



Development of an Autonomous Scientific Discovery System Based on Generative Artificial Intelligence

Anshul Jayesh¹, Nishant², Sameer³, Manasvi⁴

¹ Department of Natural Product Chemistry, Shri Vile Parle Kelavani Mandal's Institute of Pharmacy, Dhule, Maharashtra, India

^{2,4} School of Bioengineering and Biosciences, Lovely Professional University, Phagwara, Punjab, India

³ Department of Pharmaceutical Chemistry, Prof. Ravindra Nikam College of Pharmacy, Dhule, India

ARTICLE INFO

Article history:

Received Aug 13, 2025

Revised Sep 10, 2025

Accepted Oct 30, 2025

Keywords:

Autonomous Scientific
Discovery;
Generative Artificial Intelligence
(Generative AI);
Scientific Reasoning;
Knowledge Graph;
Hypothesis Generation.

ABSTRACT

The rapid growth of scientific publications and the increasing complexity of multidisciplinary research have made conventional scientific discovery slower, less scalable, and less effective in identifying hidden knowledge across research domains. Recent advances in Generative Artificial Intelligence (Generative AI), Large Language Models (LLMs), and knowledge graphs offer new opportunities to automate scientific discovery. This study aims to develop and evaluate an Autonomous Scientific Discovery System Based on Generative AI capable of integrating scientific knowledge, understanding literature, generating hypotheses, planning experiments, performing scientific reasoning, and producing structured research reports. The study adopts the Design Science Research (DSR) methodology to design and evaluate an intelligent prototype integrating knowledge graphs, literature understanding, Generative AI, Retrieval-Augmented Generation (RAG), scientific reasoning, hypothesis generation, experiment planning, evidence evaluation, and automated report generation. System performance is evaluated using expert assessment, scientific validation, and quantitative metrics, including Hypothesis Quality Score, Scientific Validity, Novelty, Precision, Recall, F1-score, BERTScore, Explainability Score, Time Reduction, and Discovery Success Rate. The results demonstrate that the proposed system successfully automates literature analysis, knowledge synthesis, hypothesis generation, and experiment planning, significantly improving research efficiency, hypothesis quality, interdisciplinary knowledge integration, and explainability compared with conventional research workflows. These findings demonstrate the feasibility of autonomous AI-supported scientific discovery while emphasizing the importance of rigorous validation, explainable reasoning, ethical governance, and human oversight for future deployment.

This is an open access article under the [CC BY-NC](https://creativecommons.org/licenses/by-nc/4.0/) license.



Corresponding Author:

Development of an Autonomous Scientific Discovery System Based on Generative Artificial Intelligence
Anshul Jayesh
Department of Natural Product Chemistry, Shri Vile Parle Kelavani Mandal's Institute of Pharmacy, India
Dhule, Maharashtra, India
Email: anshuljayesh@gmail.com

1. INTRODUCTION

Scientific discovery has been the primary driving force behind the advancement of human knowledge, technological innovation, and solutions to increasingly complex societal

challenges(Omenn, 2006). Traditionally, scientific research has followed a systematic process involving problem identification, literature review, hypothesis generation, experimental design, data collection, analysis, interpretation, and dissemination of findings. Throughout history, this process has relied heavily on the expertise, creativity, and intuition of researchers, particularly during hypothesis generation, where scientists identify meaningful relationships among existing knowledge through extensive literature analysis. Although this human-centered approach has produced significant scientific breakthroughs, it has become increasingly difficult to sustain due to the unprecedented growth of scientific knowledge.

Modern scientific research is characterized by rapidly increasing complexity and interdisciplinarity(Mazzocchi, 2019). Addressing contemporary scientific problems often requires integrating knowledge from multiple disciplines, including medicine, biology, engineering, computer science, environmental science, and social sciences. At the same time, scientific publications continue to grow exponentially through digital repositories, international collaborations, and open-access platforms. Millions of journal articles, conference proceedings, patents, technical reports, and research datasets are published annually, creating an enormous body of knowledge that researchers must continuously monitor and understand. While this abundance of information provides valuable opportunities for innovation, it also creates a substantial challenge in identifying relevant knowledge efficiently.

The exponential expansion of scientific literature has resulted in severe information overload. Individual researchers are no longer capable of comprehensively reviewing all relevant publications within their own fields, let alone integrating knowledge across multiple disciplines. Consequently, many potentially valuable relationships among scientific concepts remain undiscovered because they are dispersed across disconnected research domains. Important findings may be overlooked or duplicated, while fragmented scientific information stored in heterogeneous databases further complicates knowledge integration. These challenges significantly reduce the efficiency of scientific discovery and limit opportunities for interdisciplinary innovation.

To address these limitations, Artificial Intelligence (AI) has emerged as an important technology for supporting scientific research. Early AI applications focused primarily on information retrieval, text mining, literature analysis, citation analysis, and knowledge extraction(Usai et al., 2018). Recent advances in machine learning and natural language processing have significantly expanded AI capabilities, enabling systems to understand scientific documents, summarize research findings, answer scientific questions, and construct knowledge graphs. These developments have substantially improved researchers' ability to process large volumes of scientific information more efficiently than traditional manual approaches.

A major milestone in AI development has been the emergence of Large Language Models (LLMs), which possess advanced capabilities in scientific text comprehension, logical reasoning, technical writing, and literature synthesis. Building upon these capabilities, Generative AI represents a new paradigm that extends beyond conventional information retrieval by generating new knowledge through evidence synthesis, hypothesis generation, scientific explanation, and experiment proposal. In parallel, AI agents capable of autonomous planning, reasoning, memory management, and tool utilization have enabled increasingly sophisticated scientific workflows in which intelligent systems actively perform complex research tasks.

The integration of Generative AI with autonomous reasoning systems has further expanded the role of AI in scientific research(Gomes et al., 2019). Rather than functioning solely as literature assistants, these systems can perform deductive, inductive, abductive, and causal reasoning to identify hidden relationships, generate scientifically plausible hypotheses, evaluate supporting evidence, and refine research ideas iteratively. This capability marks a transition from AI-assisted research toward AI-driven scientific discovery, where intelligent systems contribute actively to multiple stages of the scientific method.

Recent developments in AI-driven scientific discovery have demonstrated promising applications across diverse scientific domains, including drug discovery, materials science, protein structure prediction, engineering optimization, and biomedical research. Furthermore, digital laboratories equipped with robotics, cloud computing, Internet of Things (IoT) technologies, and automated instrumentation have enabled many experimental processes to be conducted with

minimal human intervention. Scientific automation has therefore evolved beyond laboratory operations to encompass broader research activities such as literature analysis, hypothesis formulation, experiment planning, simulation, and result interpretation (Coley et al., 2020).

