsibility.

REFERNCES

Adeoti, Tosin. Book Review—*Algorithms to Live By*—Brian Christian & Tom Griffiths. *Medium*, 21 April 2021, https://tosinadeoti.medium.com/book-review-algorithms-to-live-by-by-brian-christian-tom-griffiths-4893133e0fe1, accessed in August 2025.

Atmaja, Subhanjaya Angga. "Ethical Considerations in Algorithmic Decision-Making: Towards Fair and Transparent AI Systems". *Riwayat Educational Journal of History and Humanities*, vol. 8, no. 1, Jan. 2025, pp. 620–627. https://doi.org/10.24815/jr.v8i1.44112.

BBC News. Why Are Pakistan's Christians Targeted? *BBC News*, 27 March 2018, www.bbc.com/news/world-asia-35910331, accessed in August 2025.

Buckingham, Will, et al. *The Book of Philosophy: Big Ideas Simply Explained*. DK Publishing, 2011.

Calhoun et. al. *Contemporary Sociological Theory.* Edited by Craig Calhoun, Joseph Gerteis, James Moody, Steven Pfaff, and Indermohan Virk, 4th ed., Wiley Blackwell, 2022.

Cath, Corinne. Governing Artificial Intelligence: Ethical, Legal and Technical Opportunities and Challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 376, no. 2133, Oct. 2018, pp. 1-8, https://doi.org/10.1098/rsta.2018.0080.

Christian, Brian, and Tom Griffiths. *Algorithms to Live By: The Computer Science of Human Decisions*. Henry Holt and Company, 2016.

Christian, Brian, Tom Griffiths, and Peter Norvig. *Algorithms to Live By|Brian Christian & Tom Griffiths|Talks at Google*, 2016, accessed August 2025. YouTube, https://www.youtube.com/watch?v=OwKj-wgXteo.

Datta, Teresa, et al. Tensions Between the Proxies of Human Values in AI. *arXiv,* Jan. 2022. Cornell University, https://doi.org/10.48550/arxiv.2212.07508.

Deshmukh, Suraj. Book Review: *Algorithms to Live by – The Computer Science of Human Decisions*. *Suraj Deshmukh*, 11 October 2020, accessed in August 2025, https://suraj.io/post/book-review-algorithms-to-live-by-the-computer-science-of-human-decisions/.

Dexe, Jacob, Ulrik Franke, Anneli Avatare Nou, and Alexander Rad. Towards Increased Transparency with Value Sensitive Design. *Human-Computer Interaction. HCII 2020. Lecture Notes in Computer Science, vol. 12217,* edited by Helmut Degen and Lauren Reinerman-Jones, Springer, 2020, pp. 3–15. SpringerLink, https://doi.org/10.1007/978-3-030-50334-5_1.

Eknath, Easwaran. *The Bhagavad Gita.* 2nd ed. Nilgiri Press, 2007.

Foucault, Michel. *Discipline and Punish: The Birth of the Prison*. Translated by Alan Sheridan, Vintage Books, 1977.

Ghobadi, Amir. Book Summary: *Algorithms to Live By: The Computer Science of Human Decisions by Brian Christian and Tom Griffiths. Medium*, 14 September 2024, accessed in August 2025, https://medium.com/@amirghobadi/book-summary-algorithms-to-live-by-the-computer-science-of-human-decisions-by-brian-christian-37b452760768.

Gritsenko, Daria, and Matthew Wood. Algorithmic Governance: A Modes of Governance Approach. *Regulation & Governance,* vol. 16, no. 1, Nov. 2020, pp. 45-62, Wiley Online Library, https://doi.org/10.1111/rego.12367.

Human Rights Watch. India: Increased Abuses Against Minorities, Critics: Violence, Discriminatory Policies Set Back Global Leadership Role. *Human Rights Watch*, 11 January 2024, accessed August 2025, https://www.hrw.org/news/2024/01/11/india-increased-abuses-against-minorities-critics.

Humphreys, Julian. Book Review: *Algorithms to Live By: The Computer Science of Human Decisions* by Brian Christian and Tom Griffiths. *Philosophy of Coaching: An International Journal*, vol. 1, no. 1, 2016, pp. 122–24. http://dx.doi.org/10.22316/poc/01.1.10.

Institute of Medicine, Committee on Understanding and Eliminating Racial and Ethnic Disparities in Health Care, Board on Health Sciences Policy. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care***.** Edited by Adrienne Y. Stith, Alan R. Nelson, and Brian D. Smedley, National Academies Press, 2009.

Johnsen, Maria. *AI Misalignment*. Mara Johnsen, 2024.

Kierkegaard, Soren. *The Concept of Anxiety: A Simple Psychologically Orienting Deliberation on the Dogmatic Issue of Hereditary Sin (Kierkegaard's Writings, VIII).* Edited and Translated by Reidar Thomte, Princeton UP, 1980.

Kochenderfer, Mykel J., Tim A. Wheeler, and Kyle H. Wray. *Algorithms for Decision Making***.** MIT Press, 2022.

Mills, Sara. *Michel Foucault*. Routledge, 2003.

Nietzsche, Friedrich. *Philosophy and Truth: Selections from Nietzsche's Notebooks of the Early 1870s*. Edited and translated

by Daniel Breazeale, Humanities Press International, 1992.

Osasona, Femi, et al. Reviewing the Ethical Implications of AI In Decision-Making Processes. *International Journal of Management & Entrepreneurship Research*, vol. 6, no. 2, Feb. 2024, pp. 322-335, https://doi.org/10.51594/ijmer.v6i2.773.

Pagano, T. P., et al. Bias and Unfairness in Machine Learning Models: A Systematic Review on Datasets, Tools, Fairness Metrics, and Identification and Mitigation Methods. *Big Data and Cognitive Computing*, vol. 7, no. 1, 2023, pp. 1-31. MDPI, https://doi.org/10.3390/bdcc7010015.

Rabinow, Paul, editor. *The Foucault Reader: An Introduction to Foucault's Thought*. Penguin Books, 1991.

Rahula, Walpola. *What the Buddha Taught.* Revised ed., Grove Press, 1974.

Sekh, Arif Ahmed, and Dilip K. Prasad. *Gender and Diversity Policy in AI: Strategies, Metrics, and Case Studies*. World Scientific Publishing Co. Pte. Ltd., 2025.

Selbst, Andrew D., et al. Fairness and Abstraction in Socio-technical Systems. Jan. 2019, pp. 59-68, ACM Digital Library, https://doi.org/10.1145/3287560.3287598.

Shakespeare, William. *The Complete Works of William Shakespeare*. Geddes & Grosset, 2002.

The Bhagavad Gita: Based on HH Sri Raghavendra Teertha's Gita Vivruti and Lectures by HH Sri Vidyasagara Madhava Teertha. Compiled by Dr. Giridhar Boray, Tirumala Tirupati Devasthanams, 2023.

The HarperCollins Study Bible: Fully Revised and Updated. Edited by Harold W. Attridge, revised ed., HarperSanFrancisco, 1989.

The Qur'an. Translated by M. A. S. Abdel Haleem, Oxford University Press, 2004.

Veliz, Carissa. "Moral Zombies: Why Algorithms Are Not Moral Agents". *AI & Society*, vol. 36, no. 2, 2021, pp. 487–497. Springer, https://doi.org/10.1007/s00146-021-01189-x.

Zhu, Liming, et al. AI and Ethics: Operationalising Responsible AI. *arXiv* Jan. 2021, pp. 1-21. Cornell University, https://doi.org/10.48550/arXiv.2105.08867.