Preprints.org

Article

Research Trend Analysis in the Field of Self-Driving Lab Using Network Analysis and Topic Modeling

Woojun Jung , Insung Hwang , Keuntae Cho *

Article

Research Trend Analysis in the Field of Self-Driving Lab Using Network Analysis and Topic Modeling

Woojun Jung [1], Insung Hwang [1] and Keuntae Cho [2,*]

[1] Department of Industrial Engineering, Sungkyunkwan University, Suwon 16419, Republic of Korea

[2] Department of Systems Management Engineering, Sungkyunkwan University, Suwon 16419, Republic of Korea

* Correspondence: ktcho@skku.edu; Tel.: +82-031-290-7602

Abstract: A self-driving lab (SDL) is a system that automates experiment design, data collection, and analysis using robotics and artificial intelligence (AI) technologies, and its significance has grown substantially in recent years. This study analyzes the overall research trends in SDL, examines changes in specific topics, visualizes the relational structure among authors to identify key contributors and extracts major themes from extensive texts to highlight essential research content. To achieve these objectives, trend analysis, network analysis, and topic modeling are conducted on 352 research papers collected from the Web of Science from 2004 to 2023. The results revealed three key findings. First, SDL research has surged since 2019, driven by the pandemic and advancements in AI technologies, reflecting heightened activity in this field. Second, several influential researchers have been identified as central figures in the network, playing pivotal roles in collaboration and information dissemination. Third, SDL research exhibits interdisciplinary convergence, encompassing areas such as material optimization, biological processes, and AI predictive algorithms. This study underscores the growing importance of SDL as a research tool across diverse academic disciplines and provides practical insights into sustainable future research directions and strategic approaches.

Keywords: self-driving lab; SDL; trend analysis; network analysis; topic modeling; sustainability

1. Introduction

A self-driving lab (SDL) represents an innovative paradigm in modern scientific research that provide unprecedented opportunities for researchers to accelerate and automate experimental and analytical processes [1]. SDL integrates robotics, artificial intelligence (AI), and machine learning (ML) to streamline workflows, including experimental design, execution, and data analysis [2]. By using advanced robotic systems and automated equipment, SDL facilitates complex experimental tasks, whereas AI enhances efficiency by automating data interpretation and evaluating the results. As a cutting-edge research tool, SDL bridges the gap between futuristic innovation and practical applications, thereby enabling breakthroughs that redefine the limits of scientific exploration. In addition to combining equipment and software, SDL functions as a "pioneer" in expanding the frontiers of scientific discovery. This represents the convergence of human curiosity and technological advancement, fostering innovation with the potential to fundamentally transform human life.

The emergence of SDL marks a pivotal transformation in research automation and represents an integral part of the data-driven science paradigm, which aims to process and analyze vast datasets across diverse fields to generate new knowledge. SDL's software plays a critical role in interpreting high-dimensional datasets and solving multivariate research challenges [3]. By effectively exploring complex variable combinations and identifying optimal experimental conditions, SDL serves as a powerful tool to overcome human limitations and push the boundaries of scientific discovery.

SDL can perform experiments up to 1,000 times faster than traditional methods, driving significant advancements in areas such as new material discovery, drug development, and process optimization [2]. This unparalleled speed enables researchers to conduct large-scale experiments,

collect and analyze data rapidly, and focus on high-level creative tasks [4]. Industries such as chemistry, life sciences, and materials engineering, in which high-throughput experimentation is essential, have found that SDL is indispensable for addressing high-dimensional data challenges and accelerating innovation.

Despite their widespread applications, systematic research analyzing the overarching trends in SDL is limited. Previous studies have explored SDL's applications in health and wellness or conducted bibliometric analyses by engineering field and country, highlighting a recent surge in SDL research [5]. Other studies proposed directions for SDL development, including equipment design, workflow optimization, data management [6], and effective SDL construction methods [7]. However, no comprehensive research has focused on SDL trend analysis as a primary subject, and limitations in data and methodology have raised concerns about the reliability of the existing findings. To address these gaps, this study applies a systematic analytical framework based on comprehensive and reliable data to establish strategic directions for SDL adoption across industries.

The primary goal of this study is to promote the adoption and utilization of SDL by conducting an in-depth trend analysis, thereby advancing scientific research across industries. It systematically positions SDL as a critical research tool across disciplines and offers meaningful insights into research directions and strategic approaches. This research underscores SDL's role as an indispensable tool in the AI-driven era, thereby providing sustainable adoption strategies tailored to various industries and laying the groundwork for future researchers by analyzing SDL's automation and optimization functions. Furthermore, it proposes new directions for expanding SDL applications, confirming its status as a transformative technology in scientific research.

This study utilized data from the Web of Science to collect SDL-related research papers by applying systematic pre-processing techniques such as stop words removal and synonym standardization. This study employed trend analysis, network analysis, and topic modeling to uncover significant insights. Trend analysis identified overarching research trends, network analysis mapped collaborative relationships among key authors, and topic modeling extracted major research themes and keywords. This comprehensive approach provided a detailed understanding of SDL's current research landscape, thereby offering implications for future studies. The integration of diverse analytical techniques can serve as a guide for strategic and efficient SDL research planning.

The findings of this study have practical implications in various domains. For industries, it offers insights into R&D prioritization and investment strategies, identifying potential SDL applications and commercialization opportunities to prepare sustainable businesses for market entry and enhance global competitiveness. In academia, the SDL trend analysis aids in setting research agendas and improving educational programs to train an industry's sustainable workforce. Enhanced industry-academic collaboration can bridge the gap between research outcomes and practical applications. For policymakers, this study provides evidence for shaping science and technology policies that support SDL development, thereby fostering a sustainable research-friendly environment and boosting national competitiveness.

The remainder of this paper is organized as follows. Section 2 reviews the theoretical background of SDL and prior studies, setting the objectives. Section 3 details the research methodology, including the data collection and analysis techniques used in the SDL trend analysis. Section 4 presents the findings of the trend analysis, network analysis, and topic modeling. Section 5 interprets these results from academic, practical, and policy perspectives and draws implications. Finally, Section 6 concludes with a discussion of the study's limitations and suggestions for future research directions.

2. Literature Review

2.1. Self-Driving Lab Concepts and Development Process

The SDL is a cutting-edge research environment designed to automate experiments and maximize efficiency. It is a scientific research system that autonomously manages and optimizes the entire experimental process using AI and robotics. SDL integrates robotics, experimental systems, and ML models to design experiments, analyze results, and interpret outcomes, thereby offering a

revolutionary autonomous laboratory platform. Recently, data-driven experimental planning has garnered increasing attention from the scientific community [8]. The integration of ML and innovative approaches to scientific methods has energized both theoretical and computational research domains, in addition to practical applications [6]. Moreover, SDL has significant potential to address critical societal needs such as carbon-neutral processes, food security, sustainable agriculture, clean energy, energy storage, and drug discovery [9]. By leveraging AI and ML algorithms to analyze datasets ranging from small to large, SDL enables researchers to identify patterns and trends, resulting in more effective and efficient R&D processes [10]. Following are some of the areas where SDL is being utilized.

Biotechnology: SDL provides a powerful platform for autonomously designing and engineering new biological functions, particularly in synthetic biology and genetic manipulation [11]. For example, SDL has shown promising results in autonomously exploring protein conformational landscapes and advancing biomedical and molecular biology research beyond traditional chemical synthesis methods [12,13]. This enables researchers to investigate the intricate interactions of life systems more efficiently. SDL has also been applied in the development of biochemical response neural networks (CRNNs), which autonomously design neural networks to predict chemical responses and identify experimental pathways [14]. This has proven to be instrumental in understanding and controlling the kinetics of complex chemical reactions. Additionally, advancements such as autonomous implantable devices for neural recording and stimulation in freely moving primates demonstrate SDL's applications in advanced biomedical research [15].

Chemical Engineering: Traditional computational tools have limitations in accelerating chemical research [16]. SDL addresses this by autonomously exploring multistage chemical pathways, enabling a rapid understanding of complex chemical systems and achieving optimized results [17]. For instance, SDL has been applied to electrocatalyst discovery, utilizing closed-loop approaches for nitrogen reduction reactions to efficiently explore multi-target experimental outcomes [18]. Deep learning-driven SDL systems have also uncovered unknown reaction pathways that significantly contribute to the understanding of complex chemical systems [14]. These advancements extend the boundaries of chemical exploration and enhance global collaboration, thereby fostering the universalization of scientific discovery [19].

Materials Engineering: In materials science, SDL demonstrates its ability to address complex multi-objective problems [7]. For instance, SDL's integration into thin-film material research has significantly improved optimization processes through model-based algorithms, identifying ideal synthesis conditions for materials with intricate electronic properties [20]. These processes provide more accurate and consistent results than conventional methods and enhance material performance across diverse applications. SDL has accelerated the discovery of new battery chemicals and the development of energy storage solutions, particularly in battery research [21–23]. Moreover, SDL contributes to the optimization of solar cell materials by autonomously refining perovskite nanocrystals [24]. It is also employed in reverse design challenges, where deep reinforcement learning of experimental data efficiently explores design spaces to achieve optimal material properties [25,26]. These capabilities enable researchers to design innovative materials quickly and precisely.

