sciendo Journal of Social and Economic Statistics

Vol. 12, No. 2, Winter 2023
DOI: 10.2478/jses-2023-0011

CLUSTERING THE AI LANDSCAPE: NAVIGATING GLOBAL INSIGHTS FROM LEADING AI INDEXES[*]

Eduard Mihai MANTA[A*], Flavia BOGOEVICI[B]

[a)b)] Doctoral School of Cybernetics and Statistics, The Bucharest University of Economic Studies, Romania.

Abstract

This study develops a scorecard validation model for evaluating key Artificial Intelligence (AI) indexes, aiming to provide a comprehensive framework for assessing the multifaceted nature of AI development. Focusing on four significant AI indexes and one AI report from 2021 to 2023, the research employs both expert judgment and advanced text mining techniques, including k-means clustering. This dual approach facilitates a detailed examination of AI indexes, highlighting their strengths, weaknesses, and overall market comprehensiveness. The findings contribute to understanding the AI sector's evolution, offering critical insights for policy formulation and strategic decision-making in AI. Acknowledging the inherent subjectivity in the evaluation process and potential data biases, the paper suggests future research avenues, including cross-sectoral and regional analyses of AI trends and a deeper exploration of ethical considerations in AI. This study serves as a valuable resource for stakeholders navigating the complex AI landscape, providing a structured method for comparing and understanding AI advancements.

Keywords: AI, text analysis, scorecard validation, AI indexes, content analysis
JEL Classification: C4, C45, J01, M19, O33.

[*] Corresponding author, **Flavia Bogoevici** – flavia.bogoevici@yahoo.com

*****Acknowledgements: The research study has been elaborated within the Data Science Research Lab for Business and Economics of the Bucharest University of Economic Studies, within the project ID 585 PERFECTIS, entitled Increasing institutional performance through the development of the infrastructure and research ecosystem of transdisciplinary excellence in the socio-economic field, contract number and date: 42PFE of 30.12.2021 and the project Analysis of the sources of uncertainty regarding the forecasting of the national economic environment's evolution in the context of recent global socio-economic shocks (INCERTEC 2023). Causefinder:causality in the Era of Big Data and AI and its applications to innovation management.

1. Introduction

Artificial Intelligence (AI) has emerged as a transformative force, reshaping various sectors from healthcare to finance. As AI continues to advance, comprehending its multifaceted development becomes increasingly essential. The complexity inherent in AI's evolution encompasses technological advancements, market trends, policy implications, and ethical considerations. This paper endeavours to elucidate these dimensions by systematically clustering insights derived from leading AI indexes. Such an approach is instrumental in providing a holistic view of the AI landscape, a task that has become critical in the context of rapid technological change.

The main objective of this paper is the building of a scorecard validation method for evaluating the most representative AI indexes in the market from a qualitative standpoint by giving scores from 1 (lowest) to 5 (highest) to different drivers that are aiming to find the most comprehensive AI index in the market using both a subject matter expert judgement and content analysis. Using this approach, the paper aims to standardize a framework to assess various dimensions of AI indexes, ensuring a balanced and comprehensive evaluation. Moreover, the paper enhances content analysis of the AI indexes methodologies, including word clouds, word networks, correlation networks, and topic modelling to obtain a qualitative insight allowing for a deeper understanding of the textual data within these indexes by helping in interpreting the context, themes, and nuances that a subject expert might miss only be reading and evaluating. By constructing clusters using k-means clustering method, the manuscript is adept at uncovering patterns and trends within the AI sector, offering a nuanced understanding of the interplay and evolution of its various components. Such an analytical approach is needed for anticipating future trajectories and developments within AI, a task of significant interest to both practitioners and scholars.

The implications of this study extend beyond academic exploration, playing a critical role in informing policy and strategic decisions. As AI technologies increasingly permeate societal and economic spheres, the necessity for informed decision-making in policy and strategy formulation becomes increasingly evident. This research contributes to this imperative by providing insights that can guide the development of policies and strategies, thus fostering responsible and effective AI advancement at national and international levels.

By adopting this multifaceted approach, the study seeks to demystify the AI landscape, presenting an amalgamation of empirical rigor and practical relevance. The findings are poised to serve as an asset for a diverse array of stakeholders, including researchers, policymakers, and industry leaders, who are navigating the dynamic and intricate terrain of AI development.

This work aims to provide answers to the following research questions: RQ1. How the AI indexes can be classified? RQ2.What are the most often occurring terms in the content of AI indexes? RQ3: What is the most associated pair of terms in the AI indexes documents? RQ4.

What are the AI Indexes scorecard values? RQ5. What are the most significant themes in the market AI indexes? RQ6: How AI indexes may be grouped? RQ7: What is the best performing AI index in the market using the scorecard methodology?

This paper is divided into five sections. The opening section is devoted to briefly demonstrating the topic's importance. The following section is devoted to theoretical considerations in the subject. The third section is devoted to presenting data and the working methods, while the next section highlights the empirical results. The key conclusions and discussions are presented at the end of the study.

2. Literature review

Digitalization and automation are two interconnected concepts closely related to new emerging technologies. The following represent key relationships between digitalization, automation, and new emerging technologies: digitalization and automation as facilitators of emerging technologies: Digitalization and automation are often considered important catalysts for the adoption and implementation of new emerging technologies. They can create the necessary foundation to fully exploit the potential of these technologies by integrating them into business processes and society in general (Bonnet and Westerman, 2020).

The synergy between digitalization and emerging technologies is also important because digitalization, such as the use of data, cloud computing, and advanced connectivity, can provide the necessary foundation for the development and implementation of emerging technologies, such as Artificial Intelligence (AI), the Internet of Things (IoT), blockchain, and virtual/augmented reality (AR). These emerging technologies, in turn, can enhance digitalization by providing innovative solutions and creating new business opportunities (Pozdnyakova et al., 2019).

An important aspect is the synergy between automation and emerging technologies for innovation: digitalization and automation can pave the way for innovation by integrating emerging technologies into existing processes. For example, integrating AI and data analysis into automation systems can bring new levels of intelligence and efficiency. This synergy between automation and emerging technologies can generate new business models and true transformations in various industries (Warhurst and Hunt, 2019).

There is a mutual impact of digitalization, automation, and emerging technologies: digitalization and automation can stimulate the adoption and development of emerging technologies, while emerging technologies, in turn, can accelerate the process of digitalization and automation. These three elements influence each other and can create a virtuous circle of innovation and development (Policarpo et al., 2021).

