The Journal of Supercomputing
https://doi.org/10.1007/s11227-024-06247-x

Evaluating the latest trends of Industry 4.0 based on LDA topic model

Ozcan Ozyurt[1] · Hakan Özköse[2] · Ahmet Ayaz[3]

Accepted: 16 May 2024
© The Author(s) 2024

Abstract
This study employs the Latent Dirichlet allocation method, a topic modeling technique, to reveal hidden patterns in Industry 4.0 research. The dataset comprises 8584 articles published in the Scopus database from 2011 to the end of 2022. The analysis categorized the articles into 12 distinct topics. The three most prominent topics identified are "Smart Cyber-Physical Systems," "Digital Transformation and Knowledge Management" and "Data Science in Energy," respectively. The findings from this topic modeling provide a comprehensive overview for researchers in the field of Industry 4.0, offering valuable insights into current trends and potential future research directions.

Keywords Industry 4.0 · Topic modeling · Latent Dirichlet allocation · Trend analysis · Text mining

1 Introduction

Industry 4.0 aims to create smarter, more efficient and flexible production systems through the integration of advanced technologies into the manufacturing and industrial sectors [1]. The term Industry 4.0 was first introduced in Germany in 2011 and has since been adopted worldwide as a concept for the future of manufacturing [2]. Industry 4.0 is characterized by the integration of cyber-physical systems (CPS), the Internet of Things (IoT), big data analytics, cloud computing and artificial intelligence (AI) into industrial processes. These technologies enable the creation of the

✉ Ahmet Ayaz
ahmetayaz@ktu.edu.tr

[1] Department of Software Engineering, Faculty of Technology, Karadeniz Technical University, 61830 Trabzon, Türkiye

[2] Department of Management Information Systems, Faculty of Economics and Administrative Sciences, Bartın University, 74110 Bartın, Türkiye

[3] Digital Transformation Office, Karadeniz Technical University, 61080 Trabzon, Türkiye

Published online: 24 May 2024

smart factory, where machines and processes can communicate with each other in real time, increasing efficiency, productivity and quality [3].

The application areas of Industry 4.0 are vast and include many sectors such as manufacturing, logistics, healthcare, energy and transportation. In the manufacturing sector, Industry 4.0 technologies have been used to automate production processes [4], reduce waste and downtime [5] and enable mass customization [6]. In logistics, Industry 4.0 has been used to optimize supply chain management, reduce costs and increase efficiency [7–9]. In healthcare, Industry 4.0 is being used to improve patient care and outcomes using IoT devices and big data analytics [10, 11].

Industry 4.0 has the potential to transform the manufacturing and industrial sectors by increasing productivity, reducing costs and improving quality through scientific research [12]. Since the emergence of the Industry 4.0 concept in 2011, there has been a significant increase in the number of studies in various research fields, including engineering, computer science and business. This increase in the number of studies in the field of Industry 4.0 has led to the emergence of descriptive research that provides a comprehensive view of the field. The increase in Industry 4.0 research necessitates and emphasizes the importance of conducting descriptive research to provide a comprehensive view of the field. For this reason, various studies such as content analysis, systematic review and bibliometric analysis have been conducted in the literature on Industry 4.0 research [13, 14].

The research is structured as follows: Initially, it discusses the literature related to Industry 4.0 and the significance of the study. Subsequently, the topic modeling method used in the study is described, and information about the topic modeling process is provided. The following section presents and discusses the findings of the study. In conclusion, the paper addresses the outcomes, limitations and recommendations of the research.

1.1 Literature review

Descriptive studies provide valuable insights into the field and are conducted with great frequency. There is a wide range of topics covered, especially in bibliometric literature reviews. A common theme in the research is Industry 4.0 [15–17] and its impact in various fields such as manufacturing [18–20], supply chain management [21], construction [22, 23] and logistics [24]. These reviews aim to provide a comprehensive overview of research trends, key concepts and the development of the field over time. Some of the specific topics addressed include the role of lean principles in Industry 4.0 [25, 26], the impact of Industry 4.0 on sustainability [27, 28] and circular economies [29, 30], machine learning [31] and the application of AI in manufacturing [32] and barriers to the adoption of Industry 4.0 technologies [33]. These bibliometric studies were used to identify research gaps, potential areas of collaboration and inform future research directions.

While bibliometric analyses provide a broad overview of existing research, they often lack insights based on semantic content analysis. They have limitations when it comes to providing a deeper understanding of the literature [34]. Supplement bibliometric analysis with topic modeling analysis can achieve a more

comprehensive analysis. Topic modeling, a machine learning technique, is an effective method to automatically analyze large collections of scientific articles in a systematic way [35–38]. Topic modeling can go beyond bibliometric analysis to reveal themes, research interests and trends in a field more comprehensively [39, 40]. Therefore, topic modeling analysis is a tool for understanding and uncovering the research landscape in any field [41–43].

When the literature is examined, few studies using topic modeling method on Industry 4.0 research draw attention. Jang et al. [44] used LDA and centrality analysis techniques to identify the main themes of Industry 4.0-related internet news in Korea. The article analyzed the impact of Industry 4.0 on the Korean economy and the uses of this technology. Janmaijaya et al. [45] used Latent Dirichlet allocation (LDA) and clustering-based theme identification techniques to identify the main themes in Industry 4.0 literature. The article analyzed the keywords and themes of different studies on Industry 4.0. Mazzei and Ramjattan [46] presented a systematic review of machine learning topics for Industry 4.0 using deep learning-based topic modeling. This systematic review aims to offer an overview by assembling a corpus of 45,783 pertinent papers from Scopus and Web of Science and analyzing it using BERTopic. The paper discussed different aspects of Industry 4.0 and evaluated the potential of using different machine learning techniques to address the use cases of this technology. Previous studies on topic modeling in the context of Industry 4.0 have mainly focused on specific countries, different topic modeling techniques and machine learning for Industry 4.0.

1.2 Importance of the study and the problem statement

Industry 4.0 is a relatively new area of research and therefore there is much to learn about its potential impact. This study of Industry 4.0 and uses an innovative approach of semantic analysis based on topic modeling for a comprehensive review of the field and its work. By examining the trend of research on Industry 4.0, it is possible to identify themes and focal points in the field, understand the approach of different disciplines and identify potential applications. It is also useful to help bridge the gap between academia and industry, gain a global perspective and identify opportunities for international collaboration. Overall, research on the trend of research papers on Industry 4.0 is important to understand the direction of the field, identify key themes and areas for future research and understand the practical applications of this emerging technology. This study aims to uncover trends in Industry 4.0-related articles. To achieve this, the study surveyed the scientific literature on Industry 4.0 using articles indexed in the Scopus database from the past to the present. The study aimed to answer the following research questions in order to describe all studies in detail and to reveal research interests and trends.

RQ1: What are the prominent topics in articles published in the field of Industry 4.0?

RQ2: How do the prominent topics in articles published in the field of Industry 4.0 change over time?

RQ3: What is the relationship between industry 4.0 and related topics? Which topics are most similar or most different?

2 Materials and methods

This study aimed to perform a topic modeling analysis to uncover hidden semantic patterns in large text sets of Industry 4.0-related literature. Topic modeling is a probabilistic method used to uncover hidden semantic patterns, called topics, in a collection of unstructured documents. These topics capture the essence of the semantic structure of documents. The approach is based on the idea that certain words occur more frequently in a document because of their association with a particular topic [47]. Topic modeling reveals these underlying semantic clusters by identifying groups of words that tend to co-occur in a document. This process involves calculating the probability distribution of each topic and the topic distribution per document, as well as the topic assignments per word in each document [48]. Many different topic modeling algorithms are available for text mining and natural language processing research, such as Latent Dirichlet allocation (LDA), hierarchical Latent Dirichlet allocation (HLDA), hierarchical Dirichlet process (HDP), non-negative matrix factorization (NMF), Dirichlet multinomial regression (DMR), dynamic topic model (DTM) and correlated topic model (CTM) [49]. While algorithms such as NMF, CTM and DMR encounter challenges in determining the optimal number of topics through conventional consistency scores, newer models like HDP and HLDA offer automated mechanisms to ascertain the ideal topic count [49]. In contrast, the LDA model permits manual adjustment of the topic quantity, an iterative process that enhances the precision of topic number estimation and their semantic consistency [50]. This flexibility, coupled with robust methods for computing coherence scores, solidifies LDA's position as a preferred choice across various scholarly fields for its efficacy in semantic content analysis of extensive text collections [51]. Therefore, LDA is frequently preferred in many researches and applications [47]. LDA aims to determine the semantic content of a document by analyzing its hidden semantic structures [52]. It assigns words in a document to random variables and clusters them based on a recurrent probability process using a Dirichlet distribution [48]. As an unsupervised learning approach, LDA does not require any labeling or training set, which makes it efficient to analyze large document collections within a given text corpus to reveal semantic patterns [47]. Figure 1 shows the mathematical and graphical representation of LDA.

As depicted in Fig. 1, LDA treats each document as a mixture of topics, where each topic is characterized by a distribution over words. The generative process for each document begins with the selection of topic proportions from a Dirichlet distribution with hyperparameter α. For each word in the document, a topic $Z_{d,n}$ is then sampled from these topic proportions θ_d. Subsequently, a word $W_{d,n}$ is drawn from the chosen topic's word distribution, which is itself sampled from a Dirichlet distribution governed by parameter β. The joint distribution of the topics and words can be formally expressed as $p(\beta_{1:K},\theta_{1:D},Z_{1:D},W_{1:D};\alpha,\eta)$ and factored into the product of the prior and conditional distributions of each variable in the model.