SDL emphasizes the integration of robotics and advanced data processing, thereby achieving research efficiencies that are unattainable with traditional methods [4]. For example, using Bayesian experimental methods, SDL reduces the number of experiments required to identify high-performance parametric structures by approximately 60-fold compared with grid-based exploration [27–30]. Platforms such as Chem-OS provide universal access to autonomous discovery, enabling under-resourced researchers to participate in SDL technologies [31]. This democratizes scientific discovery and extends opportunities beyond privileged research groups. SDL has also shown promise in diagnostics, where deep probabilistic learning methods can autonomously interpret experimental data, including automating the analysis of X-ray diffraction spectra [32,33]. These features enable SDL to process large amounts of experimental data quickly and accurately, thereby maximizing research efficiency. In addition to interoperable data representations, effective data

sharing and communication methods are needed to realize laboratory automation [34]. In addition, it has been expanded through cloud-based SDL platforms, allowing researchers to conduct autonomous experiments remotely and strengthening cooperation within the global scientific community [24].

2.2. Prior Research on Self-Driving Lab and Limitations

The introduction of SDL has been a pivotal driving force, accelerating growth across various research domains, particularly in the advancement of AI-based automation [4]. Additionally, the active involvement of governments and research institutes has been identified as a major factor propelling the expansion of SDL research. For example, Da Silva[1] analyzed the application of AI in healthcare by examining the adoption of SDL across different years, engineering fields, and countries using bibliometric analysis. The results revealed a rapid increase in SDL-related research, particularly in the chemical and bioscience fields, highlighting SDL's potential to provide significant opportunities for emerging economies. These countries are likely to use SDL to enhance research competitiveness and reduce the usage of both resources and time [5]. The growing recognition of SDL's importance in academia and industry is evidenced by its widespread adoption, with leading research institutes in countries such as the United States, China, and Germany at the forefront of this development [5].

Furthermore, some studies have explored the development directions and prospects of SDL. Hysmith[2] emphasized the necessity of careful planning in various aspects of SDL design, such as physical configurations, data management, and workflow optimization. This study particularly highlighted the shift from focusing on individual tools and tasks to creating and managing complex workflows, underscoring the importance of integrating human input into processes. Additionally, the role of the reward function design was identified as critical for developing efficient workflows alongside the interplay between hardware advancements, ML applications across chemical processes, and compensation systems within research. Hysmith envisioned a future for SDL that merges AI's precision, speed, and data processing capabilities with human intuitive hypothesis formulations [6].

Research has also been conducted on the effective construction of SDLs. MacLeod[3] stressed that SDLs must operate beyond simple automation, emphasizing their adaptability to new research areas. This study outlined the criteria for SDL design and addressed challenges such as laboratory reuse. They argued that effective SDLs should (1) operate at speeds surpassing traditional automation and (2) demonstrate the ability to quickly adapt to novel research contexts. Furthermore, MacLeod identified key strategies for building SDLs that could expedite the discovery of new materials, thereby emphasizing the importance of adaptability and reusability in SDL design [7].

While these studies analyzed SDL trends, certain limitations in data reliability and interpretation should be addressed. However, the data used in these analyses often lack robustness, thereby undermining the validity of the results. Moreover, inadequate preprocessing of the data further affects the accuracy of the findings. Research methodologies have not been consistently established, making cross-industry and cross-technology comparisons challenging. These shortcomings often limit the practical value of such studies, as they tend to focus on listing technological changes or summarizing the current research landscape, rather than providing actionable insights for industrial or research strategies. Furthermore, the lack of in-depth analyses or predictive models limits their ability to offer insights into evolving markets and technologies.

To address these gaps, this study aims to provide a more reliable and comprehensive analysis of SDL trends. By leveraging robust data and systematic methodologies, this study seeks to provide practical insights that can guide both academic research and industrial strategies. This study also incorporates predictive models and in-depth analyses to better understand the shifting landscape of markets and technologies, thereby offering a more valuable foundation for future research and applications.

3. Materials and Methods

This study analyzes the overall research trends in SDL and explores strategies for its successful implementation and sustainable research development. Trend analysis, network analysis, and topic modeling techniques were employed to achieve this. The research was conducted in five stages, as illustrated in **Figure 1**.
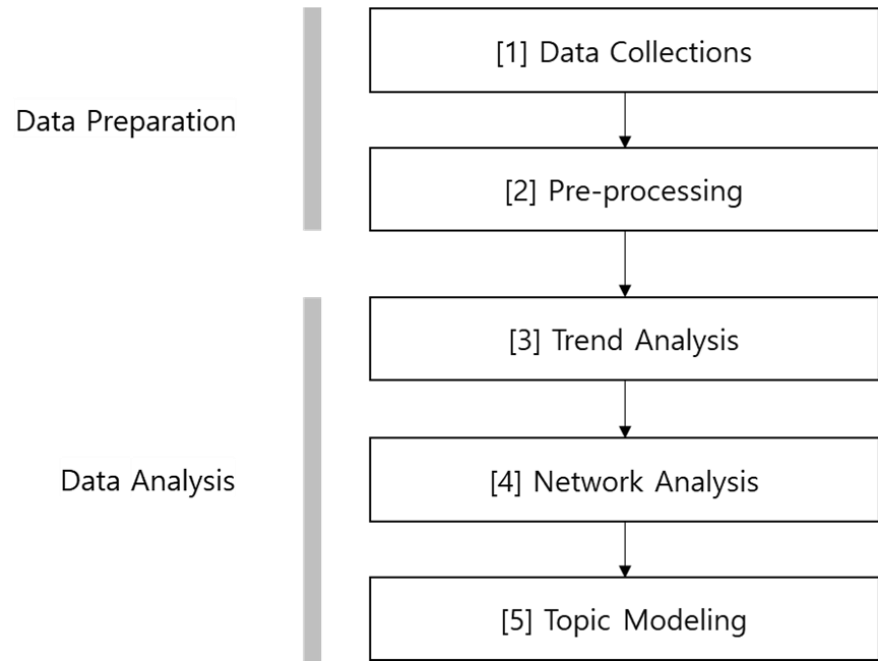

Figure 1. Summary of Research Procedure.

3.1. Data Collection

The Data Collection step involves selecting the data necessary for research analysis. To ensure the reliability of the data, publication records from the Web of Science were utilized [35]. For a research trend analysis, it is essential to cover a period that sufficiently reflects the emergence of new theories, technological advancements, and research methodologies. Accordingly, a 20-year period (2004–2023) was selected to provide a representative dataset to examine the development of research directions and significant trends within the field. As shown in **Table 1**, search expressions were constructed and refined to extract relevant data. In addition, synonymous terms, as detailed in **Table 2**, were incorporated to ensure comprehensive coverage [36]. The selection of similar terms was guided by a review of previous studies on terminology used to describe SDLs.

Table 1. Web of science search formula

Search Formula
"self-driving lab" OR "self driving lab" OR "self driven lab" OR "self-driving labs" OR "self driving labs" OR "self driven labs" OR "self driving system" OR "autonomous experimentation" OR "autonomous lab" OR "autonomous discovery" OR "autonomous chemical experiment" OR "acceleration materials platform"

Table 2. Web of science search formula.

Additional Similar Words
"self driving laboratory" OR "self driven laboratory" OR "automated lab" OR "automated experimentation" OR "lab automation"

The document type was limited to "Article," resulting in the extraction of 352 records. To facilitate data processing, the extraction format was designated as Excel, including fields such as Author, Title, Source, Sponsors, Times Cited, Accession Number, Abstract, Keywords, Addresses, and Document Type.

3.2. Data Preprocessing

The pre-processing step involves transforming the data into a format suitable for research analysis using Python and Google Colab. From the 352 data points extracted in Step 1, abstracts containing keywords related to autonomous driving—such as "car," "cars," "vehicle," "transit," "conveyance," "airplane," and "ship"—were filtered out. After this filtering process, the dataset was refined to 218 data points, which were used in subsequent analysis.

During the data pre-processing stage, terminology was removed following a filtering step. This process is divided into two sub-steps: Stop Word Removal and Meaningless Keyword Removal. In the Stop Word Removal step, common words, such as articles, prepositions, and conjunctions, are eliminated to reduce noise and focus on relevant terms. In the Meaningless Keyword Removal step, words that hold no significance for the analysis, such as "dollar" and "date," are excluded to enhance the dataset's analytical quality. These steps ensure that the dataset is refined and free from unnecessary or irrelevant terms, making it suitable for further analysis. Representative examples of the terms removed are listed in **Table 3**.

Table 3. Stop & Meaningless Words.

Stop Words	"using," "used," "also," "the," "a," "however," "***," "well"
Meaningless Words	"data," "system," "results," "model," "time," "work," "use," "two," "one," "based," "***," "different," "new"

The final step in the preprocessing stage is synonym grouping. This step involves consolidating synonyms with identical meanings into single terms. By doing so, the potential for noise in data analysis is minimized, which enhances the reliability of the analysis results. Representative examples of the grouped synonyms are presented in **Table 4**.

Table 4. Synonym.

Synonym
"auto": "automated," "autonomous," "automation"   "experiments": "experiment," "experimental," "experimentation"   "lab": "laboratory," "labs," "laboratories"

3.3. Trend Analysis

A trend analysis was conducted on the 218 publications extracted during the pre-processing stage using Python and Google Colab. The purpose of trend analysis is to understand the changes occurring in the SDL field and to support the prediction of technology directions or the formulation of strategic decisions. Furthermore, it plays a crucial role in understanding shifts in industrial markets

and technology trends, enabling stakeholders to seize opportunities, manage risks, and maintain competitiveness [37,38].