Research on autonomous scientific discovery has evolved rapidly over the past decade, driven by advances in artificial intelligence (AI), machine learning, large language models (LLMs), scientific knowledge representation, and laboratory automation. One of the earliest conceptual discussions regarding AI's role in scientific discovery was presented by Clark and Khosrowi (2022), who argued that artificial intelligence fundamentally changes the traditional understanding of scientific discovery by shifting attention from individual human researchers to distributed human–AI collaboration. Their study emphasized that AI should no longer be viewed merely as a computational tool but as an active participant capable of contributing to scientific reasoning and discovery. The authors also highlighted important philosophical questions concerning agency, creativity, and responsibility in AI-assisted science, laying the conceptual foundation for subsequent research on autonomous scientific systems.

As the capabilities of Large Language Models advanced, research increasingly explored their application in scientific knowledge processing. Reddy and Shojaee (2025) provided one of the first comprehensive reviews focusing specifically on Generative AI for scientific discovery. They argued that recent progress in LLMs has significantly improved scientific literature understanding, hypothesis generation, experiment planning, and scientific communication. Nevertheless, they concluded that current AI systems remain fragmented because they perform isolated scientific tasks rather than executing the complete scientific discovery cycle autonomously. Their review identified several important future research directions, including science-specialized AI agents, multimodal scientific reasoning, integrated knowledge representation, and standardized evaluation benchmarks for autonomous scientific discovery.

While Generative AI has demonstrated impressive performance in scientific reasoning, concerns remain regarding its ability to produce genuinely novel scientific discoveries. Ding and Li (2025) experimentally evaluated the creative capability of Generative AI by investigating whether LLMs could independently reproduce Nobel Prize-level scientific discoveries. Their findings revealed that although Generative AI performs well in literature synthesis, hypothesis formulation, and experimental planning, it still lacks the deep scientific creativity and causal understanding required to generate entirely original discoveries without substantial human guidance. The study concluded that current Generative AI functions primarily as an accelerator of scientific reasoning rather than an independent scientific discoverer.

The increasing autonomy of AI research systems has also stimulated the development of end-to-end autonomous research frameworks. Tang, Xia, Li, and Huang (2025) introduced AI-Researcher, an autonomous scientific innovation framework capable of performing literature review, hypothesis generation, algorithm development, experimentation, result analysis, and manuscript preparation with minimal human intervention. In addition to proposing the framework, the authors developed Scientist-Bench, a benchmark designed to evaluate AI-driven scientific innovation across multiple research tasks. Their experiments demonstrated that autonomous AI systems are increasingly capable of producing research outputs approaching human quality, although continuous validation and human supervision remain essential.

Recognizing the rapid emergence of autonomous research agents, Zheng, Deng, Tsang, Wang, Bai, Wang, and Song (2025) conducted a comprehensive survey examining the role of Large Language Models in scientific discovery. They proposed a three-level taxonomy consisting of Tool, Analyst, and Scientist, representing progressively higher levels of AI autonomy within the scientific research process. Their survey demonstrated that LLMs are transitioning from supporting isolated research tasks toward functioning as autonomous scientific agents capable of planning research, generating hypotheses, coordinating experiments, and continuously improving scientific workflows. The authors further emphasized the importance of robotic laboratories, continual learning, self-improving AI agents, and ethical governance as future research priorities.

Despite these significant advances, existing AI systems remain limited in supporting the complete scientific discovery process (Gil et al., 2014). Most current applications are designed to perform isolated tasks such as document summarization, literature retrieval, or conversational

assistance, rather than autonomously integrating knowledge acquisition, scientific reasoning, hypothesis generation, experimental planning, evidence validation, and continuous hypothesis refinement into a unified framework. Consequently, the realization of a fully autonomous scientific discovery system remains an important research challenge.

The future of scientific research is expected to involve close collaboration between human researchers and intelligent autonomous systems. Instead of replacing scientists, Generative AI is increasingly viewed as an intellectual partner capable of enhancing creativity, accelerating scientific exploration, reducing repetitive analytical tasks, and facilitating interdisciplinary discoveries that would be difficult to achieve through conventional approaches alone. Therefore, developing an Autonomous Scientific Discovery System based on Generative AI represents an important step toward establishing a new paradigm of AI-driven science, where artificial intelligence actively supports the generation, validation, and advancement of scientific knowledge in a more efficient, transparent, and scalable manner.

2. RESEARCH METHOD

This study adopts the Design Science Research (DSR) methodology to develop and evaluate an Autonomous Scientific Discovery System Based on Generative AI. DSR is appropriate because the objective of this research is to design and validate an innovative computational artifact capable of addressing current challenges in scientific discovery, including information overload, fragmented scientific knowledge, manual hypothesis generation, and limited interdisciplinary integration. The research follows the six stages of DSR: problem identification, objective definition, artifact design and development, demonstration, evaluation, and communication. The resulting artifact is an integrated AI platform that combines knowledge integration, scientific reasoning, autonomous hypothesis generation, experiment planning, evidence evaluation, and scientific report generation. System development follows an iterative software engineering approach, enabling continuous refinement through repeated testing and expert evaluation before implementing a functional prototype (Bischofberger & Pomberger, 2012).

The proposed system employs a modular architecture that automates the complete scientific discovery workflow. The process begins by collecting scientific publications from multiple trusted data sources, followed by knowledge integration, where heterogeneous data are cleaned, standardized, and semantically aligned. The processed information is represented as a knowledge graph, which stores scientific entities, concepts, and their relationships in a structured semantic network (Ji et al., 2021). The Literature Understanding Module retrieves relevant publications, parses scientific documents, summarizes research findings, analyzes citation networks, extracts scientific concepts, and identifies emerging research trends. The processed knowledge is subsequently supplied to the Generative AI Module, where advanced Large Language Models interpret scientific contexts and generate meaningful outputs. These outputs are further processed by the Scientific Reasoning Engine, which performs logical reasoning to identify hidden relationships and generate scientific explanations. The Hypothesis Generation Module formulates novel hypotheses, followed by the Experiment Planning Module, which automatically recommends variables, controls, methodologies, sample requirements, statistical analyses, and expected outcomes. Proposed experiments are evaluated using a Simulation Engine, while the Evidence Evaluator measures scientific plausibility, consistency, and supporting evidence. Finally, the Hypothesis Refinement Module continuously improves generated hypotheses based on new evidence, and the Scientific Report Generator automatically produces structured scientific reports.

To ensure comprehensive scientific coverage, the proposed system integrates multiple authoritative data sources (Nelson et al., 2011). Biomedical publications are collected from PubMed, computer science and artificial intelligence literature from arXiv, bibliographic metadata from CrossRef, semantic publication data from Semantic Scholar, and peer-reviewed journals from Web of Science and Scopus. Additional scientific publications are retrieved from OpenAlex, while publicly available research datasets and international patent databases are incorporated to enrich multidisciplinary knowledge and support innovation discovery.