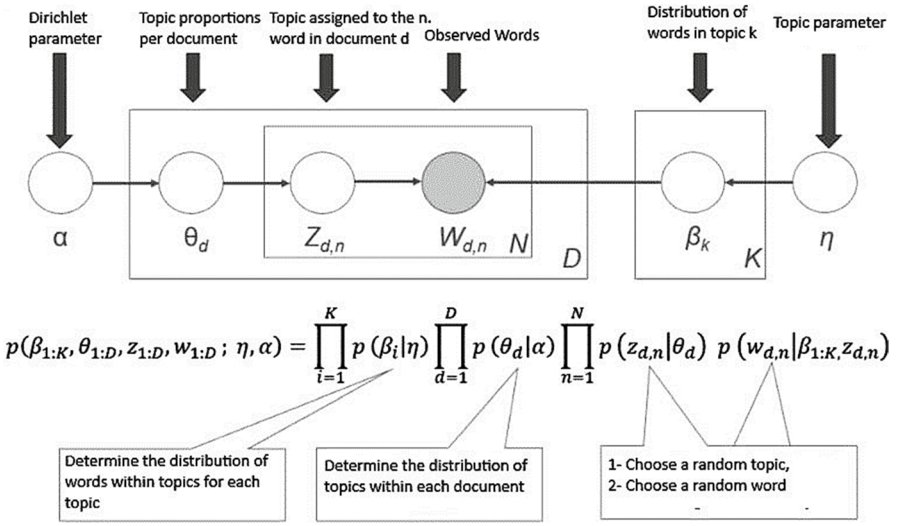


Fig. 1 LDA mathematical and graphical representation [53]

This factorization facilitates the use of inference techniques such as collapsed Gibbs sampling to estimate the hidden topic structure by iteratively sampling the posterior distributions of the topic assignments based on all other current assignments and data observations [47, 48]. This structured approach allows LDA to efficiently manage large corpora by abstracting the main themes as topics represented as distributions over words, thus providing a powerful tool for exploring hidden thematic patterns in text.

2.1 Search strategy and study selection

To collect a comprehensive set of articles relevant to the scope of this study, we used the Scopus database, which includes more than 7000 publishers from around the world [54]. Scopus is the largest abstract and citation database of peer-reviewed literature, including scientific journals, books and conference proceedings. Scopus is widely used in literature review studies [35, 37, 55]. Therefore, the researchers found the Scopus database sufficient for their study. To include the maximum number of articles, Scopus was searched for title, abstract and keywords. The researchers reviewed the literature and consulted two domain experts to create the query.

To maximize the number of articles included in our dataset, we searched for the term "Industry 4.0" in the title, abstract and keywords of journal articles (i.e., research and review articles) written in English and published from the past to the end of 2022. The following query was used as the search strategy:

TITLE-ABS-KEY ("industry 4.0") AND (LIMIT-TO (PUBSTAGE, "final")) AND (LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "re"))

AND (LIMIT-TO (SRCTYPE, "j")) AND (EXCLUDE (PUBYEAR, 2023)) AND (LIMIT-TO (LANGUAGE, "English"))

We performed this query on February 22, 2023, and retrieved a total of 8584 journal articles from Scopus. We added the title, abstract and author keywords of these articles to our dataset for LDA analysis.

2.2 Pre-processing

Datasets are collected from various sources that may have different formats. Therefore, the data must be pre-processed to ensure consistency in the format of the dataset. This involves removing unnecessary information from the collected data. Pre-processing helps to remove noisy, irrelevant and unwanted data from any corpus and thus improves the quality of the dataset. To achieve this, several steps are taken to remove noisy words and characters, resulting in a more accurate and acceptable corpus [56]. The Python NLTK library was used for data preprocessing. Natural language toolkit (NLTK) is a popular Python library for natural language processing tasks [57].

Initially, all texts were converted to lowercase to ensure uniformity. Web links and publisher information within the dataset were removed to eliminate irrelevant metadata. Subsequently, word tokenization was performed to break down the text into individual words.

Following tokenization, we removed English stop words (such as "and", "is", "or", "the", "a", "an", "for", etc.), numerical expressions, punctuations and symbols to reduce noise in the data. Additionally, generic words frequently found in academic texts—such as "article", "paper", "research", "study" and "copyright"—were filtered out to avoid the inclusion of terms that do not contribute to the formation of semantically meaningful topics [58]. Finally, the resulting words need to be lemma. Lemmatization takes the context into account and transforms the lemma words into more meaningful base words. This step aims to remove inflected words and create a lexical form of a word [59].

In order to identify high-frequency terms within the industry 4.0 corpus, an N-gram model at the word level was utilized on the previously stemmed texts. This model considers unigrams (single words), providing a comprehensive view of term co-occurrence patterns within the corpus. Finally, each article in the dataset was transformed into a word vector using the "bag of words" method, which facilitates a numerical representation of the words in the corpus [48]. These vectors were then used to create a document term matrix (DTM), providing a suitable numerical matrix form for conducting the topic modeling analysis [60]. This matrix is pivotal for the subsequent application of statistical models to discern the latent topics within the corpus. All these preprocessing steps were performed, and the resulting corpus was used to create the final corpus for LDA analysis.

2.3 Data analysis and fitting topic modeling

The implementation of the LDA algorithm was carried out through the Gensim library in Python [61]. First, optimal values were chosen for the parameters that enable the optimization of the model. Default values ("Symmetric", "Symmetric") were set for α, which determines the distribution of topics in the documents, and β, which determines the distribution of words in the topics. Symmetric LDA is a variation of the LDA model and assumes that a given distribution of parameters is symmetric. This assumes that each document belongs to each theme with a similar probability. That is, it considers that documents contribute equally to each theme.

An iterative and heuristic process was applied for model fitting [34]. To empirically determine the ideal number of topics (K) in LDA-based topic modeling analysis, a model was created for all K values between 10 and 40, and consistency values were calculated for each K value. The consistency value was taken into account in determining the appropriate number of topics in LDA analysis. The consistency value closest to 0.7 was considered optimal [47]. As a result of the analysis, a model with 12 topics was selected as shown in the number of topics and consistency value graph in Fig. 2.

With topic modeling analysis, topics and the terms that make up the topics were created. LDA's term ranking process starts by first determining the probability that documents belong to topics. The probability that each document belongs to each topic is calculated. Then, the probabilities of the terms contained in each topic are determined. That is, for each topic, the probabilities of terms associated with that topic are calculated. Terms representing topics were generated according to their representation rates and then used to label and name topics [35, 36]. The Python pyLDAvis [62] library was used for these operations. Overall, pyLDAvis is a useful tool to help understand and interpret the results of topic modeling and can

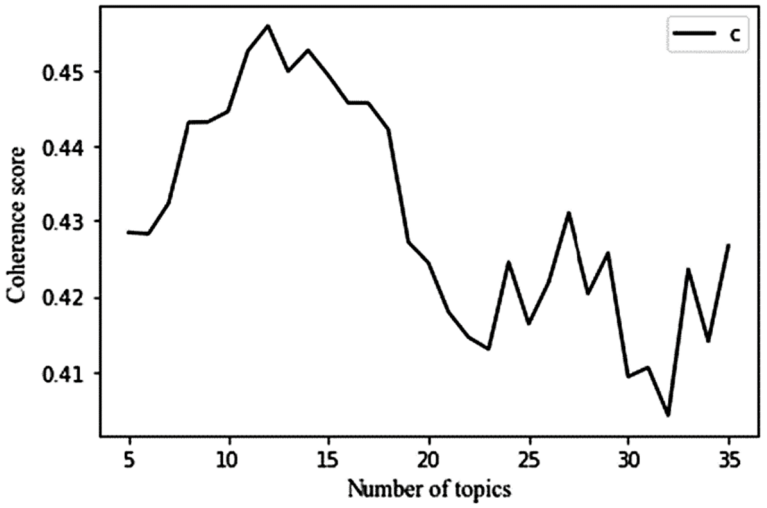


Fig. 2 Number of topics-coherence score graph

help identify patterns and trends in large text datasets. Topics were named by the researchers with input from two domain experts who reviewed and finalized topic names. In addition, the percentage of each topic per document, the word distribution within each topic and the distribution of topics within all articles were calculated. Finally, the top 15 terms with the highest frequency representing each of the 12 topics were identified.

The percentage change and acceleration values presented in this study were calculated using Microsoft Excel. Percent change values on a temporal basis are a metric used to show changes over time on a particular topic or term. These values express the change from one period to another as a percentage. Acceleration values are used to measure changes in the rate of change, which usually refers to the rate of change of rates of change. In Excel, the slope of a data series can be calculated using the SLOPE function, which shows the slope of the linear trend between time series data. The formula is as follows:

=SLOPE(y-values; x-values)

In this study, x-values represent temporal periods (years) and y-values represent quantities of terms or topics of interest in specific time periods. When performing the calculations, it is assumed that the data points are a good fit to a linear model. In this way, the slope of each topic over time was determined, and an Acc value was calculated accordingly. A positive or negative Acc value indicates an increase or decrease in the number of publications in a given topic. Graphs were created to visualize the volume and slope of each topic over time and their relationship with other topics.

