

The Dilemmas of Moral Autonomy in the Transfer of Human Consciousness to AI¹

Denis Chiriac

Moldova State University
chiriacdenis@gmail.com

Abstract

This paper examines the dilemmas of moral autonomy in the context of the transfer of human consciousness to AI, focusing on the transhumanist archetype *Homo Evolutus*. Through comparative analysis, the study explores two distinct paradigms: the secular approach, rooted in Kantian philosophy, which considers that moral autonomy can be preserved through transfer; and the religious perspective (Catholic, Protestant, and Orthodox), which argues that the human person is an essential unity of body and soul, making the transfer conceptually impossible. The study highlights three central dilemmas: the authenticity of digital moral reflection, the distribution of responsibility between the biological and digital selves, and the limits of autonomy in human-AI integration. The secular paradigm perpetuates Cartesian dualism, while the theological perspective asserts that consciousness transfer raises fundamental ontological problems that cannot be solved by secular approaches alone. The conclusion emphasizes the need for an interdisciplinary approach that includes theological aspects to clarify what it means to be a moral agent in a technologically altered world.

Introduction and Theoretical Framework

The rapid evolution of artificial intelligence and the growing possibility of “uploading” the mind have brought the question of moral autonomy to the forefront of contemporary philosophical discussions. This study focuses on *Homo Evolutus*, a transhumanist archetype that includes the image of a human being who has overcome biological limitations through fusion with AI. Although philosophical literature on mind enhancement has made progress in identifying technical and metaphysical problems (Schneider 2019), there is still no clearly defined ethical framework for considering moral autonomy in the context of the fusion of humans and AI. The relevant literature is divided into two distinct

camp: a secular approach operating within the framework of philosophical modernism, and a religious approach that interprets humans as created, embodied beings oriented toward transcendence.

The Secular Paradigm of Moral Autonomy

The secular philosophical approach to moral autonomy is rooted in Kantian philosophy, which understands autonomy as the ability of reason to independently establish laws through the categorical imperative. From this perspective, a genuine moral agent is someone who acts in accordance with principles that they can rationally universalize. Harry Frankfurt develops this tradition by arguing that autonomy consists in identifying oneself with one's second-order desires about desires (Frankfurt 1971).

This approach assumes that consciousness transfer can preserve moral autonomy as long as the hierarchical structure of the will remains unchanged. Ágnes Heller adds an important dimension by analyzing everyday life and moral choice. She emphasizes that autonomy is not only a matter of abstract rational reflection, but also the ability to make authentic choices in the specific context of everyday existence (Heller 1978). Even Heller remains within the framework of secular anthropology, which does not fundamentally dispute the possibility of separating consciousness from the biological body.

Emmanuel Levinas, although often cited in transhumanist literature, offers a subtle critique of autonomy as self-determination. For Levinas, ethics precedes ontology, and responsibility to the Other constitutes subjectivity (Levinas 1961). Secular interpretations of Levinas tend to minimize

¹Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

the Jewish theological dimension of his thought, reducing otherness to a purely philosophical category.

The secular paradigm, despite its conceptual sophistication, proceeds from the assumption that consciousness can be separated from its biological substrate without losing its essential characteristics. This dualistic assumption, though seemingly progressive, perpetuates the old Cartesian dualism of *res cogitans* and *res extensa*.

The Theological Perspective on Moral Autonomy

Catholic theology maintains that the human person is a substantial unity of body and soul, realized through action (Wojtyła 1979). Incarnation is not an accidental limitation, but a constitutive dimension of moral experience. Jacques Maritain complements this by asserting that the body is an integral dimension of human personality, not merely an instrument of the soul (Maritain 1947). Joseph Ratzinger emphasizes that human personality, created *ad imaginem Dei*, possesses an ontological dignity based on the Trinitarian relationship (Ratzinger 2000) — a dignity that cannot be artificially reproduced and presupposes participation in divine life.

Protestant theology emphasizes that the *imago Dei* is manifested through moral responsibility exercised *coram Deo*. Karl Barth asserts that human personality exists in relationship with God and others, relationships fundamental to humanity (Barth 1958). C.S. Lewis warns of reducing human nature to an object of technological manipulation (Lewis 1943), arguing that the human mind is involved with the divine Mind, making it impossible to separate consciousness from its transcendent basis. Dietrich Bonhoeffer adds that moral responsibility implies "being for others" — a dimension that cannot be algorithmized (Bonhoeffer 1995).

Orthodox theology offers the most radical critique of secular individualism through the concept of *theosis*. Dumitru Stăniloae asserts that the human person is called to participate in divine life, a process that involves incarnation as a permanent dimension of being (Stăniloae 1978). John Zizioulas develops a relational ontology in which personality is formed through relationships rather than through individual essence (Zizioulas 1985), which makes the transfer of consciousness conceptually impossible. Vladimir Lossky emphasizes that personality cannot be defined conceptually, but only experienced in relationships with God and other

people (Lossky 1976) — this apophatic point of view radically refutes any attempts to reduce personality to transmissible information.

Conclusion

The transfer of human consciousness to AI raises fundamental dilemmas regarding moral autonomy that cannot be resolved by secular approaches alone. A comparative analysis of the secular paradigm and the religious perspective shows that this problem is not only technical but also ontological in nature. The secular paradigm perpetuates Cartesian dualism and assumes that consciousness can be separated from its biological substrate without losing its essential characteristics. In contrast, the theological perspective asserts that the human person is an essential unity of body and soul, and that embodiment is a constitutive dimension of moral experience. The study highlights three central dilemmas: the authenticity of digital moral reflection, the distribution of responsibility between the biological and digital selves, and the limits of autonomy in the integration of humans and AI. The lack of a solid ethical foundation for addressing these issues underscores the need for a holistic, interdisciplinary approach that also includes theological aspects. Resolving the ethical problems of consciousness transfer requires ongoing dialogue and collaboration to clarify what it means to be a moral agent in a technologically altered world.

References

- Barth, K. 1958. *Church Dogmatics III/1: The Doctrine of Creation*. Edinburgh: T&T Clark.
- Bonhoeffer, D. 1995. *Ethics*. Minneapolis: Fortress Press.
- Journal of Consciousness Studies 17(9–10): 7–65.
- Frankfurt, H. 1971. Freedom of the Will and the Concept of a Person. *Journal of Philosophy* 68(1): 5–20.
- Heller, Á. 1978. *A Theory of Feelings*. Assen: Van Gorcum.
- Levinas, E. 1961. *Totalité et infini: Essai sur l'extériorité*. La Haye: Martinus Nijhoff.
- Lewis, C.S. 1943. *The Abolition of Man*. Oxford: Oxford University Press.
- Lossky, V. 1976. *The Mystical Theology of the Eastern Church*. Crestwood: St. Vladimir's Seminary Press.
- Maritain, J. 1947. *La Personne et le bien commun*. Paris: Desclée de Brouwer.
- Ratzinger, J. 2000. *Introduction to Christianity*. San Francisco: Ignatius Press.
- Schneider, S. 2019. *Artificial You: AI and the Future of Your Mind*. Princeton: Princeton University Press.
- Stăniloae, D. 1978. *Teologia Dogmatică Ortodoxă*, Vol. 1. București: IMBOR.
- Wojtyła, K. 1979. *The Acting Person*. Dordrecht: D. Reidel Publishing Company.
- Zizioulas, J. 1985. *Being as Communion*. Crestwood: St. Vladimir's Seminary Press.