THE FOUNDATIONS OF AI POLICY: A PHILOSOPHICAL CRITIQUE OF UNESCO RECOMMENDATION ON THE ETHICS OF ARTIFICIAL INTELLIGENCE

Ikechukwu Bartholomew Ekemezie, Ph.D
Department of Philosophy
Nnamdi Azikiwe University, Awka
Ib.ekemezie@unizik.edu.ng

Anthony Ugochukwu Nwokoye
Department of Philosophy
Nnamdi Azikiwe University, Awka
au.nwokoye@unizik.edu.ng

Abstract
The rapid proliferation of artificial intelligence systems across societal domains has triggered an urgent need for robust policy frameworks capable of governing these technologies while preserving fundamental human values. It has become necessary to conduct a critical philosophical examination of the foundations underlying contemporary AI policy, especially as current frameworks suffer from insufficient ontological grounding and inadequate attention to the metaphysical distinctions between human and artificial agency. Drawing on the moderate realist tradition of Aristotle and Thomas Aquinas, as well as Kantian deontology associated with Immanuel Kant and contemporary virtue ethics, this analysis demonstrates that effective AI governance requires the prior resolution of foundational questions concerning human dignity, moral agency, and the limits of algorithmic delegation. An examination of the UNESCO Recommendation on the Ethics of Artificial Intelligence shows its feasibility in addressing core AI challenges, including algorithmic bias, opacity, and accountability gaps. However, ethical principles—particularly human dignity, non-delegable agency, and technological subsidiarity—must serve as the normative bedrock upon which policy mechanisms are constructed, rather than as supplementary considerations appended to primarily technical frameworks. Thus, sustainable AI governance demands philosophical clarity about what distinguishes human beings from artificial systems, and policies that fail to acknowledge this distinction risk undermining the very values they purport to protect.
Keywords: artificial intelligence, AI policy, philosophical foundations, human dignity, ethics of AI, algorithmic governance

Introduction

Artificial intelligence has emerged as one of the most transformative technologies of the twenty-first century, which reshapes diverse sectors as healthcare, finance, criminal justice, and education. The AI incidents database maintained by the AI, Algorithmic, and Automation Incidents and Controversies (AIAAIC) repository contains over 3,000 documented incidents as of May 2024, illustrating the significant challenges associated with AI deployment across sectors (Madanchian & Taherdoost, 2025). These range from discriminatory outcomes produced by biased algorithms to opaque decision-making processes that elude meaningful oversight and accountability. In response, governments, international organizations, and professional bodies have promulgated numerous frameworks, guidelines, and regulations intended to govern AI development and use in ethically defensible ways (Smuha, 2025).

Yet beneath the surface of this regulatory activity lies a set of profound philosophical questions that are frequently overlooked or inadequately addressed. What, fundamentally, is the relationship between human beings and artificial systems? On what basis can we claim that humans possess dignity and rights that machines do not? What limits should constrain the delegation of human decision-making to algorithmic processes? These questions are not merely academic curiosities; they have direct implications for how policies are formulated, implemented, and evaluated.

The central argument advanced is that sustainable AI governance requires philosophical clarity about what distinguishes human beings from artificial systems. Policies that fail to acknowledge this distinction, or that treat it as irrelevant to regulatory design, risk undermining the very values they purport to protect. As Ramos-Zaga (2026) argues, "the exclusivity of human moral judgment, the imposition of boundaries on algorithmic delegation, and the adoption of technological subsidiarity should guide regulatory frameworks" intended to safeguard domains reserved for human agency within increasingly automated environments.