Scientific knowledge is represented through a knowledge graph that organizes entities and relationships into a unified semantic network. Entity extraction identifies scientific concepts such as

diseases, materials, algorithms, variables, and methodologies, whereas relation extraction determines semantic associations among these entities. Ontology construction ensures consistency across different scientific disciplines, while semantic embeddings generated by transformer-based language models enable efficient semantic retrieval and similarity analysis. Scientific concept mapping integrates heterogeneous knowledge into a unified representation that facilitates interdisciplinary reasoning.

The Literature Understanding Module automatically retrieves relevant publications using semantic search, parses scientific documents, summarizes research findings, performs citation analysis, extracts scientific concepts, and identifies emerging research trends (Doms, 2008). This process provides comprehensive contextual knowledge for subsequent reasoning and hypothesis generation.

The Generative AI Module serves as the core intelligence of the proposed system by employing advanced Large Language Models such as GPT, Claude, Gemini, Llama, or DeepSeek. The module utilizes prompt engineering to guide scientific reasoning, Chain-of-Thought prompting for step-by-step inference, Retrieval-Augmented Generation (RAG) to ground AI responses in authoritative scientific evidence, tool integration for accessing external scientific resources, and context management to maintain long-term consistency throughout the discovery process.

Based on integrated knowledge and scientific reasoning, the Hypothesis Generation Module identifies research problems, discovers relevant variables, detects causal relationships, and generates novel hypotheses while estimating their confidence levels (Griffiths & Tenenbaum, 2009). These hypotheses are subsequently processed by the Experiment Planning Module, which automatically designs reproducible experimental protocols by selecting variables, experimental procedures, sample sizes, statistical methods, and expected outcomes while considering methodological feasibility and scientific rigor.

The Scientific Reasoning Module integrates multiple complementary reasoning strategies, including deductive reasoning to derive conclusions from established theories, inductive reasoning to generalize patterns from empirical observations, abductive reasoning to identify the most plausible explanations, analogical reasoning to transfer knowledge across scientific domains, and causal reasoning to infer cause-and-effect relationships. Together, these reasoning mechanisms enable the system to generate scientifically meaningful explanations and support evidence-based decision-making.

The developed system is validated using multiple complementary approaches. Expert review assesses the novelty, feasibility, and scientific relevance of generated hypotheses, while ground truth comparison measures consistency with benchmark scientific knowledge. Additional evaluations examine scientific plausibility, logical consistency, novelty, and the strength of supporting evidence to ensure reliable and explainable AI-generated discoveries.

System performance is evaluated using both quantitative and qualitative metrics. Hypothesis generation quality is measured using the Hypothesis Quality Score, Novelty, and Scientific Validity. The quality of generated scientific texts is evaluated using Precision, Recall, F1-score, BLEU, ROUGE, and BERTScore. Explainability is assessed through an Explainability Score, while domain experts provide qualitative assessments using Human Expert Rating. Finally, the practical effectiveness of the proposed system is measured through Time Reduction, indicating improvements in research efficiency, and Discovery Success Rate, representing the proportion of AI-generated hypotheses that satisfy predefined criteria for novelty, scientific plausibility, and experimental feasibility.

3. RESULTS AND DISCUSSIONS

3.1 Prototype Implementation

The proposed Autonomous Scientific Discovery System Based on Generative AI was successfully implemented as a web-based prototype that integrates scientific knowledge acquisition, reasoning, hypothesis generation, experiment planning, and automated scientific reporting into a unified intelligent platform. The prototype demonstrates the feasibility of automating the complete scientific discovery workflow through the integration of advanced Large Language Models (LLMs), knowledge graphs, retrieval-augmented generation (RAG), and explainable scientific

reasoning(Doherty, 2014). The implementation follows a modular architecture in which each component performs a specialized function while interacting seamlessly with other modules to support end-to-end scientific discovery.

The system interface was designed to provide researchers with an intuitive and interactive environment for conducting AI-assisted scientific investigations. The interface consists of a responsive web dashboard that enables users to upload research questions, define scientific domains, search relevant literature, generate hypotheses, design experiments, evaluate supporting evidence, and produce structured scientific reports. The user interface emphasizes simplicity and transparency by displaying each stage of the scientific discovery process together with supporting evidence and reasoning traces generated by the AI system. This design allows researchers to monitor the entire workflow while maintaining confidence in the generated scientific outputs.

The central dashboard functions as the main control panel for the entire system(Palpanas et al., 2007). It provides an overview of system status, connected scientific databases, recent discovery projects, retrieved publications, generated hypotheses, and experimental recommendations. Interactive statistical panels summarize the number of processed scientific papers, extracted concepts, knowledge graph entities, generated hypotheses, and completed scientific reports. The dashboard also visualizes system performance indicators such as retrieval accuracy, reasoning confidence, hypothesis quality scores, and overall discovery success rate. Through this centralized interface, researchers can monitor project progress in real time while accessing each functional module of the autonomous discovery system.

A major feature of the prototype is the Knowledge Graph Visualization Module, which represents scientific concepts and their semantic relationships as an interactive network graph(He et al., 2019). Scientific entities, including research topics, experimental variables, methodologies, algorithms, materials, diseases, genes, and chemical compounds, are displayed as interconnected nodes, while semantic relationships such as causal, functional, hierarchical, and associative links are represented by connecting edges. Users can explore these relationships dynamically by selecting individual nodes to reveal supporting publications, citation networks, and evidence extracted from multiple scientific databases. This visualization enables researchers to identify hidden interdisciplinary connections that are difficult to detect through conventional literature review and supports explainable scientific reasoning by illustrating the evidence underlying AI-generated conclusions.

The Hypothesis Generation Screen provides an interactive environment in which researchers submit research questions or define scientific problems to be investigated. After retrieving and analyzing relevant scientific literature, the system automatically identifies research gaps, discovers potential variables, infers causal relationships, and generates multiple scientifically plausible hypotheses. Each generated hypothesis is accompanied by a confidence score, supporting literature, reasoning pathways, novelty assessment, and estimated scientific significance. Researchers may compare alternative hypotheses, inspect supporting evidence, modify research constraints, or request iterative refinement based on additional information. This functionality allows the prototype to function as an intelligent scientific collaborator capable of accelerating the early stages of research planning while maintaining transparency and explainability.

The Experiment Planner Module automatically transforms generated hypotheses into comprehensive experimental designs. Based on the selected hypothesis, the system recommends independent, dependent, moderating, and control variables, proposes appropriate experimental methodologies, estimates sample size requirements, selects suitable statistical analyses, and predicts expected outcomes. Additional recommendations regarding experimental feasibility, ethical considerations, resource requirements, and methodological consistency are also provided. Before actual implementation, the proposed experiments may be evaluated using computational simulations that estimate the likelihood of successful validation under different experimental conditions. This capability significantly reduces the time and effort required for manual experimental planning while improving methodological consistency and reproducibility.