One of the main technologies underlying the development of a digital enterprise status is Robotic Process Automation (RPA), or Automation of Processes through Robotics. This

technology assists in configuring software by creating robots that automate repetitive, rule-based processes, mimicking human actions (Automation Anywhere, 2022). According to a study conducted by Statista, the global RPA market will exceed 13 billion dollars by 2020 ("Robotic process automation (RPA) market 2030," 2023).

Automations can be characterized in two main ways: supervised and unsupervised. On one hand, supervised automations involve the presence of a person in the process, who assists the robot in executing steps (for example, contributing at a certain stage of the process, analysing data for validation and accuracy), leading to clear results and improved process productivity. On the other hand, unsupervised automations do not require human interactions, these types of robots can be programmed to execute a process autonomously (Dilmegani, 2023). The most important benefits of RPA technology are the improvement of productivity, minimization of errors, reduction of employee work hours and costs, contributing to the acceleration of companies' digitalization and the valorisation of strategic jobs. (UiPath White Paper - Identifying the Real Business Outcomes of Automation, 2023).

The effects of digitalization and automation on the economic environment can be assessed using Leontief's input-output model, which has the advantage of capturing intersectoral dependencies and the flow of goods and services within the economy. Thus, automation and the use of RPA technology can lead to changes in the production structure of an economy. By automating repetitive and routine tasks, companies can increase their productivity and reallocate resources towards higher value-added activities. Additionally, they can improve their sustainable practices through digitalization (Quille et al., 2023). It is important to note that the Leontief Model can contribute to the assessment of the impact of these changes on sectoral production, employment, and added value.

Another effect of implementing automation can lead to labour displacement, as certain tasks performed before the implementation of RPA technology are now automated. This can lead to changes in employment in various sectors and modifications in skill requirements. The Leontief Model, supported by empirical data, can provide information about the potential size and distributional effects of labour displacement caused by automation. Numerous company departments have implemented RPA techniques through which they have significantly reduced their FTEs (Full Time Equivalents), allowing them to focus on high-value-added tasks (reports and analysis within complex processes). Considering the current economic conditions, where the population is continuously decreasing and aging, employees with more skills will shift towards high value-added tasks, while those who adapt more slowly will have the opportunity to acquire new skills. Among the services and industries that require new competencies are RPA, AI, AR (Anagnoste, 2017).

A specific example of benefits in terms of improved working time, accuracy, and hours saved is the adoption of UiPath's software robots by the company BAT. These robots deliver results in a much faster and more qualitative way: the company managed to save over 230,000

working hours through automation and has not reported any errors since the implementation of RPA. For one process, three robots were developed, automating the work of about 10 employees. Considering that an employee works 40 hours per week, the robots helped save 400 hours of work per week (UiPath, 2022).

The implementation of sustainable solutions in organizations is a crucial factor for their prosperity. RPA is an effective way to improve sustainability practices within companies. It is well known that since the emergence of the term 'sustainability' in business, many companies have transitioned to adopting sustainable practices, such as green energy, reducing CO2 emissions, etc. As a result, their social, economic, and environmental objectives have improved significantly. Companies aiming to reach the pinnacle of sustainability need to consider measures in three major areas, specifically economic performance, ecological preservation, and social equity (Patrício et al., 2023).

Living in Industry 4.0, the main challenges are related to implementation and management, aiming to achieve economic, social, and environmental objectives. As the authors highlight in the article 'Manufacturing and Management Paradigms, Methods and Tools for Sustainable Industry 4.0 -Oriented Manufacturing Systems,' a wide range of methods and tools lead to the promotion of sustainability in Industry 4.0. Additionally, in the authors' view, the paradigms of the manufacturing and management field ('M&M') contribute to the discovery of new business models, supporting the circular economy and risk management (Varela et al., 2022).

It has been demonstrated that 66% of organizations are already executing or planning to apply Artificial Intelligence for sustainability purposes, for example, to improve the efficiency of renewable energy, predicting the necessary energy production in the coming days. Moreover, AI can also help manage waste more efficiently. A study published in 'Nature Communications' attests that 79% of the Sustainable Development Goals (SDG) can be achieved through this technology, which supports the circular economy and the intelligent construction of the city (Vinuesa et al., 2020).

In the current era of Industry 4.0, business strategies should be developed wisely to fulfill the durability of issues, concerning not only economic but also social and environmental concerns, by simultaneously encouraging excellence in organizational efficiency and transformational initiatives, thus contributing to the sustainable outcomes of businesses. Change and organizational transformation are now mandatory, regarding employees, their skills, and culture, to have proper management. For most industries, I4.0 management is a crucial issue and should also be realized considering the factors of production, directly or indirectly, to improve their performances (Sagina, 2022).

Among the key elements of sustainable production and management in the current I4.0 context are additive manufacturing or 3D printing, which promotes collaborative processes and practices among stakeholders in the network, alongside exponential technologies and advanced processes, high-performance computing, and disruptive technologies (autonomous and

collaborative robots, advanced mechatronics, micro and nano manufacturing, supercomputing, advanced interfaces, virtual and augmented reality, and digital twins). There are also other approaches, methods, and techniques relevant in I4.0, for example, based on artificial intelligence, such as machine learning and deep learning, pattern recognition, blockchain, etc. In addition, business information, big data, and data analytics from the specific field of data science are essential pillars of I4.0 (Varela et al., 2022).

Therefore, digitalization and automation are closely linked to new emerging technologies. They can facilitate the adoption and implementation of these technologies, and, in turn, emerging technologies can enhance digitalization and automation. The interaction between digitalization, automation, and emerging technologies can bring significant benefits in terms of innovation, efficiency, and transformation in various sectors and industries.

The validation of AI-based systems is a complex and multifaceted challenge. This complexity arises from unique qualities inherent to AI systems, such as their ability to adapt or change behaviour in response to environmental changes, often referred to as the Oracle Problem. This makes it difficult to predetermine the desired outcomes for AI systems. Furthermore, the validation of AI systems extends beyond traditional software validation due to the scale and diversity of the intended environments and contexts of use. These challenges highlight the need for research focusing on the validation of artificially intelligent systems and the development of suitable tools to assist in this process (Myllyaho et al., 2021).

