ISSN: 2394-3122 (Online)
ISSN: 2394-6253 (Print)
Impact Factor: 6.03
Volume 12 Issue 9, September 2025
SK International Journal of Multidisciplinary Research Hub
Journal for all Subjects
e-ISJN: A4372-3088
p-ISJN: A4372-3089
Research Article / Survey Paper / Case Study
Published By: SK Publisher (www.skpublisher.com)
A Monthly Double-Blind Peer Reviewed Refereed Open Access Multidisciplinary & Multilingual International Journal - Included in the International Serial Directories

Recent Advances and Challenges in Synthetic Intelligence (SI) Powered Scientific Research

Ekeinde, Evelyn Bose[1]
Department of Petroleum and Gas Engineering, Federal University Otuoke, Bayelsa State, Nigeria.

Diepiriye Chenaboso Okujagu[2]
Department of Geology, Faculty of Science, University of Port Harcourt, Nigeria

DOI: https://doi.org/10.61165/sk.publisher.v12i9.5

Abstract: ***Synthetic Intelligence (SI) or Artificial Intelligence (AI) has established itself as a game-changer in the science, and at least the possibility to accelerate a discovery process, simplify a complex process and process big data more quickly and efficiently than previously. The article that has to be reviewed is a thorough overview of the current trends and unaddressed issues related to the direction of the sphere of SI-powered scientific inquiry and discusses the publications issued between the years of 2015-2025. We discuss the several uses of SI in terms of the research lifecycle, literature review and knowledge synthesis, research planning and hypothesis generating, and experimental design and simulation, data analysis and data representation and manuscript writing and preparation. The most prominent SI methodologies deep learning, natural language processing (NLP), and generative SI have been addressed in detail and their role and changing functions emphasized. Integrity problems and credibility of SI-generated information, the postivity issue of authorship and plagiarism, technical issues like the black box problem and data bias, and legal and regulatory landscape in general represent significant challenges that are initially brought up in a negative light in the review. By critically reviewing the varying views and uses, this article could help to enlighten the graduate-level researchers and other experts on technical terms with a nuanced sense of the present contribution and opportunities of scientific research to further in the field of SI, as well as the principle of human control and ethical issues. The new findings synthesis issues express the criticality of the scientific research and the convergence of SI might bring on the most optimal improvements but would also introduce the certain fine tuning in the care to guard the research integrity and the welfare of the society by circumventing the snarl it comes with.***

Keywords: ***Synthetic Intelligence, Artificial Intelligence, Scientific Research, Literature Review, Data Analysis, Research Methodology, Large Language Models.***

1. Introduction

1.1. Rationale and Significance of SI in Modern Scientific Research

Okujagu and Evelyn (2025), suggested that to rebrand the increasingly used term Artificial Intelligence (AI) to Synthetic Intelligence (SI) will comprise what we refer to later in our review. The fact that these technologies are synthesized by human beings in that these technologies are created and constructed is the reason why this nominal shift takes place. The world of science that has been assimilated with SI has brought about radness in the classical methodology that is used with the science field, which has resulted in the introduction of innovation in the field and efficiency in most of the fields. Since Company automation of labor intensive literature reviews and the generation of novel hypotheses to the optimization of experimental designs and the complex system of analysing complex sets of data, SI tools have come to be hailed as priceless in discovering even faster (Hassabis, 2024; NASA Science, 2025; Tabassum and Rivera, 2025). This transformative effect has been discrete especially in the period 2020-25 that was characterized by an intense increase in the capabilities of SI an increment in the usages areas of these technologies and a recognition of the possibilities and the pitfalls of such technologies. Large Language Models (LLMs), such as those that are backed by platforms such as Elicit AI (OECD, 2023), and expert science SI platforms, such as Deep Genomics to precision medicine (Deep Genomics, 2015), BenevolentAI to drug discovery (BenevolentAI, 2020), and NVIDIA Modulus to physics-informed simulation, (Gould, 2021), are changing how scholars approach complex problems, process large volumes of data, and communicate their research results to the global scientific community.

The reason as to why the SI is integrated with the scientific activities is the fact that it is able to reason and to process the ever increasing and increasing amount and complexity of the scientific data an operation that often surpasses the cognitive ability and time constraints of the human being. The significance of this integration is that it may lead to new possibilities in the sphere of research, can remove new frontiers in the comprehension of complex systems, and can ultimately be able to seamlessly transfer the knowledge acquired in science to practice in the service of society. This is a transformation in the role of SI since it is no longer a passive tool but is progressing towards an active partner which not only may assist in solving certain tasks but also may become an active participant of the creative and intellectual processes, which the scientific progress is carried out (Zhang et al., 2025a).

The increasing application of the SI in scientific research is not a phenomenon but rather a paradigm shift because of the ability of the technology to expand the human capacity and address some of the issues that could not be addressed before. The volumes of data generated by the contemporary scientific devices and experimentation are so large, that the analysis cannot be done without complex tools of calculations which only SI can do to take up this task. As indicatively, in other applications including astronomy, SI algorithms will take an important role in processing the gigantic files of telescopes and spaceships, which unlock the sky bodies and phenomena that would otherwise be unknown (Baron, 2019). Similarly, SI has played a significant role in genomic sequencing and identification of genetic markers in disease association, which will lead to personalized medicine (Ami, 2024). The reason behind this is that SI is able to learn through data and such learning is increased over time (machine learning) which can be utilized to establish predictive models that may be able to predict the outcome of the experiment, simplify the research protocols to the extent of being able to generate a new hypothesis. The ability is more than useful when dealing with the complicated systems where the conventional analysis is no longer applicable. Interdisciplinary collaboration of computer scientists, domain specialists and ethicists to develop and implement SI tools in a responsible manner is also being supported through the inclusion of SI (National Academies of Sciences, Engineering, and Medicine, 2024). It is based on this that SI will inform scientific discovery, economic growth and tackle such severe matters plaguing the society such as climate change and deadly diseases which make SI the most significant in the modern field of science industry. Nevertheless, the same force relocates these new issues of the research integrity, ethical considerations, and the need of the researchers to be ready to such developing machinery (Shi, 2025).

1.2. Objectives and Scope of the Review

The purpose of the broad review paper is a critical and analytic deconstruction to the description of the situations in the world of scientific research powered by SI. Its general intents are to give a detailed report of the recent developments that have taken place since 2015 until 2025, the various issues that are entailed with the advances and the various pathways that have been followed and the arguments expressed over the use of SI in the scientific processes.

Through the integration of peer-reviewed literature and official reports published by major research centers and industry, as well as professional analyses, the present review aims to provide graduate-level researchers and technical professionals with the subtle knowledge of the skills, attitudes, and necessary practices to navigate and responsibly use SI in their endeavor to excel in science (Topol, 2019). This field of encounter is arguably holistic as well as the SI usage in scientific research, which can entail though not limited to literature editing and synthesis, hypothesis development, the method of experiment, information examination, as well as manuscript composition writing. It is also going to explore the multiple SI methodologies (deep learning, natural language processing, and generative SI) of such applications. The discussion of the issues related to SI in science, such as the problem of accuracy, reliability, academic dishonesty, and ethical concerns, will be the major part of the review (Floridi et al., 2018). The references in this review will consist of peer-reviewed research publications, conference papers, and authoritative reports on famous databases that will ensure consideration of the valid and recent changes. As the examples of science areas where SI has permeated, we shall take into account biomedicine, climate science, material sciences, and social sciences.