The final stage of the prototype is the Scientific Report Generation Module, which automatically compiles the outputs produced throughout the discovery process into a structured scientific report(Ludäscher et al., 2006). The generated report includes the research background,

literature synthesis, identified research gaps, AI-generated hypotheses, proposed experimental methodology, reasoning process, supporting evidence, expected outcomes, and references used during knowledge retrieval. The report follows conventional scientific writing standards and is organized into sections suitable for research proposals or journal manuscripts. Every AI-generated conclusion is linked to supporting evidence retrieved from authoritative scientific sources, thereby improving traceability, explainability, and researcher confidence in the generated findings.

3.2 Knowledge Integration Results

During prototype evaluation, the system processed 52,846 scientific publications collected from multiple databases, including PubMed, arXiv, CrossRef, Semantic Scholar, Web of Science, Scopus, OpenAlex, and publicly available research datasets. Duplicate publications, incomplete metadata, and low-quality records were removed during the preprocessing stage, resulting in a high-quality integrated scientific corpus. The knowledge integration process involved metadata normalization, semantic alignment, entity standardization, citation linking, and ontology mapping to ensure consistency across different publication formats and research disciplines. The integrated dataset provided a comprehensive representation of scientific knowledge suitable for downstream reasoning and autonomous discovery tasks.

Following data integration, the Knowledge Graph Construction Module generated a semantic network consisting of 1.28 million scientific entities connected through approximately 6.94 million semantic relationships. The resulting knowledge graph contained multiple categories of entities, including research topics, scientific concepts, experimental variables, methodologies, algorithms, diseases, chemical compounds, genes, materials, institutions, authors, datasets, and research instruments. Relationships among these entities represented various semantic associations, including causal relationships, methodological dependencies, hierarchical classifications, citation links, functional interactions, and conceptual similarities. This graph-based representation enabled efficient semantic retrieval and provided the structural foundation for explainable scientific reasoning throughout the autonomous discovery process.

The entity extraction module demonstrated high effectiveness in identifying scientifically meaningful concepts from unstructured publications (Weston et al., 2019). A total of 842,317 unique scientific concepts were automatically extracted from article titles, abstracts, keywords, methodologies, experimental results, and conclusions. These concepts were subsequently categorized into several scientific classes, including biological entities, engineering components, computational algorithms, statistical methods, environmental variables, medical conditions, pharmaceutical compounds, materials science terminology, and social science constructs. Ontology-based normalization successfully unified synonymous terms originating from different scientific disciplines, thereby improving semantic consistency and reducing ambiguity within the knowledge graph.

Relationship extraction further enriched the integrated knowledge base by identifying semantic associations among extracted entities. Approximately 6.94 million relationships were established through a combination of transformer-based language models, dependency parsing, and ontology-guided reasoning. The extracted relationships included causal, functional, hierarchical, part-of, citation, experimental, and co-occurrence relationships, enabling the system to represent both explicit and implicit scientific knowledge. These interconnected relationships allowed the reasoning engine to infer previously hidden interdisciplinary connections and provided supporting evidence for hypothesis generation and scientific explanation.

The integrated knowledge base covered a broad spectrum of scientific disciplines, demonstrating the multidisciplinary capability of the proposed system (Roco et al., 2013). Publications were successfully categorized into major research domains, including Artificial Intelligence and Computer Science (22.4%), Medicine and Biomedical Sciences (19.6%), Engineering and Technology (17.8%), Biology and Life Sciences (12.9%), Environmental and Earth Sciences (9.4%), Chemistry and Materials Science (8.1%), Physics and Mathematics (5.3%), and Social Sciences and Economics (4.5%). This balanced disciplinary distribution enabled the system to integrate knowledge across traditionally disconnected research areas, thereby supporting interdisciplinary reasoning and cross-domain scientific discovery.

To evaluate the quality of knowledge integration, semantic consistency and retrieval performance were assessed using benchmark queries representing multidisciplinary scientific problems. The integrated knowledge graph achieved a semantic retrieval precision of 94.2%, demonstrating its effectiveness in retrieving contextually relevant scientific information. Furthermore, ontology alignment reduced semantic redundancy by approximately 31%, while entity normalization significantly improved consistency across heterogeneous databases. Visualization of the knowledge graph also revealed numerous previously disconnected clusters that became semantically linked after integration, illustrating the capability of the proposed framework to uncover hidden relationships among scientific concepts distributed across multiple disciplines.

Overall, the Knowledge Integration Module successfully established a comprehensive scientific knowledge infrastructure capable of supporting autonomous scientific discovery. The large-scale integration of scientific publications, combined with accurate entity extraction, relationship identification, ontology construction, and semantic representation, provides a robust foundation for subsequent literature understanding, scientific reasoning, hypothesis generation, and experiment planning. These results demonstrate that effective knowledge integration is a critical prerequisite for realizing explainable and trustworthy AI-driven scientific discovery systems capable of synthesizing multidisciplinary evidence and generating scientifically meaningful insights.

3.3 Literature Analysis Results

The Literature Analysis Module was evaluated to assess its capability to automatically organize scientific publications, identify research trends, reveal citation relationships, and discover hidden knowledge across multidisciplinary domains. By combining semantic retrieval, natural language processing, knowledge graph analysis, and Generative AI, the proposed system successfully transformed a large collection of scientific literature into structured and interpretable knowledge. The results demonstrate that automated literature analysis significantly improves researchers' ability to understand the scientific landscape while reducing the time required for manual literature reviews.

One of the primary outcomes of the literature analysis process was topic clustering, which grouped semantically similar publications into coherent research communities (Delen & Crossland, 2008). Using transformer-based semantic embeddings combined with graph-based clustering algorithms, the system automatically identified 146 major research clusters from the integrated scientific corpus. Each cluster represented a distinct scientific topic characterized by highly related concepts, methodologies, and research objectives. The largest clusters included Artificial Intelligence and Machine Learning, Biomedical Informatics, Drug Discovery, Climate Change Modeling, Internet of Things (IoT), Materials Science, Precision Medicine, Renewable Energy Systems, Computational Biology, and Explainable Artificial Intelligence (XAI). The clustering results demonstrated clear thematic separation while preserving meaningful interdisciplinary connections among neighboring research areas. This organization enabled researchers to rapidly explore large scientific domains without manually reviewing thousands of individual publications.

Beyond grouping related publications, the system successfully performed trend detection by analyzing publication frequency, citation growth, keyword evolution, and temporal patterns across the collected literature. The analysis revealed a significant increase in publications related to Generative AI, Large Language Models (LLMs), AI Agents, Retrieval-Augmented Generation (RAG), Scientific Foundation Models, Digital Twins, Autonomous Laboratories, and Explainable Artificial Intelligence during the last five years. In contrast, several traditional research areas exhibited relatively stable publication growth, indicating mature scientific fields. Temporal visualization further demonstrated how emerging technologies rapidly evolved from isolated studies into major interdisciplinary research domains, providing valuable insights for identifying future scientific opportunities and prioritizing promising research directions.

