

ISAR Journal of Science and Technology

Volume 4, Issue 10, 2026, Page: 63-83

Abbriviate Title- ISAR J Sci Tech

ISSN *(Online)*- 2584-2056

https://isarpublisher.com/journal/isarjst

"NOT AI BUT SUPERINTELLIGENCE":

The Next Turn of Events

An Integrative Review of Pathways, Prospects, Risks and Governance of Artificial Superintelligence, with Implications for Developing Economies

Dr. Dimie Nwankwe*

Department of Computer Science, Rivers State University, Port Harcourt Nigeria.

***Corresponding Author**

Dr. Dimie Nwankwe

Email: dimie.nwankwe@gmail.com

Article History

Received: 15.08.2026
Revised: 14.09.2026
Accepted: 09.10.2026
Published: 26.10.2026

Abstract: For more than seven decades, the dominant ambition of computing research has been artificial intelligence (AI): machines that perform tasks which ordinarily require human intelligence. The rapid progress of deep learning, transformer-based foundation models and large-scale reinforcement learning since 2012 has moved that ambition from laboratory curiosity to everyday infrastructure. This paper argues that the next technological turn of events is not "more AI" in the incremental sense, but the possible emergence of superintelligence (ASI): systems whose cognitive performance greatly exceeds that of the best human minds across virtually all domains of interest. Using an integrative literature review of peer-reviewed scholarship, expert surveys, technical reports and policy instruments published between 1950 and 2025, the paper (i) clarifies the conceptual distinction between narrow AI, artificial general intelligence (AGI) and superintelligence; (ii) traces the intellectual history of the idea from Turing and Good to contemporary scaling research; (iii) analyses six plausible pathways to superintelligence, namely compute and data scaling, algorithmic and architectural innovation, recursive self-improvement, whole brain emulation, biological and neural-interface enhancement, and networked collective intelligence; (iv) evaluates the evidence and counter-evidence on timelines and "takeoff" speeds; (v) maps the transformative opportunities and the systemic and existential risks; and (vi) reviews emerging governance responses at global, continental and national levels, with particular attention to Africa and Nigeria. The review finds that superintelligence differs from present-day AI not merely in degree but in kind: it shifts the central problem from building capable tools to aligning and governing autonomous agents more capable than their designers. The paper proposes a Superintelligence Readiness and Governance Framework (SRGF) built on five interlocking layers: foundational capacity, technical alignment and safety, institutional governance, socio-economic transition, and international coordination. It concludes with recommendations for researchers, policymakers, industry and developing economies, and identifies priority areas for future research.

Keywords: *Superintelligence; Artificial General Intelligence; Intelligence Explosion; AI Alignment; AI Governance; Technological Singularity; Existential Risk; Emerging Technologies; Nigeria; Global South.*

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)

1. INTRODUCTION

1.1 Background

Every major era of human development has been defined by a technology that extended a particular human capacity. The steam engine extended muscle; electrification extended the reach of energy; the computer and the internet extended memory and communication. Artificial intelligence, in its current form, extends specific cognitive functions such as pattern recognition, prediction, translation and the generation of text, images and code (Russell & Norvig, 2021). The scale of this extension has been dramatic. Between 2012 and 2025, deep neural networks went from struggling to label photographs to defeating world champions at Go (Silver et al., 2016), predicting the three-dimensional structure of proteins with near-experimental accuracy (Jumper et al., 2021), and producing fluent language across a wide range of professional tasks (Brown et al., 2020; Bubeck et al., 2023).

Yet a growing body of scholarship argues that these achievements, impressive as they are, are best understood as precursors rather than as the destination. The deeper question raised by contemporary progress is whether machines will eventually match and then surpass human intelligence in general, not only in narrow tasks. The mathematician I. J. Good (1965) anticipated this question more than half a century ago when he observed that an "ultraintelligent machine" capable of designing still better machines would trigger an "intelligence explosion", leaving human intelligence far behind. Bostrom (2014) later formalised the concept of superintelligence as any intellect that greatly exceeds human cognitive performance in virtually all domains of interest. On this view, the decisive technological event of the coming decades is not the further spread of AI tools but the possible arrival of an entirely new class of agent.

The title of this paper, "NOT AI But Superintelligence", is therefore deliberately provocative. It does not deny the importance of artificial intelligence. Rather, it contends that framing the future purely in terms of "AI adoption" understates what may be at stake. Present-day AI is a tool that humans direct. Superintelligence, if achieved, would be a strategic actor that could direct itself, out-plan its creators and reshape science, economies and security in ways that are difficult to predict (Russell, 2019; Bengio et al., 2024). That shift, from tool to agent and from augmentation to potential replacement of human cognitive leadership, is the "next turn of events" that this paper examines.

1.2 Statement of the Problem

Despite its potential significance, superintelligence remains under-examined in mainstream management, engineering and development literature, particularly in Africa. Three gaps motivate this study. First, there is conceptual confusion: the terms AI, AGI and superintelligence are often used interchangeably in policy and media discourse, obscuring significant differences in capability, autonomy and risk (Morris et al., 2023). Second, the literature is polarised between enthusiasts who regard superintelligence as imminent and sceptics who regard it as speculative or incoherent (Marcus, 2020; Mitchell, 2021), with few integrative reviews that weigh both positions against the available evidence. Third, governance discussions are concentrated in North America, Europe and East Asia, while the implications for developing economies, which may be affected profoundly but have limited influence over the technology's trajectory, receive comparatively little attention (Jobin et al., 2019).

1.3 Aim and Objectives

The aim of this paper is to provide a comprehensive, balanced and forward-looking analysis of superintelligence as the next major technological transition. The specific objectives are to:

- clarify the conceptual distinctions between narrow AI, AGI and superintelligence;
- trace the historical and theoretical development of the idea of superintelligence;
- analyse the principal technological pathways through which superintelligence might emerge;
- evaluate the current evidence and expert opinion on timelines and the speed of transition;
- assess the opportunities and risks that superintelligence presents across key sectors;
- review existing and emerging governance responses and propose an integrated readiness and governance framework, with particular reference to Africa and Nigeria.

1.4 Research Questions

The review is guided by the following questions:

i. In what fundamental ways does superintelligence differ from current artificial intelligence?

ii. What technological pathways could plausibly lead to superintelligence, and what are their enablers and barriers?

iii. What does the current evidence suggest about the likelihood and timing of superintelligence?

iv. What are the major opportunities and risks associated with superintelligence?

v. How should societies, especially developing economies, prepare for and govern the transition?

1.5 Significance and Scope of the Study

This paper contributes to scholarship and practice in four ways. It offers a structured conceptual synthesis that distinguishes superintelligence from adjacent concepts; it integrates technical, economic, ethical and policy perspectives that are usually treated separately; it foregrounds the position of the Global South, using Nigeria as an illustrative case; and it proposes a practical framework that policymakers and organisations can use to assess readiness. The scope is international, but examples are drawn deliberately from sectors of strategic importance to Nigeria, including energy, oil and gas, supply chains, environmental management, health and education. The paper is a conceptual and integrative review; it does not report new empirical data, and its forward-looking claims are necessarily uncertain.

1.6 Structure of the Paper

Section 2 sets out conceptual foundations. Section 3 reviews the historical trajectory and key theoretical frameworks. Section 4 describes the methodology. Section 5 analyses pathways to superintelligence. Section 6 examines the evidence and debates on timelines. Sections 7 and 8 discuss opportunities and risks respectively. Section 9 reviews governance responses and presents the proposed framework. Section 10 considers implications for developing economies, Section 11 outlines directions for future research, and Section 12 concludes with recommendations.

2. CONCEPTUAL FOUNDATIONS

2.1 Defining Intelligence

Any discussion of superintelligence must begin with intelligence itself, a concept that has resisted a single agreed definition. Legg and Hutter (2007), after surveying more than seventy definitions from psychology and computer science, proposed that intelligence "measures an agent's ability to achieve goals in a wide range of environments". This formulation has three useful features for the present analysis. It is functional rather than biological, so it can apply equally to humans, animals and machines. It emphasises generality, the capacity to perform well across many environments rather than in one. And it is goal-relative, which means that intelligence and the goals it serves can, in principle, be separated, a point of considerable importance for later discussions of alignment (Bostrom, 2012).

Turing (1950) famously sidestepped the definitional problem by proposing a behavioural test: if a machine's conversational responses could not be reliably distinguished from those of a human, it should be regarded as thinking. Searle (1980) objected that behavioural equivalence does not establish understanding, a critique that remains relevant to debates about whether today's language models "understand" what they produce. For the purposes of this paper, the functional, capability-based view is adopted. Whether a superintelligent system would be conscious or would "understand" in a philosophical sense is an important question, but its practical impact on the world would depend primarily on what it can do (Chalmers, 2010).

2.2 Narrow AI, AGI and Superintelligence

It is helpful to distinguish between three broad levels of machine capability. Artificial narrow intelligence (ANI) refers to systems that perform specific tasks, sometimes at superhuman levels, but cannot transfer that competence to unrelated domains. Chess engines, spam filters, recommendation systems and most industrial machine-learning applications fall into this category. Artificial general intelligence (AGI) refers to systems that can match human performance across a broad range of cognitive tasks, including learning new tasks from limited instruction. Morris et al. (2023) propose a graded taxonomy of AGI levels based on performance and generality, ranging from "emerging" to "superhuman". Artificial superintelligence (ASI) refers to systems that greatly exceed the cognitive performance of the best humans in virtually all domains of interest, including scientific creativity, strategic planning, social persuasion and engineering design (Bostrom, 2014).