The Question of Ontological Status

Any coherent AI policy must rest upon some understanding of what artificial intelligences are and how they relate to human beings. This is not merely a technical question but a fundamentally philosophical one, concerning the ontological status of artificial systems and their capacities for consciousness, intentionality, and

moral agency. The tradition of moderate realism, extending from Aristotle through Thomas Aquinas to contemporary philosophers such as Oderberg (2007) and MacIntyre (1981), provides a framework for addressing these questions. On this view, living beings are characterized by substantial unity, that is, they exist as unified wholes whose parts are ordered toward the actualization of the being's essential capacities. Human beings possess not merely biological unity but also conscious interiority, genuine intentionality, and the capacity for rational self-determination. These features are not accidental properties that could be replicated in any sufficiently complex substrate; they are rooted in the kind of being that humans are.

Aristotle's De Anima establishes the foundational principle that "the soul is the form of a living body" (Durrant, 2015); the principle of organization that makes a living being the kind of thing it is. For human beings, this form includes the capacities for intellect and will that enable rational thought and free choice. Thomas Aquinas develops this insight further, arguing that the human soul, as the form of the body, is both the principle of biological life and the seat of intellectual powers that transcend purely material explanation (Aquinas, 1947). Contemporary artificial intelligence systems, however sophisticated, lack this substantial unity. They are aggregates of components designed to produce certain outputs, not unified substances whose parts are intrinsically ordered toward the flourishing of the whole. As Ramos-Zaga (2026) observes, artificial systems lack the "conscious interiority and genuine intentionality" that characterize human persons. This is not a temporary limitation that future technological advances will overcome; it is a consequence of the fundamental difference between organisms and artifacts.

Consciousness, Qualia, and the Explanatory Gap

The philosophical examination on consciousness raises further questions about the capacities of artificial systems. Thomas Nagel's influential article "What Is It Like to Be a Bat?" argues that consciousness is characterized by subjective experience "there is something it is like to be that organism" (Nagel, 1974). This subjective dimension, or qualia, presents what Joseph Levine terms the "explanatory gap" between physical processes and conscious experience (Levine, 1983). While complete physical detail the neural processes underlying perception can be described, yet this description leaves unanswered the question of why these processes should be accompanied by subjective experience at all. Chalmers (1997) has termed this the "hard problem of consciousness," distinguishing it from the "easy problems" of explaining cognitive functions. The easy problems concern how systems process information, integrate inputs, and generate outputs. The hard problem concerns why any of this processing should be accompanied by subjective awareness. For artificial systems, even those that perfectly simulate human conversational abilities, there is no evidence that they possess such subjective awareness. As the philosopher John Searle argued in his Chinese Room thought experiment, "syntax is insufficient for semantics," in other words, the manipulation of symbols according to rules does not suffice for genuine understanding of what those symbols mean (Searle, 1980).

This is not to claim that artificial systems could never possess consciousness. Some philosophers, including Eric Schwitzgebel and Mara Garza, argue that it is possible that future AIs might be "psychologically similar to natural human beings in consciousness, creativity, emotionality, self-conception, rationality, and so on" (Schwitzgebel & Garza, 2015). They contend that what matters for moral status is psychological and social properties, not the particular substrate silicon or meat that implements them (Schwitzgebel & Garza, 2015). On this view, if an artificial being possessed genuine consciousness, it would warrant moral consideration equal to that of human beings. The ontological status of AI systems remains an open philosophical question with profound implications for how they should be treated. Policy frameworks that avoid this question by focusing solely on functional outputs risk begging fundamental questions about the nature of the entities they seek to govern.

Moral Agency and Responsibility

A related set of questions concerns moral agency; that is, the capacity to act in ways that can be evaluated as right or wrong and for which the agent can be held responsible. Traditional conceptions of moral agency require capacities that artificial systems lack, that is, the ability to grasp moral reasons, to deliberate about alternative courses of action, and to choose freely among them. Immanuel Kant's moral philosophy provides the most influential framework for understanding moral agency. For Kant, the capacity for rationality grounds human dignity and the respect we owe to persons: "Act in such a way that you treat humanity, whether in your own person or in the person of any other, never merely as a means, but always at the same time as an end" (Kant, 1997, p. 38). Rational beings are ends in themselves, never to be treated merely as means. This status follows from their ability to act according to the representation of laws, to govern their conduct by principles they can rationally endorse (Kant, 1997). Artificial systems, lacking this capacity for rational self-governance, cannot be moral agents in the Kantian sense.

