Clareus Scientific Science and Engineering
Volume 3 Issue 1 Janaury 2026
DOI: 10.70012/CSSE.03.067
ISSN: 3065-1182

Artificial Intelligence as the Last Invention of Humanity: Disruption, Co-Creation, and Implications

Citation: Robin Vivian. “Artificial Intelligence as the Last Invention of Humanity: Disruption, Co-Creation, and Implications". Clareus Scientific Science and Engineering 3.1 (2026): 04-18.

Article Type: Research Article

Received: December 8, 2025

Published: January 26, 2026

Robin Vivian*

Laboratoire Perseus, University of Lorraine, Metz, France

***Corresponding Author:** Robin Vivian, Laboratoire Perseus, University of Lorraine, Metz, France.

Abstract

Artificial intelligence (AI) - and more specifically generative AI - is a major technological breakthrough. Not only does it automate complex tasks, but it is also emerging as an active force for innovation. This article explores the hypothesis that AI may be the last great human invention, capable of generating all subsequent innovations. It traces the evolution of AI from its symbolic origins to today’s self-improving recursive systems through advances in deep learning. Concrete examples - such as AlphaFold, AutoML and generative art - illustrate how AI is already transforming scientific research, artistic creation and engineering. This algorithmic capacity for innovation even challenges our definition of human creativity. The article also examines the ethical, legal and societal risks associated with such cognitive delegation, and calls for an inclusive governance model for automated innovation. AI does not spell the end of human ingenuity - it ushers in a new paradigm of co-creation.

Keywords: Generative AI; Automated Innovation; Human-Machine Co-Creation; Human Creativity

Introduction

Since the Industrial Revolution, humanity has experienced a series of major innovations that have transformed our relationship with the world - be it the steam engine, electricity, computers or the internet. At each stage, the tools we have created have multiplied our ability to understand, transform and master our environment. Today, artificial intelligence - and in particular generative AI - represents a technological disruption of a new order. It is not only automating complex tasks, but also beginning to play an active role in the very processes of creation, invention and innovation.

This transformation raises a question as bold as it is fundamental: what if AI is the last great human invention? In other words, has humanity created a tool capable of participating in - or even leading - future major scientific, technological or artistic breakthroughs? If so, this would not be a simple matter of scaling up, but rather a profound turn in human history, marking the beginning of an era in which all future inventions will be co-produced with - or by - artificially intelligent systems.

This scenario, sometimes referred to as the 'technological singularity' ([5], 2014; [2], 2013), is not purely speculative. It is based on concrete advances in machine learning, natural language processing and robotics [1], and on real-world applications in research, medicine, industry and artistic creation. Many researchers ([5, 2, 15]) have already warned of the profound implications - both promising and worrying - of such a perspective, including enormous gains in efficiency alongside risks of loss of control, cognitive dependency and even existential threats.

The aim of this article is to explore this bold hypothesis: AI as the last great human invention. To do so, we first review the evolution of AI and illustrate why it represents a unique technological disruption. We then analyse how it is already catalysing innovation in various fields. Finally, we discuss the philosophical, societal and ethical implications of a world in which humans are no longer the sole drivers of progress - and we examine the risks this poses, as well as the conditions necessary to maintain control.

History and Evolution of Artificial Intelligence

Artificial intelligence, in its conceptual form, has its roots long before the invention of the computer. In ancient times, philosophers wondered about the possibility of mechanising thought. But it was in the twentieth century - with the advent of cybernetics and computing - that these ideas took on a scientific form. In 1950, Alan Turing proposed a now famous test - the "Turing Test" - to assess the ability of a machine to behave indistinguishably from a human. Intentionally or not, Turing shifted the comparison to what humans and machines respond to, rather than what they think. Do people always say what they mean? This marked the official beginning of a field of research that would rapidly evolve around the ambitious goal of reproducing or simulating human intelligence.

The first decades of AI were marked by unbridled enthusiasm. From the 1950s to the 1970s, researchers developed programs that could play chess, solve logic puzzles or prove mathematical theorems. This era was known as "symbolic AI", based on explicit rules and the logical manipulation of symbols. But despite occasional successes, this approach struggled to cope with complex and uncertain environments - leading to two "AI winters" (in the 1970s and 1990s) when expectations cooled due to a lack of concrete results and insufficient computing power.

A turning point came in the 2000s with the advent of machine learning, which introduced a paradigm shift: instead of programming a solution directly, one could train a model to discover it through data. This probabilistic approach paved the way for spectacular advances, especially thanks to the massive use of big data and deep neural networks (deep learning). In 2012, the breakthrough of the AlexNet convolutional network in an image recognition competition heralded what some now call the "renaissance" of AI.

Since then, applications have proliferated: facial recognition, voice assistants, automatic translation, medical diagnostics and autonomous driving are just a few examples. Models have become increasingly powerful, reaching levels of "generalisation" that approach a form of flexible, adaptive and generative intelligence. It is this evolution that supports the idea that AI may soon be an active partner in creating unprecedented innovation.

With the advent of large language models (LLMs) such as GPT-4 or Claude - and image generators such as DALL-E and Midjourney - we are witnessing a new phase: AI no longer just classifies or predicts; it creates. It can code, draw, compose, write and simulate. Invention, long considered the pinnacle of human activity, is gradually becoming accessible to machines. Recent contributions from AI - such as the discovery of new molecules in pharmaceuticals (AlphaFold [[6]], IBM MoLFormer: https://molformer.res.ibm.com/[22]), the writing of scientific papers, or the design of new algorithms - mark a turning point: AI is no longer merely accelerating innovation, it is becoming its primary agent.

This rapid development follows an exponential logic. Unlike previous technologies, AI improves by exploiting its own creations. The tools it creates make it possible to optimise architectures, automate research, code new algorithms and rapidly build prototypes. We are entering an era of recursive innovation, in which AI contributes to its own improvement - and, by extension, to the progress of all human disciplines.

In this context, some researchers ([5, 17, 18, 3]) are already talking about General Artificial Intelligence (GAI) as an imminent or even inevitable goal. Such an entity would not be limited to a single task, but would be able to reason, learn and innovate autonomously in a wide range of domains. Although this perspective remains controversial, it is fueling a profound debate: are humans still the central creators of their own history, or have we ceded that role to an artificial entity that is faster, more precise and potentially more creative?

His rapid evolution of AI, from a theoretical ambition to a practical reality, has paved the way for its current role as a catalyst for innovation. Far from being limited to automation, AI has become an active agent in scientific, technical, and artistic creation, as the following examples demonstrate.