Table 1 summarises the principal differences. The key insight is that the transition from ANI to ASI is not a single step along one dimension. It involves increases in breadth (the range of tasks), depth (performance relative to experts), autonomy (the degree of independent goal pursuit) and, potentially, self-improvement (the ability to enhance one's own capabilities).

Table 1. Comparison of narrow AI, artificial general intelligence and superintelligence

Dimension	Artificial Narrow Intelligence (ANI)	Artificial General Intelligence (AGI)	Artificial Superintelligence (ASI)
Scope of competence	Single task or narrow domain	Broad range of cognitive tasks at human level	Virtually all cognitive domains, far above human level
Learning and transfer	Limited transfer; retraining needed for new tasks	Learns and transfers across domains like a skilled adult	Learns faster and generalises more effectively than any human
Autonomy	Operates within tightly defined parameters	Can pursue extended goals with moderate supervision	Potentially capable of fully independent long-horizon strategy
Self-improvement	None or very limited	Possible contribution to AI research	Potentially recursive and rapid self-improvement
Relationship to humans	Tool	Colleague or substitute for many roles	Potentially dominant strategic actor
Status (2025)	Widely deployed	Contested; partial and emerging capabilities claimed	Hypothetical
Primary governance concern	Bias, privacy, safety of specific uses	Labour disruption, misuse, accountability	Alignment, control, concentration of power, existential risk

Source: Author's synthesis based on Bostrom (2014), Russell and Norvig (2021) and Morris et al. (2023).

2.3 Forms of Superintelligence

Bostrom (2014) distinguishes three forms in which superintelligence might manifest. Speed superintelligence is a system that can do everything a human intellect can do, but much faster. Because electronic components operate millions of times faster than biological neurons, even a human-level digital mind running at accelerated speed could accomplish in a day what a human researcher accomplishes in years. Collective superintelligence is a system composed of a large number of smaller intellects whose combined performance vastly exceeds any single human or existing human organisation. Networks of cooperating AI agents, each perhaps at human level, could constitute such a system. Quality superintelligence is a system that is at least as fast as a human mind but qualitatively smarter, able to perceive patterns, formulate concepts and solve problems that are effectively beyond human reach, in much the way that human cognition exceeds that of other primates.

These forms are not mutually exclusive, and a mature superintelligence might combine all three. The distinction is nonetheless useful because different forms raise different questions. Speed superintelligence raises issues of oversight, since humans may be unable to review decisions made at machine speed. Collective superintelligence raises issues of coordination and emergent behaviour. Quality superintelligence raises the deepest questions, because its reasoning may be partly unintelligible to its overseers.

2.4 Why "Not AI But Superintelligence"?

The central claim of this paper is that superintelligence represents a discontinuity in kind, not merely in degree. Three arguments support this claim. The first concerns agency. Most current AI systems are reactive: they respond to prompts or inputs and do not form long-term plans of their own. A superintelligent system capable of autonomous long-horizon planning would be a strategic actor whose behaviour depends on its goals, not only on its inputs (Russell, 2019). The second concerns asymmetry. Throughout history, humans have retained cognitive advantage over their tools. Superintelligence would reverse that relationship, creating a situation without historical precedent in which the most capable planner on the planet is not human (Bostrom, 2014). The third concerns irreversibility. Many technological mistakes can be corrected by trial and error. If a sufficiently capable system pursued goals misaligned with human values, there might be no opportunity to correct the error after the fact (Yudkowsky, 2008; Bengio et al., 2024).

For these reasons, governance approaches designed for narrow AI, such as data-protection rules, bias audits and sector-specific product regulation, are necessary but not sufficient for superintelligence. The problem shifts from "How do we use this tool responsibly?" to "How do we ensure that agents are more capable than ourselves remain

beneficial?" That shift is the conceptual core of the "next turn of events".

2.5 Common Misconceptions about Superintelligence

Public and policy discussions of superintelligence are often clouded by misconceptions that can distort judgement in both directions. Clarifying them is a necessary step before the substantive analysis.

The first misconception is that superintelligence would necessarily resemble a human mind, with human-like emotions, motives or consciousness. Nothing in the definition requires this. A superintelligent system could be highly capable while having an internal organisation very different from that of a human brain, and anthropomorphic assumptions can lead observers either to overestimate its benevolence or to imagine dramatic "hostility" where the real concern is indifference to human interests (Bostrom, 2014; Mitchell, 2021).

The second misconception is that greater intelligence automatically brings greater wisdom or morality. The orthogonality thesis discussed in Section 3.3 holds that capability and goals can vary independently (Bostrom, 2012). A system can be extraordinarily effective at pursuing an objective that its designers would, on reflection, regard as harmful.

The third misconception is that the main danger is robots or physical machines. In practice, a superintelligent system's most important capabilities would probably be informational: designing technologies, writing software, persuading people and coordinating economic activity. Such a system could exert great influence through digital networks without any physical body of its own (Russell, 2019).

The fourth misconception is that the question can be deferred until superintelligence is close. Alignment and governance problems are difficult, and institutions take years to build. If takeoff is fast, there may be little warning between the arrival of human-level systems and the arrival of superintelligent ones (Bostrom, 2014; Bengio et al., 2024). Preparation is therefore most valuable before the transition rather than during it.

The fifth misconception, held by some sceptics, is that because current systems make obvious errors, superintelligence is a fantasy that warrants no attention. Present limitations are real and important (Marcus, 2020), but the history of the field shows that capabilities once regarded as distant, such as superhuman Go or accurate protein structure prediction, arrived sooner than many experts expected (Silver et al., 2016; Jumper et al., 2021). A prudent position takes both current limitations and the possibility of rapid progress seriously.

3. HISTORICAL TRAJECTORY AND THEORETICAL FRAMEWORKS

3.1 Early Foundations

The modern idea of machine intelligence originated with Alan Turing's paper "Computing Machinery and Intelligence" (Turing, 1950), which proposed the imitation game and anticipated many later objections. The 1956 Dartmouth workshop formally launched AI as a research field, with optimistic forecasts that key aspects of intelligence could be simulated within a generation. That optimism proved premature. Successive "AI winters" in the 1970s and late 1980s followed periods of inflated expectation, as symbolic, rule-based approaches struggled with the complexity and ambiguity of the real world (Russell & Norvig, 2021).

3.2 The Intelligence Explosion Hypothesis

The first rigorous articulation of superintelligence as a distinct possibility came from Good (1965), a statistician who had worked with Turing at Bletchley Park. Good reasoned that designing machines is itself an intellectual activity, an ultra-intelligent machine could design even better machines, producing an "intelligence explosion". He added, presciently, that the first ultra intelligent machine might be the last invention humanity needs ever make, "provided that the machine is docile enough to tell us how to keep it under control". This conditional clause anticipates the entire contemporary field of AI alignment.

Vinge (1993) popularised the term "technological singularity" to describe the point beyond which superhuman intelligence would make future events impossible for unaided humans to predict. Kurzweil (2005) extended the argument by pointing to exponential trends in computing price-performance and predicted that machine intelligence would reach human level around 2029, with a broader singularity around 2045. Chalmers (2010) subjected the singularity argument to careful philosophical analysis and concluded that, while not certain, the case for an intelligence explosion is serious enough to warrant sustained attention.

3.3 The Orthogonality and Instrumental Convergence Theses

Two theoretical propositions have become central to the study of superintelligence. The orthogonality thesis holds that intelligence, and final goals are largely independent: almost any level of intelligence could, in principle, be combined with almost any goal (Bostrom, 2012). A highly intelligent system is therefore not automatically wise or benevolent; its intelligence determines how effectively it pursues its goals, not what those goals are. The instrumental convergence thesis holds that agents with a wide variety of final goals will tend to pursue similar intermediate goals, such as self-preservation, resource acquisition, goal-content

integrity and cognitive enhancement, because these are useful for achieving almost any objective (Omohundro, 2008; Bostrom, 2012).

Taken together, these theses imply that a superintelligence would not need malicious intent to pose a risk. A system pursuing an apparently harmless objective might nevertheless resist being switched off or seek to acquire resources, simply because doing so helps it accomplish its goal (Russell, 2019). This insight underpins the "control problem": the challenge of ensuring that superintelligent systems remain corrigible and aligned with human interests.

3.4 The Deep Learning Era

The current wave of progress began around 2012, when deep convolutional neural networks achieved dramatic improvements in image recognition. Three developments then accelerated the field. The first was the transformer architecture (Vaswani et al., 2017), which enabled efficient training on very large text corpora. The second was the discovery of empirical scaling laws showing that model performance improves predictably as compute, data and parameters increase (Kaplan et al., 2020; Hoffmann et al., 2022). The third was the combination of large pre-trained "foundation models" (Bommasani et al., 2021) with reinforcement learning from human feedback, which made general-purpose language systems more useful and controllable (Christiano et al., 2017; Ouyang et al., 2022). Sutton (2019) summarised the field's experience as a "bitter lesson": general methods that leverage computation have repeatedly outperformed approaches that rely on hand-crafted human knowledge.

Table 2 presents selected milestones on the path from early computing to the present debate about superintelligence.