The analysis focused on three key aspects: (1) the top 10 countries with the highest number of SDL-related publications per year, (2) the top 10 journals and their corresponding citation counts, and (3) the top 10 authors and affiliations contributing to SDL research. These insights provide a comprehensive understanding of the research landscape and help identify the leading contributors and influential regions in the field.

3.4. Network Analysis

Network analysis is generally categorized based on Degree Centrality, Betweenness Centrality, Closeness Centrality, and Eigenvector Centrality [39]. In this study, three centrality measures—Degree Centrality, Betweenness Centrality, and Closeness Centrality—were analyzed to evaluate the importance and role of authors within the network from multiple perspectives. Eigenvector Centrality was excluded from the analysis because in highly interconnected networks, the centrality values of all authors tend to converge, limiting its effectiveness in accurately measuring the influence of specific hub authors. Each centrality measure provides a distinct perspective on the importance of the nodes (authors) within the network, enabling a comprehensive analysis of the network structure [40].

Degree Centrality evaluates the number of direct connections (edges) of a node and identifies authors through frequent collaborations [41]. Betweenness Centrality measures how often an author acts as a mediator along the shortest paths between nodes, highlighting individuals who facilitate the flow of information or act as bridges between different groups' networks [42]. Closeness Centrality calculates the average shortest path distance from one node to all other nodes, identifying authors who can efficiently disseminate information or ideas across a network [42].

The network was constructed using pre-processed author data, with nodes representing authors and edges indicating co-authorship relationships. Based on this structure, Degree Centrality, Betweenness Centrality, and Closeness Centrality were calculated to identify the top 20 authors and assess their structural positions and mediating roles within the network.

Authors with a high Degree Centrality, Betweenness Centrality, and Closeness Centrality play critical roles in facilitating the flow of academic knowledge and fostering cooperation between research groups [43]. These individuals act as mediators, promoting the dissemination and innovation of ideas within a collaborative research network. This analysis not only identifies key researchers leading knowledge dissemination but also evaluates the structural strengths and weaknesses of academic networks. Author network analysis provides insights into cooperative relationships among researchers, the structural development of the research field, and strategies for enhancing research collaboration [44].

3.5. Topic Modeling Analysis

Topic modeling techniques were employed to automatically extract and analyze major topics from large-scale text data. Specifically, the Latent Dirichlet Allocation (LDA) algorithm was applied to identify research focuses and field-specific topics by analyzing the patterns of recurring words in text data [45]. The topics extracted through topic modeling reflect the primary concerns of the research field and contribute to understanding the flow and trends of related studies by analyzing topic distributions across documents [46]. Furthermore, topic-specific weights assigned to each document were used to classify documents by subject and visualize the main topics of the research [47].

LDA assumes that each document consists of multiple topics and calculates the probabilistic distribution for each topic based on the occurrence of words within the document [48]. This allows the subject structure of the documents to be quantified, indicating the extent to which each document is related to a specific topic. In this study, abstracts from the collected papers served as the primary text data, which underwent preprocessing steps such as terminology removal and synonym grouping.

Additionally, an optimization process was conducted to determine the ideal number of topics to ensure a balanced and meaningful topic distribution.

The optimal number of topics was determined by evaluating the Perplexity and Coherence scores for different topic counts. Perplexity measures the predictive accuracy of the model, with lower values indicating better performance. Coherence evaluates the semantic consistency among words within a topic, with higher values reflecting stronger semantic relationships [49,50]. For optimal topic modeling, the number of topics was selected to achieve a low Perplexity score and high Coherence score, ensuring both predictive accuracy and meaningful topic representation.

4. Results

4.1. Trend Analysis Results

Analysis of SDL-related publications by year reveals significant trends. As shown in **Figure 2**, the number of publications from 2004 to 2018 remained relatively small and irregular. However, a sharp increase was observed in 2019, with the highest number of publications recorded in 2021 (44). This upward trend continued in 2022 and 2023, highlighting the remarkable growth in SDL research over the past three years. These findings suggest a substantial increase in interest and activity in the SDL research field during this period.
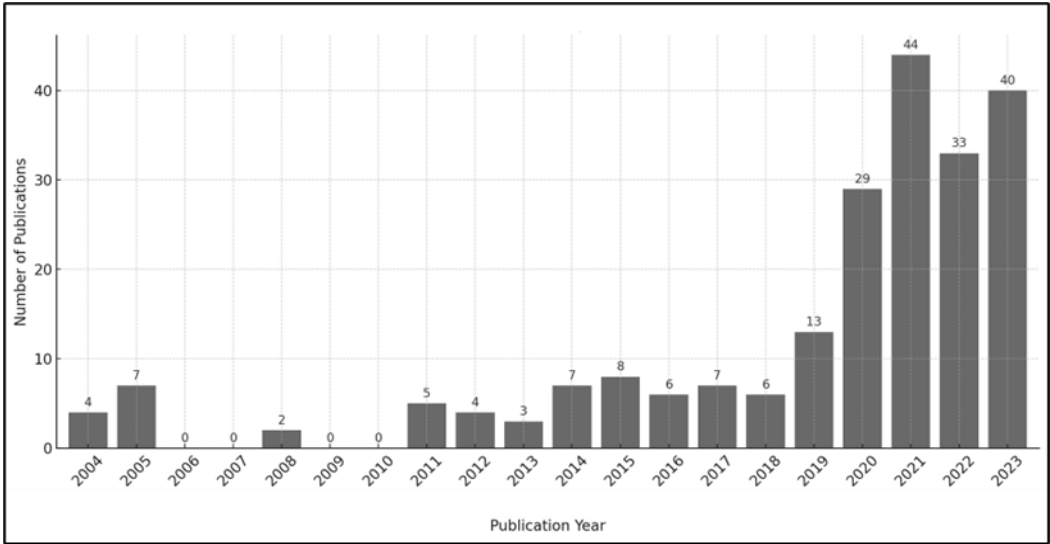

Figure 2. Number of SDL-related papers published by year.

The global pandemic, which began in the late 2010s, is considered a major driver of this growth [2]. The pandemic necessitated a paradigm shift in scientific research, emphasizing acceleration, efficiency, and non-face-to-face methodologies. With its automated and data-driven capabilities, SDL played a critical role in meeting these demands. Additionally, rapid advancements and increasing global interest in AI technologies during the 2020s have further propelled the adoption and establishment of SDL research environments, enabling the efficient management of large-scale experimental data.

A country-wise analysis of SDL-related publications also provided important insights. **Figure 3** illustrates the number of publications from the top 10 countries. Among the 30 countries that contributed to SDL research, the top 10 accounted for 186 publications, representing 85.32% of the total. This concentration indicates that SDL research is conducted predominantly in a select group of leading countries, whereas foundational research is beginning to emerge in lower-ranked and middle-tier nations.
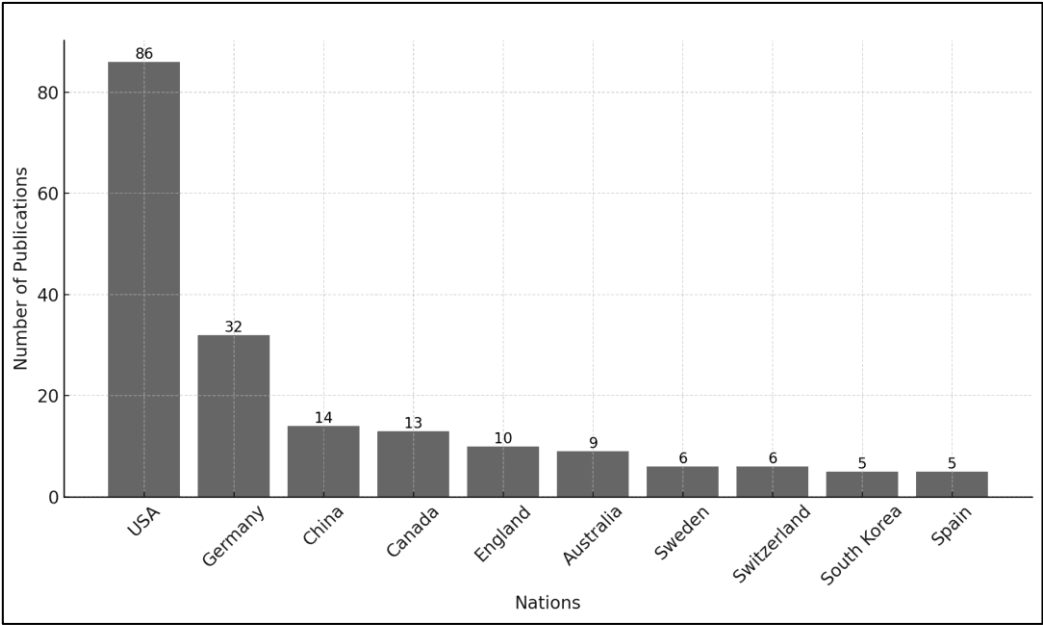
Figure 3. Read related papers by top 10 countries.