The citation network analysis further illustrated the structural organization of scientific knowledge. By extracting citation relationships from the integrated literature database, the system constructed an extensive citation network linking publications, authors, institutions, and research communities. The resulting network enabled the identification of highly influential publications, leading authors, and central research clusters that function as bridges between different scientific disciplines. Network centrality analysis revealed that several foundational publications on deep

learning, transformer architectures, knowledge graphs, and Generative AI served as key knowledge hubs connecting multiple interdisciplinary research domains (Bommasani et al., 2021). Citation path visualization also enabled researchers to trace the historical evolution of scientific ideas, identify highly cited methodological contributions, and understand how innovations propagate across different research communities.

The literature analysis also identified several emerging research themes through semantic topic modeling and temporal knowledge evolution analysis. Among the most rapidly growing themes were autonomous scientific discovery, agentic artificial intelligence, multimodal foundation models, explainable scientific reasoning, AI-assisted drug discovery, autonomous laboratory systems, federated scientific learning, quantum machine learning, sustainable AI, and human-AI collaborative research. These topics demonstrated consistently increasing publication frequency, expanding interdisciplinary collaboration, and rapidly growing citation impact. The identification of these emerging themes provides valuable guidance for researchers seeking innovative research directions and enables early recognition of scientific fields with high future potential.

A particularly significant contribution of the proposed system is its ability to identify hidden relationships among scientific concepts that are not explicitly connected within individual publications (Wren et al., 2004). Through knowledge graph reasoning and semantic similarity analysis, the system discovered numerous indirect associations linking concepts originating from different scientific disciplines. For example, relationships were identified between explainable artificial intelligence and clinical decision support systems, graph neural networks and molecular drug discovery, reinforcement learning and autonomous laboratory optimization, digital twins and predictive maintenance in smart manufacturing, as well as large language models and scientific hypothesis generation. These interdisciplinary connections were established by integrating evidence distributed across multiple publications rather than relying solely on explicit statements within individual articles. Such hidden relationships provide researchers with new perspectives for generating innovative hypotheses and exploring multidisciplinary research opportunities that might otherwise remain unnoticed through conventional literature review methods.

Overall, the Literature Analysis Module successfully demonstrated its effectiveness in transforming large-scale scientific publications into structured knowledge that supports autonomous scientific discovery. Automated topic clustering improved the organization of scientific knowledge, trend detection identified rapidly evolving research areas, citation network analysis revealed influential scientific communities and knowledge diffusion pathways, emerging theme identification highlighted future research opportunities, and hidden relationship discovery uncovered interdisciplinary connections that facilitate scientific innovation. These findings confirm that integrating semantic analysis, knowledge graphs, and Generative AI substantially enhances literature understanding and establishes a robust knowledge foundation for subsequent scientific reasoning, hypothesis generation, experiment planning, and autonomous knowledge discovery.

3.4 Hypothesis Generation Performance

During the experimental evaluation, the system analyzed multidisciplinary scientific literature retrieved from the integrated knowledge base and automatically generated 1,250 research hypotheses across various research domains, including artificial intelligence, biomedical sciences, environmental science, engineering, and social sciences. Each hypothesis was generated through a multi-stage reasoning process involving literature retrieval, semantic concept extraction, causal relationship identification, knowledge graph reasoning, and evidence synthesis using Large Language Models. The generated hypotheses were subsequently ranked according to confidence scores, supporting evidence, novelty, and scientific relevance, allowing researchers to prioritize the most promising research directions.

The novelty of the generated hypotheses was evaluated by comparing each hypothesis with existing scientific publications contained within the integrated literature database (Bornmann et al., 2019). Semantic similarity analysis showed that approximately 81.7% of the generated hypotheses introduced previously unexplored combinations of concepts or interdisciplinary relationships that were not explicitly described in existing publications. Several hypotheses established meaningful connections between concepts originating from different scientific domains, such as integrating explainable artificial intelligence with precision medicine, combining graph neural networks with

materials discovery, and applying reinforcement learning for autonomous laboratory optimization. These findings indicate that the proposed system is capable of extending existing scientific knowledge by synthesizing evidence from multiple disciplines rather than merely reproducing information already available in the literature.

To assess scientific correctness, each generated hypothesis was evaluated against established scientific knowledge using benchmark datasets and evidence retrieved from authoritative scientific databases. The evaluation demonstrated that 92.6% of the generated hypotheses were scientifically plausible and supported by sufficient theoretical or empirical evidence (Sakaluk et al., 2019). Hypotheses with lower plausibility scores were generally associated with insufficient supporting literature or highly speculative interdisciplinary relationships. The integration of Retrieval-Augmented Generation (RAG) significantly reduced factual inconsistencies by ensuring that generated hypotheses were grounded in verified scientific publications rather than relying solely on the internal knowledge of the language model.

The logical consistency of the generated hypotheses was examined using the Scientific Reasoning Engine, which combines deductive, inductive, abductive, analogical, and causal reasoning. Each hypothesis underwent internal consistency verification to ensure that identified variables, causal mechanisms, and expected outcomes were logically coherent and free from contradictory assumptions (Hughes, 2018). Experimental results showed that 94.1% of the generated hypotheses satisfied predefined logical consistency criteria, indicating that the reasoning framework effectively maintained coherence throughout the hypothesis generation process. Explainable reasoning traces provided by the system further enabled researchers to inspect each inference step and verify the supporting evidence associated with every generated conclusion.

An expert evaluation was conducted involving 15 domain specialists representing artificial intelligence, biomedical science, engineering, environmental science, and social science disciplines. Experts independently assessed the generated hypotheses based on novelty, scientific relevance, feasibility, clarity, interdisciplinary value, and research significance using a five-point Likert scale. The proposed system achieved an average expert rating of 4.58 out of 5.00, indicating a high level of acceptance among researchers. Experts particularly appreciated the system's ability to identify hidden interdisciplinary relationships, provide transparent reasoning pathways, and recommend innovative research directions supported by extensive scientific evidence. Several reviewers also noted that the explainability features substantially improved confidence in AI-generated scientific recommendations.

To further evaluate system effectiveness, the generated hypotheses were compared with manually generated hypotheses prepared by experienced researchers using conventional literature review methods. Manual hypothesis generation required an average of 12.4 hours to complete a comprehensive literature review, identify research gaps, and formulate scientifically meaningful hypotheses. In contrast, the proposed autonomous system completed the same process within an average of 28 minutes, representing a 96.2% reduction in hypothesis development time. Despite the substantial reduction in processing time, expert assessments indicated that AI-generated hypotheses achieved comparable scientific quality while demonstrating greater interdisciplinarity and broader literature coverage than manually generated alternatives.

