SUPERINTELLIGENCE

Why Humanity Would Risk Extinction To Create Its Successor

Pontsho B. Mokoena

ABSTRACT

Purpose: Artificial intelligence (AI) as a discipline is not a new concept, its history dates back to the 1950's. We have seen the sharp rise in the prominence of AI over just the last eighteen months alone because of the wide availability of large datasets, sophistication of algorithms and the advancements in computing power and hardware which presented as limitations before (Russell et al.,2021). These perceived limitations had led researchers and scientists to have a view that the evolution of AI would be steady and measured and that AI evolving beyond what we have today, being artificial narrow intelligence (ANI) into its successors (artificial general intelligence and artificial superintelligence) was likely to occur circa 2045 and 2075 respectively. The recent advancements have however caused researchers and scientists to restate the coming into being of both AGI and ASI to be much sooner, possibly before 2030 (Bostrom, 2014; Kurzweil, 2005). Artificial general intelligence is when machine intelligence mirrors human intelligence to par. The concerns that would be expected regarding AGI would be data protection issues, anti-monopoly regulations and ultimately, the control and harbouring of knowledge to a limited few groups, persons or governments (European Commission, 2021; OECD, 2019). Artificial superintelligence on the other hand is when machine intelligence exceeds human intelligence across every domain. The concerns that would be expected regarding ASI would be centralised control of AI, blocking of open source AI and ultimately, extinction risks. The purpose of this study is to explore the psychology and behavioural factors driving the pursuit of ASI, despite its known risks. Simplistically, why human beings, having an appreciation of the potential harm that superintelligence could create, would invest so much resources, time, money, brain power of some of the smartest humans on the planet to expedite the formation of a superintelligence, when the same machine could be the cause of the cessation of the human race. Artificial Superintelligence (ASI) represents both the pinnacle of human achievement and the edge of our self-destruction. Rationally, we understand that such power could render us obsolete, even extinct. But psychologically, the craving for control, recognition, and legacy outpaces that caution.

Research Questions: The key research questions framing the study are:

1.Can superintelligence outwit human intelligence?

2.If superintelligence can indeed outwit human intelligence, why would humanity risk extinction to create its successor?

Methodology: A thematic analytical approach is adopted to interpret the results of the study survey. The data was collected across 30 participants using a cross sectional method of varying ages, industries and AI knowledge.

Implications: The findings demonstrate that humanity is not oblivious to the risks of a superintelligence, but rather that humanity's desire and craving for ambition, power and control far outweighs humanity's ability to rationalise the fear of extinction and being obsolete (Yudkowsky, 2008).

Keywords: Superintelligence, AI Power, AI Control, Ambition Paradox, Human Extinction

Paper Type: Extended Abstract of a Research Study

REFERENCES

Bostrom, N. (2014) Superintelligence: Paths, dangers, strategies. Oxford: Oxford University Press.

European Commission. (2021) Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act). Brussels: European Commission.

Kurzweil, R. (2005) The singularity is near: When humans transcend biology. New York: Viking.

Russell, S. and Norvig, P. (2021) Artificial intelligence: A modern approach. 4th edn. Harlow: Pearson Education.

Yudkowsky, E. (2008) 'Artificial intelligence as a positive and negative factor in global risk', in Bostrom, N. and Ćirković, M. (eds.) Global catastrophic risks. Oxford: Oxford University Press, pp. 308–345.

SUPPLEMENTARY INFORMATION

The full paper, covering the literature review, methodology, analysis, findings and implications can be requested directly from the author.

AUTHOR

Dr. Pontsho B. Mokoena is a South African Chief Risk and Insurance Officer, Academic Researcher and Thought Leader who specialises in the fields of applied mathematics and data analytics. She holds a Doctoral Degree in Business Administration specialising in Predictive Risk Management from Paris School of Business (France), a Master's Degree in Actuarial Science from the University of Leicester (UK), an Advanced Insurance Programme with the University of South Africa (SA) and a Bachelor's Degree in Insurance & Risk Management and Business Finance from the University of the Witwatersrand (SA). She has recently completed an Artificial Intelligence Programme with Oxford University, SAID Business School (UK).

She has 20 years corporate experience across different industries occupying senior leadership and board roles in the specialised fields of enterprise risk management, short term and long-term insurance, employee benefits, healthcare, investment advisory, analytical modelling and more recently artificial intelligence. In her role as the Founder and Executive Director of KHAUTA Risk Advisory, she is responsible for leading and setting the strategic direction of the boutique consulting and advisory firm which brings together two decades of corporate experience, technical acumen, and strategic insight to help organisations anticipate, manage, and transform risk into opportunity. She can be contacted via her LinkedIn Account: Dr_Pontsho_B_Mokoena