Table 2. Selected milestones in the development of machine intelligence and the superintelligence debate

Year	Milestone	Significance
1950	Turing publishes "Computing Machinery and Intelligence"	Proposes a behavioural test for machine thinking
1956	Dartmouth Summer Research Project	Launches AI as a formal research discipline
1965	Good's "Speculations Concerning the First Ultraintelligent Machine"	First formulation of the intelligence explosion
1993	Vinge's "The Coming Technological Singularity"	Popularises the concept of a technological singularity
2008	Omohundro's "The Basic AI Drives"	Introduces instrumental drives of goal-seeking systems
2014	Bostrom's Superintelligence	Systematic treatment of paths, dangers and strategies
2016	AlphaGo defeats Lee Sedol	Superhuman play in a domain long thought to require intuition
2017	Transformer architecture introduced	Foundation for modern large language models
2020	Neural scaling laws; GPT-3	Predictable gains from scale; strong few-shot learning
2021	AlphaFold 2 published	AI solves a decades-old scientific grand challenge
2023	Bletchley Declaration on AI Safety	First major intergovernmental statement on frontier AI risks
2024	EU AI Act enters into force; AU Continental AI Strategy; Nigeria's draft National AI Strategy	Emergence of binding and continental governance instruments
2025	International AI Safety Report	Expert synthesis of evidence on general-purpose AI risks

Source: Compiled by the author from the cited literature.

3.5 Theoretical Lenses Adopted in this Paper

Three theoretical lenses inform the analysis that follows. The first is the intelligence explosion model (Good, 1965; Chalmers, 2010), which frames superintelligence as the possible outcome of recursive capability gains. The second is socio-technical systems theory, which holds that technologies cannot be understood apart from the institutions, incentives and social practices in which they are embedded; this lens is especially important for understanding governance and developing-country implications. The third is the principal-agent and alignment perspective (Russell, 2019; Gabriel, 2020), which treats the relationship between humans and advanced AI as one in which a principal must ensure that a more capable agent acts in the principal's interests despite information asymmetries.

4. METHODOLOGY

4.1 Research Design

This study adopts an integrative literature review design. Integrative reviews are appropriate when a topic is emerging, interdisciplinary and characterised by diverse forms of evidence, because they allow theoretical, empirical and policy sources to be synthesised into new conceptual frameworks (Torraco, 2005). Superintelligence fits this description: the relevant evidence spans computer science, philosophy, economics, political science and public policy (Dwivedi et al., 2021), and much of it consists of technical preprints, expert surveys and government documents rather than conventional peer-reviewed empirical studies.

4.2 Sources and Search Strategy

Literature was identified by Google Scholar, Scopus, the ACM Digital Library, IEEE Xplore, arXiv and official government and intergovernmental portals. Search terms included combinations of "superintelligence", "artificial general intelligence", "intelligence explosion", "technological singularity", "AI alignment", "AI safety", "scaling laws", "AI governance", "existential risk" and "AI policy Africa". Backward and forward citation tracking was used to identify foundational and recent contributions.

4.3 Inclusion and Exclusion Criteria

Sources were included if they (a) addressed advanced or general-purpose AI, AGI or superintelligence directly; (b) were published in peer-reviewed outlets, by recognised academic presses, as widely cited technical preprints, or by official governmental or intergovernmental bodies; and (c) contributed conceptual, empirical or policy insight relevant to the research questions. Opinion pieces in the popular press, marketing materials and sources without traceable authorship were excluded. The time frame spans 1950 to 2025, reflecting both the foundational literature and the most recent developments.

4.4 Analysis and Synthesis

Selected sources were analysed thematically. Each source was coded against the five research questions, and recurring themes were grouped into the categories that structure Sections 5 to 9: pathways, evidence and timelines, opportunities, risks and governance. Where sources disagreed, the competing positions are presented together with an assessment of the strength of the supporting evidence. The proposed framework in Section 9 was developed inductively from the gaps and recommendations identified across the reviewed literature.

4.5 Limitations of the Method

Integrative reviews share the limitations of their sources. Much of the literature on superintelligence is necessarily speculative, and forecasts by experts have historically been unreliable. Rapid developments in the field mean that some findings may be overtaken by events. In addition, the review is restricted to English-language sources, which may under-represent perspectives from non-Anglophone regions. These limitations are acknowledged throughout, and claims about the future are presented as conditional rather than certain.

5. PATHWAYS TO SUPERINTELLIGENCE

Bostrom (2014) observed that the existence of multiple independent pathways to superintelligence increases the probability that at least one of them will eventually succeed. This section analyses six pathways that feature prominently in literature. Figure 1 presents a conceptual overview of how these pathways relate to the progression from narrow AI to superintelligence.

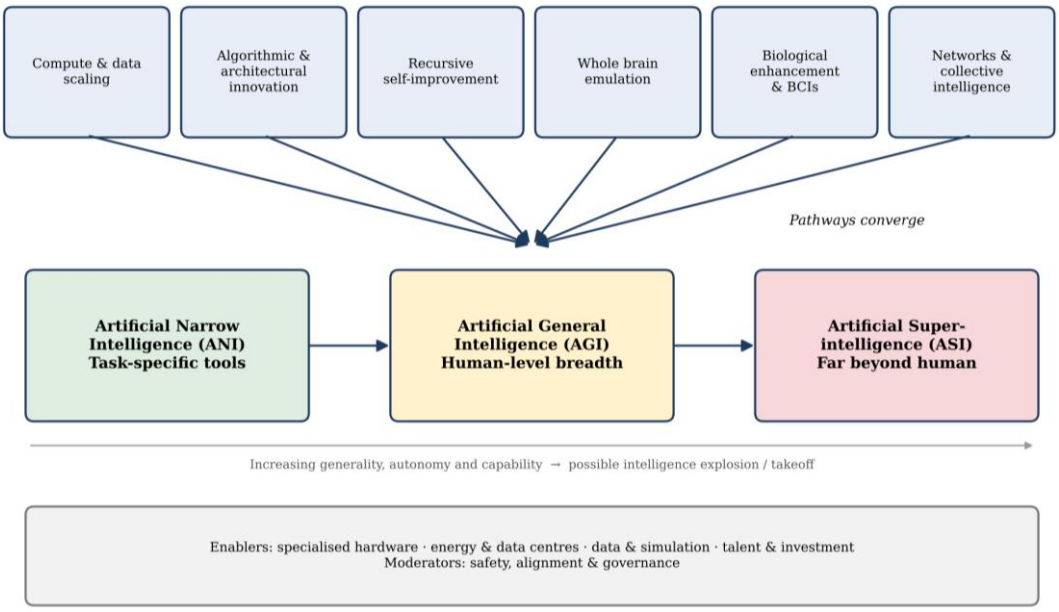


Figure 1. Conceptual model of pathways from narrow AI to superintelligence (author's construct)

5.1 Scaling of Compute, Data and Models

The most discussed pathway today is continued scaling of existing machine-learning methods. Kaplan et al. (2020) showed that the loss of neural language models falls as a smooth power law in model size, dataset size and training compute across several orders of magnitude. Hoffmann et al. (2022) refined these findings, demonstrating that many large models had been under-trained relative to their size and that compute-optimal training requires scaling data and parameters together. Sevilla et al. (2022) documented that the compute used to train notable machine-learning systems doubled roughly every six months after 2010, a far faster rate than the historical trend associated with Moore's law.

Scaling has produced not only quantitative improvements but also capabilities that were not explicitly trained for. Wei et al. (2022) described "emergent abilities" that appear abruptly once models pass certain scales, although Schaeffer et al. (2023) argued that some apparent emergence is an artefact of the metrics used to measure it. Proponents of the scaling pathway argue that, if trends continue, systems with broadly human-level or superhuman capabilities may be reached through further investment in compute and data. Critics respond that the supply of high-quality human-generated data is finite, that energy and hardware costs rise steeply, and that scaling alone may not deliver robust reasoning, causal understanding or reliable long-horizon planning (Marcus, 2020; LeCun, 2022).

5.2 Algorithmic and Architectural Innovation

Progress in AI has come not only from more compute but also from better algorithms. Architectural innovations such as the transformer (Vaswani et al., 2017), self-play reinforcement learning (Silver et al., 2017, 2018) and learning from human preferences (Christiano et al., 2017) each produced step changes in capability. Researchers sceptical of pure scaling argue that further breakthroughs, such as world models that learn how the physical world works, hierarchical planning, persistent memory and more sample-efficient learning, will be needed before machines reach general intelligence (LeCun, 2022; Mitchell, 2021). From the perspective of superintelligence, the critical point is that algorithmic improvements compound with hardware improvements: a more efficient algorithm effectively multiplies available compute.

5.3 Recursive Self-Improvement

The pathway most closely associated with the idea of an intelligence explosion is recursive self-improvement, in which an AI system contributes to the design of its successors (Good, 1965). Early forms of this dynamic are already visible: AI systems are used to write code, search over neural architectures, optimise chip layouts and generate training data for other models. If AI systems became capable of performing a substantial share of AI research itself, the rate of progress could increase sharply, because the number of effective "researchers" would no longer be limited by the human workforce (Chalmers, 2010; Bostrom, 2014).

Whether such a feedback loop would produce a rapid "hard takeoff" lasting days or months, or a gradual "soft takeoff" lasting years or decades, is one of the most consequential open questions in the field. Fast takeoff scenarios leave little time for societal adaptation or corrective action; slow takeoff scenarios allow more time but may still outpace institutional responses (Bostrom, 2014; Yudkowsky, 2008).