3 Results and discussion

3.1 Identifying the topics (RQ1)

In LDA-based topical modeling analysis, researchers conducted experiments on different numbers of topics. Eventually, they decided to use 12 topics in their study. The reason for choosing this number is that when the number of topics is increased too much, low-frequency topics may be formed, leading to moving too far away from the research field. Conversely, reducing the number of topics may result in the risk of overlooking major topics in the research field. Therefore, the researchers determined that 12 topics would be the optimal number for their analysis. After deciding on the number of topics, they obtained the terms related to each topic to be named first and calculated the total frequency percentage of these terms. Then, they consulted with three experts in the field to determine appropriate topic names, resulting in the creation of a Table 1. The related terms for each topic were listed in order of density.

The resulting Table 1 contains the names of the topics, their related terms and the frequency percentages for each term. The related terms for each topic are listed in order of their density, which means that the most frequently occurring terms for each topic appear at the top of the list.

Table 1 Topics, topic terms and frequencies

Topics	Topic terms	f (%)
Smart cyber-physical systems	system | manufacturing | production | smart | industrial | physical | process | cyber | technology | factory | digital | product | model | twin | internet	50,06
Digital transformation and knowledge management	digital | technology | innovation | transformation | management | knowledge | company | business | industrial | future | adoption | technological | process | organization | development	17,75
Data science in energy	energy | learning | machine | data | model | sensor | system | monitoring | maintenance | network | process | deep | neural | condition | detection	15,68
Sustainable supply chain management	manufacturing | chain | supply | technology | management | performance | lean | data | system | company | sustainability | production | logistics | framework | circular	6,71
Internet of things	data | network | iot | industrial | internet | system | thing | application | communication | security | smart | device | iiot | blockchain | cloud	3,24
Manufacturing system design	system | robot | process | assembly | algorithm | design | manufacturing | human | task | problem | reality | solution | machine | optimization | application	3,11
Education 4.0 in engineering	education | learning | student | engineering | construction | skill | university | development | educational | technology | training | teaching | industrial | higher | revolution	1,53
Small and medium-sized enterprises (SMEs)	model | business | smes | enterprise | process | manufacturing | maturity | performance | small | medium | innovation | factor | company | data | sustainable	0,50
Sustainable development	development | technology | economy | industrial | sustainable | country | economic | sector | revolution | social | impact | sustainability | covid | society | health	0,49
Human–robot ınteraction	human | robot | safety | work | industrial | collaborative | job | collaboration | technology | competency | skill | employee | production | development | future	0,37
Quality management	quality | management | factor | system | design | readiness | implication | process | digitalization | data | knowledge | methodology | tqm | industrial | human	0,37
Decision-making process	process | technology | manufacturing | decision | fuzzy | barrier | product | criterion | food | model | production | making | tool | material | quality	0,19

When Table 1 is examined, the "smart cyber-physical systems" topic ranks first with a percentage of 50.06. "Digital Transformation and Knowledge Management" and "Data Science in Energy" are other topics that stand out in terms of intensity. Understanding how these 12 topics relate to related terms is important for the intelligibility of the study. In this respect, information about the relevant titles is given in order of density.

According to Varadarajan et al. [63], "smart cyber-physical systems are large-scale software-intensive and pervasive systems, which by combining various data sources and applying intelligence techniques, can efficiently manage real-world processes and offer a broad range of novel applications and services". Moreover, smart cyber-physical systems include anything smart like cities, manufacturing, production, etc. [64].

Although digital transformation and knowledge management are two separate topics, they have started to be used together with Industry 4.0. Digital transformation can be defined as a process that aims to improve an entity by triggering significant changes to its properties through combinations of information, computing, communication and connectivity technologies [65]. On the other hand, according to the Koenig [66], "knowledge management (KM) is a business process that formalizes the management and use of an enterprise's intellectual assets. KM promotes a collaborative and integrative approach to the creation, capture, organization, access and use of information assets, including the tacit, uncaptured knowledge of people". Currently, some studies use digital transformation and knowledge management together. Some of the topics covered in these studies include the public sector [67], property management [68], the economy [69], impact on organizations [70] and SMEs [71].

Data science is a field that brings together various disciplines such as mathematics, statistics and computer science. Combining the knowledge and skills from these disciplines provides meaningful results that users can understand [72]. Data science studies are used in various fields, including energy, for different improvements. For instance, they can be used to address the energy crisis [73], forecast energy usage [74] and predict load [75]. Naturally, data science methods such as neural networks and deep learning are effectively used in such studies.

Sustainable supply chain management involves integrating environmentally and financially viable practices into the complete supply chain lifecycle, from product design and development to material selection [76].

The Internet of Things (IoT) is a network infrastructure that enables objects to communicate with each other via the internet. This allows objects with specific tasks to exchange information among themselves and with system users [43].

A manufacturing system is the organization and processing of relevant elements to produce a physical product, a service or information [77]. Manufacturing system design, on the other hand, concerns the function, structure and form of workpieces in the manufacturing process [78]. Additionally, it contributes to the creation of equipment selection, job design and standardization [79].

According to Atasoy [80], Education 4.0 can be explained as "the period in which the education settings integrate ICTs to develop instructional, pedagogical and technological processes. It also improves operational processes through new

learning and teaching methods, innovative solutions to current and future challenges in society". Especially in engineering education, Education 4.0 is being effectively used and studies are being conducted [81, 82].

Small and medium-sized enterprises (SMEs) are economic entities defined in regulations as micro-enterprises, small businesses or medium-sized enterprises. Such businesses should have between 10 and 250 employees, and their total turnover should not exceed 50 million euros [83]. SMEs contribute to the progress of entrepreneurship and innovation and prevent monopolization [84].

Sustainable development is the pursuit of progress that satisfies our current requirements while also safeguarding the potential of forthcoming generations to satisfy their own necessities [85]. Sustainable development initiatives are being carried out in various fields, such as health [86, 87], transportation [88, 89], agriculture [90] and energy [91].

Human–robot interaction refers to "interactions between humans and robots by means of cooperation in a task or assisting one another" [92]. Human–robot interaction is interconnected with various scientific fields, including artificial intelligence, robotics and psychology [93].

Quality management involves effectively managing, planning and organizing processes to continuously improve them while maximizing customer satisfaction with the lowest total cost [94].

Forgionne & Russell [95] described the decision-making process as "the process of developing a general problem understanding, formulating the problem explicitly, evaluating alternatives systematically and implementing the choice".

3.2 Examining the topics by dividing them into periods (RQ1)

To show how topics changed over time, data were divided into five periods, each consisting of two years. The reason for not taking data on an annual basis is that the amount of data in the early years is low. Table 2 shows the distribution of themes according to the relevant periods.

Upon examining the Table 2, "Smart Cyber-Physical Systems" is the topic with the most studies in all periods. Fifty percent of the studies were conducted under this title. "Digital Transformation and Knowledge Management" and "Data Science in Energy" are the most studied topics after "Smart Cyber-Physical Systems". The interest in the topics of "Digital Transformation and Knowledge Management", "Data Science in Energy" and "Sustainable Supply Chain Management" has increased with the 2019–2020 period. Likewise, while there was an increase in interest in the "Education 4.0 in Engineering" title in 2019–2020, a slight decrease in interest was observed in the 2021–2022 period. While there were no studies on "Sustainable Supply Chain Management", "Internet of Things" and "Manufacturing System Design" in the first period, studies were conducted on these topics from the 2015–2016 period onwards. In addition, while there were no studies on "Education 4.0 in Engineering", "Small and Medium-sized Enterprises (SMEs)", "Sustainable Development", "Human–robot Interaction", "Quality Management" and

Table 2 Distribution of topics according to periods

Topics	2013–2014	2015–2016	2017–2018	2019–2020	2021–2022	Total	Acc
Smart cyber-physical systems	14	130	591	1449	2113	4297	551,7
Digital transformation and knowledge management	1	7	98	480	938	1524	234,7
Data science in energy	1	10	69	397	869	1346	212,3
Sustainable supply chain management	0	5	38	173	360	576	88,8
Internet of things	0	7	21	91	159	278	40,2
Manufacturing system design	0	1	21	92	153	267	39,7
Education 4.0 in engineering	0	0	5	67	59	131	18,5
Small and medium-sized enterprises (SMEs)	0	0	4	16	23	43	6,2
Sustainable development	0	0	2	15	25	42	6,5
Human–robot ınteraction	0	0	4	11	17	32	4,5
Quality management	0	0	1	9	22	32	5,3
Decision-making process	0	0	2	7	7	16	2,1
Total	16	160	856	2807	4745	8584	

"Decision-making Process" in the 2013–2014 and 2015–2016 periods, studies on these topics have also been conducted since 2017–2018 period.

On the other hand, the slope value in the table represents the average annual growth of each subject. This value shows how much the number of publications of a topic has increased annually over a certain period. In a way, it can be used to measure the value of interest. For example, the slope value of the topic "Intelligent Cyber-Physical Systems" is "551.7". This shows that the subject is becoming increasingly popular and has aroused serious interest among researchers in the field of Industry 4.0.

On the other hand, the slope value of the Decision-making Process topic is "2.1'" This shows that the annual publication increase is quite low. This may indicate that the topic is less researched than other topics and receives less attention among Industry 4.0 researchers.

In general, slope values can be used as an indicator to evaluate the research trend of each topic in the field of Industry 4.0. Higher slope values may indicate topics that are growing faster and are considered more important; low slope values may reflect less important or popular topics.

In this context, when the relevant table is examined, it is seen that the topics with the highest slope values are Smart Cyber-Physical Systems, Digital Transformation and Knowledge Management and Data Science in Energy.