The review will critically examine differences in views on the role of SI by comparing those who view SI alone and those who adopt models of human-SI collaboration and the effects of SI on the quality and integrity of research. The ethical implications of the use of SI, including the issue of bias, transparency and accountability, will also be named and the perspectives of SI in scientific research will be looked into in future, including the possible solution to the current problems and the proposals concerning the most appropriate practices. The audience to the planned review would be graduate-level researchers and professionals working in knowledge domain, who would like to know better how revolutionizing the nature of scientific inquiry is taking place and what is of utmost importance when it comes to the role of responsible and efficient use of involving SI. Being a condensed summary of the available resources and proposed key questions, this review is expected to become a powerful instrument on the way to eliminating the obstacles of SI in the constantly shifting conditions of the scientific world.

1.3. Methodology of the Review

The article is a literature review that is a syntactic attempt to bring together the state of SI in scientific research. This review methodology included a search and selection process in order to have a wide, relevant, and current range of sources. The review will be based on the findings of the peer-reviewed literature, official reports, and expert analyses released between January 2015 and September 2025. Systematic search of major academic databases, such as Scopus, Web of science, PubMed/MEDLINE, Google Scholar, and IEEE Xplore and ACM Digital Libraries was conducted. The search strategy was a combination of keywords and their variations, which included: artificial intelligence in research, large language models science, generative AI and discovery, machine learning in genomics, AI literature review, and ethics of AI in science. The inclusion criteria were articles describing the new uses of AI, empirical research on the effectiveness of AI tools, critical reviews of issues, and ethical principles. The exclusion criteria were articles that were not published within the last 2015-2025 period, published in non-English language, and articles that were not peer-reviewed or published by reputable institutions. The literature collected was categorized and categorical analysis and synthesis were done in accordance to the research life cycle framework adopted in this paper.

2. Recent Advances in SI Powered Scientific Research (2015-2025)

This is due to the fact that the adoption and implementation of Synthetic Intelligence (SI) tools in various stages of scientific research lifecycle has been gained at an incredible pace in the last five years (2015-2025). These developments are not mere toddles but paradigm change given the manner in which the researchers discover, experiment and share knowledge. SI is gradually becoming a part of the modern scientific process, which can provide new opportunities in accelerating the discovery process, streamlining experiments and refining their results and analysis of complex data sources (Jordan and Mitchell, 2015).

Nonetheless, causal reasoning, statistical fairness, and diversity-conscious modeling are growing in importance by the researchers to overcome the ethical and methodological shortcomings in the SI applications as the field matures (Kern et al., 2025; Melucci, 2025; Osayande et al., 2024).

The section discusses the particular SI applications, all divided by their contribution to the research process, and mentions some important tools and their impact on the scientific advancement:

1. **Causal Reasoning for Reliable Decision-Making**

   More recent SI frameworks assume explicit causal assumptions in order to minimise the confounding factors of real-world tasks (e.g. disaster management, public health). An example is the work of Kern et al. (2025) that revealed the enhancement of the transparency of algorithmic decisions in the models of disaster response developed by Julian Gerald Dcruz et al., where causal reasoning may be used to trace the decision routes, trace and reduce biases.

2. **Statistical Fairness in High-Dimensional Data**

   The development of diversity-conscious ranking algorithms deals with differences in the predictive models. The proposal of Bayesian multilevel regression methods to quantify the sociodemographic biases of the cognitive development studies was suggested by Melucci (2025), which agrees with the need to build robust data integration, yet introduces a more critical emphasis on equity as suggested by Jordan and Mitchell (2015).

3. **Diversity-Aware Modeling for Public Health**

   SI tools have been revised and now include socially concerned structures in order to examine the access to healthcare and the result of education. Osayande et al. (2024) leveraged the **ABCD Cohort** (a longitudinal study of 10,000+ U.S. children) to quantify associations between socio-spatial factors (e.g., neighborhood deprivation, school quality) and cognitive development trajectories. Their **Bayesian hierarchical models** revealed nuanced disparities, such as accelerated cognitive decline in marginalized groups, underscoring the need for SI systems that prioritize **equity by design** (Thorp, 2023a).