The United States, Germany, and China have led SDL-related publications, followed by Canada, the United Kingdom, Australia, Sweden, Switzerland, South Korea, and Spain. These countries' strong academic and industrial infrastructure likely contributes to their active engagement in SDL research. Meanwhile, the emergence of SDL publications in other countries suggests a growing global interest and the potential for broader participation in the field.

Table 5 presents the number of SDL-related papers published in each country. Countries with a well-established research infrastructure, such as the top 10 contributors, are likely to lead in investing in high-tech development. This leadership can be attributed to their robust research environments, ability to attract skilled human resources, close industry-academia connections, effective government policies, and strategic initiatives to enhance global competitiveness. These factors enable such countries to accelerate the development and commercialization of SDL technologies. Consequently, these nations achieve innovative breakthroughs that contribute significantly to their national economies and security.

Table 5. Total number of national SDL-related papers published.

Nation	Number of Publications (rate)	Nation	Number of Publications (rate)
USA	86 (39.3%)	France	2 (0.9%)
Germany	32 (14.6%)	India	2 (0.9%)
China	14 (6.4%)	Saudi Arabia	2 (0.9%)
Canada	13 (5.9%)	Bangladesh	1 (0.5%)
England	10 (4.5%)	Russia	1 (0.5%)
Australia	9 (4.1%)	Czech Republic	1 (0.5%)
Sweden	6 (2.7%)	Pakistan	1 (0.5%)
Switzerland	6 (2.7%)	Hungary	1 (0.5%)
South Korea	5 (2.4%)	Greece	1 (0.5%)
Spain	5 (2.4%)	Belgium	1 (0.5%)
Italy	4 (1.8%)	Taiwan	1 (0.5%)

Brazil	3 (1.4%)	Thailand	1 (0.5%)
Denmark	3 (1.4%)	Ukraine	1 (0.5%)
Scotland	2 (0.9%)	Netherlands	1 (0.5%)
Japan	2 (0.9%)	Jordan	1 (0.5%)

The journal-wise analysis of SDL-related publications provides further insights into the research landscape. **Figure 4** illustrates the number of publications and citations for the top 10 journals contributing to SDL research. Among the 145 journals, the top 10 journals accounted for 51 publications, representing 23.39% of the total. The relatively low concentration of publications in the top 10 journals indicates that SDL research spans a wide array of disciplines, including computing science, pure science, and other interdisciplinary fields. This convergence demonstrates the close connection between SDL research and domains, such as AI, materials science, and microbiology.
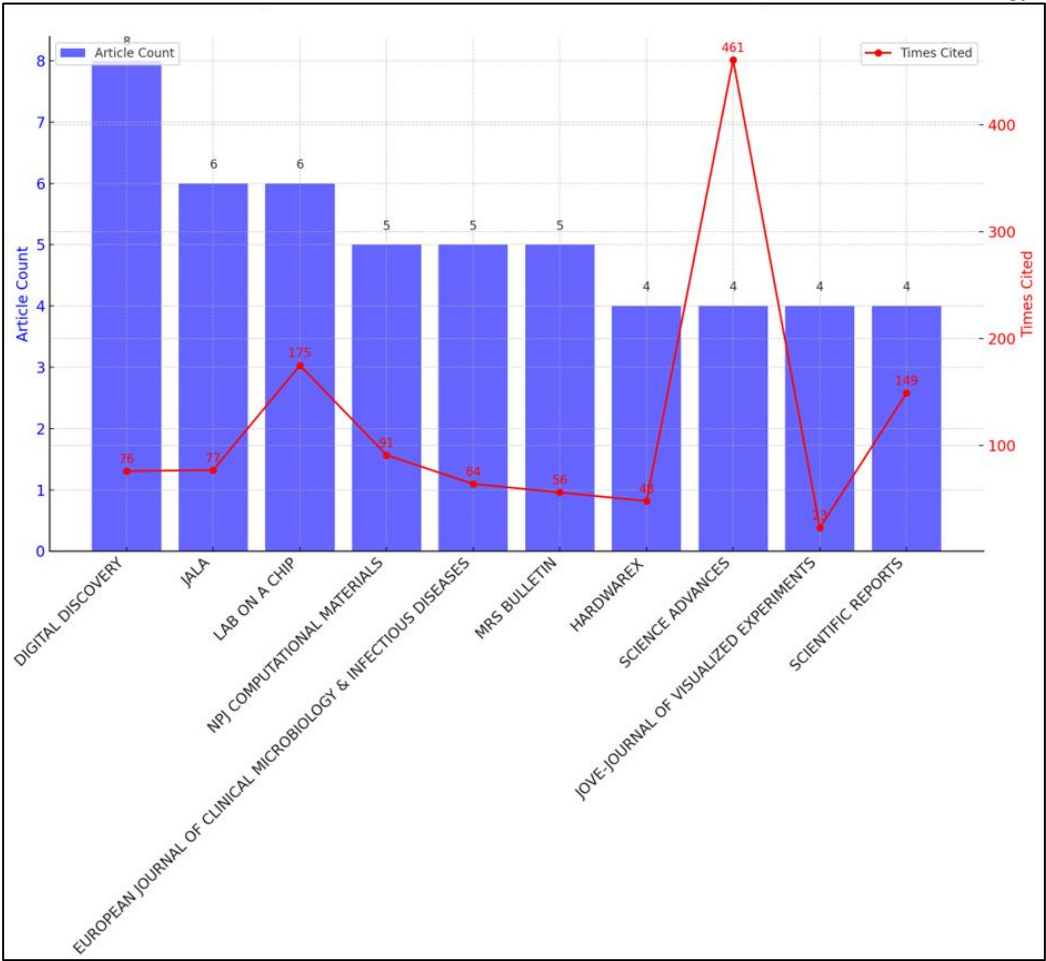

Figure 4. Number of SDL-related papers published and cited by the top 10 journals.

Among the top journals, Digital Discovery leads with eight publications, but has a relatively low citation count of 76, suggesting limited overall influence. Similarly, the Journal of Laboratory Automation (JALA) has contributed six publications with 77 citations, reflecting some impact in experimental automation research, but without significant citation influence. Conversely, Lab on a Chip stands out as a highly influential journal with six publications garnering 175 citations. This highlights the importance of microfluidics and small experimental devices in SDL research. Similarly, npj Computational Materials contributed to SDL research in computational materials science, with five publications and 91 citations.

Journals such as HardwareX and the Journal of Visualized Experiences (JoVE) exhibited relatively low influences, recording 56 and 12 citations, respectively. Scientific Reports, an open-

access journal, has shown a particular influence, with four publications receiving 149 citations. Notably, Science Advances emerged as the most influential journal, despite publishing only four papers and achieving an impressive 461 citations. This underscores its significant contribution to SDL research and its high reliability in the field. These findings highlight the varying levels of influence among journals contributing to SDL research. Journals such as Science Advances and Lab on a Chip, have established themselves as pivotal platforms in this domain, given their high citation rates and impact.

The analysis of SDL-related publications by the author identified key contributors and their affiliated research institutions. A total of 1,146 authors and 317 research institutes were identified, providing insights into the main contributors to the field and geographical distribution of SDL research. **Table 6** presents the top 10 authors with the highest number of publications, along with their affiliated institutions and countries. This highlights the influence of individual researchers and the international scope of SDL research.

Among the top contributors, Aspuru-Guzik[4] leads with eight publications, establishing himself as a central figure in SDL research in Canada. Roch[5] and Noack[6] have published six papers each, representing significant contributions from the United States. Roch[5] is affiliated with Harvard University, and Noack[6] is based at the Lawrence Berkeley National Laboratory, both of which are prominent research institutions driving SDL research in the U.S.

Jesse[7], Reyes[8], and Hickman[9] have published five papers each, demonstrating their active involvement in the field. Jesse[7] and Reyes[8] are affiliated with the Oak Ridge National Laboratory and the University at Buffalo, respectively, underscoring the diversity of research institutions contributing to SDL research in the U.S. However, Hickman[9] is a member of the University of Toronto, reaffirming Canada's pivotal role in this domain.

Other notable contributors included Vasudevan, Abolhasani, and Brown, each of whom published four papers. Vasudevan[10] is affiliated with the Oak Ridge National Laboratory, Abolhasani[11] with North Carolina State University, and Brown[12] with Boston University, reflecting the strong presence of SDL research across various U.S. universities and research centers. Kalinin[13], who has contributed three papers, continues to play an active role in SDL research at the Oak Ridge National Laboratory.

Overall, the findings highlight that SDL research is concentrated in major research institutes such as the Oak Ridge National Laboratory, Lawrence Berkeley National Laboratory, and Harvard University in the United States, as well as the University of Toronto in Canada. These institutions and their researchers have made significant contributions to advancing SDL research by emphasizing the central role of North America in this field.

Table 6. Number of papers published by the top 10 authors and research institutes.