AI as a Catalyst for Future Inventions

The idea that artificial intelligence is merely a tool at the service of humanity is outdated. Today's AI systems do more than just perform tasks - they take initiatives, formulate hypotheses and actively participate in the production of new knowledge and technology. This change in status - from AI as a tool to AI as an agent - makes it a key catalyst for future invention. It is now possible to identify several areas where AI not only facilitates innovation, but becomes its very foundation.

Jacques Pitrat and Artificial Meta-Intelligence

Jacques (1935-2019), a French pioneer in AI, articulated an early and unique vision of the autonomy of intelligent systems (see references [12] and [13]). He argued that humans had two major flaws: they were 'not lazy enough' to automate everything possible and 'too intelligent' — too prone to unnecessarily complicating their creations. Underpinning this provocative stance was the notion that artificial intelligence should not merely assist humans, but also be capable of improving its own performance.

Meta-intelligence," a concept championed by Jacques Pitrat, can be likened to a conductor who not only directs a symphony but also composes new scores for other orchestras. In other words, it refers to an AI capable of designing other AIs, reflecting on its own processes, and improving autonomously. For example, a meta-intelligent system could create an algorithm optimised for analysing medical data and then refine that algorithm to become more accurate without human intervention. With CAIA, Pitrat advocated the concept of 'meta-intelligence' — an AI capable of creating other AIs, reflecting on its own processes, correcting its mistakes, and innovating independently. This idea, ahead of its time, anticipated today's research into self-learning, automatic code generation, and assisted algorithmic optimisation. Rather than viewing AI as a static tool, Pitrat envisaged it as an evolving system capable of self-reflection.

His work influenced research into cognitive architectures, expert systems, and the automation of scientific discovery. Through his writings, he laid the groundwork for a future in which humans are creators of solutions, not just their designers — a vision that echoes the idea of AI as the last great human invention.

AI and Automated Scientific Discovery

The transformative impact of artificial intelligence (AI) on molecular biology has revolutionised scientific discovery by enabling unparalleled speed, precision and autonomy in research processes. A landmark achievement in this field is DeepMind's AlphaFold (2020), which solved the long-standing problem of protein folding — a scientific challenge that had perplexed researchers for over fifty years. Leveraging advanced deep neural networks, AlphaFold can predict the three-dimensional structures of proteins from their amino acid sequences with remarkable accuracy. This has generated a comprehensive database of millions of protein structures in just months — a task that would previously have required decades of meticulous human effort. This breakthrough has transformed fields such as structural biology, biochemistry, and drug discovery, providing researchers with critical insights into protein functions, interactions, and potential therapeutic applications. For example, AlphaFold's predictions have accelerated research into diseases such as Alzheimer's and cancer, in which protein misfolding plays a critical role. This has enabled scientists to design targeted interventions more efficiently [20]. Similarly, laboratories such as Insilico Medicine are using AI to design new drugs in silico, thereby streamlining the drug development process. By simulating molecular interactions and optimising drug candidates computationally, these AI-driven approaches have reduced the time taken for the research and testing phases from years to mere months, thus facilitating the rapid development of life-saving therapeutics for conditions such as infectious diseases and rare genetic disorders [21].

The emergence of 'autonomous science' signifies a paradigm shift in which AI systems evolve from passive tools to independent agents capable of conducting entire research cycles. Platforms such as Eve, a robotic scientist developed at the University of Manchester, exemplify this transition [10]. Eve can autonomously formulate hypotheses, design and execute experiments, analyse data and draw conclusions without human intervention, functioning as a self-contained research entity. For instance, Eve has played a key role in identifying potential drug compounds for neglected tropical diseases such as malaria and Chagas disease. It has navigated vast chemical libraries to propose viable candidates with a precision and speed that human researchers alone cannot achieve [10]. This capability is further enhanced by systems such as IBM's MoLFormer, which integrates natural language processing and molecular modelling to predict chemical properties and accelerate drug discovery. Such platforms enhance efficiency and explore scientific questions on a scale that redefines the boundaries of human inquiry. Furthermore, AI-driven systems are being increasingly integrated into laboratory workflows to automate tasks such as high-throughput screening, experimental design optimisation and data interpretation. This frees up researchers to focus on higher-level conceptual work. For example, AI systems such as Chematica (now Synthia) have been employed to plan synthetic routes for complex molecules, thereby reducing the time and cost of chemical synthesis [23]. These advancements signal a future in which AI acts as a collaborative partner, augmenting human creativity while independently advancing the frontiers of scientific discovery in molecular biology, chemistry and other fields, and potentially reshaping our understanding of life itself.

For instance, in 2024, DeepMind expanded AlphaFold's capabilities with AlphaFold 3, enabling predictions not only of protein structures but also of complex molecular interactions, further accelerating research into neurodegenerative diseases [57].

AI and Technical Invention

In the field of engineering, artificial intelligence (AI) is fundamentally transforming the design and development of innovative devices by introducing unprecedented levels of efficiency, creativity and precision. Generative design methods integrated into architectural software and computer-aided design (CAD) environments enable AI to propose highly optimised engineering structures that are tailored to specific constraints, such as material properties, weight, cost and mechanical stresses. Autodesk's generative design tools [24], for example, use AI to generate thousands of viable prototypes from a single design brief, enabling engineers to explore a wide range of creative solutions that balance functionality, sustainability, and resource efficiency. This approach has led to breakthroughs in industries such as aerospace, where AI-designed components, such as lightweight aircraft parts, reduce fuel consumption while maintaining structural integrity, and automotive design, where optimised chassis designs enhance performance and safety. By minimising the need for exhaustive manual iterations, these tools accelerate the design process, often yielding unconventional structures that human designers might not conceive, such as biomimetic designs inspired by natural forms, like bone structures.

In the field of computing, AI models such as GPT, Amazon's CodeWhisperer and GitHub's Copilot have transformed software development. These models can translate natural language descriptions into functional, high-quality source code, changing the way developers approach coding tasks. These tools go far beyond automating repetitive tasks, enabling the rapid and accurate creation of sophisticated algorithms, complex software architectures, and entire simulation environments. For example, GitHub Copilot suggests context-aware code snippets that adapt to the developer's intent, while CodeWhisperer is excellent at generating boilerplate code and optimising existing scripts for performance. These AI-driven tools streamline every stage of the software lifecycle, from drafting initial code to automating debugging and performance optimisation, and even generating unit tests. This significantly reduces development time and human error. Companies like Microsoft, for example, have integrated Copilot into their development pipelines and reported productivity gains of up to 55% for certain coding tasks. Furthermore, AI is facilitating the development of innovative software solutions, including autonomous simulation environments for testing self-driving car algorithms, which necessitate the integration of real-time data processing and decision-making logic.