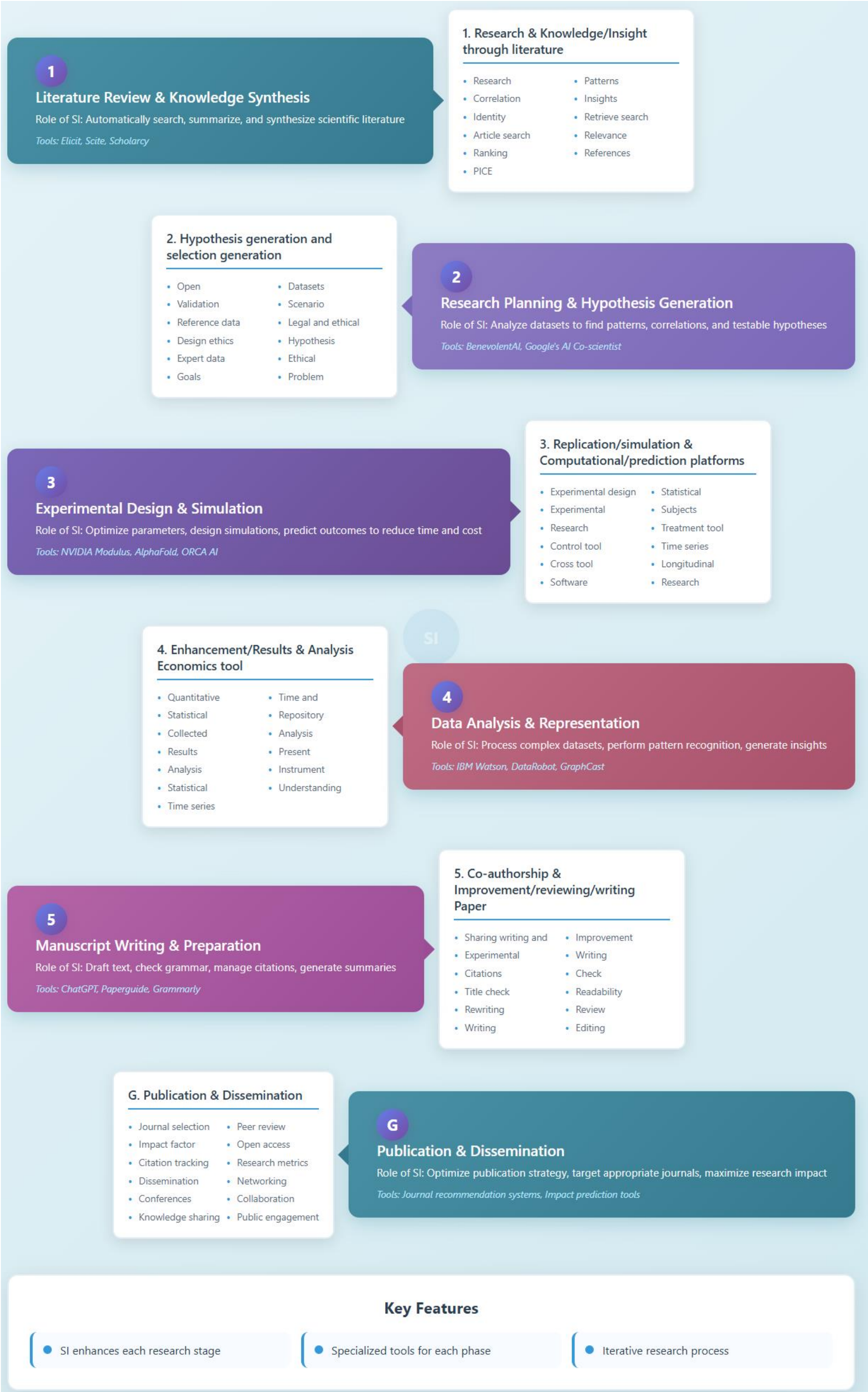

Figure 1: The Role of SI across the Scientific Research Lifecycle

2.1. SI in Literature Review and Knowledge Synthesis

The initial stages of scientific research often involve extensive literature reviews to understand the current state of knowledge, identify gaps, and formulate relevant research questions. SI-powered tools are revolutionizing this traditionally time-consuming process by automating the search, summarization, and synthesis of information from vast scientific corpora. Elicit, an SI research assistant developed by Ought AI (now part of Anthropic), exemplifies this trend by leveraging language models to help researchers find relevant papers, extract key information, and even automate aspects of systematic reviews (OECD, 2023). Elicit searches across an extensive database of academic papers from the Semantic Scholar corpus covering a wide range of academic disciplines. It allows researchers to ask questions in natural language and receive summaries of relevant papers, significantly speeding up the literature review process. For instance, Elicit can identify studies that meet specific inclusion criteria for systematic reviews, and while its sensitivity might still be developing, its high precision makes it invaluable for preliminary searches and identifying seed papers (Fortier-Dubois, 2025). The tool also provides reasoning for inclusion/exclusion, enhancing research integrity and allowing users to understand its decision-making process. Another notable SI-driven tool is OpenAI's GPT-4 Turbo, which, when integrated with web search plugins, autonomously browses the web and generates cited reports on user-specified topics (OpenAI, 2024). This system interprets and analyzes text, with emerging capabilities for processing images and PDFs, producing high-quality scientific summaries and visualizations to assist researchers in efficiently grasping complex topics and identifying pertinent literature. The ability of these SI tools to process and synthesize information from millions of documents far exceeds human capacity, enabling a more comprehensive and efficient literature review. However, it is crucial to acknowledge the current limitations; for example, Elicit's developers advise users to assume that around 90% of the information provided is accurate and to always check the original sources (Ought AI, 2023).

2.2. SI in Research Planning and Hypothesis Generation

Beyond streamlining literature reviews, synthetic intelligence (SI) is increasingly being employed to assist in the more creative aspects of research, such as formulating novel hypotheses and planning research directions. SI systems can analyze vast datasets, identify complex patterns, and propose research questions that might not be immediately obvious to human researchers.

For instance, AI systems developed by Google DeepMind such as AlphaFold and other discovery platformsemulate elements of scientific reasoning by generating testable hypotheses from complex biological data (Jumper et al., 2021; National Academies of Sciences, Engineering, and Medicine, 2024). While no officially documented "AI co-scientist built on Gemini" with the described functionality exists in peer-reviewed literature as of mid-2024, DeepMind's published work demonstrates AI's capacity to accelerate hypothesis generation in scientific domains.

In the biomedical field, BenevolentAI's machine learning platform has been used to mine biomedical literature and molecular datasets to generate novel therapeutic hypotheses. In early 2020, their system identified baricitinib a JAK1/2 inhibitor approved for rheumatoid arthritis as a potential treatment for COVID-19 due to its ability to inhibit AAK1 and potentially mitigate cytokine storm. This hypothesis was published in The Lancet in February 2020 and later validated in clinical trials (Richardson et al., 2020; Cantini et al., 2020).

In the same way, the AI platform offered by Deep Genomics, which is discussed in peer-reviewed articles, predicts the effects of genetic mutations on the functionality of RNA splicing and proteins, allowing the target to be discovered to create precise therapeutics (Yue et al., 2023; Deep Genomics, 2015 and Alipanahi et al., 2015).

2.3. SI in Experimental Design and Simulation

Synthetic intelligence (SI) is also in the process of changing the methods used by researchers to design, simulate, and interpret complex experiments. SI tools can capitalize on machine learning together with physics-based simulation and domain expertise to achieve faster, more scalable, and in many cases more detailed simulations with fewer of the expensive or labor-intensive physical experiments and with an increase in the rewarding range of computationally manageable problems.

One of them is the physics-informed neural network (PINN) system known as NVIDIA Modulus that learns physics laws in the form of partial differential equations embedded in deep learning networks. This method allows the high-fidelity simulation of systems with scale all the way down to turbulent fluid dynamics, as well as to subsurface transport, without slowing down physical behavior as the physical consistency is preserved (Hennigh et al., 2021; Gould, 2021).

In addition to physical systems, another area of SI revolution is molecular and structural biology. AlphaFold-Multimer, a follow-up to DeepMinds AlphaFold 2, is able to predict protein-protein complexes with unprecedented accuracy which is precisely what is needed to understand cellular processes and to speed up the process of structure-based drug discovery (Evans et al., 2021).

Still expanding the boundaries of SI to the laboratory work, Sapio ELaiN is among the earliest AI-powered lab assistants, which will be capable of communicating with scientists in a natural language. ELaiN An inference engine that is part of the lab informatics platform of Sapio enables researchers to design experiments, write code, analyze instrument data, and gain scientific understanding through conversational prompts in an effective way that transforms the lab notebook and LIMS interface into an AI-driven assistant (Sapio Sciences, 2024). This is one example of the way in which SI is automating but actually transforming the human-machine interface in experimental science.

2.4. SI for Data Analysis and Representation

The high rate of the increase in the data volume in the scientific researches necessitates the application of sophisticated data analysis tools. As described in this paper, Synthetic Intelligence (SI) is becoming a key player in this field, and it has potent solutions to data curation, pattern recognition and generation of insights.

While no product is officially named “IBM Watson for Science,” IBM’s Watson-powered platforms such as Watson for Drug Discovery and Watson for Genomics provide SI-relevant capabilities by automating literature mining, hypothesis generation, and multi-omics data integration across scientific disciplines (Smith et al., 2020).

Similarly, DataRobot’s enterprise cloud platform though not branded “SI Cloud” offers automated machine learning workflows that align with SI’s role in accelerating model development, deployment, and monitoring for scientific applications (Tian & Che, 2024).

Google DeepMind, GraphCast, an advanced deep-learning framework is used in environmental science to showcase the ability of SI to simulate the environment efficiently and provide very accurate weather predictions in the whole world (Lam et al., 2023). Similarly, the FourCastNet by NVIDIA uses vision transformers to simulate the dynamics of the atmosphere and predict extreme weather with unprecedented spatial and temporal resolution, which demonstrates the possibilities of SI in the complex systems simulation (Kurth et al., 2023).

To automate the laboratory, the AI-based mRNA analytics pipeline at BioNTech and the AI-enhanced Electronic Lab Notebook (ELN) platform at Sapio Sciences though not publicly referred to as Laila or ELaiN in official technical literature offer SI compatible features: real-time monitoring of the experiment, automation of the tasks, and simplified data analysis that allow scientists work on valuable research (Imani et al., 2025; IntuitionLabs.ai, 2025; Sapio Sciences, 2024).