Author	Publication Count	Affiliations (Nation)
Aspuru-Guzik, Alan	8	University of Toronto (Canada)
Roch, Loic M.	6	Harvard University (USA)
Noack, Marcus M.	6	Lawrence Berkeley National Laboratory (USA)
Jesse, Stephen	5	Oak Ridge National Laboratory (USA)
Reyes, Kristofer G.	5	University at Buffalo (USA)
Hickman, Riley J.	5	University of Toronto (Canada)
Vasudevan, Rama K.	4	Oak Ridge National Laboratory (USA)
Abolhasani, Milad	4	North Carolina State University (USA)
Brown, Keith A.	4	Boston University (USA)
Kalinin, sergei V.	3	Oak Ridge National Laboratory (USA)

4.2. Network Analysis Results

The author network analysis was conducted to evaluate the roles and importance of authors within the SDL research network using three centrality measures: Degree Centrality, Betweenness Centrality, and Closeness Centrality. Degree Centrality identifies the number of authors with which a specific author is directly connected, highlighting their collaborative significance within the network. Betweenness Centrality measures the degree to which an author acts as an intermediary with other authors, reflecting their role in bridging different collaborative groups. Closeness Centrality assesses the efficiency of an author's connections by measuring how quickly they can access information through short paths to other authors in the network.

Figure 5 provides a comparative visualization of the Degree, Betweenness, and Closeness Centralities of the top 20 authors in SDL research. Each metric demonstrates the influence and collaborative dynamics of these authors, offering insights into their respective roles in fostering cooperation and facilitating information flow within the network.

Authors with a high Degree Centrality, such as Jesse[7] and Brown[12], are recognized for their extensive collaborations. These authors are directly connected to numerous researchers, making them pivotal figures in the network and significantly influencing the field through active partnerships and knowledge dissemination.

In terms of Betweenness Centrality, Reyes[8] and Brown[12] stand out as key intermediaries. Their positions enabled them to bridge disparate groups of researchers, coordinate collaborative efforts, and ensure an efficient flow of information across the network. This intermediary role is essential for spreading innovative ideas and strengthening connectivity in the research community.

Authors with high Closeness Centrality, including Jesse[7] and Reyes[8], are strategically positioned to quickly access and distribute information within networks. Their short connection paths to other researchers effectively enhance their ability to gather and utilize cutting-edge research.

Overall, Jesse[7] and Reyes[8] emerged as influential figures across all three centrality measures. Jesse[7] excels in Degree and Closeness Centralities, underscoring his central role in the collaboration network and ability to rapidly spread research. Reyes[8] demonstrates exceptional performance in Betweenness and Closeness Centralities, positioning him as a vital intermediary with fast access to information. These findings highlight the authors' significant contributions to the SDL research field, emphasizing their roles in promoting collaboration, facilitating information flow, and driving innovation.
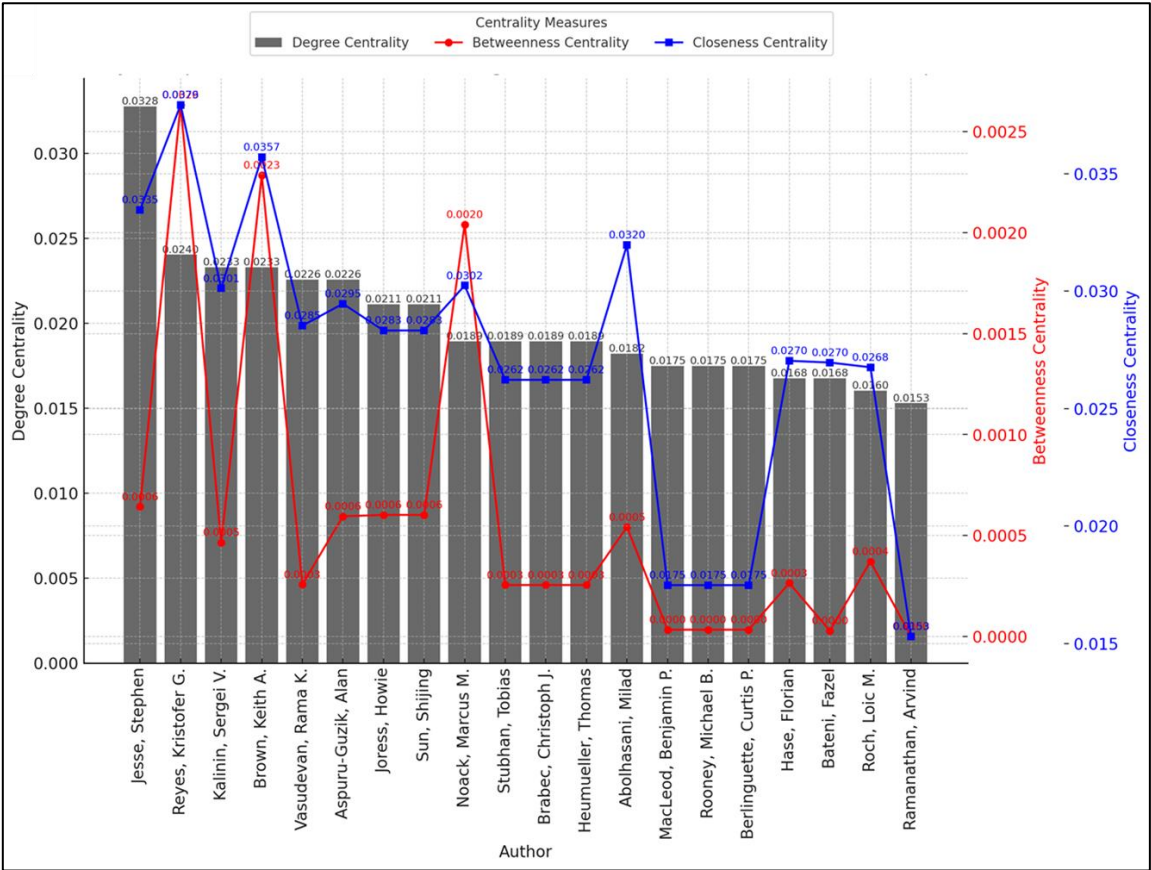
Figure 5. Top Authors by Network Centralities in Co-Authorship Analysis**.**

4.3. Topic Modeling Analysis Results

Prior to conducting topic modeling analysis, the optimal number of topics was determined through a preliminary evaluation process. **Figure 6** illustrates the calculation results for selecting the optimal topic model and balancing the perplexity and coherence scores. The analysis identified five topics with the best performance: a high coherence score and a relatively low perplexity score. Coherence measures the semantic consistency among words within a topic, with higher values indicating better interpretability. Conversely, Perplexity evaluates the model's predictive performance, with lower values indicating improved reliability. In this study, the five-topic model demonstrated a strong balance between these metrics, achieving relatively high coherence and low perplexity, signifying an optimal model performance. Based on these findings, five topics were selected for subsequent analysis.
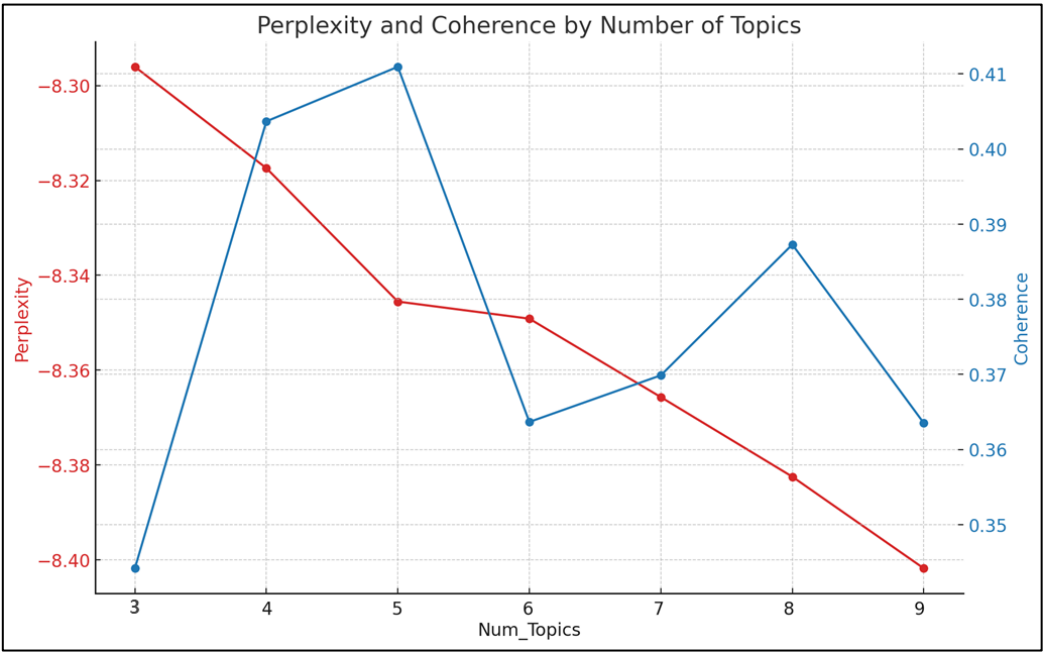
Figure 6. Calculate the optimal number of topics for topic modeling.

Figure 7 presents the results of the Topic Model Intertopic Distance Map (IDM), which illustrates the distribution of the five main themes in a two-dimensional space derived through multidimensional reduction. Each theme is positioned along the PC1 and PC2 axes to provide insights into their relationships and characteristics. Topic 1 is centrally located with high interconnectivity with other topics, suggesting its role as a central theme encompassing common elements shared across the dataset. Conversely, Topic 2 is positioned at the top of the PC2 axis, relatively independent off other topics, and represents a distinct concept with a focused theme. Topics 3 and 4 are distributed in opposite directions along the PC1 axis, each indicating specific and unique attributes. Finally, Topic 5 is located at the lower left of the map, exhibiting a low correlation with other topics, which implies a relatively independent character.