The field of AI in healthcare provides valuable insights into the development and validation of AI-based prediction models. For instance, the extraction of keywords and subsequent mapping to central topics pertinent to AI Prediction Model (AIPM) development phases demonstrates the importance of considering algorithmic bias, transparency, openness, and interpretability. Each of these factors plays a crucial role in the validation of AI systems, especially in the healthcare domain. The consideration of these overarching themes throughout the AIPM lifecycle is crucial for ensuring the robustness and reliability of AI applications (de Hond et al, 2022)

The literature suggests that the process of validating AI systems is not only intricate but also compounded by the shortage of supporting tools. While some tools exist, like Simulink for cyber-physical systems simulations, and SOCRaTEs for generating test oracles for such simulations, finding suitable tools remains a challenge. This highlights a gap in the domain-specific or recent availability of tools and frameworks that support AI system validation. The deficiency in supporting tools underscores the importance of developing methodologies and tools that can cater to the unique requirements of AI systems validation, especially in diverse domains (Myllyaho et al., 2021).

Furthermore, the literature underscores the necessity of validating complete AI systems rather than just their machine learning (ML) models. This comprehensive approach to validation is critical as it recognizes that even less accurate learning models can outperform

more refined ones when validation metrics are more appropriately chosen. It points towards the inadequacy of solely focusing on ML model validation and emphasizes the importance of considering the entire AI system's functioning in practical scenarios. This holistic approach to AI system validation is supported by guidelines and reports that seek to address problems in the realm of ML, stressing the need for more knowledge on online testing and system-wide validation (Gao et al, 2019)

Additionally, the scoping review on guidelines and quality criteria for AI-based prediction models in healthcare offers insights into the multidisciplinary nature of AI validation. It highlights the importance of algorithmic bias, transparency, openness, and interpretability in AI model development and validation. The review emphasizes that addressing these factors, particularly in the healthcare sector, is vital for obtaining trust and acceptance by end users. This multidisciplinary approach, involving technical, methodological, and medical expertise, is essential for developing robust and reliable AI systems. The consideration of these overarching topics throughout the AI development lifecycle ensures a comprehensive validation approach, integrating various perspectives and expertise (Kalra and Paddock, 2016).

In sum, these insights from the literature provide a foundational understanding for constructing a scorecard methodology aimed at validating AI indexes. The methodology should encompass a comprehensive approach, considering not only the technical aspects of AI systems but also addressing algorithmic bias, ensuring transparency, and promoting interpretability. This multidisciplinary and holistic approach is key to developing robust and reliable methodologies for AI validation, particularly in fields like healthcare where the implications of AI systems are significant and far-reaching (Koopman and Wagner, 2016; Menghi et al., 2016)

3. Data and methodology

The primary aim of this study is to develop a comprehensive evaluation framework for Artificial Intelligence (AI) indexes, employing both expert judgement and an innovative cluster analysis approach. This research meticulously scrutinizes four pivotal AI indexes – the Artificial Intelligence Index-2023, The Global AI Index-2023, Government AI Readiness Index-2022, and AI Watch Index-2021 – along with an in-depth examination of the State of AI Report-2021. These indexes and the report were selected for their relevance and impact within the AI sector, representing a diverse range of perspectives and metrics in the field.

Data for this analysis was extracted from the official publications and websites of the entities responsible for these indexes, ensuring the authenticity and reliability of the information. The study encompasses a critical period in AI development, focusing on the years 2021 to 2023, a timeframe pivotal for observing recent advancements and trends in AI.

The content analysis conducted on these indexes and the report was comprehensive, delving into their methodologies and descriptive elements. This approach was instrumental in uncovering the underlying themes and topics prevalent in these documents. By grouping similar themes into clusters, the study provides a nuanced understanding of the information landscape within these AI indexes. This methodical approach, coupled with the exclusive focus on English-language documents, ensures a coherent and in-depth evaluation of the current state and evolution of AI indexes. This paper aims to standardize and streamline the evaluation of AI indexes, offering valuable insights and benchmarks for stakeholders in the AI community.

This research yields two primary outcomes: firstly, an extensive analysis of unstructured text drawn from both these indexes and an additional AI report, deemed pertinent to the study, and secondly, a specialized scorecard tailored to evaluate four distinct AI indexes using both expert judgement and text analysis results. The scorecard was developed through a rigorous review process, aimed at validating the selected indexes by assigning scores based on specific criteria. Concurrently, for the text mining aspect of the research, the R software platform was employed to execute statistical analysis. This dual approach, encompassing both qualitative and quantitative methods, ensures a comprehensive assessment of the AI indexes and the report, providing a multi-dimensional understanding of the current AI landscape.

In the figure below there can be noticed the flow of the data, from importing the files to generating the results, comprising the text analysis technique.

Figure no 1. Text analysis stages

Source: author's processing using SmartArt

This paper begins with the critical task of data preparation, presenting its own set of unique challenges. The technique chosen for analysis significantly influences the reliability, precision, and results of a content analysis study. Similarly, preparing texts for analysis requires critical decision-making that can impact the validity, results, and accuracy of a text analysis method (Crone et al., 2006). The primary processes we identify include content importing, text processing and feature engineering, creating the document-term matrix (DTM), and applying filtering and weighting to the DTM.

In most computer-based content analysis methods, entire texts are broken down into smaller, more specific textual elements like words or word combinations. Many text analysis algorithms enhance their computational efficiency and accuracy by normalizing variables and excluding 'stopwords' like prepositions, conjunctions, pronouns, and verbs, which are deemed irrelevant and thus removed before analysis.

The normalization process involves standardizing words to a consistent format. This is crucial when determining if different spellings represent the same concept. It also reduces the amount of language data, with a simple technique like converting all text to lowercase being a common normalization method.

Common English nouns such as 'he/she/it' often don't contribute meaningfully to the content of a text. Excluding such phrases reduces data volume, lessens processing load, and typically enhances performance and accuracy.

In the bag-of-words model, the DTM is a prevalent tool for representing a content corpus (a collection of texts). The DTM is a matrix where columns represent words, rows signify documents, and cells indicate the frequency of each word in each document. This approach allows for the analysis of data using matrix and vector algebra, effectively translating language into numerical form.

Word clouds visually represent the most frequent words in a text, excluding stopwords like pronouns and conjunctions. A bigram, a type of n-gram where n=2, consists of two consecutive words. Bigrams are widely used in computational linguistics, cryptography, and speech recognition, facilitating straightforward statistical text analysis.

The correlation network illustrates the most frequently occurring words, while the word network reveals commonly co-occurring word pairs.