Hans Jonas, in The Imperative of Responsibility, extends this analysis to technological contexts. He argues that the expanded power of technology creates new ethical responsibilities while also raising questions about the moral status of technological artifacts: "Technology has become a burden that we must bear—a burden that imposes on us a new kind of responsibility, one that we cannot escape" (Jonas, 1984, p. 92). Jonas maintains

that responsibility presupposes a subject capable of responding to claims; a capacity that belongs only to beings with the kind of interiority that makes them vulnerable to harm and capable of flourishing (Jonas, 1984).

Contemporary virtue ethics, as developed by Alasdair MacIntyre, emphasizes the role of practices, traditions, and communities in shaping moral character (MacIntyre, 1981). Moral agency, on this view, is not an isolated capacity but emerges through participation in social practices that cultivate virtues. Artificial systems, lacking the capacity for such participation and the kind of narrative unity that characterizes human lives, cannot be moral agents in the sense required for virtue ethics. These philosophical considerations have direct implications for AI policy. If artificial systems are not moral agents, then responsibility for their outputs must ultimately rest with the human beings who design, deploy, and use them. Policies that obscure this responsibility by treating AI systems as quasi-autonomous actors risk creating accountability gaps that neither deter harmful conduct nor provide redress for those harmed.

The Significance of Human Dignity

The concept of human dignity serves as a foundational principle in international human rights law and in many philosophical accounts of moral status. The Universal Declaration of Human Rights opens by affirming "the inherent dignity and equal and inalienable rights of all members of the human family" (United Nations, 1948). This dignity is not earned through the possession of particular capacities; it is inherent in virtue of being human. For Kant (1997), dignity is the intrinsic worth that belongs to rational beings, who can never be priced or replaced: "In the kingdom of ends everything has either a price or a dignity. What has a price can be replaced by something else as its equivalent; what on the other hand is raised above all price and therefore admits of no equivalent has a dignity." This conception grounds prohibitions on treating persons merely as means and requires that we respect their autonomy and capacity for self-determination.

In the context of AI policy, human dignity imposes constraints on how artificial systems may be developed and deployed. Systems that manipulate human behavior without transparency, that make decisions affecting human lives without explanation, or that reduce persons to data points to be optimized, all risk violating human dignity. As the OCED (2019) emphasizes, AI systems should be developed in ways that "protect dignity, autonomy, and fundamental human rights." On the other hand, the principle of non-delegable agency, derived from considerations of dignity, holds that certain decisions cannot be transferred to artificial systems without undermining human personhood. Decisions involving moral judgment, legal culpability, or profound implications for human welfare require the exercise of distinctly human capacities that machines lack. Ramos-Zaga (2026) argues that regulatory frameworks must preserve "spheres of inalienable human agency" within increasingly automated environments.

The UNESCO Recommendation on the Ethics of Artificial Intelligence

The UNESCO Recommendation on the Ethics of Artificial Intelligence, adopted by all 193 member states in 2021, is a landmark document (UNESCO, 2024). As the first-ever global standard of its kind, it represents a significant attempt to govern the development and use of AI on an international scale. From a philosophical perspective, its value lies not just in what it says, but in how it frames the entire problem of AI governance. The Recommendation's cornerstone, as stated on the UNESCO page, is "the protection of human rights and dignity"(UNESCO, 2024). This is a profoundly important starting point. Philosophically, it grounds AI ethics not in abstract calculations of utility or efficiency, but in the inherent worth of every human being. This commitment echoes the Kantian tradition, which argues that rational beings possess dignity, that is, an intrinsic value that has no price and admits no equivalent (Kant, 1997). By placing dignity at the center, the Recommendation asserts that human beings are ends in themselves. They cannot be treated merely as data points to be processed or as means to optimize a system. This has direct implications: AI systems must serve humanity, not the other way around. The emphasis on "human oversight" flows directly from this. If a machine makes a final decision about a person's life without a human in the loop, that person is effectively being treated as an object, and their dignity is violated.