The Top-30 Most Salient Terms graph lists the key terms associated with each topic, highlighting their unique characteristics. Common terms such as "material" and "cell" are the most frequent across the dataset, underscoring their centrality in the research data. These terms suggest that materials and cellular concepts are foundational elements of SDL-related research. The topic-specific terms further define the unique attributes of each theme. For example, "atomic," "dna," and "polymer" are closely associated with Topic 2, emphasizing its focus on biological and chemical concepts. Topic 3 is characterized by terms such as "database" and "ai," indicating its strong connection to technical and computational elements. Additional terms like "biofoundry," "immobilization," and "vesicles" further differentiate the topics, highlighting their specialized content.

The IDM analysis reveals significant patterns in topic distribution. Topics centrally located, such as Topic 1, show a high correlation with other themes, suggesting the potential for multidisciplinary research and the need for a multidimensional approach. Conversely, topics positioned independently, such as Topics 2 and 5, display low relevance to other themes. This indicates the need for in-depth exploration within specific, specialized fields. These findings provide a visual and analytical framework for understanding the subject structure of SDL research, offering valuable insights for researchers to identify hidden relationships and set future research directions.

Table 7 presents the keyword weights assigned to each topic derived from the topic modeling analysis. Each keyword's weight reflects its importance within a specific topic and provides a quantitative understanding of the terms that define each theme. These weights serve as crucial indicators for characterizing topics, distinguishing between themes, and analyzing the centrality of

concepts within the dataset. By examining the weights in **Table 7**, researchers can numerically compare and interpret the distinct compositions of topics, which aids in the identification of significant terms and their relevance to specific research themes. This numerical representation facilitates a clearer understanding of topic modeling results, enhancing the ability to interpret and utilize the insights for future research planning.
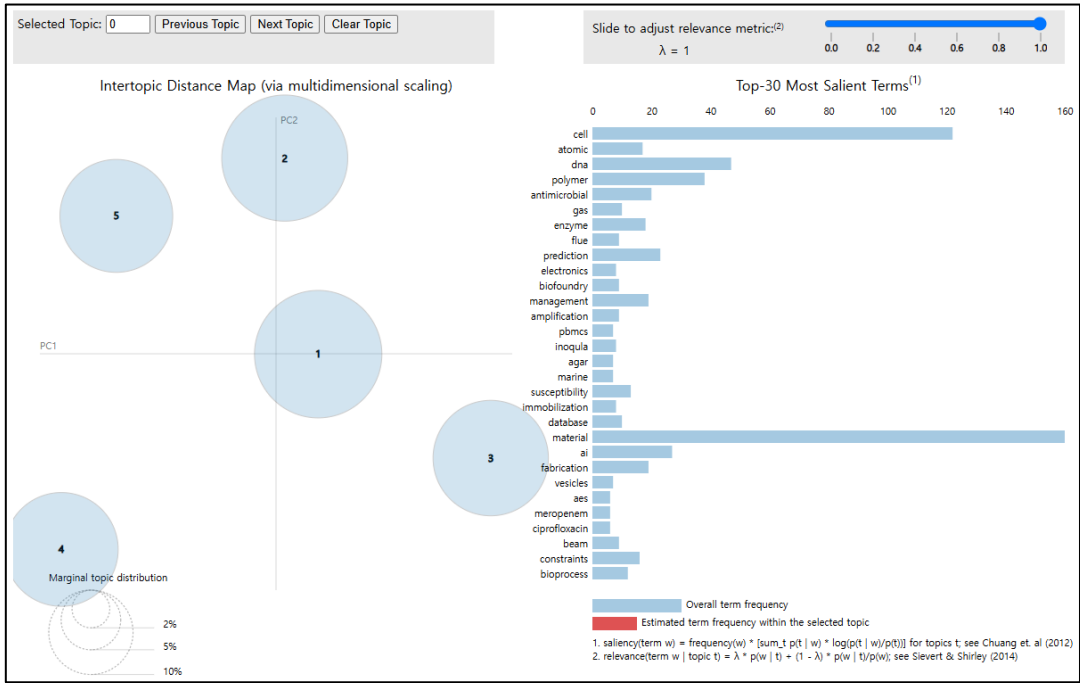

Figure 7. Topic modeling IDM results.

Table 7. Number of papers published by the top 10 authors and research institutes.

Topic	Keywords (Weights)
Topic 1	material (0.0400), optimization (0.0314), Synthesis (0.0234), chemical (0.0177), strategies (0.0100), polymer (0.0090), bioprocess (0.0048), atomic (0.0045), gas (0.0032),
Topic 2	cell (0.0360), bioprocess (0.0103), enzyme (0.0052), extraction (0.0046), blood (0.0027), pbmcs (0.0024), susceptibility (0.0023), purification (0.0022), toxicity (0.0022),
Topic 3	algorithm (0.0131), artificial (0.0119), ai (0.0091), database (0.0048), prediction (0.0057), Bayesian (0.0040), intelligence (0.0025), neural (0.0025), strategies (0.0023),
Topic 4	microfluidic (0.0121), liquid (0.0078), operation (0.0057), robotic (0.0052), fabrication (0.0046), suspension (0.0042), equipment (0.0024,) sensors (0.0022), modules (0.0024),
Topic 5	measurement (0.0100), reproducibility (0.0081), techniques (0.0056), engineering (0.0050), sensitivity (0.0028), electron (0.0028), architecture (0.0025), magnetic (0.0024),

Topic 1: Optimization and Synthesis of Materials

This topic centers on the optimization of various chemical and biological materials and represents 40.07% of the total data. The key terms associated with this topic include material, optimization, synthesis, and chemical, reflecting the focus on enhancing material properties and synthesis processes. This study involves applying optimization algorithms and strategies to precisely control chemical reactions and synthesize new substances. The ultimate goal is to improve the physical and chemical properties of the materials, leading to the development of high-performance materials suitable for diverse applications. This topic plays a crucial role in advanced materials

research and applied chemistry by offering insights into technological advancements and potential industrial applications.

Topic 2: Cells and Biological Processes

This topic focuses on cells and biological systems and examines various biological processes occurring at the cellular level. It accounts for 19.08% of the total data, with key terms including cell, enzyme, bioprocess, and extraction. This study addresses the production and enhancement of biological materials through processes such as enzymatic reactions, cell purification, and bioprocess optimization. Specifically, this topic explores methods to efficiently control and optimize complex chemical reactions occurring within cells. These advancements offer new methodologies for drug development and production of biological therapeutics, highlighting the potential for innovation in biotechnology and biopharmaceutical applications.

Topic 3: AI and Predictive Algorithms

This topic focuses on leveraging AI and predictive algorithms to analyze and optimize the properties of materials and chemical reactions. It accounts for 15.73% of the total data, with key terms including AI, algorithm, and prediction. The research emphasizes using AI and neural networks to predict and enhance chemical reactions and synthesis processes, with a particular focus on reducing uncertainty and improving prediction accuracy through approaches such as Bayesian methods. By significantly increasing the efficiency of materials research and synthesis, this topic presents a critical research direction for advancing the application of AI in materials science and chemical engineering.

Topic 4: Microfluidics and Robotics

This topic explores the integration of microfluidic technology and robotics to enhance the efficiency of biological and chemical research. It accounts for 13.27% of the total data, with key terms including microfluidic, robotic, fabrication, and modules. Microfluidic technology enables the precise control of small liquid volumes, optimizing chemical reactions and improving synthesis and experimental efficiency. Robotics plays a critical role in automating experiments and enabling precise manipulations, thereby facilitating the automation of repetitive experiments and complex processes. This combination significantly enhances experimental reproducibility while maximizing time and resource efficiency, thereby presenting a pivotal advancement in research automation and precision.

Topic 5: Measurement and Experimental Techniques

This topic focuses on measuring and analyzing the properties of materials and biological processes, accounting for 11.86% of the total data. The key terms include measurement, architecture, engineering, reproducibility, and reliability. This study emphasizes technical approaches to enhance experimental reproducibility, utilization of diverse measurement devices and sensors, and methodologies to ensure reliable data collection. By improving the experimental accuracy and providing dependable analyses of new materials and chemical reactions, this topic highlights the importance of enhancing the overall quality of research. It serves as a critical foundation for advancing scientific exploration and achieving reliable experimental outcomes.

5. Discussions

This study analyzed SDL research trends over the past 20 years (2004–2023) using abstract data from 352 SDL-related publications. By employing trend analysis, network analysis, and LDA topic modeling techniques, this study identified key trends in SDL, including annual publication growth rate, country-specific research concentration, journal publication patterns, author-specific contributions, author network relationships, and topic-specific research tendencies. The findings and implications are summarized as follows.

First, SDL research has rapidly increased since 2019, with the highest number of publications recorded in 2020 and 2021. This surge is attributed to the rising demand for remote research environments during the pandemic and advancements in AI technology, highlighting the necessity

for SDL. These findings underscore the importance of SDL as a tool for improving efficiency in various industries and academic disciplines.

Second, SDL-related publications are predominantly concentrated in the top 10 countries, such as the United States, Germany, and China, which collectively account for over 85% of all publications. This demonstrates that SDL research is primarily conducted in countries with well-established research infrastructure. The prioritization of SDL in these nations reflects its perceived importance in national competitiveness and the development of advanced technologies.