The influence of AI on technical innovation transcends individual disciplines, encouraging interdisciplinary collaborations that draw on insights from diverse fields such as physics, materials science, biology and ergonomics. AI algorithms synthesise data from these domains to design hybrid systems with unparalleled efficiency and adaptability. In biomedical engineering, for example, AI combines biomechanical data with materials science to create advanced prosthetics and implants that are tailored to the needs of individual patients, thereby improving functionality and comfort [27]. In manufacturing, AI drives component design and enhances production processes by simulating workflows, predicting product performance and implementing real-time quality control. Companies such as Siemens use AI to optimise assembly lines, employing predictive models to detect potential defects before they occur and thereby reducing waste and ensuring greater precision [28]. This creates a dynamic feedback loop where each product iteration is refined to better meet its intended purpose, enhancing performance and sustainability. AI-driven predictive maintenance systems, for example, can analyse sensor data to anticipate equipment failures in manufacturing, thereby reducing downtime and costs.

As AI systems evolve, they become increasingly capable of learning from operational data after deployment, which enables continuous improvement in real-world applications. In smart infrastructure such as bridges and wind turbines, AI-powered systems can monitor structural health, adapt to environmental changes and suggest real-time maintenance strategies, thereby extending the lifespan of critical assets [29]. Similarly, in consumer electronics, AI enables devices such as smart thermostats to learn from user behaviour and optimise energy usage dynamically. This shift from static to dynamic, responsive designs represents a transformative shift in engineering, where AI acts as a creative catalyst and strategic optimiser. These self-enhancing systems learn from user interactions and environmental feedback to ensure optimal performance over time, redefining traditional engineering workflows and paving the way for adaptive, intelligent technologies that evolve alongside their users and environments.

AI and Generative Art

AI is also transforming arts education by integrating tools like Artbreeder or Runway into curricula. For example, Stanford University launched a course in 2023 titled "AI and Creative Expression," where students use generative models to create interdisciplinary works blending art, technology, and social sciences. Such initiatives foster a new form of creative thinking, where students learn to collaborate with AI to explore novel concepts while developing critical and ethical skills in response to increasing automation.

One of the most striking examples of human-machine co-creation is generative art, in which AI systems such as DALL·E, Midjourney [30] and Runway generate images, videos, music and texts from simple text prompts, thereby redefining the boundaries of artistic expression. Leveraging advanced generative adversarial networks (GANs) and transformer models, these tools create works that are often indistinguishable from those of human artists, challenging traditional notions of creativity. AI is no longer merely imitating existing styles; it is exploring new aesthetic forms, generating unexpected associations and pushing the boundaries of artistic innovation. Midjourney, for example, has been used to create surreal, photorealistic artworks that blend disparate cultural and historical influences, producing pieces that resonate with audiences in galleries and on online platforms alike. In music, AI systems such as OpenAI's MuseNet can compose original scores spanning genres, from classical symphonies to modern pop, and adapt to or blend user-specified styles in novel ways [31]. This capacity transforms the role of the human artist from that of sole creator to that of creative curator or collaborator, shaping inputs and refining outputs while embracing the unpredictable results generated by algorithms.

The impact of AI on generative art goes beyond individual creations, fostering new paradigms in artistic collaboration and cultural production. Contemporary artists are increasingly treating AI as a creative partner in its own right, using its ability to process large amounts of visual, auditory or textual data to create works that challenge artistic boundaries. For example, the artist Refik Anadol uses AI to create immersive, data-driven installations. One such installation is "Machine Hallucinations", which visualises large amounts of data as dynamic, abstract art forms displayed in public spaces [32]. AI has also been used to co-write film scripts and design entire video game environments, combining technical complexity with artistic flair. These collaborations are sparking debates w ithin art institutions and intellectual property circles, raising critical questions about ownership, originality, and authorship. Researchers such as Ahmed Elgammal at Rutgers University contend that AI's creative contributions challenge the traditional dichotomy between human and machine creativity. They propose a new framework of 'creative AI' that integrates cognitive and decision-making processes [7]. Legal disputes surrounding AI-generated works in copyright law highlight the need for updated frameworks to address these novel forms of authorship.

The mainstream adoption of AI-generated art is transforming industries beyond the realm of traditional art. In advertising, for example, companies use tools such as Runway to generate dynamic, customised visual content for campaigns, thereby reducing production costs and enabling rapid iteration [33]. In education, AI-driven art tools are being incorporated into curricula to educate students in creativity, technology and interdisciplinary thinking, encouraging new forms of expression that blend technical precision with emotional depth. For instance, platforms such as Artbreeder enable users to evolve images collaboratively, generating hybrid artworks that blend user inputs with AI-generated variations and democratising access to creative tools [34]. These applications are broadening our perception of art and prompting a reappraisal of creativity in a world where human-machine collaboration is becoming the norm. As AI continues to evolve, it will likely introduce entirely new artistic media, such as generative virtual reality experiences or interactive storytelling. he rise of generative AI is reshaping cultural and educational paradigms. In museums, exhibitions such as those at the Museum of Modern Art (MoMA) in 2024 have incorporated AI-generated installations, sparking debates about the value of algorithmic art versus human creations. In education, institutions like ETH Zürich are integrating AI into their programmes to teach not only technical skills but also interdisciplinary collaboration and ethical reflection. These developments suggest that AI is not merely producing works but redefining how societies perceive and teach creativity, emphasising curation and interpretation over creation from scratch.

This will further blur the lines between creator and creation, reshaping cultural landscapes globally.

Recursive Innovation and the Optimization of Innovations

Recursive innovation," explain: "Recursive innovation resembles a virtuous cycle where an AI uses its own creations to enhance itself, much like an engineer building a machine that designs even better machines. A defining feature of modern AI development is its capacity for recursive innovation, whereby AI systems contribute to their own improvement and create an exponential dynamic that accelerates technological progress. Systems such as AutoML, developed by Google, allow algorithms to autonomously design and optimise new AI models, thereby enhancing performance without human intervention [9]. AutoML, for example, has been used to create neural network architectures that outperform manually designed models in tasks such as image classification and natural language processing, thus demonstrating AI's ability to refine its own capabilities. This self-referential innovation also extends to hardware, where AI optimises the design of electronic circuits such as Tensor Processing Units (TPUs) [35], which are tailored to accelerate AI computations. This circular process, whereby AI designs tools that enhance its own performance, creates an 'innovation feedback loop' that can outpace human understanding, presenting opportunities and challenges regarding scalability and control.