A common criticism of AI ethics guidelines is that they are full of good intentions but lack practical application; a gap between abstract "soft ethics" and concrete action (Floridi, 2018). UNESCO highlights that the Recommendation attempts to bridge this gap through its "extensive Policy Action Areas." This is a significant philosophical and practical strength. With the identification of specific areas like data governance, gender, education, and health, the framework acknowledges that ethical principles are not one-size-fits-all. What fairness means in an algorithm used for hiring is different from what it means in one used for medical diagnosis. When policymakers are required to translate values into action in these distinct contexts, it becomes a move toward what philosophers call "applied ethics." It forces a consideration of the real-world consequences of AI, moving beyond pure theory. For example, applying the principle of fairness to the "environment and ecosystems" area pushes us to consider not just AI's social impact, but its material impact on the planet.

However, despite its strengths, few issues need to be addressed. The concepts of "transparency and fairness" are presented as fundamental, but their meaning is contested. In philosophy, "fairness" can be understood in many ways: as equal treatment, as equitable outcomes, or as procedural justice. The Recommendation does not

and perhaps cannot, at a global level commit to a single definition. This leaves a critical gap for policymakers: whose definition of fairness should guide regulation? An algorithm that treats everyone the same; formal equality might still produce discriminatory outcomes because of existing social inequalities. The Recommendation's reliance on these terms without deeper specification means the philosophical hard work is delegated to those who implement it. Also, the Recommendation emphasizes human rights as the framework. Habiba (2023) argue that a human rights framework is inherently anthropocentric as it places all value on the human. It struggles to address the moral status of the AI systems themselves. If AI were to ever achieve a form of consciousness or sentience, as thinkers like Schwitzgebel and Garza (2015) speculate, a purely human-rights-based framework would be ill-equipped to handle the moral claims of the AI itself.

The Role of Ethical Principles in Informing AI Policy

Human Dignity as Foundational

Human dignity, understood in the Kantian tradition as the intrinsic worth of rational beings, grounds several specific constraints on AI development and deployment. Firstly, AI systems must respect the autonomy of those they affect, providing meaningful opportunities for consent and opting out. Secondly, they must be transparent in ways that enable affected individuals to understand and contest decisions. Third, they must not treat persons merely as means to others' ends, however beneficial those ends may be. The principle of human oversight, which appears in virtually all AI ethics frameworks, derives from dignity. If decisions affecting human lives were made entirely by machines without human involvement, those affected would be treated as objects to be processed rather than as persons deserving respect. The EU AI Act's requirement for human oversight of high-risk systems reflects this concern (European Union, 2024). However, human oversight must be meaningful to satisfy dignity's demands. A human operator who merely rubber-stamps algorithmic recommendations without genuine understanding or discretion does not restore dignity. As Ramos-Zaga (2026) argues, "the exclusivity of human moral judgment must guide regulatory frameworks" certain decisions require the exercise of capacities that only humans possess.

The concept of non-delegable agency identifies domains of human activity that cannot be transferred to artificial systems without undermining human personhood (Braun, 2025). These include decisions involving moral responsibility, legal culpability, and profound implications for human welfare. Consider criminal justice, when algorithms predict recidivism or recommend sentences, they are not merely providing information but shaping decisions that affect liberty (Carlson, 2017). If such decisions were fully automated, who would be responsible for their outcomes? The algorithm cannot be held accountable; its creators and operators may be distant from the decision and lack information about its context. The result is an accountability gap that undermines the moral intelligibility of the justice system.