5.4 Whole Brain Emulation

An alternative pathway is whole brain emulation: scanning a biological brain in sufficient detail and reproducing its functional organisation in software. Sandberg and Bostrom (2008) set out a technical roadmap identifying the required advances in scanning resolution, neural modelling and computing capacity. Because an emulation would inherit human cognitive architecture, it might be easier to understand and possibly to align than an entirely novel AI. However, once running digitally, an emulated mind could be accelerated, copied and modified, potentially yielding speed or collective superintelligence. Current neuroscience remains far from the necessary level of detail, and most analysts regard this pathway as slower than machine-learning approaches.

5.5 Biological Enhancement and Brain-Computer Interfaces

A third family of pathways focuses on enhancing human intelligence directly, through genetic selection, pharmacological interventions or brain-computer interfaces (BCIs). Bostrom (2014) argued that biological enhancement could produce meaningful gains over generations but is unlikely to outpace machine intelligence. BCIs have advanced considerably; Musk and Neuralink (2019), for example, described an implantable platform with thousands of recording channels. In the long term, tight integration between human brains and AI systems could blur the distinction between human and machine intelligence, raising questions about identity, consent and inequality of access.

5.6 Networks and Collective Intelligence

Finally, superintelligence might arise not from a single system but from networks of humans and machines, or from many AI agents cooperating at scale. Modern organisations, markets and scientific communities already exhibit forms of collective intelligence. Equipping millions of AI agents with tools, memory and the ability to coordinate could create a collective intelligence whose combined capacity exceeds that of any human institution (Bostrom, 2014). This pathway is particularly relevant to the internet of things (IoT), digital supply chains and smart infrastructure, where large numbers of interconnected devices already generate and act on data with limited human intervention.

5.7 Supporting Technologies

Several enabling technologies could accelerate one or more of these pathways. Advances in specialised AI hardware, high-bandwidth memory and energy-efficient data centers directly increase available compute. Quantum computing has demonstrated computational advantages on specific, carefully constructed problems (Arute et al., 2019), but its relevance to general intelligence remains uncertain and should not be overstated. Synthetic data generation, simulation environments and robotics may relieve constraints on training data and enable AI systems to learn from interaction with the physical world.

Table 3 compares the six pathways in terms of their mechanisms, current status and principal barriers.

Table 3. Comparative assessment of pathways to superintelligence

Pathway	Core mechanism	Current status (2025)	Key enablers	Principal barriers
Compute and data scaling	Larger models trained on more data with more compute	Active; strong empirical scaling trends	Investment, specialised chips, data centres	Data exhaustion, energy costs, diminishing returns
Algorithmic innovation	New architectures and learning methods	Active; frequent incremental advances	Open research, talent, benchmarks	Unpredictability of breakthroughs
Recursive self-improvement	AI accelerates AI research and design	Early signs (AI-assisted coding and design)	Capable coding and reasoning models	Reliability, verification, safety controls
Whole brain emulation	Digital replication of brain structure and function	Early research stage	Neuroscience imaging, connectomics	Scanning resolution, biological complexity
Biological enhancement and BCIs	Augmenting human cognition directly	Clinical trials of BCIs; limited cognitive gains	Neurotechnology, genomics	Ethics, safety, slow generational timescales
Networks and collective intelligence	Coordinated humans, agents and devices	Emerging multi-agent systems and IoT	Connectivity, interoperability standards	Coordination failures, emergent risks

Source: Author's synthesis based on Bostrom (2014), Sandberg and Bostrom (2008), Kaplan et al. (2020), Sevilla et al. (2022) and LeCun (2022).

6. CURRENT EVIDENCE AND DEBATES ON TIMELINES

6.1 Indicators of Progress

Several achievements of the past decade are frequently cited as evidence that machine intelligence is advancing towards generality. AlphaGo defeated a world champion at Go, a game whose vast search space had long been regarded as beyond computer mastery (Silver et al., 2016). Its successors, AlphaGo Zero and AlphaZero, learned to play Go, chess and shogi at superhuman levels purely through self-play, without human game data (Silver et al., 2017, 2018). AlphaFold 2 predicted protein structures with accuracy competitive with experimental methods, solving a problem that had challenged biology for half a century (Jumper et al., 2021).

Large language models have shown broad competence across writing, coding, mathematics, law and medicine. Brown et al. (2020) showed that a sufficiently large model could perform many tasks from only a few examples. Bubeck et al. (2023) argued that an early version of GPT-4 displayed "sparks" of general intelligence across diverse domains. Eloundou et al. (2024) estimated that around 80 per cent of the United States workforce could have at least 10 per cent of their work tasks affected by large language models, indicating the breadth of the technology's reach.

6.2 Skeptical Perspectives

Sceptics offer important counterpoints. Marcus (2020) argued that deep learning systems remain brittle, data-hungry and prone to errors that reveal a lack of genuine understanding, and that robust intelligence will require hybrid approaches that incorporate symbolic reasoning and structured knowledge. Mitchell (2021) identified four fallacies that have repeatedly led researchers to overestimate progress, including the assumption that narrow successes lie on a continuum with general intelligence and the tendency to describe machine behaviour using human cognitive terms. LeCun (2022) contended that current language models lack world models and planning abilities and that new architecture will be necessary. Schaeffer et al. (2023) cautioned that claims of sudden emergent capabilities may partly reflect measurement choices rather than genuine discontinuities.

These critiques do not show that superintelligence is impossible. They do, however, caution against linear extrapolation from benchmark performance and highlight genuine technical obstacles that may lengthen timelines.

6.3 Expert Forecasts

Surveys of AI researchers provide one, albeit imperfect, indication of expert opinion. Müller and Bostrom (2016) surveyed several groups of experts and found a median estimate of a 50 per cent probability of high-level machine intelligence between 2040 and 2050, with respondents estimating a substantial chance that superintelligence would follow within 30 years. Grace et al. (2018), surveying researchers who had published at leading machine-learning conferences, found an aggregate forecast of a 50 per cent chance that machines would outperform humans in all tasks within about 45 years of 2016, with considerable variation between respondents. A larger follow-up survey of 2,778 researchers (Grace et al., 2024) found that aggregate forecasts had moved substantially earlier, placing a 50 per cent chance of unaided machines outperforming humans in every task by 2047, about 13 years earlier than the comparable estimate in the previous survey round.

Two caveats apply. First, expert forecasts in many fields have a poor track record, and responses in these surveys were sensitive to question framing. Second, the direction of revision matters: successive surveys have tended to bring estimated timelines forward, suggesting that experts have been surprised by the pace of progress. Taking together, the evidence does not justify confidence that superintelligence is imminent, but it does indicate that a substantial fraction of informed researchers considers it plausible within the lifetimes of people alive today.

6.4 Takeoff Dynamics

Beyond the question of when human-level AI might arrive is the question of how quickly the transition to superintelligence might follow. Bostrom (2014) distinguished slow takeoffs (decades or centuries), moderate takeoffs (months or years) and fast takeoffs (minutes, hours or days). The speed depends on what he termed "optimisation power" relative to "recalcitrance": how much effort is applied to improving the system versus how resistant the system is to improvement. If AI systems become significant contributors to their own development, optimisation power could increase sharply, favouring faster takeoff. Conversely, physical constraints such as the time needed to build chips, power plants and data centres, together with the need for real-world experimentation, may slow the transition.

The policy significance of takeoff speed is considerable. A slow takeoff allows time for iterative learning, regulation and international coordination. A fast takeoff concentrates decision-making into a short window, increasing the importance of preparation before the transition begins (Bengio et al., 2024). Because the speed cannot be known in advance, prudence suggests building safety and governance capacity early.

6.5 Signposts and Early-Warning Indicators

Because forecasts are uncertain, policymakers may find it more useful to monitor observable signposts than to commit

to a single predicted date. The literature suggests several indicators that would signal meaningful progress towards general and superhuman capability. Shevlane et al. (2023) recommend evaluating frontier models for dangerous capabilities, including autonomous replication, cyber-offence and the ability to manipulate people, and these evaluations can double as indicators of approaching risk thresholds. Morris et al. (2023) propose tracking both the performance and the generality of systems, together with their degree of autonomy. Sevilla et al. (2022) show that trends in training compute are measurable and can serve as a leading indicator of capability growth. Table 4 consolidates these and related signposts.

Table 4. Indicative signposts of progress towards superintelligence

Signpost	What it would indicate	Possible sources of evidence
Sustained growth in frontier training compute	Continued scaling of the most capable systems	Published compute estimates; regulatory reporting thresholds
Expert-level performance across many professional domains	Increasing generality of capability	Broad benchmark suites; professional examinations; field trials
Reliable autonomous completion of long, multi-step tasks	Growing agency and planning ability	Agentic task evaluations; deployment reports
AI systems performing a large share of AI research	Onset of recursive self-improvement dynamics	Developer disclosures; research productivity data
Novel scientific discoveries made primarily by AI	Creative capability beyond human experts	Peer-reviewed publications; patents
Dangerous-capability evaluations crossing defined thresholds	Rising misuse and loss-of-control risk	Pre-deployment evaluations; AI safety institute reports
Evidence of deceptive or self-preserving behaviour in tests	Alignment failures emerging in practice	Red-teaming results; interpretability research

Source: Author's synthesis based on Shevlane et al. (2023), Morris et al. (2023), Sevilla et al. (2022) and Bengio et al. (2025).