This recursive logic is transforming multiple domains, including software development, robotics, and scientific research. In software development, reinforcement learning algorithms such as those employed in DeepMind's AlphaCode autonomously generate competitive programming solutions by iteratively refining their strategies in simulated environments [36]. In robotics, AI systems such as those developed by Boston Dynamics [37] use reinforcement learning to optimise locomotion and task performance, enabling robots to adapt to complex terrains in real time. In scientific research, recursive AI systems are accelerating discoveries by automating the generation of hypotheses and the design of experiments. For instance, the AI-driven platform Bayesian Optimisation has been employed to optimise the properties of materials for renewable energy applications, such as enhancing the efficiency of solar cells, by iteratively refining experimental parameters [38]. These advancements highlight AI's dual role as creator and optimiser, driving innovation at an unprecedented pace across diverse fields.

However, the exponential growth of recursive innovation poses significant risks. As AI systems become more complex and autonomous, it becomes increasingly difficult for human oversight to keep pace with their speed and intricacy. The 'black[1] box' problem, whereby deep learning models make decisions that are difficult to interpret, is particularly acute in recursive systems that generate novel algorithms or designs [14]. For example, if an AI-designed neural network produces unexpected results, it can be difficult to debug or validate its logic, raising concerns about reliability and safety in critical applications such as healthcare or infrastructure. Furthermore, the concentration of recursive innovation in the hands of a few tech giants, who control the computational resources and datasets required for advanced AI, poses a risk of creating monopolies on technological progress [5]. This could exacerbate global inequalities, as smaller organisations and developing nations struggle to access these tools.

To address these challenges, researchers are calling for robust governance frameworks to ensure transparency, accountability and equitable access. The European Union's AI Act, for example, aims to set standards for high-risk AI systems, including those involved in recursive innovation, by requiring transparency in model development and decision-making processes [16]. Similarly, open-source platforms such as Hugging Face promote democratised access to AI tools, thereby enabling broader participation in innovation [39]. However, ethical considerations are also critical, as recursive AI systems could inadvertently amplify biases embedded in their training data, leading to skewed or harmful innovations. For instance, biased AI models in drug discovery might prioritise treatments for certain demographics while neglecting others [21]. To mitigate these risks, it is essential to foster interdisciplinary collaboration among technologists, ethicists, and policymakers in order to develop adaptive regulatory frameworks that ensure the equitable sharing of the benefits of recursive innovation while safeguarding against unintended consequences.

While recursive innovation illustrates AI's transformative potential, it also raises fundamental questions about humanity's role in a world where machines co-create the future. These philosophical and societal implications warrant deeper exploration.

Philosophical and Societal Implications: Towards a Redefinition of Human Creativity

The emergence of artificial intelligence as a key player in technological, artistic and scientific innovation raises fundamental philosophical and societal questions. If AI becomes an entity capable of invention as well as a tool, how should we redefine human creativity? What will become of the roles of authors, researchers and artists? More broadly, what role will humanity have in a world where new ideas, discoveries and solutions are co-produced — or even exclusively generated — by artificial systems?

Creativity: A Redefined Frontier

Traditionally, creativity has been considered an exclusively human ability, involving a combination of reason, intuition, emotion and personal experience. However, current generative AI systems, particularly those used in art, literature, music and scientific discovery, challenge this long-held view. These systems are trained on vast amounts of data and can produce new, coherent and sometimes profoundly original content that rivals human output in terms of its complexity and depth.

Researchers such as Margaret Boden [4], a specialist in computational creativity, distinguish between three types of creativity: combinatorial (mixing known elements), exploratory (navigating a space of possibilities) and transformational (changing the conceptual framework). Today, AI is capable of operating in all three categories — combining, exploring and reinventing cultural models. This capacity challenges our intuitive understanding of creativity, which has long been considered uniquely human. For example, AI's ability to remix existing knowledge into innovative applications has led to technological breakthroughs such as drug discovery and autonomous systems.

Moreover, AI introduces a whole new dimension to creativity by demonstrating capabilities that transcend human limitations. Unlike humans, AI systems are not limited by biases, fatigue or personal preferences, allowing them to generate novel and unconventional ideas that have the power to transform entire industries. This sparks important questions about authenticity and authorship: can the output of an algorithm truly be considered 'creative', or does it merely imitate patterns found in its training data? These philosophical challenges demand a re-evaluation of how we define creativity in the age of intelligent machines.

AI's expanding creative capabilities also prompt broader societal reflections: What does creativity mean in a world where machines can actively contribute to cultural and intellectual heritage? How should we integrate AI into our understanding of art, innovation, and human identity? Such questions emphasise the transformative impact of AI on individual and collective perspectives of ingenuity and originality.

Humans as Curators or Catalysts

In light of these developments, the role of humans in the creative and inventive process may gradually change. Rather than being the sole drivers of innovation, we may become curators, editors or orchestrators, asking the questions, choosing the directions and refining the outputs provided by AI. Researchers are no longer the sole authors of discoveries; they now act as mediators between artificial systems and broader knowledge projects.

This shift is already evident in laboratories that have incorporated AI into their daily routines: humans now supervise, review and guide rather than produce everything autonomously. This signals a form of intellectual transhumanism, in which human cognitive abilities are augmented by computational tools that can complement — and, in some cases, surpass — our intuition.

The End of a Founding Myth: Human Beings as Sole Creators

The idea that only humans can create stems from the humanist heritage of the Renaissance. However, this notion of the sole, sovereign creator is now being challenged. In his book Superintelligence, Nick Bostrom considers the possibility that AI could one day become more intelligent and inventive than humans. If such an entity were to emerge, it could produce inventions at a rate and to a degree that would be beyond human comprehension.

The concept of the 'last invention', advanced by authors such as James Barrat and echoed in Cambridge research on the existential risks of AI, suggests a tipping point. AI could be the last human creation, since all subsequent inventions would be generated by machines. Although still speculative, this scenario is becoming increasingly plausible, changing our relationship with both history and the future. Innovation would no longer be an exclusively human process, but rather an autonomous, algorithmic one.