Content analysis at the academic level can uncover significant insights into word relationships by identifying if certain terms frequently follow others or appear together in the same publications. This study utilized the correlation network to explore word relationships in scientific papers. Co-occurrences occurring at least 10 times and with a correlation exceeding 0.7 were emphasized.

Content analysis primarily relies on document collections from scientific journals to categorize them into discernible natural groups. Topic modelling, an unsupervised classification method for such texts, identifies natural clusters of elements, akin to qualitative data clustering. Latent Dirichlet allocation (LDA) is a common method for creating a topic model. It views each document as a mix of topics and each topic as a mix of words. LDA mathematically estimates the mixture of components for each topic and the mix of topics characterizing each document (Roberts et al., 2014; Blei et al., 2003). This approach allows for overlapping content in documents, more closely resembling natural language use.

The manuscript utilizes k-means clustering technique, which are particularly adept at identifying underlying patterns and trends in the artificial intelligence (AI) sector. These methods are effective in analysing text data, as they can systematically organize and interpret large volumes of textual information, which is often unstructured and complex.

In the context of this research, a scorecard is conceptualized as an evaluative instrument, systematically designed to facilitate the assessment and comparative analysis of various Artificial Intelligence (AI) indexes. The primary function of the scorecard is to apply a uniform

set of evaluative criteria across different AI indexes, thereby ensuring that each index is subject to an equitable and standardized assessment. This standardization is imperative, as it ensures the consistency and validity of the comparative analysis, rendering the comparisons both equitable and substantively meaningful. This approach is documented by researchers with focus especially as a credit risk tool (Glenon et al., 2008; Beque et al., 2017; Huang and Scott, 2017; Siddiqi, 2017).

The scorecard incorporates a multitude of criteria or indicators, thereby enabling a multi-faceted analysis of each AI index. This multi-dimensional approach is critical, as it allows for the examination from score 1 (lowest) to 5 (highest) of each index across various aspects (documented in the annex), encompassing 14 primary drivers along with the score rationale and some questions for the validation prospective purpose. The main areas in which the AI indexes are evaluated consist of definition and purpose, scaling and units of measurement, data source and collection, calculation methodology, weighting, handling of missing data, temporal consistency, geographical coverage, validation and benchmarking, transparency, sensitivity analysis, feedback look, stakeholder engagement and lastly, ethical considerations. The starting point of defining the framework was the European Commission paper about tools for constructing composite indicators (Nardo et. al, 2005). Such an inclusive evaluation framework is essential for providing a holistic appraisal of each index, elucidating not only the strengths but also the potential areas of improvement.

By incorporating both quantitative and qualitative measures, the paper offers a balanced view of the indexes. Quantitative data might include numerical indicators, while qualitative insights could come from expert opinions or textual analysis. By providing a clear set of criteria and a scoring mechanism, the scorecard method can be used for benchmarking AI indexes against each other. This helps in ranking or categorizing them based on their performance in various aspects. The comparative analysis facilitated by the scorecard can highlight gaps in the current AI landscape, such as areas lacking sufficient development or attention. It can also point out opportunities for growth and improvement in various domains of AI.

4. Empirical results

In table 1 it can be noticed a set of descriptive statistics of the corpus, which contains a table with types, tokens and sentences. The index with the highest value of unique words (types) is “The Global AI Index”(852), followed by “AI Watch Index” (633). Moreover, the index with the most used words (tokens) is still “The Global AI Index” (2728), the following being again “AI Watch Index” (2347). Regarding the sentences, the highest values are again held by the two indexes, more exactly 93 and 58.

Table no. 1: Corpus descriptive statistics

Text	Types	Tokens	Sentences	Number
AIIndex	500	1378	54	01
AIWatchIndex	633	2347	58	02
GovernmentAIReadinessIndex	287	620	19	03
StateofAIReport	59	84	5	04
TheGlobalAIIndex	852	2728	93	05

Source: author's processing using R.

"The Global AI Index" ranks countries based on capacity for AI, measuring the levels of investment, innovation and implementation and presenting the actual international view on areas such as talent, research and development, infrastructure, operating environment, talent, government strategy and commercial ventures. In addition, the current edition of the index explores per-capita indicators, reflecting the degree of intensity of AI development across countries.

The indicators present in "AI Watch Index" explore the European areas of strength, together with those requiring attention in the AI domain. The index also contains a structured set of quantitative indicators regarding EU's performance and position across different AI dimensions, relevant for policymaking.

In the Figure 2 there can be visualized the most frequent words from the analysed indexes. The most frequent words in the analysed documents are "AI", "data", "intelligence", "index", "research", "global", "chapter", "development", "performance", "countries".

Figure no 2. Unigram word cloud

Source: author's processing using R

The Figure 3 measures the co-occurrence of words through the intensity of the lines that are linking each pair of two words. The central nodes in the network are "technology",

"development", "future", ''business'', ''economic'', ''performance''. The pairs being formed are: ai-regulation-business-impact-performance-field-research-data-world-countries-index.

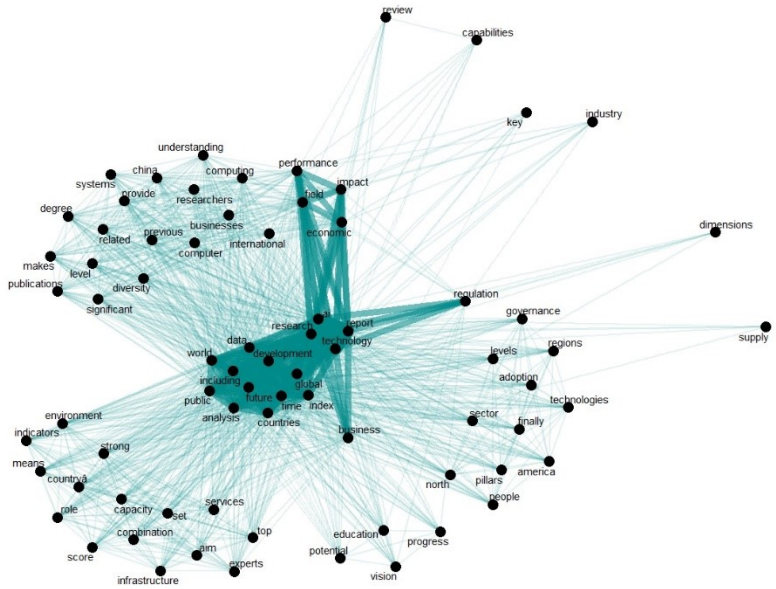

Figure no 3. Word pair network

Source: author's processing using R

To create the word correlation network, a 0.8 threshold was used. The most encountered combinations of words are global-analysis-index-countries-public-data-including-global-development, environment-score-indicators-capacity-services-infrastructure and geographical-activity-indicator-comparative.