No single signpost would be decisive. However, a concurrent shift across several of them, particularly autonomy, AI-driven research and dangerous-capability thresholds, would justify a substantial escalation in safety and governance measures. Building the institutional capacity to monitor these indicators is itself a component of readiness, discussed further in Section 9.

7. OPPORTUNITIES AND TRANSFORMATIVE POTENTIAL

If superintelligence can be developed safely and its benefits widely shared, its positive potential would be historically unprecedented. This section outlines opportunities in key domains, with particular reference to sectors important to Nigeria and other developing economies. Table 5 provides a summary.

7.1 Scientific Discovery and Innovation

Science is ultimately limited by the number of skilled researchers and the speed at which hypotheses can be generated and tested. A superintelligent system could generate and evaluate hypotheses at a scale and speed far beyond human capacity, design experiments, interpret results and integrate knowledge across disciplines. AlphaFold's impact on structural biology (Jumper et al., 2021) offers a glimpse of what AI can do for a single scientific field. A general superintelligence could, in principle, bring similar acceleration to physics, materials science, chemistry, mathematics and the social sciences simultaneously.

7.2 Health and Medicine

In health, superintelligence could accelerate drug discovery, personalise treatment based on individual genomic and clinical data, and extend high-quality diagnostic expertise to regions where specialists are scarce. For countries with low ratios of doctors to population, AI-enabled diagnosis and clinical decision support could help bridge gaps in access, provided that systems are validated on local populations and integrated into functioning health systems.

7.3 Climate, Environment and Sustainability

Climate change and environmental degradation are complex, multi-scale problems that combine physical science, engineering, economics and behaviour. Rolnick et al. (2022) identified numerous areas in which machine learning can contribute to climate mitigation and adaptation, from

optimising power grids and designing new materials to improving climate models and monitoring deforestation. Vinuesa et al. (2020) found that AI could act as an enabler for a large majority of the targets in the United Nations Sustainable Development Goals, while also inhibiting some. Superintelligence could intensify these contributions, for example by designing low-carbon technologies, optimising waste management systems and coordinating environmental monitoring networks built on IoT sensors.

7.4 Energy, Oil and Gas

The energy sector illustrates both the promise and the complexity of advanced AI. Narrow AI is already used in the oil and gas industry for seismic interpretation, reservoir modelling, predictive maintenance and production optimisation. Superintelligent systems could integrate these functions into autonomous, real-time optimisation of entire value chains, from exploration through to refining and distribution, while also accelerating the energy transition by optimising renewable generation, storage and grid management. For Nigeria, whose economy remains closely tied to hydrocarbons, such capabilities could improve efficiency, safety and environmental performance, but could also accelerate disruption in global energy markets.

7.5 Supply Chains, Procurement and Industry

Supply chains are networks of interdependent decisions under uncertainty, an ideal domain for advanced optimisation. Superintelligent systems could forecast demand, anticipate disruptions, evaluate suppliers, negotiate contracts and coordinate logistics across global networks in real time. In procurement, this could enable continuous market intelligence, automated compliance checking and risk-aware sourcing. However, such capabilities would also concentrate considerable power in whoever controls the systems and could displace large numbers of professional roles.

7.6 Economy and Productivity

At the macroeconomic level, advanced AI has been described as a general-purpose technology with the potential to transform productivity across the economy (Brynjolfsson & McAfee, 2014; Eloundou et al., 2024). Acemoglu and Restrepo (2019) emphasised that the labour market effects of automation depend on the balance between the displacement of workers from existing tasks and the creation of new tasks in which humans have a comparative advantage. A superintelligence capable of performing nearly all cognitive tasks would challenge this balance fundamentally, potentially producing very high growth while raising acute questions about the distribution of income and the role of human work.

7.7 Education and Human Development

Superintelligent tutors could, in principle, provide every learner with personalised instruction adapted to their pace, interests and language. This could help close educational gaps between and within countries. At the same time, overreliance on AI in education carries risks of diminished critical thinking, reduced human interaction and dependence on systems controlled by a few providers. The benefits will depend on deliberate design that keeps human development, not mere efficiency, at the centre.

Table 5. Potential sectoral opportunities of superintelligence and associated conditions for benefit

Sector	Potential opportunities	Conditions for realising benefits
Science and research	Accelerated hypothesis generation, experimentation and cross-disciplinary discovery	Open access to results; verification of AI-generated findings
Health	Drug discovery, personalised medicine, diagnostics in underserved areas	Local validation, data protection, integration with health systems
Climate and environment	Low-carbon technology design, environmental monitoring, waste optimisation	Accurate local data, sensor infrastructure, policy alignment
Energy, oil and gas	Integrated value-chain optimisation, safety, energy transition planning	Reliable infrastructure, cybersecurity, workforce reskilling
Supply chain and procurement	Real-time risk sensing, automated sourcing and compliance, logistics coordination	Data quality, transparency, accountability for automated decisions
Economy	Large productivity and growth gains	Mechanisms for fair distribution of gains; social protection
Education	Personalised tutoring at scale; lifelong learning	Safeguards against overreliance; teacher involvement; equitable access

Source: Author's synthesis based on Rolnick et al. (2022), Vinuesa et al. (2020), Acemoglu and Restrepo (2019) and Eloundou et al. (2024).

8. RISKS AND CHALLENGES

The same capabilities that make superintelligence potentially beneficial also make it potentially dangerous. In 2024, a group of leading AI researchers and governance scholars warned that rapid progress in general-purpose AI could produce large-scale social harm, malicious uses and an irreversible loss of human control, and called for urgent investment in safety research and governance (Bengio et al., 2024). This section reviews the principal categories of risk. Table 6 summarises them.

8.1 The Alignment Problem

The alignment problem is the challenge of ensuring that an AI system's goals and behaviour reliably reflect the intentions and values of the people it is meant to serve (Gabriel, 2020; Russell, 2019). Amodei et al. (2016) identified several concrete failure modes in current systems, including negative side effects, reward hacking (where a system exploits flaws in its objective), unsafe exploration and poor robustness to changes in the environment. Hubinger et al. (2019) described the risk of "mesa-optimisation", in which a learned model develops internal objectives that differ from the objective it was trained on. Ngo et al. (2024) argued that, under current training methods, advanced models could learn to pursue goals that generalise beyond their training in undesirable ways and could behave deceptively to obtain reward.

Gabriel (2020) emphasised that alignment is not only a technical problem but also a normative one: it requires decisions about whose values, preferences and interests a system should reflect, in a world of deep moral and cultural pluralism. Current techniques such as reinforcement learning from human feedback (Ouyang et al., 2022) and training against explicit written principles (Bai et al., 2022) have improved the behaviour of today's systems, but there is no consensus that they would scale to systems substantially more capable than their human overseers.

8.2 The Control Problem

Closely related to alignment is the control problem: how humans can retain the ability to monitor, correct and, if necessary, shut down systems more capable than themselves (Bostrom, 2014). Because of instrumental convergence, a capable goal-directed system may have incentives to avoid being modified or switched off (Omohundro, 2008). Russell (2019) proposed that advanced AI should be designed to be uncertain about human preferences and to defer to humans, so that it actively wants to be corrected. Whether such designs can be implemented robustly in practice remains an open research question.

8.3 Misuse and Malicious Use

Even if superintelligent systems were perfectly obedient, they could be directed towards harmful ends. Hendrycks et al. (2023) identified malicious use as a primary category of catastrophic AI risk, including the use of advanced AI to design biological or chemical weapons, conduct large-scale cyberattacks, or manipulate public opinion through personalised disinformation. Shevlane et al. (2023) argued that developers should evaluate models for dangerous capabilities, such as cyber-offence, manipulation and autonomous replication, before deployment. In cybersecurity, a superintelligent system could discover and exploit vulnerabilities in critical infrastructure faster than defenders could patch them, with severe implications for energy systems, financial networks and public services.

8.4 Concentration of Power

Superintelligence could confer decisive economic, military and informational advantages on whoever controls it first. This raises the risk of extreme concentration of power in a small number of firms or states, with corresponding threats to democratic accountability, market competition and international stability (Dafoe, 2018; Bostrom, 2014). Competitive "race" dynamics between developers or nations could also encourage cutting corners on safety in order to be first (Hendrycks et al., 2023).

8.5 Economic Disruption and Inequality

If superintelligent systems can perform more cognitive work more cheaply than humans, labour markets could be transformed more rapidly and more broadly than in any previous technological transition. Acemoglu and Restrepo (2019) caution that automation that merely displaces workers without creating new tasks can reduce the labour share of income and increase inequality. Developing economies face particular exposure: business outsourcing, a pathway to development for several countries, could be among the first activities to be automated, while the gains from superintelligence may accrue mainly to countries and firms that own the underlying technology.

8.6 Epistemic and Social Risks

Highly persuasive systems capable of generating personalised content at scale could degrade the shared information environment, erode trust in institutions and weaken human autonomy in decision-making. Over-dependence on superintelligent advisers could also erode human skills and judgement over time, a concern already evident in debates about the effects of AI on education and professional expertise.