3.3 Examining the percentage changes of the topics according to the periods (RQ2)

In this part of the study, the aim was to illustrate how each topic changed over time. To achieve this, the percentage of work done on each topic in each period was presented. For example, the work done on "Smart Cyber-Physical Systems" accounted for 0.33% of the total work in the 2013–2014 period, 3.03% in the 2015–2016 period, 13.75% in the 2017–2018 period, 33.72% in the 2019–2020 period and the remaining 49.71% in the 2021–2022 period. This way, we can see how the work on each topic developed in each period. Finally, acceleration values are shown in Table 3 to provide an overall picture of how each topic evolved over time.

According to the Table 3, all topics have shown substantial development in the 2021–2022 period. Only the topic of "Education 4.0 in Engineering" appears to have been more concentrated in the 2019–2020 period. Moreover, the percentage of work on the "Decision-making Process" was equal in both the 2019–2020 and 2021–2022 periods. The volume ratios given in Table 3 are visualized in Fig. 3.

Figure 3 shows that there has been a significant increase in the number of article publications between 2013 and 2022, especially between 2019 and 2022. The topics with the highest increase include Smart Cyber-Physical Systems, Decision-making Process, Human–robot Interaction, Education 4.0 in Engineering, Small and Medium-sized Enterprises (SMEs), Internet of Things, Manufacturing System Design, Digital Transformation and Knowledge Management, Sustainable Supply

Table 3 Percentage change and acceleration values of the topics on a temporal basis

Topics	2013–2014 (%)	2015–2016 (%)	2017–2018 (%)	2019–2020 (%)	2021–2022 (%)	Acc
Smart cyber-physical systems	0,33	3,03	13,75	33,72	49,17	12,84
Decision-making process	0,00	0,00	12,50	43,75	43,75	13,13
Human–robot ınteraction	0,00	0,00	12,50	34,38	53,13	14,06
Education 4.0 in engineering	0,00	0,00	3,82	51,15	45,04	14,12
Small and medium-sized enterprises (SMEs)	0,00	0,00	9,30	37,21	53,49	14,42
Internet of things	0,00	2,52	7,55	32,73	57,19	14,46
Manufacturing system design	0,00	0,37	7,87	34,46	57,30	14,87
Digital transformation and knowledge management	0,07	0,46	6,43	31,50	61,55	15,40
Sustainable supply chain management	0,00	0,87	6,60	30,03	62,50	15,42
Sustainable development	0,00	0,00	4,76	35,71	59,52	15,48
Data science in energy	0,07	0,74	5,13	29,49	64,56	15,77
Quality management	0,00	0,00	3,13	28,13	68,75	16,56

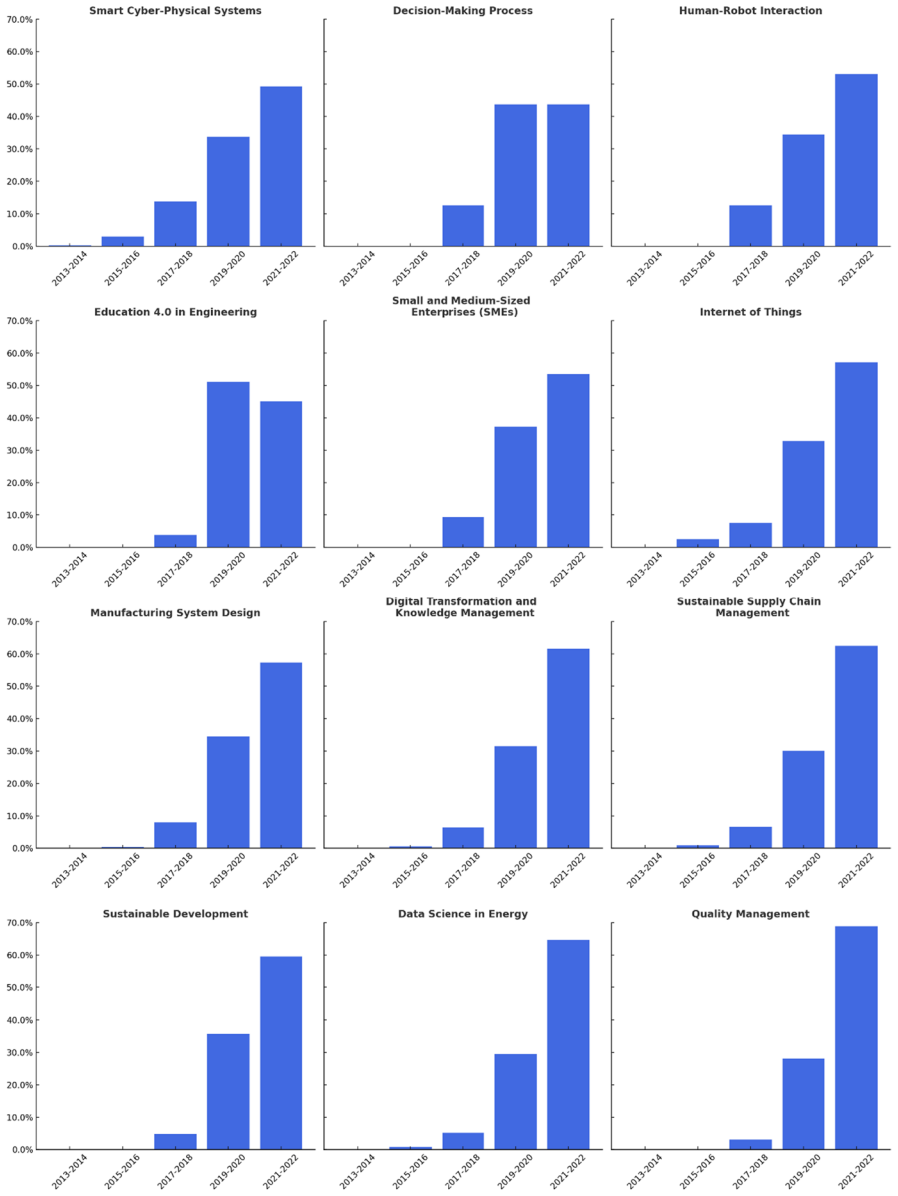


Fig. 3 Percentage change values of the topics

Chain Management, Sustainable Development, Data Science in Energy and Quality Management.

When the acceleration values are examined, "Quality Management", "Data Science in Energy" and "Sustainable Development" stand out as the topics with the highest acceleration, respectively. Studies on these topics will cover more

ground over time, and scientific research may be more directed toward these topics. On the other hand, the interest in "Smart Cyber-Physical Systems", "Decision-making Processes" and "Human–robot Interaction" may decrease compared to other topics. It is also worth noting that there are not very significant differences between the acceleration values. In this context, visuals related to the three titles with the highest and lowest acceleration values are given below to increase understanding.

As can be seen in Fig. 4, there are not very significant differences in the acceleration values of the titles. The field is open to development in all topics.

3.4 Examining the change of topics relative to each other depending on the periods (RQ2)

In this phase of the study, the researchers analyzed how the different topics related to each other and determined the percentage of research that focused on each topic annually. For example, during the 2021–2022 period, the breakdown was as follows: 44.53% of the studies were on "Smart Cyber-Physical Systems", 0.15% on "Decision-making Process", 0.36% on "Human–robot Interaction", 0.46% on "Quality Management", 0.48% on "Small and Medium-sized Enterprises (SMEs)", 0.53% on "Sustainable Development", 1.24% on "Education 4.0", 3.35% on "Internet of Things", 3.22% on "Design and Manufacturing Systems", 7.59% on "Sustainable Supply Chain Management", 18.31% on "Data Science in

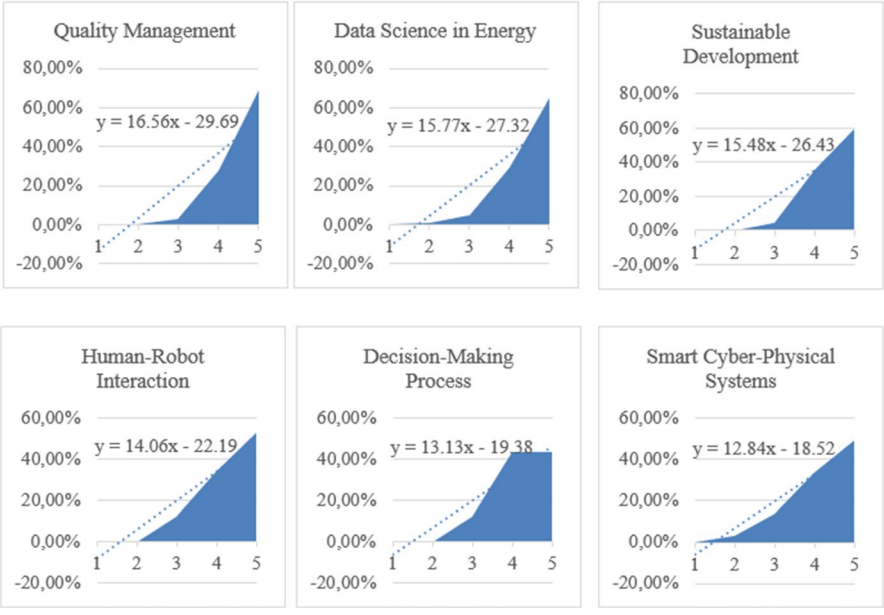


Fig. 4 Headings with the highest and lowest acceleration values

Energy" and 19.77% on "Digital Transformation and Knowledge Management". The relevant data are provided in detail in Table 4, allowing for a more effective analysis of changes between topics.

