John Lennox:
2084: Artificial Intelligence and the Future of Humanity
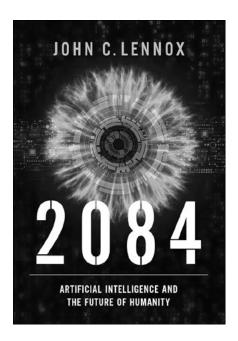
Reviewed by Philipp Koebe

In a world dominated by atheistic scientists and believers in technology, John Lennox presents a differentiated view of artificial intelligence and superintelligence with regard to universal divine characteristics in his work *2084: Artificial Intelligence and the Future of Humanity* . This critique of an atheistic worldview uses biblical analogies and evidence to introduce numerous new arguments into the discussion and broaden the perspective on AI.

John Lennox is a British mathematician, philosopher and Christian apologist, as well as an emeritus fellow of mathematics and the philosophy of science at the University of Oxford. He has published numerous books in which he explores the dialogue between science and faith and argues for the compatibility of science with a Christian worldview. His life and work reflect a unique combination of scientific excellence and deep-rooted faith. In contrast to other works on the subject of AI or superintelligence, Lennox focuses on religious-spiritual arguments, while other protagonists highlight Enlightenment rationalism or longtermism. This is certainly the main added value of this book in a discussion on this topic that has been going on for years.

The initial sections of the manuscript commence by delineating the potential and also the challenges intrinsic to artificial intelligence (AI) and superintelligence. Chapter 1 functions as a preamble to the discourse, introducing for the first time the historical oeuvre of George Orwell, which concurrently serves as an allegory for the book's title (11-13). In this context, explicit allusions are made to the conceivable hazards that a manifestation of AI might pose, concurrently outlining the author's overarching argumentative framework. Lennox underscores Orwell's *1984* (published in 1949) as a seminal work, positing mass surveillance and media information control as pivotal elements of the debate, contending that these are "ideas that nowadays increasingly come up in connection with developments in artificial intelligence (AI)" (13). Lennox further engages in a comparative analysis of dystopian fiction with Aldous Huxley's *Brave New World* and cites Neil Postman's insights from *Amusing Ourselves in Death* (12). Additionally, Lennox draws upon Dan Brown's novel *Origin* to depict a dystopian future, reinforcing his thematic exploration (13).

Chapters 2 and 3 pivot towards inquiries regarding the origin and trajectory of humanity. In pursuit of this, the author delves into a multitude of recent philosophical works so as to scrutinise the essence of AI. Key focal points include the works of Yuval Noah Harari, specifically *Sapiens* (2011) and *Homo Deus* (2015), with recurrent references to perspectives articulated by John Gray and Ray Kurtzweil throughout the monography. Lennox adeptly weaves an analytical narrative, incorporating arguments that substantiate his theses while concurrently addressing those he approaches with a more critical lens.

Chapters 4 and 5 of the monograph delineate diverse facets of weak artificial intelligence and elucidate its transformative impact on human lives. Emphasis is placed on authoritative studies in AI research, particularly those scrutinising the evolving landscape of employment (56-61). The text delves into the functionalities of digital assistants, AI applications in the medical domain, autonomous vehicle technology, and automated marketing (56-61). Lennox adeptly engages the reader through a highly informative and accessible introduction, effectively immersing them in the subject matter. The fifth chapter assumes a critical stance, scrutinising the precarious dimensions inherent in AI utilisation. Lennox scrutinizes extant models, such as the social credit system (which uses AI-facial recognition to assign citizens a social credit score based upon their behaviour and assumed trustworthiness), and issues encompassing data collection, manipulation, and surveillance via social networks and emerging technologies (68-71).The central focus of Chapter 6 is on transhumanism, with an exhaustive exploration of Harari's perspectives. Lennox critically examines Harari's assertions regarding the aspirational realisation of goals such as "a serious bid for human immortality" (86) or the pursuit of "ensuring global happiness" (87) through AI. Chapter 7 revisits the perils associated with AI governance, probing historical antecedents derived from authoritarian events. Chapters 8 to 10 pivot around an examination of human perspectives, particularly the diverse interpretations of morality and ethical values within religious contexts. Lennox posits that moral rationalism becomes "not liveable" in the context of AI integration, drawing parallels with the biblical narrative of the 'Fall of Man'. The concluding chapters (chapters 11-13) intricately interweave analogies to biblical passages, extensively quoting and contextualising them within the thematic discourse. These references, particularly enlightening for individuals without a Christian background, introduce novel arguments into the overarching discussion.