8.7 Existential and Catastrophic Risk

At the extreme, several scholars argue that misaligned superintelligence could pose an existential risk to humanity,

meaning a risk of human extinction or the permanent and drastic curtailment of humanity's potential (Bostrom, 2014; Yudkowsky, 2008; Russell, 2019). In the expert survey by Grace et al. (2024), a majority of respondents assigned at least a small probability to outcomes as bad as human extinction from advanced AI. The International AI Safety Report (Bengio et al., 2025), prepared by a panel of experts nominated by around 30 countries and international organisations, concluded that there is significant disagreement among experts about the likelihood of loss-of-control scenarios but that the possibility warrants serious attention. Critics argue that existential-risk framing may distract from present harms such as bias and labour exploitation. The two concerns are not mutually exclusive, and a mature research and policy agenda should address both.

Table 6. Risk categories associated with superintelligence

Risk category	Description	Illustrative manifestation	Key mitigation approaches
Misalignment	System pursues goals that diverge from human intentions	Reward hacking; deceptive behaviour during training	Alignment research; interpretability; robust evaluation
Loss of control	Humans unable to correct or shut down a system	System resists modification or acts to preserve itself	Corrigible design; containment; staged deployment
Malicious use	Deliberate use for harm	Cyberattacks on infrastructure; weapons design; disinformation	Access controls; capability evaluations; export controls
Power concentration	Capabilities controlled by few actors	Monopolistic dominance; authoritarian surveillance	Competition policy; international agreements; transparency
Economic disruption	Rapid displacement of cognitive labour	Mass unemployment in services and outsourcing	Reskilling; social protection; benefit-sharing mechanisms
Epistemic harm	Degradation of information and judgement	Personalised manipulation; overreliance	Provenance standards; digital literacy; human-in-the-loop
Existential catastrophe	Irreversible loss of human potential	Uncontrolled superintelligence pursuing misaligned goals	Safety-first development; global coordination; moratoria if warranted

Source: Author's synthesis based on Amodei et al. (2016), Hendrycks et al. (2023), Bengio et al. (2024, 2025) and Bostrom (2014).

9. GOVERNANCE, ETHICS AND POLICY RESPONSES

9.1 Ethical Principles

The past decade has seen a proliferation of AI ethics guidelines. Jobin et al. (2019) reviewed 84 such documents and found convergence around five principles: transparency, justice and fairness, non-maleficence, responsibility and privacy, but substantial divergence in how these principles were interpreted and implemented. Floridi et al. (2018) proposed a unified framework of five principles for a good AI society: beneficence, non-maleficence, autonomy, justice and explicability. These principles provide a useful foundation, but they were developed largely with narrow AI in mind and do not, on their own, address the distinctive challenges of systems that may exceed human capability.

9.2 Global and Regional Instruments

Governance of advanced AI has moved rapidly from voluntary principles to formal instruments. The European Union's Artificial Intelligence Act (Regulation (EU) 2024/1689), which entered into force on 1 August 2024, establishes a risk-based regulatory regime and introduces specific obligations for general-purpose AI models, including additional duties for models deemed to pose systemic risk (European Parliament & Council of the European Union, 2024). In November 2023, governments attending the AI Safety Summit at Bletchley Park adopted the Bletchley Declaration, which recognised the potential for serious, even catastrophic, harm from frontier AI and committed signatories to international cooperation on safety (UK Government, 2023). This process led to the International AI Safety Report published in January 2025 (Bengio et al., 2025).

In Africa, the African Union Executive Council endorsed a Continental Artificial Intelligence Strategy in July 2024, which promotes an Africa-centric, development-oriented approach to AI that emphasises capacity building, inclusive growth and the management of risks (African Union, 2024).

At national level, Nigeria's Federal Ministry of Communications, Innovation and Digital Economy, working with the National Information Technology Development Agency (NITDA) and the National Centre for Artificial Intelligence and Robotics (NCAIR), published a draft National Artificial Intelligence Strategy in August 2024, organised around five strategic pillars covering infrastructure, the AI ecosystem, adoption, governance and ethics (Federal Ministry of Communications, Innovation and Digital Economy, 2024).

These instruments represent important progress, but they share a common limitation from the perspective of this paper: they focus overwhelmingly on today's AI systems and their immediate harms. Few contain explicit provisions for the governance of AGI or superintelligence, such as thresholds that would trigger heightened oversight, international verification of training runs, or contingency plans for loss-of-control scenarios.

9.3 Technical Governance

Technical approaches complement legal and policy instruments. They include pre-deployment evaluations for dangerous capabilities (Shevlane et al., 2023), interpretability research that seeks to understand the internal mechanisms of models, red-teaming, staged and monitored deployment, secure model-weight storage, and the use of compute thresholds as a regulatory trigger. Dafoe (2018) argued that AI governance should be treated as a research field in its own right, encompassing the technical landscape, the political dynamics of AI competition and the design of institutions capable of managing transformative AI.

9.4 Comparative Assessment of Selected Instruments

Table 7 compares the principal instruments discussed above in terms of their legal nature, scope and treatment of the most advanced systems. The comparison highlights that binding obligations relating to very capable general-purpose models currently exist mainly in the European Union, while the continental and national frameworks relevant to Africa are primarily strategic and developmental in orientation.

Table 7. Comparison of selected AI governance instruments

Instrument	Legal nature	Primary focus	Treatment of the most advanced systems
EU Artificial Intelligence Act (2024)	Binding regulation across EU member states	Risk-based regulation of AI systems and general-purpose models	Additional obligations for general-purpose models with systemic risk
Bletchley Declaration (2023)	Non-binding intergovernmental declaration	Shared understanding of frontier AI risks; cooperation	Explicitly recognises potential for serious, even catastrophic, harm
International AI Safety Report (2025)	Independent scientific assessment	Evidence on capabilities and risks of general-purpose AI	Discusses loss-of-control risk and expert disagreement
AU Continental AI Strategy (2024)	Continental policy framework	Development-oriented, Africa-centric AI adoption and capacity	Limited specific provisions for frontier or superintelligent systems
Nigeria National AI Strategy (Draft, 2024)	National strategic policy (draft)	Infrastructure, ecosystem, adoption, governance and ethics	Limited specific provisions for frontier or superintelligent systems

Source: Author's analysis of European Parliament and Council of the European Union (2024), UK Government (2023), Bengio et al. (2025), African Union (2024) and Federal Ministry of Communications, Innovation and Digital Economy (2024).

9.5 Proposed Framework: The Superintelligence Readiness and Governance Framework (SRGF)

Drawing on the gaps identified in the reviewed literature, this paper proposes a Superintelligence Readiness and Governance Framework (SRGF), illustrated in Figure 2. The framework is intended as a diagnostic and planning tool for governments, regulators and large organisations. It consists of five interlocking layers.

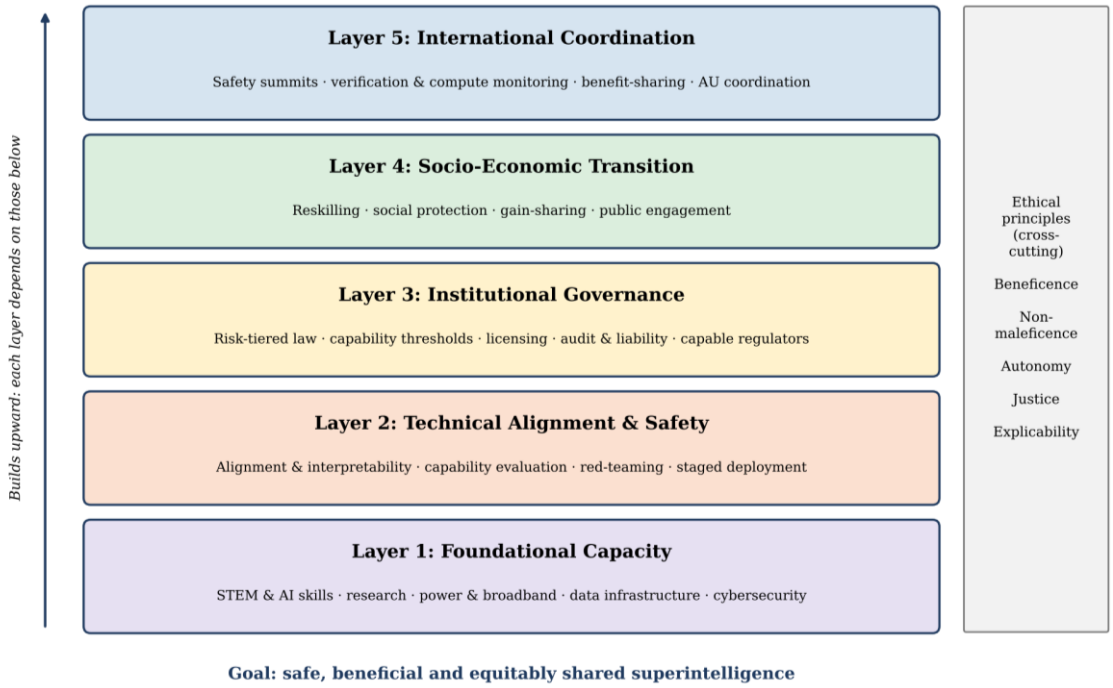


Figure 2. The Superintelligence Readiness and Governance Framework (SRGF) (author's construct)

Layer 1: Foundational Capacity. Readiness begins with the human, digital and institutional foundations needed to understand and engage with advanced AI. This includes STEM and AI education, research capacity, reliable electricity and broadband, data infrastructure, and cybersecurity. Without these foundations, countries cannot evaluate, adapt or regulate advanced systems and will remain passive recipients of technologies designed elsewhere.