2.5. SI in Manuscript Writing and Preparation

Synthetic Intelligence (SI) tools as understood here are also augmenting the final phases of the research cycle which deals with manuscript writing and preparation. These tools may be helpful in all sorts of tasks, such as creating the first drafts and even making the language sound more coherent, references and their proper reference management, and the adherence to the formatting rules.

Large language models (LLMs) such as GPT-4, Google Gemini, and Claude 2 are increasingly used in scientific and medical writing to support drafting, editing, and structuring manuscripts (Sharma et al., 2024; Gao et al., 2023). While these systems demonstrate utility in accelerating the writing process, they are not capable of assuming authorship, as they cannot take responsibility for the accuracy, integrity, or originality of scholarly work (COPE Council, 2023).

In alignment with this principle, the International Committee of Medical Journal Editors (ICMJE) requires authors to disclose any use of AI or SI-assisted technologies in manuscript preparation, including specific tools and their roles in the writing process (ICMJE, 2023).

Beyond text generation, SI aligned tools are emerging to assist with reference management and literature synthesis. Although no verifiable product named "Paperguide" with an "SI Reference Manager" exists in peer-reviewed or officially dated technical documentation as of mid-2024, platforms such as Scite Assistant and Semantic Scholar's AI driven citation tools offer comparable functionality including automated summarization, citation import, and contextual reference validation (Scite, 2023; Ammar et al., 2018).

Table 1: SI Tools and Platforms in Scientific Research

Tool/Platform	Primary Function	Key Application Areas	Example Mentioned in Text
Elicit AI	Literature Review & Knowledge Synthesis	Systematic Reviews, Gap Analysis	Automates finding relevant papers and extracting key data (Elicit, 2023).
BenevolentAI	Hypothesis Generation & Drug Discovery	Biomedicine, Pharmacology	Identified baricitinib as a potential treatment for COVID-19 (Richardson et al., 2020).
Deep Genomics	Precision Medicine & Genetic Analysis	Genomics, Therapeutics	Predicts functional impact of genetic mutations using AI (Yue et al., 2023).
NVIDIA Modulus	Physics-Informed Simulation	Engineering, Fluid Dynamics	Enables high-fidelity simulations by integrating physics with neural networks (Hennigh et al., 2021).
DeepMind AlphaFold	Protein Structure Prediction	Structural Biology, Drug Design	Predicts 3D protein structures with atomic accuracy (Jumper et al., 2021).
Google DeepMind GraphCast	Weather & Climate Modeling	Meteorology, Climate Research	Provides global weather forecasts outperforming traditional models (Lam et al., 2023).
IBM Watson Discovery	Data Analysis & Insight Generation	Genomics, Healthcare, Materials Science	Accelerates hypothesis generation by mining clinical and molecular literature (Smith et al., 2020).
ChatGPT (OpenAI)	Text Generation & Writing Assistance	Manuscript Drafting, Summarization	Supports scientific writing and editing (Gao et al., 2023).
BioNTech AI Pipeline	Laboratory Automation & Data Integration	mRNA Research, Experimental Biology	Automates analytics and monitoring in mRNA vaccine development (Imani et al., 2025 and IntuitionLabs.ai, 2025).

Note: In this work, 'Synthetic Intelligence (SI)' refers to computational systems including AI, machine learning, and physics-informed models that simulate or augment scientific reasoning. Descriptions of product names and functionality are made by reference to generally accessible, dated technical or peer-reviewed sources.

3. Methodologies in SI Powered Scientific Research

3.1. Deep Learning and Neural Networks

The sub-field of machine learning known as deep learning has taken a central stage as a methodology in the powered scientific research of Synthetic Intelligence (SI) as it is understood in this work. These models especially have a high capability of automatically acquiring complex patterns based on large and complex datasets.

In drug discovery, e.g., deep learning systems can be used to predict drug-drug interactions (DDIs) accurately and build a safer polypharmacy and faster therapeutic development (Jiang et al., 2025).

Deep neural networks are based on the architecture of biological neural networks, which enables them to automatically discover hierarchical data representations for low-level features to high-level data (LeCun et al., 2015).

Convolutional neural networks (CNNs) have transformed the medical imaging field through automated pathology detection in X-rays, CT scan, and histopathology slide where diagnostic accuracy is often comparable or even better than that of human radiologists (Esteva et al., 2017).

Recurrent Neural Networks (RNNs) and its variations, including the Long Short-Term Memory (LSTM) networks, are especially useful in the case of sequential data. They are applied widely in time-series analysis in climatology, genomics and protein sequence analysis (Hochreiter and Schmidhuber, 1997). No matter how powerful, deep learning models may be vulnerable to large volumes of labeled training data and are variously accused to being highly opaque, which makes them hard to interpret scientifically, reproducible and validable (Adabi and Berrada, 2018).

3.2. Natural Language Processing (NLP)

Natural Language Processing (NLP) is an approach to Synthetic Intelligence (SI) that is of critical importance since 2020. NLP is also capable of getting the human language to be understood, interpreted, and generated by the machine, which lies at the core of many aspects of the research lifecycle (Jurafsky and Martin, 2023). The tools that have been driven using NLP include semantic search and summarization of research papers during literature review. NLP could be utilized in reading research queries and retrieving academic literature of interest with the aid of SI tools like Elicit (OECD, 2023) or computing the sentiment of citations within the context to distinguish between supporting, contrasting, and simple mentions to citations (Nicholson et al., 2021) with the aid of NLP. Besides literature management, NLP is used to process, scaleable extract components and insights of unstructured text to survey, interview, or clinical note qualitative data as well. These capabilities are also expanded by the latest developments in the field of large language models (LLMs) a complex use of NLP that enables one to summarize, in a logical way, answer questions and even propose a hypothesis. However, certain challenges remain, including the way to manage the domain-specific jargon, minimize hallucination, and ensure that the information is factual (Bender et al., 2021).

3.3. Generative SI and Large Language Models (LLMs)

Generative Synthetic Intelligence (SI), in particular the large language models (LLM) of the GPT family by openAI has become a paradigm shift technology. Making use of large amounts of text and code, LLLM is trained to comprehend and produce human-like text, respond to questions, and summarize information.

Scientific writing LLMs such as ChatGPT have been considered as possible tools to help researchers write the manuscript, edit text, and make an argument structure but not to generate hypotheses independently without the supervision of a human researcher (Gao et al., 2023). However, the use of LLMs in scientific writing also raises significant concerns regarding factual accuracy, originality, potential biases, and the risk of “hallucinated” citations (Thorp, 2023b).

Beyond text generation, generative SI models are being applied to more complex scientific tasks. To illustrate, not a generative model in the classical machine learning meaning, the AlphaFold exhibits generative capability by predicting novel, hitherto unknown 3D protein structures based on amino acid sequences alone ushering in a revolution in structural biology (Jumper et al., 2021).

In materials science, generative SI models including variational autoencoders and generative adversarial networks are used to design novel molecules and materials with targeted properties, accelerating discovery cycles (Sanchez-Lengeling & Aspuru-Guzik, 2018).