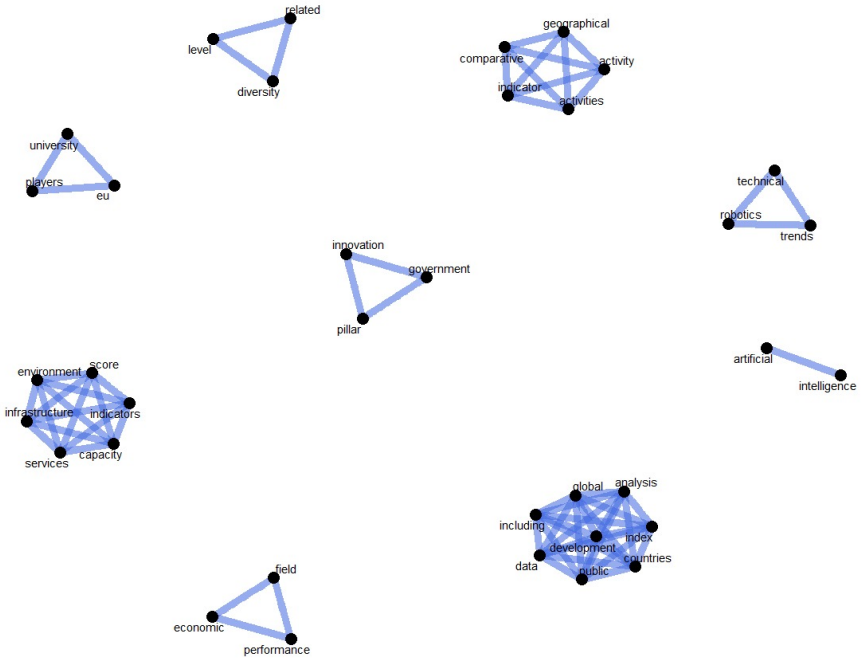

Figure no 4. Word correlation network

Source: author's processing using R

LDA is a type of unsupervised machine learning that identifies topics within a corpus of text by clustering similar words together. The Figure 5 shows the top five terms within each of the five topics that the LDA algorithm has identified as being significant within the dataset:

- Topic 1 (Red Bars): The dominant terms such as "EU," "research," and "robotics" suggest this topic is likely concerned with European Union research initiatives, particularly in the field of robotics. The inclusion of "players" and "geographical" could indicate a discussion of key stakeholders and the geographical spread or focus of robotics research within the EU context.
- Topic 2 (Yellow Bars): Terms like "countries," "government," "sector," "infrastructure," and "technology" indicate this topic pertains to governmental roles in technology infrastructure across various countries. It may reflect on how different governments approach the development of technological sectors within their national infrastructure.

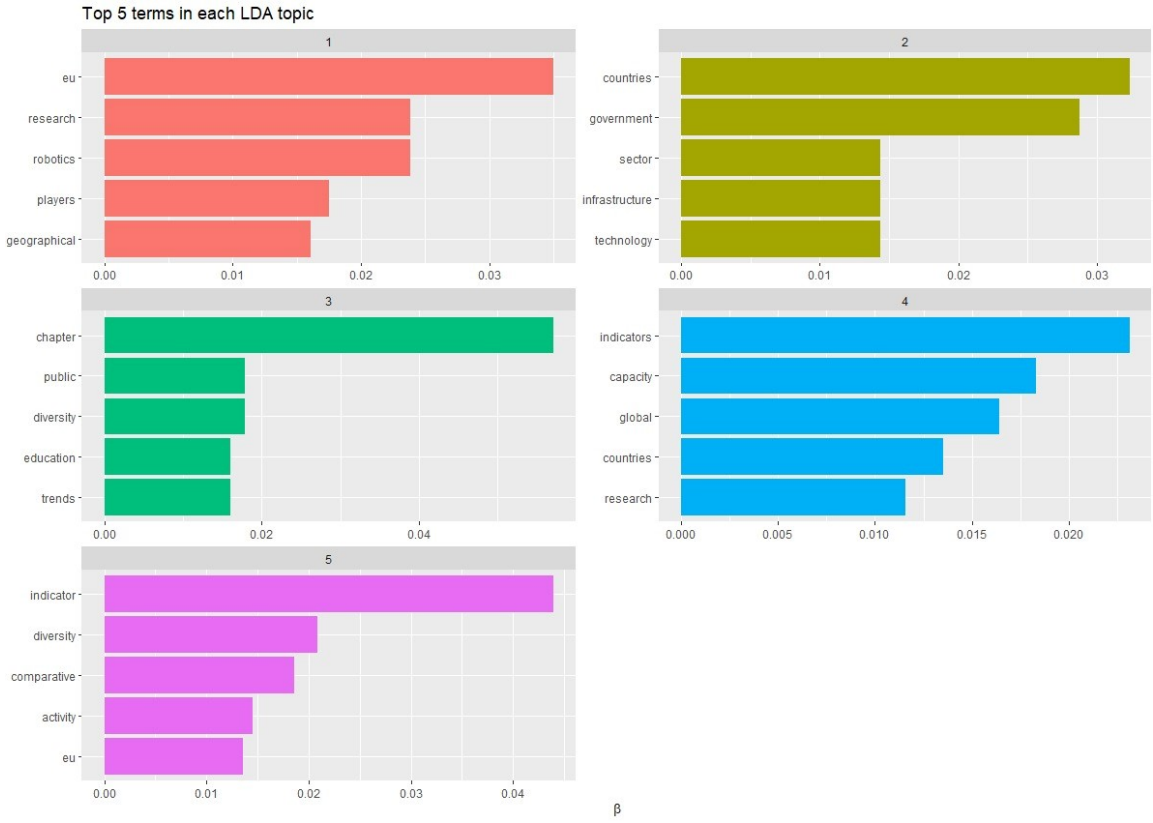

Figure no 5. Topic modelling

Source: author's processing using R

- Topic 3 (Green Bars): This topic, with words like "chapter," "public," "diversity," "education," and "trends," seems to address educational diversity and public trends. The term "chapter" may imply a focus on specific sections or themes within educational publications or reports.