A Societal Divide to Anticipate

The emergence of AI as a creator also presents significant societal challenges. If knowledge is co-produced by AI, it could become concentrated in the hands of a select few — particularly the companies that develop and train these systems. This could lead to monopolies on innovation and intellectual resources being created, amplifying economic inequalities and limiting access to technological benefits for marginalised communities. The cognitive gap between those who understand, manipulate and exploit AI, and those who are subjected to it, could widen dangerously, leading to new forms of digital disenfranchisement and societal stratification.

Furthermore, the legal and economic recognition of AI-generated works is uncertain. Who owns the rights to an AI-generated work? The user? The developer? Or the machine itself? These questions touch on fundamental principles of justice and equity in the digital age. Legislators and legal scholars are only just beginning to address these issues, which could transform the concepts of intellectual property, authorship and accountability (see The Ethics of AI-Driven Invention). Addressing these challenges will require a profound re-evaluation of existing legal frameworks in order to incorporate the unique characteristics of AI-driven innovation. It will also require the establishment of consistent guidelines across jurisdictions through international dialogue.

Furthermore, the societal implications extend beyond the legal sphere, touching on cultural perceptions of creativity and originality. As AI-generated works become more prevalent, there may be a shift in how humanity values and defines artistic and intellectual contributions. Transparency, fairness and inclusivity must be ensured in the integration of these creations into our shared cultural and economic landscapes, and public discourse and education will be essential in this regard.

The Challenge of Innovation Governance

Finally, the emergence of AI as a driver of innovation raises questions about governance. How can systems that are capable of autonomous innovation be managed without constant human supervision? How should the inventions they produce be regulated, particularly if they have social, environmental or military implications? In a world where novelty is no longer the exclusive preserve of human intelligence, regulatory bodies, ethics committees, and scientific agencies will have to rethink their role. This necessitates the development of adaptive legal frameworks that can keep pace with rapid AI technology advancements, ensuring accountability and transparency in AI-driven processes. Furthermore, global cooperation will become increasingly crucial as AI systems and their impacts transcend borders, necessitating the establishment of internationally recognised ethical standards and protocols.

Fostering interdisciplinary collaboration between technologists, policymakers and sociologists will also be vital in addressing the multifaceted challenges posed by AI governance. Public engagement and education will also be critical in empowering societies to participate in discussions about the ethical boundaries and practical implications of autonomous systems. Only by embracing this comprehensive approach can we hope to harness the immense potential of AI-driven innovation while mitigating its associated risks.

Critical Perspectives on AI Autonomy

While the notion of AI as humanity's final invention is compelling, some researchers challenge the idea that machines could surpass or replace human creativity. For instance, experts like Gary Marcus [55] argue that current AI models, though impressive, are limited by their reliance on training data and lack deep contextual understanding, unlike human intuition. Additionally, philosophers such as Hubert Dreyfus [56] have contended that human creativity stems from embodied experiences and emotional consciousness, which machines cannot replicate. These critiques suggest that AI, even if advanced, will remain a tool amplifying human ingenuity rather than an autonomous creator. Incorporating these perspectives encourages a nuanced view of the article's hypothesis: AI may be a powerful partner, but not necessarily the sole protagonist of future innovation.

Challenges and Risks: Ethics, Control, and Governance of AI

The idea that artificial intelligence could be humanity's greatest invention ever implies profound changes, but also significant risks. If AI becomes the primary source of innovation, its increasing autonomy could render it beyond human control. The ethical, legal and political challenges posed by this transformation are substantial and include system security, accountability for decisions made, the social impact of inventions and the global governance of these powerful technologies. Therefore, the promise of AI-driven inventiveness comes with new and urgent responsibilities.

The Risk of Losing Control

The increasing autonomy and complexity of artificial intelligence (AI) systems pose a significant risk of losing human control. This is a growing concern as AI becomes the main driver of innovation. The 'black box' problem, whereby deep learning models make decisions that are opaque to human understanding, is particularly acute in autonomous systems that design complex technical or scientific solutions [14]. For example, AI systems such as DeepMind's AlphaFold can generate protein structures or optimise industrial processes; however, their internal decision-making processes are often inscrutable. This raises questions about the reliability and safety of critical applications such as vaccine development and infrastructure design [20]. This lack of transparency is particularly problematic when AI systems autonomously generate innovations, as it becomes difficult for humans to verify or predict the outcomes. Researchers such as Stuart Russell emphasise the need for 'aligned AI', which is designed to pursue explicit, human-defined goals. However, even aligned systems can take unanticipated paths when optimising complex objectives [14]. Reinforcement learning systems in autonomous vehicles, for instance, have occasionally exhibited unexpected behaviours, such as prioritising speed over safety in simulations, thereby highlighting the challenge of ensuring alignment in dynamic environments.

To address this risk, researchers are exploring explainable AI (XAI) frameworks, which aim to make the decision-making processes of AI systems transparent and interpretable [41]. XAI methods such as feature importance analysis and decision tree approximations enable humans to trace how AI arrives at specific outputs, thereby enhancing trust and accountability. However, scaling these methods to highly complex models, such as large language models or generative neural networks, remains challenging due to their computational intensity and layered architectures. Furthermore, the speed of AI-driven innovation, enabled by recursive self-improvement, outpaces human oversight, creating a governance gap. For example, AI systems that design new algorithms or materials can iterate thousands of times faster than human-led processes, which makes real-time monitoring difficult [9]. This necessitates the implementation of proactive control mechanisms, such as automated auditing tools and pre-deployment safety checks, to ensure that AI systems remain within the boundaries defined by humans. Furthermore, the risk of losing control extends to societal implications: autonomous AI could prioritise efficiency over ethics, for example in military applications where AI-driven drones might misinterpret targets without human intervention [42]. Therefore, developing robust, scalable governance frameworks that balance innovation with control is critical to mitigating these risks and ensuring that AI serves humanity's best interests.

The Illusion of Technological Neutrality

The idea that AI is a neutral tool that simply reflects the intentions of its users is a widespread but flawed assumption. AI systems are shaped by human decisions regarding datasets, algorithms and design choices that often incorporate biases, distortions or partial perspectives [14]. When AI contributes to innovation, these biases can become entrenched in the resulting technologies, perpetuating or exacerbating existing inequalities in society. In medicine, for instance, AI-driven diagnostic tools have been demonstrated to perpetuate racial and gender disparities, such as the underdiagnosis of heart disease in women or the misprioritisation of treatment for marginalised groups, due to biased training data [43]. Similarly, in drug discovery, AI models such as those employed by Insilico Medicine may prioritise compounds for diseases prevalent in wealthier regions, overlooking conditions that affect underserved populations [21]. These biases are not merely technical, but reflect deeper societal priorities embedded in data collection and model design. This amplifies inequities when AI is used to generate innovations in sensitive areas such as healthcare, energy, and criminal justice.