When examining Table 4, the total coverage of "Smart Cyber-Physical Systems" in terms of publications has gradually decreased over the relevant periods. While 87.50% of the total studies were conducted in this area in the 2013–2014 period, this rate dropped to 44.53% in the 2021–2022 period. The main reason for this is the increasing number of studies conducted on other topics. On the other hand, the areas covered by the topics "Data Science in Energy", "Digital Transformation and Knowledge Management" and "Sustainable Supply Chain Management" are gradually increasing. The volumes of topic according to the periods given in Table 4 are visualized in Fig. 5. There are some notable trends in article publications between 2013 and 2022. In the field of Smart Cyber-Physical Systems, there was a high start in the period 2013–2014, but decreased in the following years. In contrast, areas such as Sustainable Supply Chain Management and Digital Transformation and Knowledge Management saw significant increases in 2019–2022. Similarly, Data Science in Energy experienced a significant increase in the 2019–2022 period. In general, there is a growing interest in digital transformation, knowledge management and sustainability. To increase clarity, Fig. 6 showing the change in momentum values of the topics is provided.

As shown in Fig. 6, the acceleration value is an important indicator to show the changes in the field. Thus, topics that may become more prominent or less effective can be identified. When considering the acceleration values, only "Smart Cyber-Physical Systems" has a negative acceleration value, while all other topics have positive acceleration values. This means that the impact of studies on "Smart Cyber-Physical Systems" will decrease over time, while the impact of other topics will increase. We can especially say that the three topics with the highest acceleration values will begin to take up more space in Industry 4.0 studies. Other topics may not be as strong, but they will continue to progress.

3.5 LDAvis (RQ3)

LDAvis was developed by Carson Sievert and Kenneth E. Shirley in 2014. "LDAvis, visualization system allows users to flexibly explore topic-term relationships using relevance to better understand a fitted LDA model" [96].

The analysis consists of two parts: The first part, which is the left side of the screen, provides a general view that answers the questions "How prevalent is each topic?" and "How do the topics relate to each other?" In this view, topics are represented by circles. The centers of the circles are calculated based on the distances between the topics and placed on a two-dimensional graph. The size of the circles indicates the density of the amount of work related to the topic. In other words, as the circle grows, the density of the amount of work related to the topic increases, while it decreases as the circle shrinks. The second part displays representations of the terms that will be used to determine what the topics are. This reveals the meanings of the topics. In the previous sections of the study, the topics were determined

Table 4 Volumes of topics by period

Topics	2013–2014 (%)	2015–2016 (%)	2017–2018 (%)	2019–2020 (%)	2021–2022 (%)	Acc
Smart cyber-physical systems	87,50	81,25	69,04	51,62	44,53	− 11,56
Decision-making process	0,00	0,00	0,23	0,25	0,15	0,05
Human–robot ınteraction	0,00	0,00	0,47	0,39	0,36	0,11
Quality management	0,00	0,00	0,12	0,32	0,46	0,12
Small and medium-sized enterprises (SMEs)	0,00	0,00	0,47	0,57	0,48	0,15
Sustainable development	0,00	0,00	0,23	0,53	0,53	0,16
Education 4.0 in engineering	0,00	0,00	0,58	2,39	1,24	0,49
Internet of things	0,00	4,38	2,45	3,24	3,35	0,56
Manufacturing system design	0,00	0,63	2,45	3,28	3,22	0,91
Sustainable supply chain management	0,00	3,13	4,44	6,16	7,59	1,82
Data science in energy	6,25	6,25	8,06	14,14	18,31	3,20
Digital transformation and knowledge management	6,25	4,38	11,45	17,10	19,77	3,98

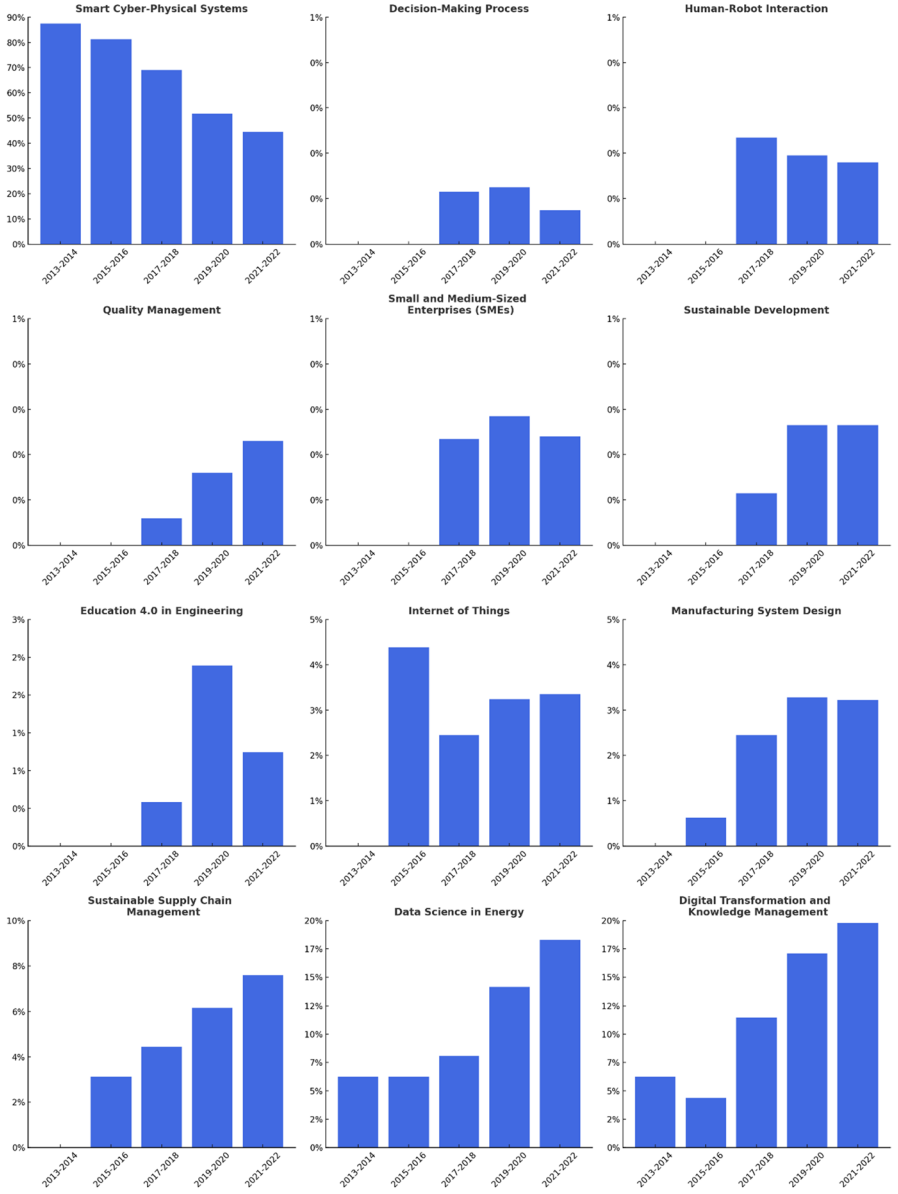


Fig. 5 The volumes of topic according to the periods

using these terms. In this part, the density of the terms within the relevant topic is also displayed by stacking them on top of each other [96].

The relevance metric value in LDA is responsible for modifying the significance assigned to a word's probability of being associated with a particular topic, as noted by Sievert and Shirley [96]. In this study, the lambda value of the relevance metric

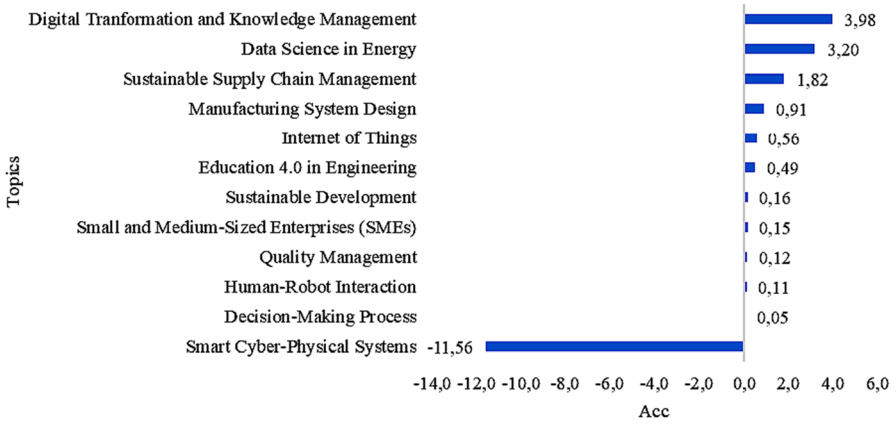


Fig. 6 The acceleration values of the topics

was set to 0.6. This was done to better assist researchers in interpreting the topics. Shrader et al. [97] stated that if λ equals 1 in a given context, then any word that emerges in the corpus is deemed relevant, irrespective of the number of topics it relates to or where it appears. The visual below displays the result of LDAvis. The left side of the visual shows the topics, while the right side displays information on the 30 most important terms in the field.