Similar considerations apply in healthcare, where treatment decisions affect patients' lives and well-being. While algorithms can assist diagnosis or recommend treatment options, final decisions should rest with human clinicians who can integrate algorithmic outputs with contextual knowledge and exercise professional judgment (Grote & Berens, 2020). The principle of technological subsidiarity; that decisions should be made at the most local level compatible with their effective resolution, supports keeping such decisions within human hands (Ramos-Zaga, 2026). Non-delegable agency does not preclude using AI as a decision-support tool. It requires, rather, that humans remain ultimately responsible for decisions with moral significance and that they retain genuine discretion to override algorithmic recommendations. This has implications for system design: AI should be designed to augment rather than replace human judgment, providing explanations that enable informed decision-making rather than opaque outputs that must be accepted on faith.

The principle of accountability requires that there be identifiable agents responsible for AI outcomes and that mechanisms exist for holding them to account (Novelli et al., 2024). This seemingly straightforward principle becomes complicated in practice due to the distributed nature of AI development and deployment. Who is responsible when an AI system produces discriminatory outcomes? The developers who wrote the code? The data scientists who selected and prepared training data? The organization that deployed the system without adequate testing? The users who acted on its recommendations? Multiple agents may share responsibility, and the causal chains connecting their actions to harmful outcomes may be long and complex.

Floridi's (2018) distinction between "hard ethics" and "soft ethics" is relevant here. Hard ethics operates at the level of regulatory standards, establishing binding requirements backed by enforcement mechanisms. Soft ethics operates within the space left by regulation, guiding conduct beyond compliance. Both are necessary, but hard ethics is essential for establishing clear lines of accountability and providing remedies for those harmed. The challenge for policy is to design accountability mechanisms that are effective without being unfair. Strict liability such as holding developers responsible for all harms caused by their systems may supress innovation and fail to distinguish between foreseeable and unforeseeable risks. Negligence standards require determining what reasonable care would have required in complex technical contexts. No-fault compensation schemes may spread risk but weaken deterrence. These difficulties do not justify abandoning accountability as a principle. They indicate, rather, that implementing accountability requires careful attention to the structure of AI development and

the distribution of knowledge and control among participants. Policies must be designed with realistic understanding of how AI systems are actually developed and deployed, not idealized assumptions about linear chains of command.

Justice requires that AI systems not perpetuate or amplify existing social inequalities and that their benefits and burdens be distributed fairly (Soni, 2025). This principle has received extensive attention in the literature on algorithmic bias, which documents numerous cases where AI systems have produced discriminatory outcomes (Mittelstadt et al., 2016). The philosophical grounding of fairness principles is contested. Utilitarian approaches would evaluate fairness in terms of overall welfare consequences, while deontological approaches would emphasize rights and entitlements regardless of consequences. Contractualist approaches would ask what principles impartial agents would agree to, while capabilities approaches would focus on whether AI systems enable all persons to develop their capacities. Despite these theoretical differences, there is broad agreement that AI systems should not discriminate on grounds such as race, gender, or disability. The challenge lies in specifying what nondiscrimination requires in practice.

Toward Philosophically Grounded AI Policy

The foregoing analysis suggests that sustainable AI governance requires philosophical clarity about the nature of the entities being governed. Policies that avoid foundational questions about human dignity, moral agency, and the distinction between persons and artifacts risk incoherence and ineffectiveness. This does not mean that policy must resolve all philosophical disputes before proceeding. Reasonable people disagree about consciousness, moral status, and related matters, and policy cannot await consensus. It does mean that policy should be explicit about its assumptions and their implications, and should be revisable as understanding evolves.

The moderate realist tradition provides one coherent framework for such clarity. On this view, human beings possess dignity because they are persons, unified substances with rational nature, while artificial systems, however sophisticated, remain artifacts lacking the interiority that personhood requires (Oderberg, 2007). This ontological distinction grounds normative conclusions: humans are ends in themselves; AI systems are tools to be used for human purposes. Alternative frameworks are possible. Schwitzgebel and Garza (2025) argue that what matters for moral status is psychological and social properties, not biological substrate. On this view, sufficiently advanced AIs might warrant moral consideration equal to humans, and policies should be designed with this possibility in mind. The critical requirement is not that everyone agree on one framework, but that frameworks be philosophically explicit rather than avoiding foundational questions.