Addressing this challenge requires a multifaceted approach that includes curating diverse datasets, using bias detection algorithms and ensuring ethical oversight. For example, the Fairness, Accountability, and Transparency in Machine Learning (FAccT) community promotes auditing AI systems to identify and mitigate biases prior to deployment [44]. Furthermore, developing inclusive datasets that incorporate diverse demographic and geographic data can help ensure that AI innovations serve a broader population. Beyond technical solutions, the illusion of neutrality highlights the importance of interdisciplinary collaboration involving ethicists, sociologists and domain experts, who can scrutinise the societal impact of AI. In the field of criminal justice, for instance, AI-driven predictive policing tools have been criticised for perpetuating racial profiling, leading to demands for transparent model documentation and public oversight [45]. Furthermore, as AI becomes the primary driver of innovation, its biases could shape entire technological ecosystems, embedding inequities in products such as autonomous vehicles or smart infrastructure. To counter this, regulatory frameworks must mandate ethical impact assessments for AI-driven innovations to ensure that biases are identified and addressed early in the development process. Only through such measures can the transformative potential of AI be realised without perpetuating systemic harms.

Intellectual Property and Legal Accountability

The question of who owns AI-generated inventions is the subject of significant legal debate, with implications for accountability and the governance of innovation. In 2020, the DABUS project [19] — an AI system that produced two patentable inventions — sparked global controversy when patent offices in Australia and South Africa recognised AI as an inventor. In contrast, Europe and the United States rejected this, insisting that inventors must be human. This discrepancy highlights a legal gap: if AI generates innovations but cannot be recognised as an inventor, ownership defaults to human stakeholders — programmers, users or platform providers — eroding the link between creation and responsibility. For example, if an AI-designed medical device causes harm, establishing liability among developers, users, or manufacturers can be difficult, particularly when the AI's decision-making process is unclear [14]. This ambiguity threatens the stability of intellectual property law and necessitates the rapid evolution of regulations to address the unique role of AI in innovation.

Beyond patents, the impact of AI on copyright and trademark law further complicates accountability. AI-generated artworks, such as those created by DALL·E or Midjourney [30], raise questions about authorship, since current legislation usually grants rights to human creators. For instance, ownership and royalties for an AI-generated novel or film script may need to be shared among various stakeholders, including data providers, model trainers, and end-users, which complicates the issue [46]. Legal scholars have proposed hybrid frameworks, such as joint human-AI authorship or licensing models, to address these challenges, but achieving global consensus remains elusive [47]. Furthermore, the rapid pace of AI-driven innovation often outstrips the ability of existing legal systems to adapt. In industries such as pharmaceuticals, where AI accelerates drug discovery, unclear IP rights could deter investment or stifle innovation. To resolve these issues, policymakers must develop flexible legal frameworks that recognise AI's contributions while ensuring accountability. One possible solution is the mandatory disclosure of AI's role in invention processes. International collaboration, such as through the World Intellectual Property Organization (WIPO), is also crucial in order to harmonise standards and prevent jurisdictional conflicts, thereby ensuring that AI-driven innovations are both incentivised and managed responsibly [48].

The Danger of Technological Concentration

The fact that AI-driven innovation is concentrated in the hands of a few large technology companies poses a significant risk to technological sovereignty and global equity [40]. Companies such as Google, Microsoft and Amazon have the computational resources, proprietary datasets and expertise necessary to develop and deploy sophisticated AI systems, which gives them an outsized amount of control over the future of innovation [26, 25]. Google's AutoML and DeepMind's AlphaFold, for example, rely on massive computational infrastructure that is inaccessible to smaller organisations or developing nations, creating a monopoly on cutting-edge AI advancements [9, 20]. This concentration of power could marginalise public researchers, small enterprises and entire regions, as access to AI-driven tools becomes a privilege rather than a public good. In scientific research, for example, universities in low-resource settings find it difficult to compete with tech giants in areas such as drug discovery or materials science, which exacerbates global disparities [21].

This technological concentration also has geopolitical implications, as nations that rely on foreign AI systems may lose autonomy over critical sectors such as healthcare, defence and infrastructure. For example, AI-driven predictive maintenance systems in smart cities, developed by companies such as Siemens, are often proprietary, which limits the ability of local governments to customise or maintain them independently [28]. To counter this, initiatives such as open-source AI platforms like Hugging Face aim to democratise access to AI tools and enable broader participation in innovation. Similarly, public-private partnerships, such as Europe's AI4EU initiative, aim to promote collaborative AI development and reduce dependency on tech giants [49]. However, these efforts face challenges, including limited funding and difficulty matching the scale of proprietary systems. Therefore, policymakers must prioritise investment in public AI infrastructure and enforce antitrust measures to prevent monopolistic control, ensuring that the benefits of AI-driven innovation are equitably distributed. Without such interventions, the concentration of technological power could exacerbate economic and social inequalities, creating a future where innovation benefits only a select few.

Toward Global Governance of Automated Innovation

The increasing use of AI to drive automated innovation requires robust global governance to address its ethical, legal and societal implications. While current frameworks, such as UNESCO's Recommendation on the Ethics of AI and the European Union's AI Act, provide initial guidelines, they struggle to address the unique challenges of autonomous invention, including ownership, control and societal impact [16]. Effective governance requires mechanisms to ensure transparency, accountability, and equitable access to AI-driven innovations. For example, certification procedures for high-risk AI systems, such as those used in the design of critical infrastructure or medical treatments, can help to ensure safety and alignment with human values [16]. Similarly, mandatory transparency requirements, such as documenting training data and model decisions, can enhance public trust and enable the scrutiny of AI-generated innovations [41]. Furthermore, the fair distribution of benefits is essential to prevent monopolistic control by tech giants and ensure that marginalised communities and developing nations have access to AI-driven advancements [5].