- Topic 4 (Blue Bars): With terms such as "indicators," "capacity," "global," "countries," and "research," this topic might explore global research indicators and capacity, assessing how countries measure up to various indicators of research and development.
- Topic 5 (Purple Bars): This topic shares terms with Topic 1, like "indicator" and "EU," but also includes "diversity," "comparative," and "activity." It could be focusing on comparative studies of diversity and activity within the EU, potentially in a socio-economic or policy context.

Two clusters were identified by the k-means algorithm, as indicated in the Figure 6 by the two different symbols (dots and triangles). The positioning of the points within the plot suggests how the algorithm has grouped the data points based on their similarity. The first cluster (dots) contains "State of AI" and "Government AI Readiness", which are closely positioned, indicating that these terms or documents have similar features or content regarding the current state and governmental preparedness for AI. The second cluster (triangles) includes "AI Watch", "The Global AI", and "AI Index", which are spread out but still grouped together, indicating a thematic relation to AI monitoring, global perspectives on AI, and indexing or rating AI, but with each term having a distinct profile within this broader theme.

The first two dimensions account for 42.7% and 31.8% of the variance respectively, which suggests that they capture significant information about the dataset.

The plot shows that "State of AI", and "Government AI Readiness" are closely linked and quite distant from the other three terms. This indicates a clear differentiation in the dataset that the k-means algorithm has captured. The results of k-mean clustering suggest that the AI indexes should receive similar points based on each cluster.

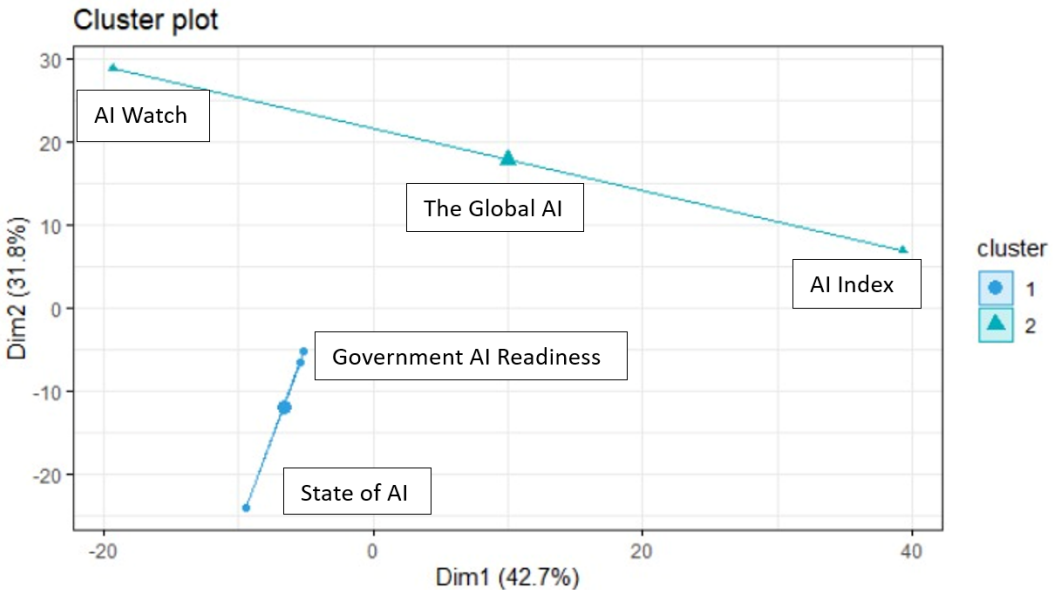

Figure no 6. K-means clustering results

Source: author's processing using R

Using the results of context analysis and the expert judgement, the detailed evaluation of the AI indexes was conducted using the scorecard approach that is presented in the annexes. The results suggest that the most comprehensive AI indexes frameworks are The Global AI Index and AI Watch Index 2021 as seen in the Figure 8. Furthermore, the lowest score received the Government AI Readiness Index of 2022. This qualitative approach of framework evaluations is commonly and consistently used especially with respect with banking and financial institutions.

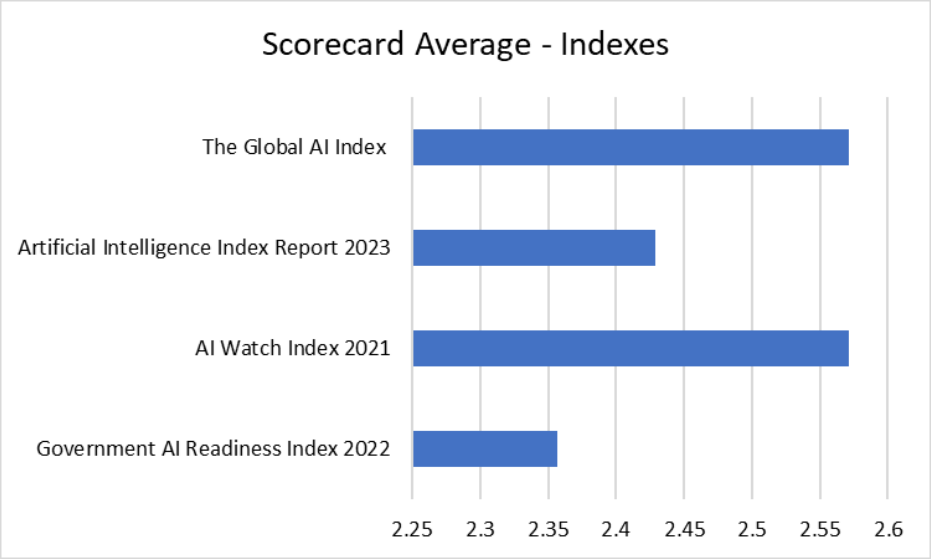

Figure no 8. Score card results

Source: author's processing using Excel

5. Conclusions and discussion

The study's comprehensive analysis of AI indexes using a novel scorecard validation model and advanced text mining techniques yields significant insights into the AI sector's current state and future trajectory. The detailed examination of AI indexes revealed that 'The Global AI Index' and 'AI Watch Index' demonstrate a higher level of comprehensiveness in their content, as evidenced by their greater diversity in unique words and total word usage. This finding aligns with their more extensive coverage and nuanced exploration of AI-related topics, such as AI development intensity across countries and the European Union's AI strengths and areas needing attention.

Our text mining analysis, encompassing word frequency, co-occurrence networks, and topic modeling, further underscores the prevalent themes and concerns within these indexes. For example, the dominance of terms like 'AI,' 'data,' and 'countries' in these documents indicates a focus on the global impact and development of AI. Additionally, the LDA topic analysis provided a deeper understanding of the thematic areas covered by these indexes, ranging from EU research initiatives to the role of governments in technological infrastructure.