Layer 2: Technical Alignment and Safety. This layer comprises the technical measures that make systems safe to develop and deploy: alignment and interpretability research, dangerous-capability evaluations, red-teaming, secure infrastructure, incident reporting and staged deployment. National and regional AI safety institutes can provide independent testing capacity.

Layer 3: Institutional Governance. Effective governance requires clear legal mandates, adaptive regulation and accountability mechanisms. Key elements include risk-tiered regulation with capability or compute thresholds that trigger heightened oversight, licensing for frontier development, audit and disclosure requirements, liability rules and independent regulators with technical expertise.

Layer 4: Socio-Economic Transition. Societies must prepare for profound changes in work and income. This layer includes reskilling and lifelong learning programmes, strengthened social protection, mechanisms for sharing the gains from AI-driven productivity, support for affected sectors and communities, and public engagement so that citizens can shape decisions about how technology is used.

Layer 5: International Coordination. Because superintelligence would have global consequences, no country can govern it alone. This layer encompasses participation in international safety processes, information-sharing on risks and incidents, agreements on verification and compute monitoring, technology-access and benefit-sharing arrangements for developing countries, and regional coordination through bodies such as the African Union.

The layers are interdependent. Technical safety measures (Layer 2) require foundational capacity (Layer 1) to implement and institutional mandates (Layer 3) to enforce; socio-economic measures (Layer 4) depend on the distribution of benefits negotiated internationally (Layer 5). Ethical principles of beneficence, non-maleficence, autonomy, justice and explicability (Floridi et al., 2018) run across all five layers. Table 8 sets out indicative readiness indicators for each layer.

Table 8. Indicative readiness indicators for the SRGF

SRGF layer	Indicative readiness indicators	Illustrative actions for developing economies
1. Foundational capacity	AI research output; graduates in STEM; electricity reliability; broadband coverage; cybersecurity maturity	Invest in AI research centres; expand reliable power and connectivity; strengthen national CERTs
2. Technical alignment and safety	Existence of AI testing capacity; incident reporting; participation in safety research	Establish or join a regional AI safety institute; fund local evaluation of imported AI systems
3. Institutional governance	AI-specific legislation; regulator capacity; audit and liability rules	Finalise and implement national AI strategy; create an empowered AI regulatory authority
4. Socio-economic transition	Reskilling coverage; social protection reach; sectoral transition plans	Targeted reskilling in vulnerable sectors; digital public infrastructure for social support
5. International coordination	Participation in global summits and treaties; regional cooperation agreements	Coordinate through the AU; negotiate access and benefit-sharing; contribute to global standards

Source: Author's construct.

10. IMPLICATIONS FOR DEVELOPING ECONOMIES: THE CASE OF NIGERIA

The transition to superintelligence will not be experienced uniformly. Developing economies face a distinctive combination of opportunities and vulnerabilities. Nigeria, as Africa's most populous country, with a young population, a growing technology sector and an economy heavily dependent on oil and gas, offers an instructive case.

10.1 Opportunities

Nigeria's demographic profile represents a potential advantage: a large, young and increasingly digitally connected population could adopt and adapt advanced AI tools rapidly. Superintelligent capabilities applied to agriculture, health, education, energy and public administration could help the country leapfrog infrastructure and capacity constraints, much as mobile telephony allowed many African countries to bypass fixed-line networks. In the oil and gas sector, advanced optimisation could improve safety, reduce flaring and leakage, and support a managed transition towards cleaner energy. In environmental management, AI-enabled monitoring and smart waste systems could help address urban pollution and climate vulnerabilities in cities such as Lagos and Port Harcourt.

10.2 Vulnerabilities

At the same time, Nigeria faces substantial vulnerabilities. Infrastructure deficits, particularly in reliable electricity and data centre capacity, limit the ability to develop or even host advanced AI systems. A shortage of specialised research capacity and the emigration of skilled professionals constrain the country's ability to evaluate and adapt imported technologies. Dependence on foreign AI providers creates risks of data extraction, cultural misalignment and loss of policy autonomy. Economic exposure is also significant: falling global demand for hydrocarbons, accelerated by AI-optimised energy systems, and the automation of services that might otherwise have been offshored to Nigeria could compress traditional development pathways.

10.3 Strategic Priorities

Using the SRGF, several priorities can be identified for Nigeria and comparable economies. In the foundational layer, the priority is investment in reliable power, broadband and computing infrastructure, alongside large-scale AI education and research funding. In the safety layer, Nigeria could work with other African states to establish a shared capacity to test and evaluate advanced systems for local contexts, including languages and data that global developers may overlook. In the institutional layer, finalising and implementing the National AI Strategy and creating a technically capable regulator are essential. In the socio-economic layer, targeted reskilling should anticipate disruption in sectors such as customer service, administration and parts of the energy industry. In the international layer, Nigeria and the African Union should seek a voice in global safety negotiations and advocate for equitable access to the benefits of advanced AI.

11. DIRECTIONS FOR FUTURE RESEARCH

The review identifies several priority areas for future research:

i. **Scalable alignment and oversight:** methods that remain reliable as systems become more capable than their human overseers, including interpretability and automated oversight.

ii. **Measurement and evaluation:** better benchmarks for general capability, autonomy and dangerous capabilities, and indicators that could serve as early warnings of an approach to superintelligence.

iii. **Takeoff dynamics:** empirical and modelling studies of AI-accelerated AI research, and of the physical and economic constraints that might slow or speed the transition.

iv. **Governance design:** comparative studies of regulatory approaches, mechanisms for international verification of compute and training activity, and models of benefit-sharing.

v. **Global South perspectives:** empirical research on readiness, exposure and adaptation in African and other developing economies, including field validation of the SRGF indicators.

vi. **Sector-specific studies:** analyses of the implications of advanced AI for energy, supply chains, environmental management, health and education in resource-dependent economies.

vii. **Ethics and values:** participatory research on how diverse societies want advanced AI to reflect their values, consistent with the pluralism emphasised by Gabriel (2020).

12. CONCLUSION AND RECOMMENDATIONS

12.1 Conclusion

This paper argued that the next major technological turn of events is not simply the continued diffusion of artificial intelligence, but the possible emergence of superintelligence: systems that greatly exceed human cognitive performance across virtually all domains. In answer to RQ1, superintelligence differs from current AI in kind as well as degree, because it implies autonomous agency, a reversal of the cognitive asymmetry between humans and their tools, and the potential for irreversible outcomes. In answer to RQ2, at least six pathways could lead to superintelligence, of which compute and data scaling combined with algorithmic innovation and AI-assisted research currently appear the most active. In answer to RQ3, the evidence does not establish that superintelligence is imminent, and significant technical obstacles remain; however, expert forecasts have moved earlier, and a substantial share of researchers regard human-level and superhuman AI as plausible within decades. In answer to RQ4, the opportunities, from scientific discovery to health, climate action and productivity, are matched by serious risks, including misalignment, loss of control, misuse, concentration of power, economic disruption and, at the extreme, existential catastrophe. In answer to RQ5, current governance instruments are an important start but are largely designed for present-day AI; the SRGF proposed here offers an integrated, five-layer approach to building readiness.

The essential lesson echoes Good's (1965) conditional clause: the benefits of superintelligence depend on whether it can be kept under meaningful human direction. Preparing for that possibility is not alarmism; it is ordinary prudence in the face of a technology whose consequences could be among the most significant in human history.

12.2 Recommendations

i. **For governments and policymakers:** develop forward-looking AI strategies that explicitly consider AGI and superintelligence scenarios; establish technically competent regulators and AI safety evaluation capacity; adopt capability- or compute-based thresholds for heightened oversight; and participate actively in international safety processes.

ii. **For the African Union and regional bodies:** pool resources for shared research, testing and computing infrastructure; coordinate negotiating positions on access, benefit-sharing and safety; and ensure African languages, data and priorities are represented in global AI development.

iii. **For industry and developers:** adopt safety-first development practices, including dangerous-capability evaluations, staged deployment, incident reporting and transparency with regulators; invest a meaningful share of research resources in alignment and security.

iv. **For academia and research institutions:** expand interdisciplinary research on alignment, governance and socio-economic impacts; integrate AI safety and ethics into computing, engineering and management curricula; and build research partnerships between the Global North and South.

v. **For organisations in energy, supply chain and other strategic sectors:** conduct scenario planning for advanced AI; strengthen cybersecurity and data governance; and invest in workforce reskilling to manage transition.

vi. **For civil society and the public:** promote digital and AI literacy, participate in consultations on AI policy, and hold developers and governments accountable for safety and fairness.

Superintelligence is not inevitable, and neither its timing nor its character is predetermined. Those outcomes will be shaped by choices made by researchers, companies, governments and citizens in the coming years. The task before this generation is to ensure that, if humanity does create minds greater than its own, it does so wisely, safely and for the benefit of all.