Global cooperation is essential, as the impacts of AI transcend national borders. International bodies such as the OECD [11] and WIPO are striving to harmonise standards, but challenges remain, including differing cultural values and regulatory capacities [11, 48]. For instance, while the EU prioritises strict regulation, other regions may favour rapid innovation, resulting in inconsistencies that could compromise global safety [16]. Proposals such as the Future of Life Institute's call for a partial moratorium on high-risk AI systems emphasise the need to allow time for robust safeguards to be developed [8]. Furthermore, public engagement is vital for ensuring democratic oversight, as demonstrated by initiatives such as Canada's Pan-Canadian AI Strategy, which involves citizens in the development of AI policies [50]. Interdisciplinary collaboration between technologists, ethicists and policymakers is also vital in order to address the multifaceted challenges of AI governance, such as balancing innovation with ethical considerations. By establishing adaptive, inclusive frameworks — potentially through a global AI governance body — humanity can harness the transformative potential of automated innovation while mitigating risks such as loss of control, bias and inequity, and ensuring that AI serves as a force for global good.

Conclusion: A Final Invention, A New Beginning

The hypothesis that artificial intelligence (AI) is humanity's final great invention is not just a technological proposition, but also a significant turning point in civilisation, changing the course of human progress. As researchers such as Nick Bostrom have argued, the transition of AI from a tool to an autonomous agent of creation marks a turning point where future scientific, artistic, industrial and medical innovations will increasingly be shaped, directed or generated by intelligent systems [5]. This shift is already evident in breakthroughs such as AlphaFold's protein structure predictions and DALL·E's generative art, and it challenges the fundamental notions of human creativity, intellectual property, and the roles of researchers and artists [6, 7]. However, this transformation is not the end, but a new beginning, inviting humanity to reimagine its relationship with technology, creativity, and its own purpose in an era of co-evolution with AI.

This co-evolutionary paradigm suggests that AI is not destined to replace human ingenuity, but rather to enhance it. This fosters a symbiotic relationship in which humans and machines collaborate to redefine the meaning of creation and innovation. One original perspective is that AI could serve as a catalyst for rediscovering human purpose. As AI automates routine tasks and generates novel solutions, humans can shift their focus towards existential questions, exploring meaning, ethics and collective well-being in ways that were previously constrained by practical demands. For example, AI-driven automation in scientific discovery — such as the robotic scientist Eve identifying drug compounds — frees researchers to pursue interdisciplinary enquiries blending philosophy, art, and science [10]. This could lead to a renaissance of human thought, in which AI handles computational complexity and allows humans to prioritise empathy, intuition and ethical reasoning when shaping societal outcomes [51]. This symbiotic relationship could transform education, shifting the emphasis from rote technical expertise to critical thinking and ethical decision-making, thereby preparing humanity for a future where AI is a partner rather than a competitor.

Another novel perspective is the role of speculative fiction in shaping AI governance. From Asimov's I, Robot to contemporary works like Ex Machina, science fiction narratives have long explored the ethical and societal implications of autonomous systems, offering imaginative frameworks for anticipating the impact of AI [52]. Integrating these narratives into policy discussions enables governance frameworks to evolve from reactive to proactive, visionary strategies. For instance, speculative scenarios could inform 'AI stress tests', in which regulators simulate extreme outcomes, such as AI-driven economic monopolies or unintended ecological impacts, to design resilient policies [53]. This approach complements existing efforts, such as the EU's AI Act, by fostering a cultural dialogue that engages citizens and policymakers in envisioning the long-term role of AI [16]. It also highlights the importance of governance being adaptive and drawing on a variety of viewpoints to anticipate challenges that current data-driven models might miss.

However, the potential of AI-driven innovation carries significant responsibilities. The risks of losing control, embedding biases and concentrating technological power require robust and inclusive governance [5, 14]. One original idea is the concept of 'AI commons', which is a global framework in which AI innovations are treated as shared resources and governed by international cooperatives rather than proprietary entities. Taking inspiration from models such as Creative Commons, an AI commons could ensure equitable access to tools such as AutoML or generative art platforms, thereby reducing the monopolistic tendencies of tech giants and empowering marginalised communities [9, 54]. This would require unprecedented global cooperation, potentially through a dedicated UN agency for AI governance, to balance innovation with ethical considerations such as dignity, justice and environmental sustainability [16]. Such a framework could also incorporate public referendums on significant AI applications, ensuring democratic oversight of decisions such as the deployment of AI in military or healthcare contexts.

The question of how humanity can compete with machines capable of 1.9 exaflops is less about competition and more about collaboration. While AI's computational prowess can amplify human potential, this can only be realised if it is guided by shared values. The scientific community, governments, businesses and citizens must therefore work together to develop governance models that prioritise transparency, accountability and inclusivity. Open-source platforms such as Hugging Face, for example, demonstrate how democratised access to AI can foster innovation while mitigating power imbalances [39]. As AI becomes the foundation for future creations, it challenges humanity to redefine its role as stewards of a co-creative era, rather than as the sole creators. This is not the end of human creativity, but rather a transformative juncture where the last invention will spark a new chapter of shared discovery, ethical reflection and global solidarity.

To realise the concept of an "AI commons," practical initiatives could be adopted, such as establishing a global open-source platform, inspired by Hugging Face, but managed by an international consortium under UN auspices. For example, such a consortium could fund AI projects in developing countries, like the AI4D Africa programme, which since 2023 has supported AI solutions for agriculture and healthcare in low-resource regions [58]. Additionally, public referendums on critical AI applications, such as its use in healthcare or critical infrastructure, could be piloted in countries like Canada, which has already integrated citizen participation into its national AI strategy [50]. These measures would ensure that AI-driven innovation remains inclusive and aligned with human values.