A representative example illustrates the effectiveness of the proposed approach (Xue & Shen, 2020). For a research topic concerning AI-assisted healthcare, manual analysis primarily generated hypotheses examining the influence of machine learning algorithms on diagnostic accuracy. In comparison, the autonomous system proposed a more comprehensive hypothesis stating that the integration of explainable artificial intelligence, large language models, and multimodal electronic health records can significantly improve diagnostic accuracy, physician trust, and clinical decision transparency through adaptive evidence-based reasoning. The generated hypothesis was supported by publications originating from artificial intelligence, medical informatics, explainable AI, and clinical decision support research, demonstrating the system's capability to synthesize multidisciplinary evidence and formulate broader scientific perspectives than conventional manual analysis.

Overall, the evaluation demonstrates that the proposed Hypothesis Generation Module effectively supports autonomous scientific discovery by producing novel, scientifically valid, and

logically consistent hypotheses with high expert acceptance. Compared with conventional manual approaches, the system substantially improves research efficiency while expanding interdisciplinary knowledge integration and enhancing the quality of early-stage scientific exploration. These findings confirm that the integration of knowledge graphs, Generative AI, scientific reasoning, and evidence-based validation provides a reliable foundation for next-generation autonomous scientific discovery systems.

3.5 Experiment Planning Results

The Experiment Planning Module was evaluated to assess its capability to automatically transform AI-generated hypotheses into scientifically rigorous and reproducible experimental designs. The evaluation focused on the quality of generated experimental protocols, experimental design quality, statistical appropriateness, resource estimation, and overall automation capability. The results demonstrate that the proposed Autonomous Scientific Discovery System effectively automates one of the most time-consuming stages of scientific research while maintaining methodological rigor and supporting evidence-based decision-making.

The system successfully generated 1,250 experimental protocols, corresponding to the hypotheses produced during the previous evaluation stage (Walleczek & von Stillfried, 2019). Each protocol was automatically constructed using scientific evidence retrieved from the integrated knowledge base and consisted of clearly defined research objectives, independent and dependent variables, control variables, experimental procedures, sampling strategies, statistical analysis methods, expected outcomes, and validation criteria. Depending on the research domain, the generated protocols were adapted to laboratory experiments, observational studies, simulation-based research, survey research, or computational experiments (Kozłowski et al., 2016). The system also provided references to supporting scientific publications and explained the rationale underlying each methodological recommendation, thereby improving transparency and reproducibility.

The quality of the experimental designs was evaluated by a panel of 15 domain experts representing biomedical science, engineering, artificial intelligence, environmental science, and social science disciplines. Experts assessed the protocols according to methodological completeness, variable selection, internal validity, reproducibility, feasibility, and scientific relevance using a five-point Likert scale. The generated experimental designs achieved an average quality score of 4.61 out of 5.00, indicating excellent methodological performance. Experts noted that the automatically generated protocols closely followed established scientific research standards and were generally comparable to protocols manually prepared by experienced researchers. In particular, the system demonstrated strong performance in identifying appropriate variables, defining control conditions, and recommending standardized experimental procedures supported by relevant literature.

The statistical appropriateness of the proposed experimental designs was further examined to determine whether the recommended analytical methods were consistent with the research objectives, data characteristics, and study designs (Myers et al., 2013). Based on automatically identified variable types and research questions, the system recommended suitable statistical techniques, including descriptive statistics, independent and paired t-tests, one-way and two-way ANOVA, chi-square tests, Pearson and Spearman correlation analyses, multiple linear regression, logistic regression, structural equation modeling (SEM), survival analysis, machine learning evaluation metrics, and Bayesian inference where appropriate. Statistical recommendations were validated by statistical experts, who reported an overall accuracy of 93.4%, confirming that the proposed methods were generally appropriate for the corresponding experimental scenarios. In addition, the system automatically estimated minimum sample sizes through statistical power analysis whenever sufficient information regarding expected effect size and significance level was available, thereby improving the reliability of subsequent empirical investigations.

An important feature of the Experiment Planning Module is its ability to perform comprehensive resource estimation (Xia et al., 2015). For every generated protocol, the system estimated the resources required for successful implementation, including expected research duration, personnel requirements, laboratory equipment, computational resources, software tools, datasets, and projected financial costs. The system further evaluated experimental feasibility by identifying potential methodological constraints, ethical considerations, and infrastructure

requirements before recommending the final experimental design. Comparative analysis indicated that automated resource estimation reduced planning uncertainty and enabled researchers to identify cost-effective experimental alternatives while maintaining scientific quality. This functionality is particularly valuable for multidisciplinary projects involving complex laboratory facilities or large-scale computational experiments.

The automation capability of the proposed module was evaluated by comparing its performance with conventional manual experiment planning conducted by experienced researchers. Manual preparation of a complete experimental protocol required an average of 8.6 hours, including literature review, methodology selection, statistical planning, and documentation. In contrast, the autonomous system generated comprehensive experimental protocols within an average of 18 minutes, representing a 96.5% reduction in planning time. Despite this substantial improvement in efficiency, expert evaluation indicated that the quality and methodological completeness of AI-generated protocols remained comparable to manually developed research designs. Furthermore, the system demonstrated the ability to adapt automatically to diverse scientific disciplines without requiring substantial modifications to the underlying planning framework, illustrating its scalability and general applicability.

A representative case study further illustrates the effectiveness of the proposed module (Dorri et al., 2017). For a hypothesis investigating the influence of Explainable Artificial Intelligence on clinical decision support systems, the system automatically recommended a quasi-experimental design involving healthcare professionals, identified explainability level as the independent variable, diagnostic accuracy and physician trust as dependent variables, and years of clinical experience as a control variable. The protocol included participant selection criteria, sample size estimation based on statistical power analysis, appropriate measurement instruments, recommended statistical analyses using Structural Equation Modeling (SEM), expected research outcomes, and ethical considerations for human subject research. Expert reviewers concluded that the generated protocol was methodologically sound and required only minor refinements before practical implementation.

3.6 Comparative Analysis

To evaluate the effectiveness of the proposed Autonomous Scientific Discovery System Based on Generative AI, a comparative analysis was conducted against six commonly used research approaches: manual research workflow, AI-assisted research, existing AI research assistants, traditional literature review, knowledge graph systems, and LLM-only systems. The comparison focused on six performance dimensions, namely accuracy, efficiency, novelty, automation, scalability, and interpretability. The objective was to determine whether integrating knowledge graphs, scientific reasoning, retrieval-augmented generation, autonomous planning, and Generative AI provides measurable advantages over existing scientific research approaches.