The application of k-means clustering methods offered a valuable perspective on the relationship and thematic similarities among the indexes. The clustering results, which grouped indexes based on content similarity, informed the scorecard evaluation, leading to a nuanced understanding of how these indexes align or differ in their focus and coverage.

The scorecard evaluation, combining these analytical insights with expert judgment, concluded that 'The Global AI Index' and 'AI Watch Index' are the most comprehensive AI indexes in the market, while the 'Government AI Readiness Index' of 2022 received the lowest score. This evaluation not only reflects the depth and breadth of these indexes but also their relevance and utility in understanding and navigating the AI landscape.

In conclusion, this research contributes a detailed and multi-faceted perspective on AI index evaluation, combining quantitative and qualitative analyses to provide a richer, more comprehensive understanding of AI's current state and potential directions. While the study acknowledges the inherent limitations of data completeness and the subjective nature of scorecard development, it lays the groundwork for future research in AI index evaluation, potentially expanding to cross-sectoral and regional comparisons and deeper ethical considerations in AI development.

References

Anagnoste, S., 2017. Robotic Automation Process-The next major revolution in terms of back-office operations improvement, in: *Proceedings of the International Conference on Business Excellence*. pp. 676–686.

Bequé, A., Coussement, K., Gayler, R. and Lessmann, S., 2017. Approaches for credit scorecard calibration: An empirical analysis. *Knowledge-Based Systems*, 134, 213-227.

Blei, D. M., Ng, A. Y. and Jordan, M. I., 2003. Latent Dirichlet Allocation. *The Journal of Machine Learning Research*, 3 , 993–1022.

Bonnet, D., Westerman, G., 2020. The New Elements of Digital Transformation. *MIT Sloan Management Review*.

Crone, S. F., Lessmann, S., and Stahlbock, R. 2006. The impact of preprocessing on data mining: An evaluation of classifier sensitivity in direct marketing. *European Journal of Operational Research*, 173(3), 781-800.

de Hond, A.A., Leeuwenberg, A.M., Hooft, L., Kant, I.M., Nijman, S.W., van Os, H.J., Aardoom, J.J., Debray, T.P., Schuit, E., van Smeden, M. and Reitsma, J.B., 2022. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. *NPJ digital medicine*, 5(1), p.2.

Dilmegani, C., 2023. ChatGPT education use cases, benefits & challenges in 2023. *AI Multiple*.

Gao, S., Kollmannsberger, S. and Meschke, G., 2019. AI in Construction Engineering and Management: Validation Challenges and Opportunities. *Journal of Construction Engineering and Management*, 145(10). DOI:10.1061/(ASCE)CO.1943-7862.0001697.

Glennon, D., Kiefer, N. M., Larson, C. E. and Choi, H. S., 2008. Development and validation of credit scoring models. *Journal of Credit Risk, Forthcoming*.

Huang, E. and Scott, C., 2007. Credit risk scorecard design, validation and user acceptance. *HBOS Bank*.

Kalra, N. and Paddock, S.M., 2016. Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability?. *Transportation Research Part A: Policy and Practice*, 94, pp.182-193. DOI:10.1016/j.tra.2016.09.010.

Koopman, P. and Wagner, M., 2016. Challenges in Autonomous Vehicle Testing and Validation. *SAE International Journal of Transportation Safety*, 4(1), 15-24. DOI:10.4271/2016-01-0128.

Menghi, C., Nejati, S., Gaaloul, K. and Briand, L. C., 2019. Generating Automated and Online Test Oracles for Simulink Models with Continuous and Uncertain Behaviors. *Proceedings of the 2019 27th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. DOI:10.1145/3338906.3338955.

Myllyaho, L., Raatikainen, M., Männistö, T., Mikkonen, T. and Nurminen, J. K., 2021. Systematic literature review of validation methods for AI systems. *Journal of Systems and Software*, 181, 111050.

Nardo, M., Saisana, M., Saltelli, A. and Tarantola, S., 2005. Tools for composite indicators building. *European Comission, Ispra*, 15(1), 19-20.

Patrício, L., Ávila, P., Varela, L., Cruz-Cunha, M.M., Ferreira, L.P., Bastos, J., Castro, H. and Silva, J., 2023. Literature review of decision models for the sustainable implementation of Robotic Process Automation. *Procedia Computer Science*, 219, pp.870-878. https://doi.org/10.1016/j.procs.2023.01.362

Policarpo, L.M., da Silveira, D.E., da Rosa Righi, R., Stoffel, R.A., da Costa, C.A., Barbosa, J.L.V., Scorsatto, R. and Arcot, T., 2021. Machine learning through the lens of e-commerce initiatives: An up-to-date systematic literature review. *Computer Science Review*, 41, p.100414.

Pozdnyakova, U.A., Mukhomorova, I.V., Golikov, V.V., Sazonov, S.P. and Pleshakov, G.G., 2019. Internet of things as a new factor of production in the conditions of digital economy. *Ubiquitous Computing and the Internet of Things: Prerequisites for the Development of ICT*, pp.1145-1151.

Quille, R.V.E., Almeida, F.V.D., Borycz, J., Corrêa, P.L.P., Filgueiras, L.V.L., Machicao, J., Almeida, G.M.D., Midorikawa, E.T., Demuner, V.R.D.S., Bedoya, J.A.R. and Vajgel, B., 2023. Performance Analysis Method for Robotic Process Automation. *Sustainability*, 15(4), p.3702. https://doi.org/10.3390/su15043702

Siddiqi, N., 2017. *Intelligent credit scoring: Building and implementing better credit risk scorecards*. John Wiley & Sons.

UiPath White Paper - Identifying the Real Business Outcomes of Automation, 2023. https://www.uipath.com/resources/automation-whitepapers/identifying-the-real-business-outcomes-of-automation (Accessed: 3 November 2023)

Roberts, M.E., Stewart, B.M., Tingley, D., Lucas, C., Leder‐Luis, J., Gadarian, S.K., Albertson, B. and Rand, D.G., 2014. Structural topic models for open‐ended survey responses. *American journal of political science*, 58(4), pp.1064-1082.

Sagina, D.M., 2022. *Pengaruh Kepemilikan Keluarga, Leverage, Dan Pengungkapan Corporate Social Responsibility (CSR) Terhadap Tax Avoidance Dengan Kepemilikan Insititutional Sebagai Variabel Moderasi* (Studi Pada Perusahaan Sektor Industri Dasar, Kimia, dan Pertambangan Yang Terdaftar di Bursa Efek Indonesia Tahun 2021) (Doctoral dissertation, Universitas Muhammadiyah Malang).