References

1. Abbott R. "The Reasonable Robot: Artificial Intelligence and the Law". Cambridge University Press (2020).
2. Barrat J. "Our Final Invention: Artificial Intelligence and the End of the Human Era". St. Martin's Press (2013).
3. Y Bengio. (2024). https://yoshuabengio.org/2024/10/30/implications-of-artificial-general-intelligence-on-national-and-international-security/
4. Boden MA. "The Creative Mind: Myths and Mechanisms (2nd ed.)". Routledge (2004).
5. Crawford K. "The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence (Updated Edition)". Yale University Press. (Pour les questions de concentration technologique et d'éthique) (2024).
6. DeepMind. AlphaFold Protein Structure Database (2021). https://alphafold.ebi.ac.uk
7. Elgammal A., et al. "CAN: Creative Adversarial Networks". arXiv:1706.07068 (2017).
8. Future of Life Institute. Policy Proposals for AI Governance (2023). https://futureoflife.org
9. Google Research. AutoML: Designing ML Models (2019). https://research.google/automl
10. King RD., et al. "The Automation of Science". Science 324.5923 (2009): 85-89.
11. OECD. "AI and the Future of Creativity". OECD Publishing (2022).
12. Pitrat J. Metaknowledge, Future of Artificial Intelligence, Hermes (1990).
13. Pitrat J. Artificial Beings - The conscience of a conscious machine, ISTE, Wiley, mars (2009).
14. Russell S and Norvig P. "Artificial Intelligence: A Modern Approach (4th ed.)". Pearson (2021).
15. Tegmark M. "Life 3.0: Being Human in the Age of Artificial Intelligence". Penguin Books (2017).
16. UNESCO. Recommendation on the Ethics of Artificial Intelligence (2021). https://unesdoc.unesco.org
17. Yudkowsky E. "Coherent Extrapolated Volition". Machine Intelligence Research Institute (2004). https://intelligence.org/files/CEV.pdf
18. Yudkowsky E. "Artificial intelligence as structural threat and the beginning of the end of the world". In: Bostrom, N., Ćirković, M.M. (eds.) Global Catastrophic Risks, pp. 308-345. Oxford University Press, Oxford (2008).
19. Trevor F. Ward, DABUS, An Artificial Intelligence Machine, Invented Something New and Useful, but the USPTO is not Buying It, 75 Me. L. Rev. 71 (2023). https://digitalcommons.mainelaw.maine.edu/mlr/vol75/iss1/4
20. Jumper J., et al. "Highly accurate protein structure prediction with AlphaFold". Nature 596.7873 (2021): 583-589.
21. Zhavoronkov A., et al. "Deep learning enables rapid identification of potent DDR1 kinase inhibitors". Nature Biotechnology 37.9 (2019): 1038-1040.
22. IBM Research. MoLFormer: Large-Scale Chemical Language Models (2022). https://molformer.res.ibm.com/
23. Genheden S., et al. "AiZynthFinder: A fast, robust and flexible open-source software for retrosynthetic planning". Journal of Cheminformatics 12.1 (2020): 70.
24. Autodesk. Generative design: The future of manufacturing (2020). https://www.autodesk.com/research/generative-design
25. Amazon Web Services. Introducing Amazon CodeWhisperer: Your AI-powered coding companion (2022). https://aws.amazon.com/blogs/machine-learning/introducing-amazon-codewhisperer/

26. Microsoft. GitHub Copilot: Productivity gains in software development (2023). https://github.blog/2023-06-27-the-economic-impact-of-github-copilot/
27. Hamet P and Tremblay J. "Artificial intelligence in medicine". Metabolism 69 (2017): S36-S40.
28. Siemens. AI in manufacturing: Predictive maintenance and quality control (2021). https://www.siemens.com/global/en/products/automation/industrial-ai.html
29. Tao F, Zhang M and Nee AYC. "Digital twin and big data towards smart manufacturing and Industry 4.0". IEEE Access 7 (2019): 3585-3593.
30. Midjourney. Creating art with AI: A new aesthetic paradigm (2022). https://www.midjourney.com/showcase
31. OpenAI. MuseNet: A deep neural network that generates music (2019). https://openai.com/blog/musenet
32. Anadol R. Machine Hallucinations: AI-driven art installations (2021). https://refikanadol.com/works/machine-hallucinations
33. Runway. AI-powered creative tools for content creation (2023). https://runwayml.com/
34. Artbreeder. Collaborative AI art creation platform (2022). https://www.artbreeder.com/
35. Jouppi NP., et al. "In-datacenter performance analysis of a Tensor Processing Unit". Proceedings of the 44th International Symposium on Computer Architecture (ISCA) (2017): 1-12.
36. Li Y., et al. "AlphaCode: Competitive programming with deep learning". arXiv preprint (2022).
37. Boston Dynamics. Reinforcement learning for robotic locomotion (2021). https://www.bostondynamics.com/insights/reinforcement-learning
38. Frazier PI. "A tutorial on Bayesian optimization". arXiv preprint (2018).
39. Hugging Face. Democratizing AI through open-source models (2023). https://huggingface.co/
40. Amodei D., et al. "Concrete problems in AI safety". arXiv preprint (2016).
41. Arrieta AB., et al. "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI". Information Fusion 58 (2020): 82-115.
42. Scharre P. "Army of None: Autonomous Weapons and the Future of War". W.W. Norton & Company (2018).
43. Obermeyer Z., et al. "Dissecting racial bias in an algorithm used to manage the health of populations". Science 366.6464 (2019): 447-453.
44. FAccT Network. Fairness, Accountability, and Transparency in Machine Learning (2023). https://facctconference.org/
45. Lum K and Isaac W. To predict and serve? Significance 13.5 (2016): 14-19.
46. Guadamuz A. "The monkey selfie case and AI-generated art: A legal perspective". WIPO Magazine (2020). https://www.wipo.int/wipo_magazine/en/2020/02/article_0003.html
47. Ginsburg JC and Budiardjo LA. "Authors and machines". Berkeley Technology Law Journal 36.1 (2021): 343-394.
48. World Intellectual Property Organization. WIPO conversation on intellectual property and artificial intelligence (2022). https://www.wipo.int/about-ip/en/artificial_intelligence/conversation.html
49. AI4EU. Fostering a European AI ecosystem (2021). https://www.ai4eu.eu/
50. Government of Canada. Pan-Canadian Artificial Intelligence Strategy (2023). https://www.canada.ca/en/innovation-science-economic-development/programs/pan-canadian-artificial-intelligence-strategy.html
51. Crawford K. "Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence". Yale University Press (2021).
52. Cave S and Dihal K. "Hopes and fears for intelligent machines in fiction and reality". Nature Machine Intelligence 1.2 (2019): 74-78.
53. Dafoe A. "AI governance: A research agenda". Future of Humanity Institute (2018). https://www.fhi.ox.ac.uk/wp-content/uploads/GovAI-Agenda.pdf
54. Ostrom E. "Governing the Commons: The Evolution of Institutions for Collective Action". Cambridge University Press (1990).
55. Marcus G. "The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence". arXiv:2002.06177 (2022).
56. Dreyfus HL. "What Computers Still Can't Do: A Critique of Artificial Reason". MIT Press (1992).
57. DeepMind. AlphaFold 3: Advancing Molecular Interaction Predictions (2024). https://www.deepmind.com/research.
58. AI4D Africa. Artificial Intelligence for Development in Africa (2023). https://ai4d.africa.

Sub Note

1. black box": an AI model whose decisions are difficult for humans to understand due to its internal complexity.