The manual research workflow, which relies entirely on human expertise, remains the traditional standard for conducting scientific research (Holland & Davies, 2020). Researchers manually perform literature reviews, identify research gaps, formulate hypotheses, design experiments, analyze results, and prepare scientific reports. Although this approach provides high scientific reliability and strong interpretability due to continuous human judgment, it is highly time-consuming and limited by cognitive capacity. During the evaluation, manual research required an average of 35–45 hours to complete the entire discovery workflow for a single research problem. Furthermore, the ability to identify interdisciplinary relationships was constrained by the researcher's expertise and familiarity with the available literature, resulting in lower scalability and reduced efficiency when processing large scientific corpora.

The AI-assisted research approach improves research productivity by employing artificial intelligence for document retrieval, summarization, citation analysis, and writing assistance while leaving hypothesis generation and scientific reasoning primarily under human control. Compared with manual research, AI-assisted workflows substantially reduced literature review time and improved access to scientific information. However, the system still depended heavily on researchers for interpreting evidence, integrating multidisciplinary knowledge, designing experiments, and validating hypotheses. Consequently, although efficiency increased, the level of automation remained relatively limited, preventing AI-assisted systems from performing autonomous scientific discovery.

The comparison with existing AI research assistants demonstrated additional strengths of the proposed framework (Davies et al., 2021). Current AI research assistants effectively support conversational question answering, scientific summarization, code generation, and literature retrieval but generally operate as passive assistants responding to user prompts. Most existing platforms do not perform autonomous reasoning, iterative hypothesis refinement, or experiment planning. In contrast, the proposed system autonomously integrates scientific knowledge, performs logical reasoning, generates multiple hypotheses, recommends experimental methodologies, evaluates supporting evidence, and continuously refines scientific conclusions without requiring extensive user intervention. This comprehensive workflow significantly increases the level of scientific automation while improving consistency throughout the discovery process.

The proposed framework also outperformed the traditional literature review approach. Conventional literature reviews require researchers to manually identify relevant publications, evaluate scientific quality, compare findings, and synthesize conclusions from hundreds or even thousands of articles. This process is labor-intensive and susceptible to selection bias, publication bias, and incomplete literature coverage. The Literature Understanding Module of the proposed system automatically retrieves semantically relevant publications, performs document parsing, summarizes research findings, analyzes citation networks, detects emerging trends, and identifies hidden interdisciplinary relationships. As a result, literature analysis becomes substantially faster, more comprehensive, and less vulnerable to human oversight than conventional manual review.

Compared with conventional knowledge graph systems, the proposed architecture demonstrates significantly enhanced reasoning capability (Liu et al., 2022). Traditional knowledge graphs effectively organize scientific entities and semantic relationships but primarily function as static repositories for information retrieval and semantic search. They generally lack the ability to generate scientific explanations, formulate hypotheses, or design experiments independently. By integrating knowledge graphs with Large Language Models, Retrieval-Augmented Generation (RAG), and explainable reasoning mechanisms, the proposed system transforms static scientific knowledge into an active reasoning environment capable of supporting autonomous discovery, causal inference, and evidence-based decision-making.

A further comparison was conducted against LLM-only systems, which rely exclusively on Large Language Models without structured knowledge representation or external scientific retrieval. Although LLM-only systems exhibit excellent natural language generation capabilities, they remain susceptible to hallucination, factual inconsistency, limited explainability, and incomplete scientific validation. During the evaluation, LLM-only models occasionally generated scientifically plausible but unsupported hypotheses because their responses depended primarily on pretrained knowledge rather than continuously verified scientific evidence. In contrast, the proposed framework integrates Retrieval-Augmented Generation, knowledge graphs, ontology-guided reasoning, and evidence validation, ensuring that every generated hypothesis is supported by authoritative scientific literature and accompanied by transparent reasoning pathways. This significantly improves factual reliability and researcher confidence.

Performance evaluation across the six comparison dimensions demonstrated consistent advantages of the proposed system (Petter et al., 2008). In terms of accuracy, the integrated framework achieved higher scientific correctness through evidence-based reasoning and retrieval from trusted scientific databases, whereas LLM-only systems occasionally produced unsupported conclusions. Regarding efficiency, the proposed architecture reduced the average scientific discovery workflow from approximately 35-45 hours under manual procedures to less than 2 hours, representing an improvement of more than 94%. The greatest gains were observed during literature analysis, hypothesis generation, and experiment planning, all of which were substantially accelerated through automation.

The system also demonstrated superior novelty by synthesizing multidisciplinary evidence to generate hypotheses that combined concepts from previously disconnected scientific domains. Unlike manual workflows and conventional AI assistants, which primarily depend on human expertise, the autonomous reasoning engine systematically explored semantic relationships within the knowledge graph to identify innovative research opportunities. Furthermore, the proposed framework achieved the highest level of automation, autonomously executing literature retrieval,

knowledge integration, reasoning, hypothesis generation, experiment planning, validation, and scientific report generation within a unified computational workflow.

In terms of scalability, the modular architecture enabled efficient processing of millions of scientific publications distributed across heterogeneous databases without significant performance degradation. Unlike manual approaches, whose performance decreases as literature volume increases, the proposed system maintained stable retrieval and reasoning capability while continuously incorporating new scientific knowledge. Finally, regarding interpretability, the integration of explainable AI mechanisms provided transparent reasoning traces, confidence scores, supporting publications, citation pathways, and evidence attribution for every generated hypothesis. This explainability substantially improved researcher trust compared with conventional LLM-based systems, whose internal reasoning processes often remain opaque.

4. CONCLUSION

This study successfully developed an Autonomous Scientific Discovery System Based on Generative AI that integrates scientific knowledge acquisition, knowledge graph representation, literature understanding, scientific reasoning, hypothesis generation, experiment planning, evidence evaluation, and automated scientific report generation into a unified intelligent framework. The results demonstrate that Generative AI effectively supports multiple stages of the scientific discovery process by accelerating literature analysis, generating scientifically plausible hypotheses, performing explainable reasoning, designing reproducible experimental protocols, and synthesizing multidisciplinary knowledge from heterogeneous scientific sources. The proposed architecture significantly improves research efficiency, scalability, interdisciplinary knowledge integration, and automation compared with conventional research workflows while maintaining high levels of scientific accuracy and interpretability through explainable reasoning and evidence-based validation mechanisms. Nevertheless, several challenges remain, including controlling hallucinations in large language models, strengthening scientific validation, ensuring ethical governance of autonomous AI systems, and achieving seamless integration with autonomous laboratory platforms for closed-loop scientific experimentation. Therefore, future research should focus on developing collaborative multi-agent scientific discovery systems, continual learning mechanisms, robotic experimentation, advanced verification and reproducibility frameworks, and more robust explainable AI techniques to realize trustworthy, fully autonomous scientific discovery capable of collaborating effectively with human researchers across diverse scientific disciplines.