The titles corresponding to the numbers in Fig. 7 are as follows: (1) "Smart Cyber-Physical Systems", (2) "Internet of Things", (3) "Digital Transformation and

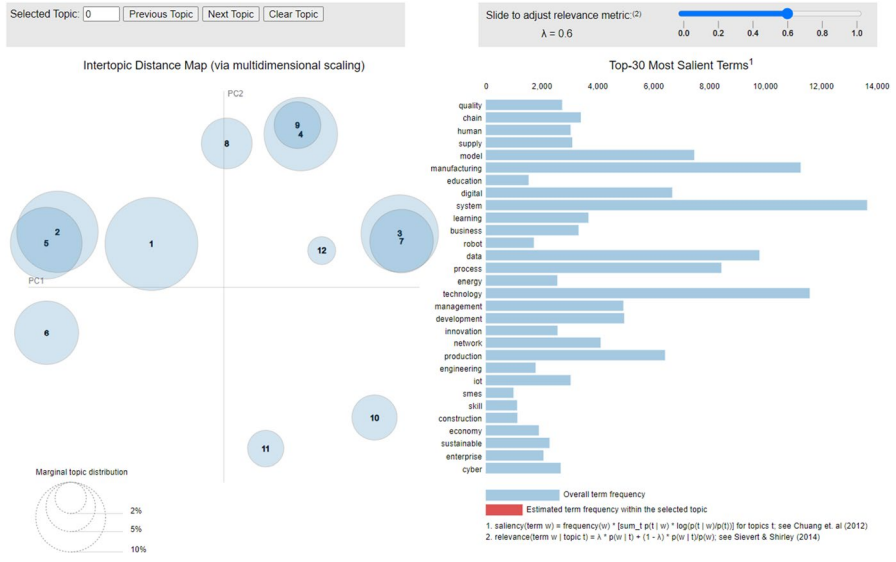


Fig. 7 The layout of LDAvis (intertopic distance map and top-30 most salient terms)

Knowledge Management", (4) "Sustainable Supply Chain Management", (5) "Data Science in Energy", (6) "Manufacturing System Design", (7) "Sustainable Development", (8) "Decision-making Process", (9) "Small and Medium-sized Enterprises (SMEs)", (10) "Education 4.0 in Engineering", (11) "Human–robot Interaction" and (12) "Quality Management". The prevalence of the topics can be seen in the visualization. The size of the circles is used to indicate the prevalence of the titles. "Marginal Topic Distribution" helps us here. "Marginal Topic Distribution" shows which words or terms the LDA model uses for a particular topic and how much weight these terms have in total. According to this information, topics which are numbered as 1, 2, 3, 4 and 5 can be considered as prominent topics in the field of industry 4.0.

The relationships between the titles can also be understood from the left side of the visual. It can be observed that (2) "Internet of Things" and (5) "Data Science in Energy" appear to overlap in PC1 negative and PC2 positive regions. When two topics overlap in the negative PC1 and positive PC2, it suggests that they possess opposing features. These topics can potentially represent distinct subjects with unique vocabularies and terminologies in the LDA model. On the other hand, (4) "Sustainable Supply Chain Management" and (9) "Small and Medium-sized Enterprises (SMEs)", and (3) "Digital Transformation and Knowledge Management" and (7) "Sustainable Development" overlap in the positive side of PC1 and PC2. If there are two topics overlapping in PC1 positive and PC2 positive, it may indicate that these topics exhibit similar characteristics and are closely related to each other. It means that regions on the intertopic distance map where similar topics are located indicate that the LDA model has learned specific topics with similar words and terms, representing different aspects of the same topic within a particular field. On the other hand, the distance between topics on the intertopic distance map reflects the dissimilarity between topics learned by the LDA model, with different terms or words representing distinct aspects of the same topic. These terms or words capture the multiple dimensions of a specific topic and help distinguish it from other topics on the intertopic distance map.

According to the analysis of the most frequent terms in the field, the top 10 terms are "system", "technology", "manufacturing", "data", "process", "model", "digital", "production", "development" and "management". These terms have been used to name the topics.

4 Conclusions

4.1 Implications of this study

The topic modeling analysis identified "Smart Cyber-Physical Systems", "Digital Transformation and Knowledge Management" and "Data Science in Energy" as the most extensively studied topics with an increasing number of publications over time.

The percentage distribution of the topics according to years and the development acceleration of these topics over time were analyzed. According to the results of the research, " Smart Cyber-Physical Systems" is the most studied topic in all periods, while interest in "Data Science in Energy", "Digital Transformation and Knowledge

Management" and "Sustainable Supply Chain Management" has increased in recent years. "Quality Management", "Data Science in Energy" and "Sustainable Development" have been identified as having the highest momentum values and are expected to attract more attention in the coming years. However, as there are not large differences between the momentum values, it can be concluded that the field has room for improvement in all topics.

To show how the topics have changed over time, the data are divided into five periods of 2 years each. With the 2019–2020 period, interest in "Digital Transformation and Knowledge Management", "Data Science in Energy" and "Sustainable Supply Chain Management" increased. While there were no studies on "Sustainable Supply Chain Management", "Internet of Things" and "Production System Design" in the first semesters, studies on these topics have started to be conducted since 2015–2016 semester.

Furthermore, analysis of the topics with the highest increase in research intensity compared to other topics reveals that "Digital Transformation and Knowledge Management", "Data Science in Energy" and "Sustainable Supply Chain Management" are the top three topics. It is worth monitoring these topics in the coming years to follow their growth and development.

With the continuous development and evolution of technology, industries are undergoing constant change and seeking to leverage these technological advancements. Naturally, they are undergoing a continual transformation process toward greater intelligence, efficiency and sustainability. As new technologies such as the internet of things, artificial intelligence, cloud computing, big data, augmented reality and virtual reality are being adopted, industrial processes operate with increased efficiency. Additionally, areas such as sustainability and data management have now become more crucial for industries. This dynamic structure brought about by technological progress presents significant opportunities for both society and factories. However, its potential to create challenges in the future, especially for members of society, should not be overlooked.

In addition, machine learning is one of the areas with high potential in industrial processes. It seems that it has become necessary to make improvements in many issues such as extracting meaningful information from industrial data, optimizing production processes, error control and quality control. In particular, machine learning techniques, together with the digitalization and automation processes that lie at the core of Industry 4.0, can enable the development of smarter and adaptive production systems in the industry and significantly increase efficiency. Additionally, it can help businesses stand out in the industry by providing a competitive advantage. Therefore, it is recommended to focus more on machine learning technologies for the successful adoption and implementation of Industry 4.0.

4.2 Limitations and future works

This study aimed to examine the Industry 4.0 field from 2013 to the end of 2022. Using topic modeling analysis, the study identified the research interests and trends of the field and revealed its current development. The innovative aspect of the study

is the use of topic modeling analysis, which takes bibliometric studies one step further. However, the study also has some limitations: First, the dataset is limited to peer-reviewed articles from the Scopus database, which excludes other databases and document types. Future work could address this limitation by expanding the corpus to include a wider range of databases and document types. Second, to ensure the quality of the research, it only considered journal articles, which may have excluded other relevant document types. Future studies could include a wider range of document types for a more comprehensive analysis. Third, the study used the widely accepted LDA algorithm for topic modeling analysis, but future comparative studies with different algorithms could provide additional insights. Fourth, periodic replication of such studies on the Industry 4.0 field's most voluminous or fastest accelerating topics or subtopics could provide valuable insights into how trends change over time. Finally, such insights could be crucial for directing future research toward areas that could potentially yield high impacts, such as specific technological integrations, policy implications or industry applications that are under-researched. This approach not only helps in advancing the understanding of Industry 4.0 but also aligns academic and practical efforts to address the most pressing and relevant issues as this industrial revolution evolves. Additionally, programs such as Biblioshiny, Vosviewer, SciMat and CiteSpace can be used to uncover significant topic headings in various fields. In our next study, we plan to conduct a comparative analysis of prominent topics in Industry 4.0 using these software packages.

Author contributions All authors have contributed equally to this work.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK).

Declarations

Conflict of interest We have no conflicts of interest to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Lasi H, Fettke P, Kemper HG, Hoffmann M (2014) Industry 4.0. Bus Inf Syst Eng 6:239–242. https://doi.org/10.1007/s12599-014-0334-4
2. Lu Y (2017) Industry 4.0: a survey on technologies, applications and open research issues. J Ind Inf Integr 6:1–10. https://doi.org/10.1016/j.jii.2017.04.005