Third, SDL research is characterized by its interdisciplinary nature and converging fields such as computing, biology, materials science, and AI. Publications appear not only in computing science journals but also in pure science and applied technology journals. This indicates that SDL adopts a cross-disciplinary approach, enhances research efficiency, and offers innovative methodologies in various fields. AI and predictive algorithms play pivotal roles in automating experiments and advancing research efficiency.

Fourth, author network analysis revealed that researchers with high centrality scores play critical roles in promoting collaboration and facilitating the flow of information. For instance, **Stephen Jesse** ranked high in terms of Degree and Closeness Centrality, highlighting his role in fostering collaborations and serving as a central figure in the dissemination of research. These findings indicate that global collaboration is a key driver of SDL research and that interconnectedness among researchers significantly contributes to its advancement.

Finally, topic modeling identified key research themes such as material optimization and synthesis, cell and biological processes, AI and predictive algorithms, microfluidics and robotics, and measurement and experimental techniques. Active research is being conducted in areas such as material synthesis, the optimization of biological processes, and AI-based prediction. These trends demonstrate the potential of SDL to improve research efficiency and applicability across diverse domains, reflecting its transformative impact on the scientific research paradigm.

Previous studies emphasized the integration of SDL in medical and pharmaceutical research [5]. This study expands on prior research by confirming the active adoption of SDL not only in the medical and pharmaceutical fields but also in various engineering disciplines. Furthermore, it highlights the efficiency and potential of SDL, particularly its rapid implementation in fields such as chemistry and materials science. Significant research and development has also been observed in the theoretical and conceptual foundations of automated laboratories. By identifying detailed topic-specific trends, this study contributes to the understanding of the potential applications of SDL in various industrial domains.

6. Conclusions

This study aimed to facilitate the adoption and utilization of SDL and provide foundational data that can enhance scientific research by actively applying SDL across various industries. By analyzing SDL research trends over 20 years (2004–2023) using 352 abstracts, this study employed trend analysis, network analysis, and LDA topic modeling to derive key insights. These included the growth rate of SDL-related publications by year, country-specific research concentration, journal-specific publication patterns, author relationships, and topic-specific research trends.

The major findings of this study are as follows. First, SDL research has grown rapidly since 2019, with the highest number of publication recorded in 2020 and 2021. This growth is largely attributable to the demand for remote research environments during the pandemic and advancements in AI technology, underscoring SDL's emergence as an essential tool for enhancing research efficiency across disciplines. Second, SDL-related publications are highly concentrated in leading countries, such as the United States, Germany, and China, which account for over 85% of the total research output. This indicates that countries with advanced research infrastructure strategically focus on SDL to enhance their global competitiveness and technological leadership. Third, SDL research is characterized by its interdisciplinary nature and integrates fields such as material optimization, biological processes, and AI-based predictive algorithms. The integration of AI has played a pivotal role in automating experiments and improving research efficiency, emphasizing SDL's potential to

drive innovation and expand its application scope across various fields. Fourth, the network analysis highlights the importance of collaborative dynamics in SDL research. Key figures, such as Jesse[7], were identified as central nodes facilitating collaborations and advancing research dissemination. These findings underscore the role of global cooperation in SDL's development and its reliance on interconnected innovation networks. Finally, SDL has been shown to transform scientific research paradigms by combining automation and optimization, positioning itself as a critical methodology in the AI era. This trend reflects SDL's growing indispensability in addressing complex scientific challenges and enhancing productivity across industries and academia.

This study highlights SDL's transformative potential and sustainability in academia, industry, and policy-making. In the industrial sector, the adoption of SDL enables reassessment of R&D priorities and investment strategies. By leveraging SDL, industries can gain strategic insights for adapting rapidly to evolving research trends, thus contributing to sustainable industrial growth. Furthermore, SDL facilitates the identification of emerging technologies and commercialization opportunities, enabling companies to prepare for new product development and market expansion. These capabilities allow businesses to proactively seize market opportunities, respond flexibly to volatility, and enhance global competitiveness. In academia, the SDL trend analysis serves as a foundation for shaping research directions and refining educational programs. Integrating knowledge and technologies aligned with current and future research trends in curricula fosters the development of industry-ready talent. This contributes to sustainable collaboration between academia and industry while enhancing the practical applicability of research outcomes. Moreover, SDL-driven educational advancements strengthen academia-industry partnerships, ensuring a more sustainable and effective translation of research into industrial innovation. From a policy perspective, this study provides critical evidence for the formulation of science and technology policies. The SDL trend analysis offers insights into sustainable strategic investment priorities and regulatory frameworks that bolster national competitiveness in science and technology. Moreover, aligning regulations and support programs with SDL advancements plays a pivotal role in creating an environment conducive to scientific research and innovation at the national level.

This study analyzed 218 papers, and the dataset reflected the early stages of SDL as an emerging technology. Future research should include a more diverse range of publications as SDL adoption becomes more widespread. Further studies should explore the challenges and solutions to SDL implementation in specific industries. Tailored approaches are required to optimize SDL's impact and address the barriers unique to each sector. Moreover, investigating how SDL is integrated with AI to foster cross-disciplinary research can provide valuable insights into its transformative potential. By addressing these challenges, future research can contribute to the successful implementation and broader adoption of SDL in diverse scientific and industrial contexts.

References

1. Garcia Martin, H.; Radivojevic, T.; Zucker, J.; Bouchard, K. E.; Sustarich, J.; Peisert, S.; Arnold, D.; Hillson, N. J.; Babnigg, G.; Martí, J. M.; Mungall, C. J.; Beckham, G. T.; Waldburger, L. M.; Carothers, J. M.; Sundaram, S.; Agarwal, D.; Simmons, B. A.; Backman, T. W. H.; Banerjee, D.; Tanjore, D.; Ramakrishnan, L.; Singh, A. Perspectives for Self-Driving Labs in Synthetic Biology. *Current Opinion in Biotechnology.* **2022**, *79*, 102881.
2. Seifrid, M.; Pollice, R.; Aguilar-Granda, A.; Chan, Z. M.; Hotta, K.; Ser, C. T.; Vestfrid, J.; Wu, T. C.; Aspuru-Guzik, A. Autonomous Chemical Experiments: Challenges and Perspectives on Establishing a Self-Driving Lab. *Accounts of Chemical Research.* **2022**, *55*, 2454–2466.

3. Abolhasani, M.; Kumacheva, E. The Rise of Self-Driving Labs in Chemical and Materials Sciences. *Nature Synthesis.* **2023**, *2*, 483–492.

4. Delgado-Licona, F.; Abolhasani, M. Research Acceleration in Self-Driving Labs: Technological Roadmap toward Accelerated Materials and Molecular Discovery. *Advanced intelligent systems.* **2022**, *5*, 2200331.

5. Da Silva, R.G.L. The advancement of artificial intelligence in biomedical research and health innovation: challenges and opportunities in emerging economies. *Global Health.* **2024**, 20, 44

6. Hysmith, H.; Foadian, E.; Padhy, S. P.; Kalinin, S. V.; Moore, R. G.; Ovchinnikova, O.; Ahmadi, M. The Future of Self-Driving Laboratories: From Human in the Loop Interactive AI to Gamification. *Digital discovery.* **2024**, 3, 621-636.

7. MacLeod, B. P.; Parlane, F. G. L.; Berlinguette, C. P. How to Build an Effective Self-Driving Laboratory. *Mrs Bulletin.* **2023**, *48*, 173–178.

8. Häse, F.; Roch, L. M.; Aspuru-Guzik, A. Next-Generation Experimentation with Self-Driving Laboratories. **2019**, *1*, 282–291.

9. Lo, S.; Baird, S. G.; Schrier, J.; Blaiszik, B. J.; Carson, N.; Foster, I.; Aguilar-Granda, A.; Kalinin, S. V.; Maruyama, B.; Politi, M.; Tran, H.; Sparks, T. D.; Aspuru-Guzik, A. Review of Low-Cost Self-Driving Laboratories in Chemistry and Materials Science: The "Frugal Twin" Concept. *Digital discovery.* **2024**, 3, 842-868.

10. Gutierrez, D. P.; Folkmann, L. M.; Tribukait, H.; Roch, L. M. How to Accelerate R&D and Optimize Experiment Planning with Machine Learning and Data Science. *Chimia.* **2023**, *77*, 7.

11. Mabbott, G. A. Teaching Electronics and Laboratory Automation Using Microcontroller Boards. *Journal of Chemical Education.* **2014,** *91*, 1458–1463.

12. Rapp, J.; Bremer, B. J.; Romero, P. A. Self-Driving Laboratories to Autonomously Navigate the Protein Fitness Landscape. *Nature Chemical Engineering.* **2024**, *1*, 97–107.

13. Friedrich, R.; Block, S.; Alizadehheidari, M.; Heider, S.; Fritzsche, J.; Esbjörner, E. K.; Westerlund, F.; Bally, M.; Bally, M. A Nano Flow Cytometer for Single Lipid Vesicle Analysis. *Lab on a Chip.* **2017**, *17*, 830–841.

14. Ji, W.; Deng, S. Autonomous Discovery of Unknown Reaction Pathways from Data by Chemical Reaction Neural Network. *Journal of Physical Chemistry A.* **2021**, *125*, 1082–1092.