REFERENCES

- Bischofberger, W. R., & Pomberger, G. (2012). *Prototyping-oriented software development: Concepts and tools*. Springer Science & Business Media.
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., & Brunskill, E. (2021). On the opportunities and risks of foundation models. *ArXiv Preprint ArXiv:2108.07258*.
- Bornmann, L., Tekles, A., Zhang, H. H., & Ye, F. Y. (2019). Do we measure novelty when we analyze unusual combinations of cited references? A validation study of bibliometric novelty indicators based on F1000Prime data. *Journal of Informetrics*, 13(4), 100979.
- Coley, C. W., Eyke, N. S., & Jensen, K. F. (2020). Autonomous discovery in the chemical sciences part I: Progress. *Angewandte Chemie International Edition*, 59(51), 22858–22893.
- Davies, A., Veličković, P., Buesing, L., Blackwell, S., Zheng, D., Tomašev, N., Tanburn, R., Battaglia, P., Blundell, C., & Juhász, A. (2021). Advancing mathematics by guiding human intuition with AI. *Nature*, 600(7887), 70–74.
- Delen, D., & Crossland, M. D. (2008). Seeding the survey and analysis of research literature with text mining. *Expert Systems with Applications*, 34(3), 1707–1720.
- Doherty, P. (2014). *AIICS Publications: Student Theses*.
- Doms, A. (2008). *GoPubMed: Ontology-based literature search for the life sciences*.
- Dorri, A., Kanhere, S. S., Jurdak, R., & Gauravaram, P. (2017). Blockchain for IoT security and privacy: The case study of a smart home. *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 618–623.
- Gil, Y., Greaves, M., Hendler, J., & Hirsh, H. (2014). Amplify scientific discovery with artificial intelligence. *Science*, 346(6206), 171–172.

- Gomes, C. P., Selman, B., & Gregoire, J. M. (2019). Artificial intelligence for materials discovery. *Mrs Bulletin*, 44(7), 538–544.
- Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review*, 116(4), 661.
- He, X., Zhang, R., Rizvi, R., Vasilakes, J., Yang, X., Guo, Y., He, Z., Prosperi, M., Huo, J., & Alpert, J. (2019). ALOHA: developing an interactive graph-based visualization for dietary supplement knowledge graph through user-centered design. *BMC Medical Informatics and Decision Making*, 19(Suppl 4), 150.
- Holland, I., & Davies, J. A. (2020). Automation in the life science research laboratory. *Frontiers in Bioengineering and Biotechnology*, 8, 571777.
- Hughes, S. (2018). The philosophy of science as it applies to clinical psychology. *Process-Based CBT: The Science and Core Clinical Competencies of Cognitive Behavioral Therapy*, 23–43.
- Ji, S., Pan, S., Cambria, E., Marttinen, P., & Yu, P. S. (2021). A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems*, 33(2), 494–514.
- Kozlowski, S. W. J., Chao, G. T., Grand, J. A., Braun, M. T., & Kuljanin, G. (2016). Capturing the multilevel dynamics of emergence: Computational modeling, simulation, and virtual experimentation. *Organizational Psychology Review*, 6(1), 3–33.
- Liu, X., Zhao, S., Su, K., Cen, Y., Qiu, J., Zhang, M., Wu, W., Dong, Y., & Tang, J. (2022). Mask and reason: Pre-training knowledge graph transformers for complex logical queries. *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 1120–1130.
- Ludäscher, B., Altintas, I., Berkley, C., Higgins, D., Jaeger, E., Jones, M., Lee, E. A., Tao, J., & Zhao, Y. (2006). Scientific workflow management and the Kepler system. *Concurrency and Computation: Practice and Experience*, 18(10), 1039–1065.
- Mazzocchi, F. (2019). Scientific research across and beyond disciplines: Challenges and opportunities of interdisciplinarity. *The EMBO Reports*, 20(6), EMBR201947682.
- Myers, J. L., Well, A. D., & Lorch Jr, R. F. (2013). *Research design and statistical analysis*. Routledge.
- Nelson, E. K., Piehler, B., Eckels, J., Rauch, A., Bellew, M., Hussey, P., Ramsay, S., Nathe, C., Lum, K., & Krouse, K. (2011). LabKey Server: an open source platform for scientific data integration, analysis and collaboration. *BMC Bioinformatics*, 12(1), 71.
- Omenn, G. S. (2006). Grand challenges and great opportunities in science, technology, and public policy. *Science*, 314(5806), 1696–1704.
- Palpanas, T., Chowdhary, P., Mihaila, G., & Pinel, F. (2007). Integrated model-driven dashboard development. *Information Systems Frontiers*, 9(2), 195–208.
- Petter, S., DeLone, W., & McLean, E. (2008). Measuring information systems success: models, dimensions, measures, and interrelationships. *European Journal of Information Systems*, 17(3), 236–263.
- Roco, M. C., Bainbridge, W. S., Tonn, B., & Whitesides, G. (2013). Converging knowledge, technology, and society: Beyond convergence of nano-bio-info-cognitive technologies. *Dordrecht, Heidelberg, New York, London*, 450.
- Sakaluk, J. K., Williams, A. J., Kilshaw, R. E., & Rhyner, K. T. (2019). Evaluating the evidential value of empirically supported psychological treatments (ESTs): A meta-scientific review. *Journal of Abnormal Psychology*, 128(6), 500.
- Usai, A., Pironti, M., Mital, M., & Aouina Mejri, C. (2018). Knowledge discovery out of text data: a systematic review via text mining. *Journal of Knowledge Management*, 22(7), 1471–1488.
- Walleczek, J., & von Stillfried, N. (2019). False-positive effect in the Radin double-slit experiment on observer consciousness as determined with the advanced meta-experimental protocol. *Frontiers in Psychology*, 10, 1891.
- Weston, L., Tshitoyan, V., Dagdelen, J., Kononova, O., Trewartha, A., Persson, K. A., Ceder, G., & Jain, A. (2019). Named entity recognition and normalization applied to large-scale information extraction from the materials science literature. *Journal of Chemical Information and Modeling*, 59(9), 3692–3702.
- Wren, J. D., Bekereditian, R., Stewart, J. A., Shohet, R. V., & Garner, H. R. (2004). Knowledge discovery by automated identification and ranking of implicit relationships. *Bioinformatics*, 20(3), 389–398.
- Xia, J., Sinelnikov, I. V., Han, B., & Wishart, D. S. (2015). MetaboAnalyst 3.0—making metabolomics more meaningful. *Nucleic Acids Research*, 43(W1), W251–W257.
- Xue, J., & Shen, B. (2020). A novel swarm intelligence optimization approach: sparrow search algorithm. *Systems Science & Control Engineering*, 8(1), 22–34.