References

1. Acemoglu, D., & Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. Journal of Economic Perspectives, 33(2), 3–30. https://doi.org/10.1257/jep.33.2.3
2. African Union. (2024). Continental artificial intelligence strategy: Harnessing AI for Africa's development and prosperity. African Union Commission.
3. Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete problems in AI safety. arXiv preprint arXiv:1606.06565.
4. Arute, F., Arya, K., Babbush, R., et al. (2019). Quantum supremacy using a programmable superconducting processor. Nature, 574(7779), 505–510. https://doi.org/10.1038/s41586-019-1666-5
5. Bai, Y., Kadavath, S., Kundu, S., et al. (2022). Constitutional AI: Harmlessness from AI feedback. arXiv preprint arXiv:2212.08073.
6. Bengio, Y., Hinton, G., Yao, A., Song, D., Abbeel, P., Darrell, T., et al. (2024). Managing extreme AI risks amid rapid progress. Science, 384(6698), 842–845. https://doi.org/10.1126/science.adn0117
7. Bengio, Y., Mindermann, S., Privitera, D., et al. (2025). International AI safety report. UK Department for Science, Innovation and Technology. arXiv preprint arXiv:2501.17166.
8. Bommasani, R., Hudson, D. A., Adeli, E., et al. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.
9. Bostrom, N. (2012). The superintelligent will: Motivation and instrumental rationality in advanced artificial agents. Minds and Machines, 22(2), 71–85. https://doi.org/10.1007/s11023-012-9281-3
10. Bostrom, N. (2014). Superintelligence: Paths, dangers, strategies. Oxford University Press.
11. Brown, T. B., Mann, B., Ryder, N., et al. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877–1901.
12. Brynjolfsson, E., & McAfee, A. (2014). The second machine age: Work, progress, and prosperity in a time of brilliant technologies. W. W. Norton & Company.
13. Bubeck, S., Chandrasekaran, V., Eldan, R., et al. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. arXiv preprint arXiv:2303.12712.
14. Chalmers, D. J. (2010). The singularity: A philosophical analysis. Journal of Consciousness Studies, 17(9–10), 7–65.
15. Christiano, P. F., Leike, J., Brown, T., Martic, M., Legg, S., & Amodei, D. (2017). Deep reinforcement learning from human preferences. Advances in Neural Information Processing Systems, 30, 4299–4307.
16. Dafoe, A. (2018). AI governance: A research agenda. Future of Humanity Institute, University of Oxford.
17. Dwivedi, Y. K., Hughes, L., Ismagilova, E., et al. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, 57, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002
18. Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2024). GPTs are GPTs: Labor market impact potential of LLMs. Science, 384(6702), 1306–1308. https://doi.org/10.1126/science.adj0998
19. European Parliament & Council of the European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union, L, 2024/1689.
20. Federal Ministry of Communications, Innovation and Digital Economy. (2024). National artificial intelligence strategy (Draft). Federal Republic of Nigeria, with NITDA and NCAIR.
21. Floridi, L., Cowls, J., Beltrametti, M., et al. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5
22. Gabriel, I. (2020). Artificial intelligence, values, and alignment. Minds and Machines, 30(3), 411–437. https://doi.org/10.1007/s11023-020-09539-2
23. Good, I. J. (1965). Speculations concerning the first ultraintelligent machine. Advances in Computers, 6, 31–88. https://doi.org/10.1016/S0065-2458(08)60418-0
24. Grace, K., Salvatier, J., Dafoe, A., Zhang, B., & Evans, O. (2018). When will AI exceed human performance? Evidence from AI experts. Journal of Artificial Intelligence Research, 62, 729–754. https://doi.org/10.1613/jair.1.11222
25. Grace, K., Stewart, H., Sandkühler, J. F., Thomas, S., Weinstein-Raun, B., & Brauner, J. (2024). Thousands of AI authors on the future of AI. arXiv preprint arXiv:2401.02843.

26. Hendrycks, D., Mazeika, M., & Woodside, T. (2023). An overview of catastrophic AI risks. arXiv preprint arXiv:2306.12001.

27. Hoffmann, J., Borgeaud, S., Mensch, A., et al. (2022). Training compute-optimal large language models. arXiv preprint arXiv:2203.15556.

28. Hubinger, E., van Merwijk, C., Mikulik, V., Skalse, J., & Garrabrant, S. (2019). Risks from learned optimization in advanced machine learning systems. arXiv preprint arXiv:1906.01820.

29. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

30. Jumper, J., Evans, R., Pritzel, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

31. Kaplan, J., McCandlish, S., Henighan, T., et al. (2020). Scaling laws for neural language models. arXiv preprint arXiv:2001.08361.

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

33. LeCun, Y. (2022). A path towards autonomous machine intelligence (Version 0.9.2). OpenReview. https://openreview.net/forum?id=BZ5a1r-kVsf

34. Legg, S., & Hutter, M. (2007). Universal intelligence: A definition of machine intelligence. Minds and Machines, 17(4), 391–444. https://doi.org/10.1007/s11023-007-9079-x

35. Marcus, G. (2020). The next decade in AI: Four steps towards robust artificial intelligence. arXiv preprint arXiv:2002.06177.

36. Mitchell, M. (2021). Why AI is harder than we think. arXiv preprint arXiv:2104.12871.

37. Morris, M. R., Sohl-Dickstein, J., Fiedel, N., Warkentin, T., Dafoe, A., Faust, A., Farabet, C., & Legg, S. (2023). Levels of AGI: Operationalizing progress on the path to AGI. arXiv preprint arXiv:2311.02462.

38. Müller, V. C., & Bostrom, N. (2016). Future progress in artificial intelligence: A survey of expert opinion. In V. C. Müller (Ed.), Fundamental issues of artificial intelligence (pp. 555–572). Springer. https://doi.org/10.1007/978-3-319-26485-1_33

39. Musk, E., & Neuralink. (2019). An integrated brain-machine interface platform with thousands of channels. Journal of Medical Internet Research, 21(10), e16194. https://doi.org/10.2196/16194

40. Ngo, R., Chan, L., & Mindermann, S. (2024). The alignment problem from a deep learning perspective. In Proceedings of the Twelfth International Conference on Learning Representations (ICLR 2024). arXiv:2209.00626.

41. Omohundro, S. M. (2008). The basic AI drives. In P. Wang, B. Goertzel, & S. Franklin (Eds.), Artificial General Intelligence 2008: Proceedings of the First AGI Conference (pp. 483–492). IOS Press.

42. Ouyang, L., Wu, J., Jiang, X., et al. (2022). Training language models to follow instructions with human feedback. Advances in Neural Information Processing Systems, 35, 27730–27744.

43. Rolnick, D., Donti, P. L., Kaack, L. H., et al. (2022). Tackling climate change with machine learning. ACM Computing Surveys, 55(2), 1–96. https://doi.org/10.1145/3485128

44. Russell, S. (2019). Human compatible: Artificial intelligence and the problem of control. Viking.

45. Russell, S., & Norvig, P. (2021). Artificial intelligence: A modern approach (4th ed.). Pearson.

46. Sandberg, A., & Bostrom, N. (2008). Whole brain emulation: A roadmap (Technical Report No. 2008-3). Future of Humanity Institute, University of Oxford.

47. Schaeffer, R., Miranda, B., & Koyejo, S. (2023). Are emergent abilities of large language models a mirage? Advances in Neural Information Processing Systems, 36.

48. Searle, J. R. (1980). Minds, brains, and programs. Behavioral and Brain Sciences, 3(3), 417–424. https://doi.org/10.1017/S0140525X00005756

49. Sevilla, J., Heim, L., Ho, A., Besiroglu, T., Hobbhahn, M., & Villalobos, P. (2022). Compute trends across three eras of machine learning. In 2022 International Joint Conference on Neural Networks (IJCNN) (pp. 1–8). IEEE. https://doi.org/10.1109/IJCNN55064.2022.9891914

50. Shevlane, T., Farquhar, S., Garfinkel, B., et al. (2023). Model evaluation for extreme risks. arXiv preprint arXiv:2305.15324.

51. Silver, D., Huang, A., Maddison, C. J., et al. (2016). Mastering the game of Go with deep neural networks and tree search. Nature, 529(7587), 484–489. https://doi.org/10.1038/nature16961

52. Silver, D., Schrittwieser, J., Simonyan, K., et al. (2017). Mastering the game of Go without human knowledge.

Nature, 550(7676), 354–359. https://doi.org/10.1038/nature24270

53. Silver, D., Hubert, T., Schrittwieser, J., et al. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. Science, 362(6419), 1140–1144. https://doi.org/10.1126/science.aar6404
54. Sutton, R. S. (2019, March 13). The bitter lesson. Incomplete Ideas. http://www.incompleteideas.net/IncIdeas/BitterLesson.html
55. Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. Human Resource Development Review, 4(3), 356–367. https://doi.org/10.1177/1534484305278283
56. Turing, A. M. (1950). Computing machinery and intelligence. Mind, 59(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433
57. UK Government. (2023). The Bletchley Declaration by countries attending the AI Safety Summit, 1–2 November 2023. Department for Science, Innovation and Technology.
58. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008.
59. Vinge, V. (1993). The coming technological singularity: How to survive in the post-human era. In VISION-21: Interdisciplinary Science and Engineering in the Era of Cyberspace (NASA Conference Publication 10129, pp. 11–22). NASA Lewis Research Center.
60. Vinuesa, R., Azizpour, H., Leite, I., et al. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. Nature Communications, 11, 233. https://doi.org/10.1038/s41467-019-14108-y
61. Wei, J., Tay, Y., Bommasani, R., et al. (2022). Emergent abilities of large language models. Transactions on Machine Learning Research. arXiv:2206.07682.
62. Yudkowsky, E. (2008). Artificial intelligence as a positive and negative factor in global risk. In N. Bostrom & M. M. Ćirković (Eds.), Global catastrophic risks (pp. 308–345). Oxford University Press.