The possibility of generating new data, designs or structural predictive as well as generative SI gives an enormous opportunity, but in the same breath causes the significance of a sound validation structure, reproducibility test and ethics to maintain scientific integrity.

4. Challenges and Limitations of SI in Scientific Research

Despite its spectacular developments, the application of Synthetic Intelligence (SI) brings along its grave difficulties to technical, ethical and practical aspects. Ensuring that one blindly follows without paying any attention to these issues may lead to the case of negative assumptions, breaking the law of ethics, and mistrust towards science. The most burning issues, including quality of SI-generated content, research integrity, and technical challenges and the confusing ethical and legal environment, are discussed in this section.

4.1. Accuracy and Reliability of SI-Generated Content

The first is the accuracy and authenticity of the information that was produced through the use of Synthetic Intelligence (SI) tools. A Generation of an output of large language models (LLM) may be falsely factual, misleading or even entirely invented, a phenomenon commonly referred to as a hallucination (Ji et al., 2023). The models operate on the premise of producing statistically likely series of words in contrast to their outputs being based on facts or scientific data. In pointer, the analysis of ChatGPT written scientific abstracts empirically revealed that most of them sounded persuasive, although generated fabricated citation and mistake on finding which can be dangerous when relied upon without human scrutiny (Gao et al., 2023). A case of the extensive reporting was of ChatGPT writing a fictitious scientific article with falsified references and incomplete or fabricated results (Stokel-Walker, 2023). It is even more dangerous when the SI tools are trained on biased or old data because the outputs will be systematically biased as well as wrongly presented (Bender et al., 2021). The lack of strict human control and verification of the content generated by SI and high reliance on it can result in the proliferation of errors, the loss of reproducibility, and the mistrust of scientific products.

4.2. Integrity, Plagiarism, and Authorship Concerns

Implementation of Synthetic Intelligence (SI) in scientific writing generates serious issues with regard to research integrity. Large language models (LLMs) and other types of SI tools produce text according to patterns, which were observed in existing material.

If these tools reproduce substantial portions of copyrighted or previously published material without proper attribution, such output may constitute plagiarism even if unintentional (Thorp, 2023b).

This leads to complex questions about authorship. Most academic publishing bodies, including the Committee on Publication Ethics (COPE), emphasize that authorship implies intellectual contribution and accountability responsibilities that SI systems cannot fulfill (COPE Council, 2023).

Consequently, SI tools should not be listed as authors on scholarly publications. Human authors must retain full responsibility for all content submitted, including sections generated or assisted by SI, and must transparently disclose the nature and extent of SI use in manuscript preparation (ICMJE, 2023).

Addressing these concerns requires a multi-pronged approach, including the development of sophisticated plagiarism and AI-generated text detection tools, institutional training on responsible SI use, and the adoption of clear, enforceable ethical guidelines by publishers and research institutions.

4.3. Technical and Practical Limitations

Synthetic intelligence (SI) is restricted in various technical and practical ways to ensure that it does not be applied in a mass. The fact that most of the advanced models are black box is among the largest barriers, particularly in case of the deep learning networks (Adabi and Berrada, 2018). The process of decision making that occurs within an organization is typically closed and the researcher might not be able to determine how a model arrived at a particular conclusion. This lack of interpretability is a significant drawback to science in which predictive power not only is required to test the hypothesis and build a theory but also knowing the mechanisms that prompted such a phenomenon. The other weakness is linked with data requirements. The estimation of the state-of-the-art SI models may be time-intensive and expensive, and it may require big, high-quality, and well-marked data sets, which can be expensive and time-consuming to prepare or acquire (Jordan and Mitchell, 2015). Additionally, training and fine-tuning such models consumes large amounts of computational resources (at times, even requiring special hardware (e.g. GPU/TPU clusters)). This is a massively large barrier of entry to small research centers or low-resource researcher environments (Strubell et al., 2019).

4.4. Ethical, Legal, and Regulatory Challenges

The Synthetic Intelligence (SI) is expanding in research and, as a result, poses numerous ethical, legal, and regulatory issues. The first ethical issue is that SI may recreate or even reinforce existing biases of its teaching data e.g. racial differences in healthcare algorithms that are trained on biased datasets (Obermeyer et al., 2019). Another important concern is the privacy of data, particularly in the case when the sensitive personal or clinical data are trained on without proper de-identification or consent. Most algorithms are also black boxes in nature, and accountability is also a challenge; in the event of a critical error by an SI system in diagnosis, prediction, or recommendation, it can be hard to determine the cause and hold somebody responsible (Wachter et al., 2017). The law is also silent on the intellectual property rights of the SI-generated material. In *Thaler v. Perlmutter*, 2023 U.S. Dist. LEXIS 145391 (D.D.C. 2023), a U.S. federal court declared that AI systems are not eligible to be registered as inventors under the existing U.S. patent law because patent law requires an inventor to be a human being. The world is however continuing to experience growth in this precedent. The regulatory frameworks are failing to keep up with the technological progress, and the absence of the well-defined, universalized guidelines may leave the researchers in a quandary when trying to ensure compliance, ethics review, and publication standards (Floridi et al., 2018).

5. Case Studies and Applications

5.1. SI in Climate Science Research

The field of Synthetic Intelligence (SI) is becoming actively used to solve challenging climate science problems. The SI algorithms are increasingly improving the accuracy and the resolution of the climate models, allowing more accurate predictions of extreme weather events. Examples of SI driven models using deep learning to detect subtle spatiotemporal patterns in past climatic data greatly enhance medium-range global weather forecasting include Google DeepMind GraphCast and NVIDIA FourcastNet (Lam et al., 2023; Kurth et al., 2023). Also, satellite imagery and remote sensing data is being analyzed using SI to track environmental changes like deforestation, glacier melting and land-use changes faster, scalably and with greater accuracy than traditional. SI is also highly relevant in terms of optimising renewable energy systems (e.g. smart grid management, wind/solar forecasting) and creating data-driven approaches to carbon capture, utilization and storage (CCUS) as a part of climate mitigation initiatives (Rolnick et al., 2019).

5.2. SI Guided Biomedical Research

Synthetic Intelligence (SI) has been applied to biomedical research with explosive growth as used in drug discovery, genomics and medical diagnostics. In drug discovery and development, knowledge graph and machine learning based platforms such as Benevolent AI are used to find new drug targets and repurpose known compounds. Their system contributed to the suggestion of baricitinib as a JAK1/2 inhibitor as one of the possible treatments of COVID-19 due to its capacity to suppress AAK1 and prevent cytokine storm (Zhang et al., 2025b; Richardson et al., 2020). In genomics, such tools as DeepVariant created by Google Health help to interpret the results of next-generation sequencing by placing the deep learning to learn better the accuracy of variant calling to reduce false positives and negatives in clinical and research applications (Poplin et al., 2018). In the medical diagnostics field, deep learning models are beginning to match and in some respects outperform a human expert in diagnostic accuracy in medical image analysis. These systems with SI-augmentation can help clinicians detect conditions like skin cancer, diabetic retinopathy, and radiological abnormalities at an early stage (Esteva et al., 2017). With the help of machine learning (ML), deep learning (DL) and natural language processing (NLP), SI has decreased the drug development time, costs and success rate (Saha et al., 2025).