3. Rojko A (2017) Industry 4.0 concept: Background and overview. Int J Interact Mob Technol 11:77–90. https://doi.org/10.3991/ijim.v11i5.7072
4. Biffl S, Musil J, Musil A, Meixner, K, Lüder, A, Rinker, F, Weyns, D, Winkler, D (2021) An ındustry 4.0 asset-based coordination artifact for production systems engineering. In: 2021 IEEE 23rd Conference on Business Informatics (CBI). IEEE, pp 92–101
5. Bányai T, Tamás P, Illés B, Stankevičiūtė Ž, Bányai Á (2019) Optimization of municipal waste collection routing: impact of industry 4.0 technologies on environmental awareness and sustainability. Int J Environ Res Public Health 16:1–26. https://doi.org/10.3390/ijerph16040634
6. Wang Y, Ma HS, Yang JH, Wang KS (2017) Industry 4.0: a way from mass customization to mass personalization production. Adv Manuf 5:311–320. https://doi.org/10.1007/s40436-017-0204-7
7. Kanimozhi Suguna S, Nanda Kumar S (2019) Application of cloud computing and internet of things to improve supply chain processes. In: Al-Turjman F (ed) Edge Computing. Springer, Cham, pp 145–170
8. Witkowski K (2017) Internet of things, big data, industry 4.0 - innovative solutions in logistics and supply Chains management. Procedia Eng 182:763–769. https://doi.org/10.1016/j.proeng.2017.03.197
9. Xie Y, Yin Y, Xue W, Shi H, Chong D (2020) Intelligent supply chain performance measurement in industry 4.0. Syst Res Behav Sci 37:711–718. https://doi.org/10.1002/sres.2712
10. Aceto G, Persico V, Pescapé A (2020) Industry 4.0 and health: internet of things, big data, and cloud computing for healthcare 4.0. J Ind Inf Integr 18:1–13. https://doi.org/10.1016/j.jii.2020.100129
11. Elhoseny M, Abdelaziz A, Salama AS, Riad AM, Muhammad K, Sangaiah AK (2018) A hybrid model of Internet of things and cloud computing to manage big data in health services applications. Futur Gener Comput Syst 86:1383–1394. https://doi.org/10.1016/j.future.2018.03.005
12. Oztemel E, Gursev S (2020) Literature review of industry 4.0 and related technologies. J Intell Manuf 31:127–182. https://doi.org/10.1007/s10845-018-1433-8
13. Zurita G, Shukla AK, Pino JA, Merigó JM, Lobos-Ossandón V, Muhuri PK (2020) A bibliometric overview of the Journal of network and computer applications between 1997 and 2019. J Netw Comput Appl 165:1–21. https://doi.org/10.1016/j.jnca.2020.102695
14. Kumar R, Rani S, Al AM (2022) Exploring the application sphere of the internet of things in industry 4.0: a review. Bibliometr Content Anal Sens 22:1–35. https://doi.org/10.3390/s22114276
15. Mariani M, Borghi M (2019) Industry 4.0: a bibliometric review of its managerial intellectual structure and potential evolution in the service industries. Technol Forecast Soc Change 149:119752. https://doi.org/10.1016/j.techfore.2019.119752
16. Muhuri PK, Shukla AK, Abraham A (2019) Industry 4.0: a bibliometric analysis and detailed overview. Eng Appl Artif Intell 78:218–235. https://doi.org/10.1016/j.engappai.2018.11.007
17. Liao Y, Deschamps F, Loures ED, Ramos LF (2017) Past, present and future of Industry 4.0: a systematic literature review and research agenda proposal. Int J Prod Res 55:3609–3629. https://doi.org/10.1080/00207543.2017.1308576
18. Ante L (2021) Digital twin technology for smart manufacturing and industry 4.0: a bibliometric analysis of the intellectual structure of the research discourse. Manuf Lett 27:96–102. https://doi.org/10.1016/j.mfglet.2021.01.003
19. Moiceanu G, Paraschiv G (2022) Digital twin and smart manufacturing in industries: a bibliometric analysis with a focus on industry 4.0. Sensors 22:1–22. https://doi.org/10.3390/s22041388
20. Teixeira JE, Tavares-Lehmann AT (2023) Industry 4.0: the future of manufacturing from the perspective of business and economics—a bibliometric literature review. Compet Rev An Int Bus J 33:458–482. https://doi.org/10.1108/CR-07-2022-0091
21. Majiwala H, Kant R (2023) A bibliometric review of a decade' research on industry 4.0 and supply chain management. Mater Today Proc 72:824–833. https://doi.org/10.1016/j.matpr.2022.09.058
22. Newman C, Edwards D, Martek I, Lai J, Thwala WD, Rillie I (2021) Industry 4.0 deployment in the construction industry: a bibliometric literature review and UK-based case study. Smart Sustain Built Environ 10:557–580. https://doi.org/10.1108/SASBE-02-2020-0016
23. Zabidin NS, Belayutham S, Ibrahim CKIC (2020) A bibliometric and scientometric mapping of Industry 4.0 in construction. J Inf Technol Constr 25:287–307. https://doi.org/10.36680/j.itcon.2020.017
24. Bigliardi B, Casella G, Bottani E (2021) Industry 4.0 in the logistics field: a bibliometric analysis. IET Collab Intell Manuf 3:4–12. https://doi.org/10.1049/cim2.12015

25. Alsadi J, Antony J, Mezher T, Maalouf M (2023) Lean and Industry 4.0: a bibliometric analysis, opportunities for future research directions. Qual Manag J 30:41–63. https://doi.org/10.1080/10686967.2022.2144785
26. Nedjwa E, Bertrand R, Sassi Boudemagh S (2022) Impacts of industry 4.0 technologies on lean management tools: a bibliometric analysis. Int J Interact Des Manuf 16:135–150. https://doi.org/10.1007/s12008-021-00795-9
27. Ejsmont K, Gladysz B, Kluczek A (2020) Impact of industry 4.0 on sustainability-bibliometric literature review. Sustain 12:1–29. https://doi.org/10.3390/su12145650
28. Ghobakhloo M, Fathi M, Iranmanesh M, Maroufkhani P, Morales ME (2021) Industry 4.0 ten years on: a bibliometric and systematic review of concepts, sustainability value drivers, and success determinants. J Clean Prod 302:1–20. https://doi.org/10.1016/j.jclepro.2021.127052
29. de Oliveira Neto GC, da Conceição SA, Filho MG (2022) How can Industry 4.0 technologies and circular economy help companies and researchers collaborate and accelerate the transition to strong sustainability? A bibliometric review and a systematic literature review. Int J Environ Sci Technol 20:3483–3520. https://doi.org/10.1007/s13762-022-04234-4
30. Hettiarachchi BD, Seuring S, Brandenburg M (2022) Industry 4.0-driven operations and supply chains for the circular economy: a bibliometric analysis. Oper Manag Res 15:858–878. https://doi.org/10.1007/s12063-022-00275-7
31. Yadav A, Garg RK, Sachdeva AK (2022) Application of machine learning for sustainability in manufacturing supply chain ındustry 4.0 perspective: A bibliometric based review for future research. In: 2022 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM). IEEE, pp 1427–1431
32. Mateo FW, Redchuk A (2022) Artificial intelligence as a process optimization driver under industry 4.0 framework and the role of IIoT, a bibliometric analysis. J Ind Integr Manag. https://doi.org/10.1142/S2424862222500130
33. Deepthi B, Bansal V (2022) Barriers to Industry 4.0 adoption: mapping the global research trends using bibliometric analysis. Int J Manag Enterp Dev 21:344–372. https://doi.org/10.1504/IJMED.2022.126563
34. Gurcan F, Ozyurt O, Cagiltay NE (2021) Investigation of emerging trends in the e-learning field using latent dirichlet allocation. Int Rev Res Open Distance Learn 22:1–18. https://doi.org/10.19173/irrodl.v22i2.5358
35. Özköse H, Ozyurt O, Ayaz A (2023) Management information systems research: a topic modeling based bibliometric analysis. J Comput Inf Syst 63:1166–1182. https://doi.org/10.1080/08874417.2022.2132429
36. Ozyurt O, Ayaz A (2022) Twenty-five years of education and information technologies: insights from a topic modeling based bibliometric analysis. Educ Inf Technol 27:11025–11054. https://doi.org/10.1007/s10639-022-11071-y
37. Ozyurt O, Ayaz A (2024) Identifying cyber security competencies and skills from online job advertisements through topic modeling. Secur J. https://doi.org/10.1057/s41284-024-00420-w
38. Ayaz A, Ozyurt O, Al-Rahmi WM et al (2023) Exploring gamification research trends using topic modeling. IEEE Access 11:119676–119692. https://doi.org/10.1109/ACCESS.2023.3326444
39. Ozyurt O, Gurcan F, Dalveren GGM, Derawi M (2022) Career in cloud computing: exploratory analysis of in-demand competency areas and skill sets. Appl Sci 12:1–15
40. Ozyurt O, Ozyurt H (2022) A large-scale study based on topic modeling to determine the research interests and trends on computational thinking. Educ Inf Technol 28:3557–3579. https://doi.org/10.1007/s10639-022-11325-9
41. Hu Y, Boyd-Graber J, Satinoff B, Smith A (2014) Interactive topic modeling. Mach Learn 95:423–469. https://doi.org/10.1007/s10994-013-5413-0
42. Kang J, Kim S, Roh S (2019) A topic modeling analysis for online news article comments on nurses' workplace bullying. J Korean Acad Nurs 49:736–747. https://doi.org/10.4040/jkan.2019.49.6.736
43. Özköse H (2023) Bibliometric analysis and scientific mapping of IoT. J Comput Inf Syst 63:1438–1459. https://doi.org/10.1080/08874417.2023.2167135
44. Jang KA, Park M, Kim WJ (2019) Topic analysis for the fourth industrial revolution using LDA and centrality analysis in Korea. In: Proceedings—International Conference on Machine Learning and Data Engineering, iCMLDE 2018. pp 108–116
45. Janmaijaya M, Shukla AK, Muhuri PK, Abraham A (2021) Industry 4.0: Latent Dirichlet allocation and clustering based theme identification of bibliography. Eng Appl Artif Intell 103:1–22. https://doi.org/10.1016/j.engappai.2021.104280