Whatever framework one adopts, human dignity must occupy a central place in AI governance. The concept of dignity captures the intuition that human beings are not mere resources to be optimized or data points to be processed, but ends in themselves with inherent worth. From dignity flow several specific requirements for AI policy. AI systems must be designed and deployed in ways that respect human autonomy, providing meaningful opportunities for consent, contestation, and opting out. Also, decisions with significant implications for human welfare must remain subject to meaningful human oversight and control. Then, AI systems must be transparent enough that affected individuals can understand and challenge decisions that affect them.

If ethical principles are to inform AI policy effectively, they must be integrated into policy design from the outset rather than appended as afterthoughts. This requires moving beyond aspirational statements of principles to specification of what principles require in concrete contexts, and to institutional arrangements that can enforce those requirements. On the other hand, UNESCO's Readiness Assessment Methodology offers another model for integration (UNESCO, 2025). By assessing institutional capacities across multiple dimensions, it enables countries to identify gaps and design targeted interventions. Its emphasis on context-specific adaptation recognizes that implementation must be tailored to local conditions rather than imposed uniformly.

What remains underdeveloped in both frameworks is attention to the political economy of AI development, that is, the ways that power, incentives, and institutional structures shape how AI is actually developed and deployed. Ethical principles will have limited impact if they cannot counteract the commercial pressures that drive cutting corners on safety, fairness, and transparency. Regulation is necessary but not sufficient for ethical AI governance. Legal requirements can establish minimum standards, create accountability mechanisms, and provide remedies for those harmed. They cannot, however, ensure that AI is developed and used in ways that promote human flourishing. This limitation has several sources. Firstly, regulation inevitably lags behind technological development; by the time harms are recognized and responses formulated, the technology may have evolved. Secondly, regulation faces enforcement challenges, particularly when violations are difficult to detect or when regulated actors are resourceful in avoiding compliance. Thirdly, regulation addresses only what is legally required, not what is ethically desirable. For these reasons, ethical AI governance requires more than regulation. It requires professional norms and standards that guide conduct beyond compliance. It requires organizational cultures that value ethical deliberation and provide space for it. It requires public education that enables informed citizens to understand and evaluate AI systems.

Conclusion

The governance of artificial intelligence presents one of the most significant policy challenges of the twenty-first century. As AI systems become more capable and more pervasive, the need for frameworks that can guide their development and deployment in ethically defensible ways becomes increasingly urgent. This research has argued that meeting this challenge requires philosophical clarity about foundational questions that current policy frameworks often avoid.

Drawing on the moderate realist tradition, the analysis has proposed that human dignity, non-delegable agency, and technological subsidiarity should guide regulatory design. Human dignity grounds requirements for respect, transparency, and meaningful oversight. Non-delegable agency identifies domains in which human judgment must remain ultimately responsible. Technological subsidiarity directs that decisions be kept at the most local level compatible with effective resolution. Current policy frameworks, such as the UNESCO Recommendation (UNESCO, 2024), have made significant progress in articulating ethical principles and proposing mechanisms for implementation; yet, they exhibit limitations that reflect insufficient attention to foundational questions.

More robust frameworks must be philosophically explicit about their grounding assumptions and normatively justified in their requirements. Stronger enforcement must include mechanisms for monitoring, auditing, and accountability that can counteract the commercial pressures that drive ethical shortcuts. The ultimate goal of AI governance should not be merely to prevent harm but to promote human flourishing—to ensure that AI systems are developed and used in ways that enable human beings to live well.