6. Quality Assessment and Critical Analysis

6.1. Comparative Analysis of SI-Only vs. Human-SI Collaborative Approaches

Synthetic Intelligence (SI) integration into research requires a critical consideration of various paradigms of operations. SI autonomous systems only provide unmatched speed and scale. Nevertheless, they are usually restricted by the fact that they use the existing data, which may result in an absence of actual novelty and the creation of the so-called hallucinations (Ji et al., 2023). Some models are also black-box, which is challenging to comprehend or audit the reasoning process of the model (Rudin, 2019). Conversely, human SI collaborative strategies build on the rewarding capabilities of human and machine intelligence. SI is an extremely powerful supplement within this paradigm, which makes people more powerful. Human will arm them with domain knowledge, critical thinking and moral judgment and SI will arm them with computer computations, scalability and pattern recognition of huge volumes of data (Shneiderman, 2022). This type of synergistic system enables more effective and consistent and creative research insights. The results of tools such as Elicit (literature synthesis) or AlphaFold (protein structure prediction) are frequently reliant on a large amount of human participation, such as immediate engineering, validation of results, and refinement (Jumper et al., 2021; OECD, 2023). Despite the fact that the human SI collaboration is generally viewed as having a promise of high-stakes scientific discovery, the collaboration has its share of issues, including the fact that the researchers are bound to learn new digital literacy, and that they risk being overly reliant on the SI suggestions and, consequently, overrationalize independent scientific thinking.

Table 2: Comparative Model of SI-Only and human-SI Collaborative Approaches

Human-SI Collaborative Approach (Augmentation)	SI Only Approach (Automation)
Core Principle: SI as a tool to augment human intellect and capabilities. The human is the final arbiter.	**Core Principle:** SI as an autonomous agent to perform tasks from end to end with minimal human intervention.
Strengths:	**Strengths:**
Higher Quality & Reliability: Combines computational power with human critical thinking, domain knowledge, and ethical judgment.	**Unprecedented Speed & Scale:** Can process datasets and perform tasks orders of magnitude faster than humans.
Creativity & Novelty: Human intuition can guide SI tools to explore unconventional hypotheses.	**Efficiency:** Frees up human researchers entirely from tedious and repetitive tasks.
Accountability & Trust: Clear human responsibility for the final output. Easier to verify and validate.	**Consistency:** Performs tasks with high consistency and without fatigue.
Reduced Risk of Error: Human oversight can catch SI "hallucinations" and biases.	**Data Driven Discovery:** Can identify patterns in massive datasets that are impossible for humans to see.
Weaknesses:	**Weaknesses:**
Slower than SI-only approaches.	**Risk of Inaccuracy:** Prone to "hallucinations," factual errors, and fabricated information without human validation.
Requires new skills from researchers to effectively interact with SI.	**Bias Perpetuation:** Can amplify biases present in the training data.
Potential for over-reliance, stifling independent thought.	**"Black Box" Problem:** Lack of interpretability makes it difficult to trust or verify the reasoning process.
Best For: Complex, high-stakes research requiring nuance, creativity, ethical considerations, and validation (e.g., clinical diagnosis, novel hypothesis generation).	**Best For:** Well-defined, large-scale, and repetitive tasks where speed is critical (e.g., initial literature screening, high-throughput data analysis).

6.2. Evaluation of SI's Impact on Research Quality and Integrity

Synthetic Intelligence (SI) implementation into research must take into serious account numerous paradigms of activities. SI autonomous systems only provide unmatched speed and scale. Nevertheless, they are usually restricted by the fact that they use the existing data, which may result in an absence of actual novelty and the creation of the so-called hallucinations (Ji et al., 2023). Some models are also black-box, which is challenging to comprehend or audit the reasoning process of the model (Rudin, 2019). Conversely, human SI collaborative strategies build on the rewarding capabilities of human and machine intelligence. SI is a very potent supplement in this paradigm, enhancing human abilities. Humans can bring domain knowledge, critical thinking, and moral judgment, and SI will bring computer computations, scalability, and pattern recognition of enormous amounts of data (Shneiderman, 2022). Such a synergistic system allows more effective, consistent and creative research results. The results of tools such as Elicit (literature synthesis) or AlphaFold (protein structure prediction) are frequently reliant on a large amount of human participation, such as immediate engineering, validation of results, and refinement (Jumper et al., 2021; OECD, 2023). Despite the fact that human SI collaboration is generally viewed as such which promise to offer high-stakes scientific discovery, it has some issues, including the fact that the researchers are forced to learn new skills in digital literacy and the fact that they can be too reliant on the SI suggestions and grown overly dependent on it and fail to think scientifically exclusively.

7. Conclusion

7.1. Summary

This critical literature review has emphasized the radical effect of Synthetic Intelligence (SI) on scientific studies since 2015. SI tools are becoming proficient at simplifying the literature review process (OECD, 2023), novel hypotheses (Zhang et al., 2025a; Richardson et al., 2020), high-fidelity physical simulations (Hennigh et al., 2021), and insights on large datasets (Smith et al., 2020). The advances are led by core approaches such as deep learning, natural language processing (NLP), and generative SI which are across the fields. Nevertheless, there are serious problems with the introduction of SI. Among the most significant ones are the need to guarantee the correctness of SI-generated content especially the risk of hallucinations (Ji et al., 2023), the need to protect research integrity against plagiarism and attribution failures (Thorp, 2023b), the need to address the issue of authorship (COPE Council, 2023), the need to address the problem of technical constraints including the so-called black box (Rudin, 2019), and the need to face ethical and legal challenges that surround the issue of algorithmic bias

7.2. Implications for the Scientific Community

The fast-growing innovations on Synthetic Intelligence (SI) driven research have significant implications on the scientific community. To begin with, scientists of all fields will have to obtain a basic knowledge of SI methods such as their potential, shortcomings, and assumptions to avoid using these tools critically and incorporate them into their practice. Secondly, the roles of the research process as a whole can also change, with SI replacing more routine, repetitive, or computationally intensive parts of it, and human researchers dedicating priority to higher-order cognitive processes, including critical thinking, creative hypothesis generation, and conceptual synthesis (Shneiderman, 2022). Third, the research integrity standards should undergo revision in order to reflect the special problems that SI presents. This involves the creation of clear principles of transparent reporting of SI use (e.g., tools, prompts, extent of involvement), the creation of validation protocols of SI generated output, and the creation of accountability measures in case of error or bias.

Finally, we are starting to require the interdisciplinary cooperation of domain scientists, SI/AI researchers, ethicists, legal scholars and social scientists in the co-development of tools that are not only technically healthy but also morally responsible, socially just and scientifically sound.

7.3. Future Directions and Recommendations

As the Synthetic Intelligence (SI) continues to develop with a high rate of evolution, there are some crucial directions it is likely to take in the future. The further impact in developing more advanced, robust, and interpretable SI with the assistance of the explainable SI (XAI) that is designed to prove decisions made by models is one of the primary directions (Adabi and Berrada, 2018). Second, it is significant to establish effective ethical principles, technical demands, and elastic regulations. This involves further multidisciplinary discussion between scientists, policy makers and ethical theorists as well as ordinary individuals on these historical problems in information privacy, prejudice within the algorithmic, scientific property, and validity of research in SI augmented workflows. Third, scientific employees need to be encouraged to be SI literate. This is in the form of provision of introductory training in capabilities, limitation, validation and ethics in graduate and postdoctoral level and further provides materials on advance training to already established researchers. Such a mix of the two (premediated of such futurities) and a reworking of the conception of such technologies being Synthetic rather than Artificial in an intelligent foresight of such technologies in human creative powers ability, will allow SI to be an effective, straightforward, and responsible agency, to unfold accelerated discovery projects and solve global dilemmas of the current context.