The overarching theme is Lennox's keen interest in the progressive trajectory of artificial intelligence, coupled with a nuanced critique of those who posit technological conquest as a substitute for divine existence. Lennox critically examines transhumanism, contending that the aspiration to supersede divine authority through AI is unfounded. He delineates clear distinctions between human

capacities and the inherent limitations of AI. Lennox underscores the necessity for a higher authority in the development of AI models, akin to the example of China's social credit system. However, he cautions against the utopian notion of omniscient control by a select few, citing historical precedents such as National Socialism and Soviet communism as cautionary tales. Lennox rejects the misguided notion of creating an all-knowing authority to govern the world through a super-intelligent computer, advocating instead for reliance on an existing divinely guided system, as referenced in numerous biblical passages which critique attempts to attain divine abilities.

The book's title *2084* draws a parallel to Orwell's novel *1984*, depicting a dystopian society wherein technological means are employed for the comprehensive monitoring, control, and sanctioning of the entire population. Throughout the book, Lennox asserts that many possibilities portrayed in dystopian novels of the last century are currently technologically feasible in certain domains. However, the absence of a nationwide social credit system in China prompts him to acknowledge that the technical feasibility and societal acceptance of such a system, particularly within the context of the populous People's Republic with 1.3 billion people, remain unverified. Lennox addresses this example at various points. It underpins his thesis that an AI cannot be given the function of a sovereign power or even simulate one. Despite technological advancements and the concomitant wield of power by authoritarian regimes, instances of initial resistance to stringent measures surfaced in China during the coronavirus pandemic. Globally, widespread real-time surveillance and restrictions to freedoms encountered robust criticism and opposition, especially in Western liberal democracies. The implementation of such a system appears improbable in the foreseeable future within these democratic frameworks.

Lennox advocates for a tempered assessment of transhumanism, contending that access to transhuman technologies will not be universally distributed. He posits Silicon Valley as the focal point for these considerations, where tech billionaires and technology-oriented atheists aspire to transcend mortality. However, Lennox suggests that this pursuit might be confined to a certain sphere of influence, potentially failing to attain universal acceptance. He points out that transhumanism will be the goal or dream of a fraction of the population and that an injustice paradigm could therefore emerge. Despite notable progress towards achiev-ing the United Nations' Sustainable Development Goals, Lennox emphasises that a substantial portion of the global population aspire to a better quality of life, free from hunger, exploitation, or subsistence, rather than pursuing immortality. The dichotomy between the pursuit of technological advancements on the one hand and the persistent, fundamental human needs on the other underscores Lennox's nuanced exploration of the societal implications of emerging technologies.

Lennox presents a somber outlook on the potential consequences of realising assumptions surrounding a world dominated by artificial intelligence. Despite extensive reference to documents, historical experiences, and risk assessments from the past century, the overly optimistic expectations of future technologies have proven to be inflated, while excessively pessimistic predictions about humanity have proven untenable. Consequently, the work *2084* can be characterised as a contribution to enhancing literacy in the future. It provides subsequent generations with inspiration, a foundational platform for discussion, and a means to critique prevailing opinions, thereby facilitating the anticipation of desirable futures and aiding decision-making processes. Lennox's arguments offer pathways of action that can play a pivotal role in strategic foresight within political, institutional, or corporate contexts.

In the realm of education, it is advocated that young individuals engage with the inherent conflicts surrounding superintelligence and the associated technologies and data processing. Furthermore, the work has the potential to contribute significantly to religious education across all denominations by fostering a harmonious relationship between technological progress and spiritual attitudes. The integration of Lennox's insights into educational curricula holds promise for cultivating a nuanced understanding of the ethical and societal implications of emerging technologies, thereby promoting a more informed and responsible engagement with the challenges posed by AI.