This requires attending not only to what AI does but also to what it is, and to what we are as the beings who create and use it. Philosophical reflection on these questions is not a luxury to be postponed until after practical problems are solved; rather, it is essential to understanding what the problems are and what their solutions might look like.

References

Aquinas, T. (1947). Summa theologica (Fathers of the English Dominican Province, Trans.). *Benziger Bros*, 1265-1274.

Braun, T. (2025). Liability for artificial intelligence reasoning technologies–a cognitive autonomy that does not help. *Corporate Governance: The International Journal of Business in Society*.

Carlson, A. M. (2017). The need for transparency in the age of predictive sentencing algorithms. *Iowa L. Rev.*, *103*, 303.

Chalmers, D. J. (1997). *The conscious mind: In search of a fundamental theory*. Oxford Paperbacks.

Durrant, M. (2015). *Aristotle's De anima in focus*. Routledge.

Floridi, L. (2018). Soft ethics: its application to the general data protection regulation and its dual advantage. *Philosophy & Technology*, *31*(2), 163-167.

Grote, T., & Berens, P. (2020). On the ethics of algorithmic decision-making in healthcare. *Journal of medical ethics*, *46*(3), 205-211.

Habiba, U. (2023). Protecting the environment with human rights: mechanism rooted in anthropocentric approach. *Human Rights in the Global South (HRGS)*, *2*(2), 106-123.

Jonas, H. (1984). *The imperative of responsibility: In search of an ethics for the technological age*. University of Chicago press.

Kant, I. (1997). Groundwork of the Metaphysics of Morals, ed. Mary Gregor. *Cambridge: Cambridge University Press*, *4*, 422-424.

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific philosophical quarterly*, *64*(4), 354-361.

MacIntyre, A. (1981). After Virtue: A Study in Moral Theory (Notre Dame, Ind. In: University of Notre Dame Press.

Madanchian, M., & Taherdoost, H. (2025). Ethical theories, governance models, and strategic frameworks for responsible AI adoption and organizational success. *Frontiers in Artificial Intelligence*, *8*, 1619029.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big data & society*, *3*(2), 2053951716679679.

Nagel, T. (1974). The Philosophical Review. *What is it Like to Be a Bat*, 435-450.

Nations, U. (1948). *Universal Declaration of Human Rights*. Retrieved March 2 from https://www.un.org/en/about-us/universal-declaration-of-human-rights

Novelli, C., Taddeo, M., & Floridi, L. (2024). Accountability in artificial intelligence: what it is and how it works. *Ai & Society*, *39*(4), 1871-1882.

OCED. (2019). *Recommendation of the Council on Artificial Intelligence*. https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449

Oderberg, D. S. (2007). *Real essentialism*. Routledge.

Ramos-Zaga, F. A. (2026). Normative principles for the legal regulation of artificial intelligence: Human dignity and non-delegable agency. *Revista Justicia & Derecho*, *9*(1), 1-22.

Schwitzgebel, E., & Garza, M. (2015). A defense of the rights of artificial intelligences. *Midwest Studies in Philosophy*, *39*(1), 98-119.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and brain sciences*, *3*(3), 417-424.

Smuha, N. A. (2025). The Cambridge handbook of the law, ethics and policy of artificial intelligence.

Soni, B. (2025). Algorithmic Justice and Social Inequality: The Sociological Impact of Artificial Intelligence on Law and Access to Justice. *Available at SSRN 5913526*.

UNESCO. (2024). *Recommendation on the Ethics of Artificial Intelligence*. Retrieved March 3 from https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence

UNESCO. (2025). *Proceedings from the Global Forum on the Ethics of AI 2025.*

. Retrieved March 4 from https://unesdoc.unesco.org/ark:/48223/pf0000396997

Union, E. (2024). *Regulation (EU) 2024/1689 on artificial intelligence (AI Act)*

Official Journal of the European Union. . Retrieved March 4 from https://artificialintelligenceact.eu/the-act/