46. Mazzei D, Ramjattan R (2022) Machine learning for industry 4.0: a systematic review using deep learning-based topic modelling. Sensors 22:1–26. https://doi.org/10.3390/s22228641
47. Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. J Mach Learn Res 3:993–1022. https://doi.org/10.1016/b978-0-12-411519-4.00006-9
48. Blei DM (2012) Probabilistic topic models. Commun ACM 55:77–84. https://doi.org/10.1145/2133806.2133826
49. Vayansky I, Kumar SAP (2020) A review of topic modeling methods. Inf Syst 94:1–15. https://doi.org/10.1016/j.is.2020.101582
50. Gurcan F, Cagiltay NE (2022) Exploratory analysis of topic interests and their evolution in bioinformatics research using semantic text mining and probabilistic topic modeling. IEEE Access 10:31480–31493
51. Gurcan F, Sevik S (2019) Business ıntelligence and analytics: an understanding of the ındustry needs for domain-specific competencies. In: 1st Int Informatics Softw Eng Conf Innov Technol Digit Transform IISEC 2019 - Proc. https://doi.org/10.1109/UBMYK48245.2019.8965457
52. Griffiths TL, Steyvers M (2004) Finding scientific topics. Proc Natl Acad Sci U S A. https://doi.org/10.1073/pnas.0307752101
53. Gurcan F, Kose C (2017) Analysis of software engineering industry needs and trends: implications for education. Int J Eng Educ 33:1361–1368
54. Scopus (2023) Content coverage guide. In: Elsevier. https://www.elsevier.com/solutions/scopus/how-scopus-works/content?dgcid=RN_AGCM_Sourced_300005030. Accessed 1 Jun 2023
55. Gurcan F, Ayaz A, Menekse Dalveren GG, Derawi M (2023) Business ıntelligence strategies, best practices, and latest trends: analysis of scientometric data from 2003 to 2023 using machine learning. Sustainability. https://doi.org/10.3390/su15139854
56. Aggarwal CC, Zhai CX (2012) An introduction to text mining. In: Aggarwal C, Zhai C (eds) Mining text data. Springer, Boston, pp 1–522
57. Bird S, Klein E, Loper E (2010) Natural language processing with python: analyzing text with the natural language toolkit. O'Reilly
58. Gurcan F, Cagiltay NE, Cagiltay K (2021) Mapping human-computer interaction research themes and trends from its existence to today: a topic modeling-based review of past 60 years. Int J Hum Comput Interact 37:267–280. https://doi.org/10.1080/10447318.2020.1819668
59. Plisson J, Lavrac N, Mladenić DD (2004) A rule based approach to word lemmatization. In: Proceedings of the 7th International Multiconference Information Society (IS'04). pp 83–86
60. Karl A, Wisnowski J, Rushing WH (2015) A practical guide to text mining with topic extraction. Wiley Interdiscip Rev Comput Stat 7:326–340. https://doi.org/10.1002/wics.1361
61. Řehůřek R, Sojka P (2011) Gensim—Statistical semantics in python. In: Lect. Notes Comput Sci
62. Mabey B (2023) pyLDAvis 3.4.1. In: https://pypi.org/. https://pypi.org/project/pyLDAvis/. Accessed 1 Jun 2023
63. Varadarajan V, Kommers P, Piuri V (2022) Preface of Special Issue on Advanced techniques and emerging trends in Smart Cyber–Physical Systems. Future Gener Comput Syst 135:299-302. https://doi.org/10.1016/j.future.2022.05.011
64. Ma L, Huo Y, Hu C, Li W (2019) Security and privacy for smart cyber-physical systems. Secur Commun Netw. https://doi.org/10.1155/2019/7045862
65. Vial G (2019) Understanding digital transformation: a review and a research agenda. J Strateg Inf Syst 28:118–144. https://doi.org/10.1016/j.jsis.2019.01.003
66. Koenig MED (2018) What is KM? Knowledge management explained. In: KMWorld. https://www.kmworld.com/Articles/Editorial/What-Is/What-is-KM-Knowledge-Management-Explained-122649.aspx. Accessed 6 Jan 2023
67. Alvarenga A, Matos F, Godina R, Matias JCO (2020) Digital transformation and knowledge management in the public sector. Sustain 12:1–24. https://doi.org/10.3390/su12145824
68. Shen Q, Hua Y, Huang Y, Ebstein R, Yu X, Wu Z (2022) Knowledge management and modern digital transformation of the property management industry in China. J Knowl Manag 26:2133–2144. https://doi.org/10.1108/JKM-04-2021-0320
69. Pavlekovskaya I, Urintsov A, Staroverova O, Nefedov Y (2018) The impact of digital transformation of the Russian economy on knowledge management processes. In: Proceedings of the European Conference on Knowledge Management, ECKM. pp 677–684
70. Erceg V, Zoranović T (2022) Knowledge management and digital business transformation. Strateg Manag 27:57–63. https://doi.org/10.5937/straman2200007e

71. Castagna F, Centobelli P, Cerchione R, Esposito E, Oropallo E, Passaro R (2020) Customer knowledge management in SMEs facing digital transformation. Sustain 12:1–16. https://doi.org/10.3390/su12093899
72. Mike K, Hazzan O (2023) What is data science? Commun ACM 66:12–13. https://doi.org/10.1145/3575663
73. Ullah S, Asif M, Ahmad S, Imdad, U, Sohaib, O (2019) Application of data science for controlling energy crises: a case study of Pakistan. In: ACM International Conference Proceeding Series. pp 60–64
74. Martínez-Álvarez F, Troncoso A, Riquelme JC (2018) Data science and big data in energy forecasting. Energies 11:1–2. https://doi.org/10.3390/en11113224
75. Fan C, Xiao F, Wang S (2014) Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. Appl Energy 127:1–10. https://doi.org/10.1016/j.apenergy.2014.04.016
76. Ferreira Jr. PCD, Machado EMM, Meiriño MJ, Quelhas, OLG, Mexas, MP (2022) COVID-19 pandemic. In: Handbook of Research on SDGs for Economic Development, Social Development, and Environmental Protection. IGI Global, pp 113–127
77. Suh NP, Cochran DS, Lima PC (1998) Manufacturing system design. CIRP Ann Manuf Technol 47:627–639. https://doi.org/10.1016/S0007-8506(07)63245-4
78. Seo Y (1999) Automated design of manufacturing systems through virtual manufacturing modeling technique. In: Proceedings - 3rd Russian-Korean International Symposium on Science and Technology, KORUS 1999. IEEE, Novosibirsk, Russia, pp 293–296
79. Cochran DS, Dobbs DC (2001) Evaluating manufacturing system design and performance using the manufacturing system design decomposition approach. J Manuf Syst 20:390–404. https://doi.org/10.1016/s0278-6125(01)80059-5
80. Atasoy R (2023) Sustainability of the school leadership in the digital era under the shadow of crisis. In: Masmoudi EA-A& M (ed) Leadership and Workplace Culture in the Digital Era. IGI Global, pp 149–168
81. Swartz B (2021) Ethics in engineering education 4.0: the educator's perspective. SAIEE Africa Res J 112:181–188
82. Miranda J, Navarrete C, Noguez J, Molina-Espinosa JM, Ramírez-Montoya MS, Navarro-Tuch SA, Bustamante-Bello MR, Rosas-Fernández JB, Molina A (2021) The core components of education 4.0 in higher education: three case studies in engineering education. Comput Electr Eng 93:1–13. https://doi.org/10.1016/j.compeleceng.2021.107278
83. Iborra M, Safón V, Dolz C (2021) The role of family ownership in survival and bouncing back: good and bad news? In: I. Management Association (ed) Research Anthology on Small Business Strategies for Success and Survival. IGI Global, pp 1117–1138
84. Robu M (2013) The dynamic and importance of smes in economy. USV Ann Econ Public Adm 13:84–89
85. Caperna A (2012) Integrating ICT into sustainable local policies. In: Silva C (ed) Handbook of Research on E-Planning: ICTs for Urban Development and Monitoring. IGI Global, pp 340–364
86. Kiani MM, Khanjankhani K, Takbiri A, Takian A (2021) Refugees and sustainable health development in Iran. Arch Iran Med 24:27–34. https://doi.org/10.34172/aim.2021.05
87. Koivusalo M (2017) Global health policy in sustainable development goals. Glob Soc Policy 17:224–230. https://doi.org/10.1177/1468018117703442
88. Shi J, Yang L, Jiao PP, CSTP (2004) Transportation structure model for sustainable development in Beijing. In: Proceedings of the World Engineers' Convention 2004, Vol C, Transportation and Sustainable Mega-Cities. pp 105–110
89. Zhao J (2012) The development of green sustainable transportation in China. In: Mackett RL, May AD, Kii M, Pan H (eds) Sustainable Transport for Chinese Cities (Transport and Sustainability). Emerald Group Publishing Limited, pp 99–117
90. Boiffin J, Hubert B, Durand N (2006) Sustainable agriculture or sustainable development? In: Livestock production and society. Wageningen Academic, pp 127–134
91. Belloumi M, Alshehry AS (2015) Sustainable energy development in Saudi Arabia. Sustain 7:5153–5170. https://doi.org/10.3390/su7055153
92. Görür OC, Erkmen AM (2019) Intention and Body-mood engineering via proactive robot moves in HRI. In: Management Association I (ed) Rapid Automation: Concepts, Methodologies, Tools, and Applications. IGI Global, pp 247–275

93. Sheridan TB (2016) Human-robot interaction: status and challenges. Hum Factors 58:525–532. https://doi.org/10.1177/0018720816644364
94. Kidd TT (2009) Handbook of research on technology project management, planning, and operations. IGI Global
95. Forgionne G, Russell S (2011) Unambiguous goal seeking through mathematical modeling. In: Adam F, Humphreys P (eds) Encyclopedia of Decision Making and Decision Support Technologies. IGI Global, pp 892–900
96. Sievert C, Shirley K (2014) LDAvis: A method for visualizing and interpreting topics. In: Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces. Association for Computational Linguistics, Stroudsburg, PA, USA, pp 63–70
97. Shrader CB, Ravenscroft SP, Kaufmann JB, Hansen K (2021) Collusion among accounting students: data visualization and topic modeling of student interviews. Decis Sci J Innov Educ 19:40–62. https://doi.org/10.1111/dsji.12226

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.