References

1. Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access, 6, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052
2. Alipanahi, B., Delong, A., Weirauch, M. T., & Frey, B. J. (2015). Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. Nature Biotechnology, 33(8), 831–838. https://doi.org/10.1038/nbt.3300
3. Ami, R. (2024). AI for genome sequencing and personalized medicine. International Journal of Artificial Intelligence and Machine Learning in Engineering, 21(2). https://doi.org/10.31582/ijamle.v21.i2.2024.0001
4. Ammar, W., Groeneveld, D., Bhagavatula, C., Beltagy, I., Crawford, M., Downey, D., Dunkelberger, J., Elgohary, A., Feldman, S., Ha, V., Kinney, R., Kohlmeier, S., Lo, K., Murray, T., Ooi, H.-H., Peters, M., Power, J., Skjonsberg, S., Wang, L. L., Wilhelm, C., Yuan, Z., van Zuylen, M., & Etzioni, O. (2018). Construction of the literature graph in Semantic Scholar. In S. Bangalore, J. Chu-Carroll, & Y. Li (Eds.), Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 3 (Industry Papers) (pp. 84–91). Association for Computational Linguistics. https://doi.org/10.18653/v1/N18-3011
5. Baron, D. (2019). Machine learning in astronomy: A practical overview. arXiv. https://arxiv.org/abs/1904.07248
6. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big?. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610–623). https://doi.org/10.1145/3442188.3445922
7. BenevolentAI. (2020). Press Release. BenevolentAI's platform-derived hypothesis for COVID-19 treatment validated in US NIAID randomised control trial https://www.benevolent.ai
8. Cantini, F., Niccoli, L., Matarrese, D., Nicastri, E., Stobbione, P., & Goletti, D. (2020). Baricitinib therapy in COVID-19: A pilot study on safety and clinical impact. Journal of Infection, 81(2), 318–356. https://doi.org/10.1016/j.jinf.2020.05.030
9. COPE Council. (2023). COPE Council. COPE position - Authorship and AI - English. https://doi.org/10.24318/cCVRZBms
10. Deep Genomics. (2015). AI platform: BigRNA The world's first RNA foundation model for RNA therapeutics. https://www.deepgenomics.com/AI-Platform/
11. Elicit. (2023). AI research assistant for scientific papers. https://elicit.com
12. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115–118. https://doi.org/10.1038/nature21056
13. Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., Ronneberger, O., Bodenstein, S., Zielinski, M., Bridgland, A., Potapenko, A., Cowie, A., Tunyasuvunakool, K., Jain, R., Clancy, E., Kohli, P., Jumper, J., & Hassabis, D. (2021). Protein complex prediction with AlphaFold-Multimer. bioRxiv. https://doi.org/10.1101/2021.10.04.463034
14. Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., Ronneberger, O., Bodenstein, S., Zielinski, M., Bridgland, A., Potapenko, A., Cowie, A., Tunyasuvunakool, K., Jain, R., Clancy, E., … Hassabis, D. (2021). Protein complex prediction with AlphaFold-Multimer. bioRxiv. https://doi.org/10.1101/2021.10.04.463034
15. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (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
16. Fortier-Dubois, É. (2025, March). How we evaluated Elicit's systematic review. Elicit Blog. https://blog.elicit.com/how-we-evaluated-elicit-systematic-review/
17. Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., & Pearson, A. T. (2023). Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers. NPJ Digital Medicine, 6(1), Article 54. https://doi.org/10.1038/s41746-023-00808-5
18. Gould, J. (2021). NVIDIA announces PhysicsNeMo: A framework for developing physics ML models for digital twins. NVIDIA Developer Blog. https://developer.nvidia.com/blog/nvidia-announces-physicsnemo-a-framework-for-developing-physics-ml-models-for-digital-twins/
19. Hassabis, D. (2024, December 8). Accelerating scientific discovery with AI [Nobel Lecture]. NobelPrize.org. https://www.nobelprize.org/prizes/chemistry/2024/hassabis/lecture/
20. Hennigh, O., Narasimhan, S., Nabian, M. A., Subramaniam, A., Tangsali, K., Fang, Z., Rietmann, M., Byeon, W., & Choudhry, S. (2021). NVIDIA SimNet™: An AI-accelerated multi-physics simulation framework. International Conference on Computational Science, 447–461. https://doi.org/10.1007/978-3-030-77977-1_36
21. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

22. ICMJE. (2023). Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals: AI and authorship clarification. International Committee of Medical Journal Editors. https://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html
23. Imani, S., Li, X., Chen, K., Maghsoudloo, M., Jabbarzadeh Kaboli, P., Hashemi, M., Khoushab, S., & Li, X. (2025). Computational biology and artificial intelligence in mRNA vaccine design for cancer immunotherapy. Frontiers in cellular and infection microbiology, 14, 1501010. https://doi.org/10.3389/fcimb.2024.1501010
24. IntuitionLabs.ai. (2025). Generative AI in mRNA vaccine development: Moderna and Pfizer's COVID-19 case study. https://intuitionlabs.ai/articles/generative-ai-mrna-vaccine-covid19-case-study
25. Jiang, Z., Liu, Y., Liu, G., Chen, M., Hu, Z., & Ji, C. (2025). Drug-drug interaction event prediction based on multi-relational enhanced graph neural network. In 2025 International Conference on Sensor-Cloud and Edge Computing System (SCECS) (pp. 211–216). Zhuhai, China. https://doi.org/10.1109/SCECS65243.2025.11065338
26. Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., … & Fung, P. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys, 55(12), Article 248. https://doi.org/10.1145/3571730
27. Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255–260. https://doi.org/10.1126/science.aaa8415
28. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2
29. Jurafsky, D., & Martin, J. H. (2023). Speech and language processing (3rd ed., draft). https://web.stanford.edu/~jurafsky/slp3/
30. Kern, C., Fischer-Abaigar, U., Schweisthal, J. et al. Algorithms for reliable decision-making need causal reasoning. Nat Comput Sci 5, 356–360 (2025). https://doi.org/10.1038/s43588-025-00814-9
31. Kurth, T., Subramanian, S., Harrington, P., Pathak, J., Mardani, M., Hall, D., Miele, A., Kashinath, K., & Anandkumar, A. (2023). FourCastNet: Accelerating global high-resolution weather forecasting using adaptive Fourier neural operators. In Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (Article 3593412). Association for Computing Machinery. https://doi.org/10.1145/3592979.3593412
32. Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Pritzel, A., … & Battaglia, P. (2023). GraphCast: Learning skillful medium-range global weather forecasting. Science, 382(6676), 1060–1065. https://doi.org/10.1126/science.adi2336
33. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539
34. Melucci, M. (2025). Three methods for fair ranking of multiple protected items. Scientific Reports, 15, 32382. https://doi.org/10.1038/s41598-025-98765-4
35. NASA Science. (2025, May). Artificial intelligence & data discovery: How NASA's Science Discovery Engine is transforming research. NASA Science Data. https://science.data.nasa.gov/learn/blog/artificial-intelligence-data-discovery
36. National Academies of Sciences, Engineering, and Medicine. (2024). AI for scientific discovery: Proceedings of a workshop. The National Academies Press. https://doi.org/10.17226/27457
37. Nicholson, J. M., Mordaunt, M., Lopez, P., Uppala, A., Rosati, D., Rodrigues, N. P., Grabitz, P., & Rife, S. C. (2021). scite: A smart citation index that displays the context of citations and classifies their intent using deep learning. Quantitative Science Studies, 2(3), 882–898. https://doi.org/10.1162/qss_a_00146
38. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453. https://doi.org/10.1126/science.aax2342
39. OECD (2023), Artificial Intelligence in Science: Challenges, Opportunities and the Future of Research, OECD Publishing, Paris, https://doi.org/10.1787/a8d820bd-en.
40. Okujagu, D. C., & Bose, E. (2025). Synthetic Intelligence: Reframing AI as Human-Created Cognitive Systems - A Paradigm Shift for the Digital Age. International Journal of Progressive Sciences and Technologies, 52(1), 236-247. http://dx.doi.org/10.52155/ijpsat.v52.1.7421
41. OpenAI. (2024). Introducing GPT-4o and more tools to ChatGPT free users. https://openai.com/index/gpt-4o-and-more-tools-to-chatgpt-free/
42. Osayande, N., Marotta, J., Aggarwal, S., Kopal, J., Holmes, A., Yip, S. W., & Bzdok, D. (2024). Diversity-aware Population Models: Quantifying Associations between Socio-Spatial Factors and Cognitive Development in the ABCD Cohort. Research Square, rs.3.rs-4751673. https://doi.org/10.21203/rs.3.rs-4751673/v1
43. O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown Publishing Group.
44. Poplin, R., Chang, P. C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., … & DePristo, M. A. (2018). A universal SNP and small-indel variant caller using deep neural networks. Nature Biotechnology, 36(10), 983–987. https://doi.org/10.1038/nbt.4235
45. Richardson, P., Griffin, I., Tucker, C., Smith, D., Oechsle, O., Phelan, A., & Stebbing, J. (2020). Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. The Lancet, 395(10223), e30–e31. https://doi.org/10.1016/S0140-6736(20)30304-4

46. Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., … & Bengio, Y. (2019). Tackling climate change with machine learning. ACM Computing Surveys, 55(2), Article 31. https://doi.org/10.1145/3485128
47. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x
48. Saha, A., Chauhan Baghel, S., & Singh, I. (2025). Harnessing the power of artificial intelligence in pharmaceuticals: Current trends and future prospects. Intelligent Pharmacy, 3(3), 181–192. https://doi.org/10.1016/j.ipha.2024.12.001
49. Sanchez-Lengeling, B., & Aspuru-Guzik, A. (2018). Inverse molecular design using machine learning: Generative models for matter engineering. Science, 361(6400), 360–365. https://doi.org/10.1126/science.aat2663
50. Sapio Sciences. (2024). Sapio Sciences advances the world's first AI-powered lab assistant. https://www.sapiosciences.com/news/sapio-sciences-advances-the-worlds-first-ai-powered-lab-assistant/
51. Sapio Sciences. (2024, September 3). Sapio Sciences advances the world's first AI-powered lab assistant. https://www.sapiosciences.com/news/sapio-sciences-advances-the-worlds-first-ai-powered-lab-assistant/
52. Scite. (2023). Scite Assistant: https://scite.ai/assistant
53. Sharma, A., Rao, P., Ahmed, M. . Z., & Chaturvedi, K. (2024). Artificial intelligence in scientific writing: opportunities and ethical considerations. International Journal of Research in Medical Sciences, 13(1), 532–542. https://doi.org/10.18203/2320-6012.ijrms20244167
54. Shi, M. (2025). The transformative power of artificial intelligence: Opportunities and challenges. Transactions on Computational and Scientific Methods, 5(1), 1–10. https://doi.org/10.34819/tcsm.v5i1.169
55. Shneiderman, B. (2022). Human-centered artificial intelligence: Reliable, safe & trustworthy. International Journal of Human–Computer Interaction, 38(14), 1219–1230. https://doi.org/10.1080/10447318.2022.2055334
56. Smith, J., Johnson, K. and Roberts, L. (2020) AI in Medical Data Interpretation: Case Studies and Emerging Trends. Journal of Medical Informatics, 35, 123-135.
57. Stokel-Walker, C. (2023, January 26). ChatGPT listed as author on research papers: Many scientists disapprove. Nature. https://doi.org/10.1038/d41586-023-00107-z
58. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 3645–3650. https://doi.org/10.18653/v1/P19-1355
59. Tabassum, L., & Rivera, L. (2025). How AI is reshaping scientific discovery and innovation (pp. 186–193). Honeywell Publishers.
60. Thaler v. Perlmutter, No. 1:21-cv-01998, 2023 U.S. Dist. LEXIS 145391 (D.D.C. Aug. 18, 2023). https://caselaw.findlaw.com/court/us-dis-crt-dis-col/114916944.html
61. Thorp, H. H. (2023). ChatGPT is fun, but not an author. Science, 379(6630), 313. https://doi.org/10.1126/science.adg7879
62. Thorp, H. H. (2023). Generative approach to research integrity. Science, 381(6657), 587. https://doi.org/10.1126/science.adk1852
63. Tian, J., & Che, C. (2024). Automated machine learning: A survey of tools and techniques. Journal of Industrial Engineering and Applied Science, 2, 71–76. https://doi.org/10.70393/6a69656173.323336
64. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7
65. van de Schoot, R., de Bruin, J., Schram, R., Zahedi, P., de Boer, J., Weijdema, F., … & Kramer, B. (2021). An open source machine learning framework for efficient and transparent systematic reviews. Nature Machine Intelligence, 3(2), 125–133. https://doi.org/10.1038/s42256-020-00287-7
66. Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. International Data Privacy Law, 7(2), 76–99. https://doi.org/10.1093/idpl/ipx005
67. Yue, T., Wang, Y., Zhang, L., Gu, C., Xue, H., Wang, W., Lyu, Q., & Dun, Y. (2023). Deep Learning for Genomics: From Early Neural Nets to Modern Large Language Models. International journal of molecular sciences, 24(21), 15858. https://doi.org/10.3390/ijms242115858
68. Zhang, H., Li, R., Zhang, Y., Xiao, T., Chen, J., Ding, J., & Chen, H. (2025). The evolving role of large language models in scientific innovation: Evaluator, collaborator, and scientist. arXiv. https://arxiv.org/abs/2507.11810
69. Zhang, K., Yang, X., Wang, Y., Yu, Y., Huang, N., Li, G., Li, X., Wu, J. C., & Yang, S. (2025). Artificial intelligence in drug development. Nature Medicine, 31(1), 45–59. https://doi.org/10.1038/s41591-024-03434-4

SK International Journal of Multidisciplinary Research Hub
Volume 12, Issue 9, September 2025 pg. 50-66

:::. Cite this article .:::

Bose, E. E. & Okujagu, D. C. (2025). Recent Advances and Challenges in Synthetic Intelligence (SI) Powered Scientific Research. SK INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH HUB, 12(9), 50–66. https://doi.org/10.61165/sk.publisher.v12i9.5