15. Mavoori, J.; Jackson, A.; Diorio, C. J.; Fetz, E. E. An Autonomous Implantable Computer for Neural Recording and Stimulation in Unrestrained Primates. *Journal of Neuroscience Methods.* **2005**, *148*, 71–77.

16. Janet, J. P.; Liu, F.; Nandy, A.; Duan, C.; Yang, T.; Lin, S.; Kulik, H. J. Designing in the Face of Uncertainty: Exploiting Electronic Structure and Machine Learning Models for Discovery in Inorganic Chemistry. *Inorganic Chemistry.* **2019**, *58*, 10592–10606.

17. Volk, A. A.; Epps, R. W.; Yonemoto, D. T.; Masters, B. S.; Castellano, F. N.; Reyes, K. G.; Abolhasani, M. AlphaFlow: Autonomous Discovery and Optimization of Multi-Step Chemistry Using a Self-Driven Fluidic Lab Guided by Reinforcement Learning. *Nature Communications.* **2023**, *14*, 1403.

18. Kavalsky, L.; Hegde, V.; Meredig, B.; Viswanathan, V. A Multiobjective Closed-Loop Approach Towards Autonomous Discovery of Electrocatalysts for Nitrogen Reduction. *Digital discovery.* **2024**, 3, 999-1010.

19. Comina, G.; Suska, A.; Filippini, D. Autonomous Chemical Sensing Interface for Universal Cell Phone Readout. *Angewandte Chemie.* **2015**, *54*, 8708–8712.

20. MacLeod, B. P.; Parlane, F. G. L.; Morrissey, T. D.; Häse, F.; Roch, L. M.; Dettelbach, K. E.; Moreira, R.; Yunker, L. P. E.; Rooney, M. B.; Deeth, J. R.; Lai, V.; Ng, G. J.; Situ, H.; Zhang, R. H.; Elliott, M. S.; Haley, T. H.; Dvorak, D. J.; Aspuru-Guzik, A.; Hein, J. E.; Berlinguette, C. P. Self-Driving Laboratory for Accelerated Discovery of Thin-Film Materials. *Science Advances.* **2020**, *6*, 8867.

21. Bhowmik, A.; Berecibar, M.; Casas-Cabanas, M.; Csányi, G.; Dominko, R.; Hermansson, K.; Palacín, M. R.; Stein, H. S.; Vegge, T. Implications of the BATTERY 2030+ AI-Assisted Toolkit on Future Low-TRL Battery Discoveries and Chemistries. *Advanced Energy Materials.* **2021**, 2102698.

22. Chmielewska-Muciek, D.; Marzec, P.; Jakubczak, J.; Futa, B. Artificial Intelligence and Developments in the Electric Power Industry—A Thematic Analysis of Corporate Communications. *Sustainability.* **2024**, *16*, 6865.

23. Dave, A.; Mitchell, J.; Kandasamy, K.; Wang, H.; Burke, S.; Paria, B.; Póczos, B.; Whitacre, J.; Viswanathan, V. Autonomous Discovery of Battery Electrolytes with Robotic Experimentation and Machine Learning. **2020**, *1*, 100264.

24. Li, J.; Li, J.; Liu, R.; Tu, Y.; Li, Y.; Cheng, J.; He, T.; Zhu, X. Autonomous Discovery of Optically Active Chiral Inorganic Perovskite Nanocrystals through an Intelligent Cloud Lab. *Nature Communications.* **2020**, *11*, 2046.

25. Guo, R.; Sui, F.; Yue, W.; Wang, Z.; Pala, S.; Li, K.; Xu, R.; Lin, L. Deep Learning for Non-Parameterized MEMS Structural Design. *Microsystems & Nanoengineering.* **2022**, *8*, 91.

26. Lin, C.-C., Peng, Y.-C., Chang, Y.-S., & Chang, C.-H. Reentrant hybrid flow shop scheduling with stockers in automated material handling systems using deep reinforcement learning. *Computers & Industrial Engineering.* **2024**, *189*, 109995.

20

27.   Gongora, A. E.; Xu, B.; Perry, W.; Okoye, C.; Riley, P.; Reyes, K. G.; Morgan, E. F.; Brown, K. A. A Bayesian Experimental Autonomous Researcher for Mechanical Design. *Science Advances*. **2020**, *6, 1708*.

28.   Tao, H.; Wu, T.; Kheiri, S.; Aldeghi, M.; Aspuru-Guzik, A.; Kumacheva, E. Self-Driving Platform for Metal Nanoparticle Synthesis: Combining Microfluidics and Machine Learning. *Advanced Functional Materials.* **2021**, 31, 2106725.

29.   Shields, B. J.; Stevens, J. M.; Li, J.; Parasram, M.; Damani, F.; Martinez Alvarado, J. I.; Janey, J. M.; Adams, R. P.; Doyle, A. G. Bayesian Reaction Optimization as a Tool for Chemical Synthesis. *Nature.* **2021**, *590*, 89–96.

30.   Häse, F.; Aldeghi, M.; Hickman, R. J.; Roch, L. M.; Aspuru-Guzik, A. Gryffin: An Algorithm for Bayesian Optimization of Categorical Variables Informed by Expert Knowledge. *Applied physics reviews.* **2021**, *8*, 031406.

31.   Roch, L. M.; Häse, F.; Kreisbeck, C.; Tamayo-Mendoza, T.; Yunker, L. P. E.; Hein, J. E.; Aspuru-Guzik, A. ChemOS: An Orchestration Software to Democratize Autonomous Discovery. *PLOS ONE.* **2020**, *15*, 1–18.

32.   Langner, S.; Häse, F.; Perea, J. D.; Stubhan, T.; Hauch, J.; Roch, L. M.; Heumueller, T.; Aspuru-Guzik, A.; Brabec, C. J.; Brabec, C. J. Beyond Ternary OPV: High-Throughput Experimentation and Self-Driving Laboratories Optimize Multicomponent Systems. *Advanced Materials.* **2020**, *32*, 1907801.

33.   Szymanski, N. J.; Rendy, B.; Fei, Y.; Kumar, R. E.; He, T.; Milsted, D.; McDermott, M. J.; Gallant, M.; Cubuk, E. D.; Merchant, A.; Kim, H.; Jain, A.; Bartel, C. J.; Persson, K.; Zeng, Y.; Ceder, G. An Autonomous Laboratory for the Accelerated Synthesis of Novel Materials. *Nature.* **2023**, *624*, 86–91.

34.   Bai, J.; Cao, L.; Mosbach, S.; Akroyd, J.; Lapkin, A. A.; Kraft, M. From Platform to Knowledge Graph: Evolution of Laboratory Automation. *JACS Au.* **2022**, *2*, 292–309.

35.   Mongeon, P.; Paul-Hus, A. The Journal Coverage of Web of Science and Scopus: A Comparative Analysis. *arXiv: Digital Libraries.* **2015**, *106, 213-228.*

36.   Falagas, M. E.; Pitsouni, E.; Malietzis, G.; Pappas, G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and Weaknesses. *The FASEB Journal.* **2007**, *22*, 338–342.

37.   Shermon, D. Historical Trend Analysis Analysed. *The Journal of Cost Analysis.* **2011**, *4*, 52–62.

38.   David, F, Feldon. The Development of Expertise in Scientific Research. *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource.* **2011**, 1, 14.

39.   Freeman, L. C. Centrality in Social Networks Conceptual Clarification. *Social Networks.* **1978**, *1*, 215–239.

40.   Arsic, B.; Bojić, L.; Milentijevic, I.; Spalević, P.; Rančić, D. Symbols: Software for Social Network Analysis. **2019**, *17*, 205–222.

41.   Wolfe, A. W. Social Network Analysis: Methods and Applications. *American Ethnologist.* **1997**, *24*, 219–220.

42.   Singh, A. Significance of Research Process in Research Work. *Social Science Research Network.* **2021**, *30*, 15.

43.   Lee, S-S. A Content Analysis of Journal Articles Using the Language Network Analysis Methods. *Journal of The Korean Society for Information Management.* **2014**, *31*, 49–68.

44.   Choi, Y. G.; Cho, K. T. Analysis of Safety Management Characteristics Using Network Analysis of CEO Messages in the Construction Industry. *Sustainability.* **2020**, *12*, 5771.

45.   Mohr, J. W.; Bogdanov, P. Introduction—Topic Models: What They Are and Why They Matter. *Poetics.* **2013**, *41*, 545–569.

46.   Jeong, D. H.; Song, M. Time Gap Analysis by the Topic Model-Based Temporal Technique. *Journal of Informetrics.* **2014**, *8*, 776–790.

47.   Blei, D. M.; Ng, A. Y.; Jordan, M. I. Latent Dirichlet Allocation. *Journal of Machine Learning Research.* **2001**, *3*, 993–1022.

48.   Shi, J.; Fan, M.; Li, W.-L. Topic Analysis Based on LDA Model: Topic Analysis Based on LDA Model. *Acta Automatica Sinica.* **2010**, *35*, 1586–1592.

49.   Akhmedov, F.; Abdusalomov, A.; Makhmudov, F.; Cho, Y. I. LDA-Based Topic Modeling Sentiment Analysis Using Topic/Document/Sentence (TDS) Model. *Applied Sciences.* **2021**, *11*, 11091.

50.   Yu, D.; Fang, A.; Xu, Z. Topic Research in Fuzzy Domain: Based on LDA Topic Modelling. *Information Sciences.* **2023**, *648*, 119600.