Varela, L., Ávila, P., Castro, H., Putnik, G.D., Fonseca, L.M.C. and Ferreira, L., 2022. Manufacturing and management paradigms, methods and tools for sustainable industry 4.0-oriented manufacturing systems. *Sustainability*, 14(3), p.1574. https://doi.org/10.3390/su14031574

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S.D., Tegmark, M. and Fuso Nerini, F., 2020. The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature communications*, 11(1), pp.1-10. https://doi.org/10.1038/s41467-019-14108-y

Warhurst, C. and Hunt, W., 2019. *The digitalisation of future work and employment: Possible impact and policy responses* (No. 2019/05). JRC Working Papers Series on Labour, Education and Technology.

https://www.automationanywhere.com/resources (Accessed: 1 November 2023)

https://www.statista.com/statistics/1259903/robotic-process-automation-market-size-worldwide/ (Accessed: 15 October 2023)

Appendix – Validation Scorecard of synthetical measurements for AI

Matrix for comparing existing synthetical measures for AI	Driver	Score rationale	Validation prospective	Government AI Readiness Index 2022	AI Watch Index 2021	Artificial Intelligence Index Report 2023	The Global AI Index
Definition and Purpose	Ensure that all indicators are defined consistently and that their purposes are clear and uniform across reports	Researching if there are sufficient foundational, theoretical concepts and if the measurement and sub-components of the framework are clearly highlighted, together with weights and indicators illustrating the dimensions of the index.	Are there detailed methodological reports existing on the official websites?	**4**	**4**	**4**	**4**
			Does the official website/institution own a dashboard?				
			How are the indicators defined across different index reports?				
			What is the intended purpose of each indicator?				
			Are there any discrepancies in definitions or purposes across indices reports?				
			Is the index built on at least three dimensions/pillars containing items?				
Scaling and Units of Measurement	Ensuring all indicators are converted to a common scale or measurement unit. This might involve normalization or standardization procedures.	Checking if there is enough information disclosed related to the measurement of the index.	Is there sufficient information disclosed related to the measurement of the index?	**3**	**4**	**4**	**4**
			What scale or unit is each indicator currently measured in?				
			Are there any challenges in converting indicators to a common scale or unit?				
			What procedures have been used for normalization or standardization?				
Data Source and Collection	Validating the sources of data used for each indicator and ensuring consistent methodologies are applied when collecting the data.	Data collection methodologies should be clearly stated and quality check for indicators should be performed, assessing the strengths and weaknesses of each indicator.	What are the primary data sources for each indicator?	**3**	**4**	**4**	**4**
			How is the data collected for each indicator?				
			Are there consistency checks in place to ensure data collection methodologies are uniform?				
			Do the variables and indicators respect the principles?				
			Does the index offer access to raw data? How exactly? They provide full access to raw indicators database or sources?				
Calculation Methodology	Adopt consistent method or formula for calculating each indicator, especially if the indices use different approaches.	Discovering methods used in calculating the indicators, with potential discrepancies.	What formula or method is used to calculate each indicator?	**3**	**4**	**4**	**2**
			Are there variations in calculation methods across indices?				

			How are discrepancies in calculation methodologies addressed?				
Weighting	If the composite indicators are derived from multiple individual metrics, ensure that the weights assigned to each metric are standardized or justified based on their relevance or importance.	Selecting the most suitable weighting and aggregation procedure referring to the theoretical framework is crucial.	How are weights assigned to each individual metric withing the composite indicators?	**3**	**3**	**2**	**4**
			Are the weights based on expert judgement, statistical methods, or another approach?				
			Is there justification provided for the chosen weighting scheme?				
			Are there any correlation issues?				
Handling of Missing Data	Establish consistent procedures for dealing with missing or incomplete data, such as imputation methods or exclusion rules.	Checking and discussing missing values and the appearance of outliers in the dataset.	Did the index take into account the existence of missing data? Is there disclosed any imputation procedure? Is there any information related to the magnitude of the amount of missing data?	**3**	**3**	**2**	**4**
			How covering is the information on the website/in the report?				
			Is there some detailed information related to the existence and treatment of potential outliers in the process of building the composite indicator?				
			How often does missing or incomplete data occur?				
			What methods are in place to handle missing data?				
			Are these methods consistently applied across indices?				
Temporal Consistency	Ensure that time periods used for data collection or reporting are consistent across indices. This might involve aggregating monthly data into quarterly or annual figures.	Defining clear time period definitions for when the data was collected or reported and checking consistency across the time periods?	What time periods are used for data collection or reporting?	**4**	**4**	**4**	**4**
			How is data from different time periods aggregated or reconciled?				
Geographical Coverage	If the indicators cover multiple regions or countries, ensure that the same geographical boundaries and definitions are used.	Checking if the geographical regions are presented with the corresponding definitions	Is there information disclosed related to the index coverage (regions/countries)?	**4**	**4**	**4**	**4**
			What geographical regions or countries are covered by the indicators?				
			Are there any discrepancies in geographical boundaries or definitions across indices?				
Validation and Benchmarking	Compare standardized indicators against benchmarks or other	Checking if accurate and reliable standards are met and clearly stated.	Against which benchmarks or standards are the indicators compared?	**1**	**1**	**1**	**1**

	recognized standards to validate their accuracy and reliability.		How often this validation exercise carried out?				
Transparency	Clearly document the standardization process, including any decisions made, methodologies adopted, and assumptions used.	Assessing the existence of clear communication for users regarding the standardization process.	Is there documentation available that explains the standardization process?	**1**	**1**	**1**	**1**
			How are decisions, methodologies, and assumptions communicated to users?				
Sensitivity Analysis	Test how sensitive the standardized indicators are to changes in the underlying data or assumptions. This can help in understanding the robustness of the indicators.	Checking if the results are sensitive to alternative ways of measuring the treatment and if the uncertainty is measured in inputs or outputs. Improving transparency by using sensitivity analysis. Identifying the sources of uncertainty in the development of the indicator.	Have any sensitivity tests been conducted on the indicators?	**1**	**1**	**1**	**1**
			What were the results and implications of these test?				
Feedback Loop	Allow for periodic revies and revisions of the standardization process, especially as new data becomes available or as methodologies evolves.	Checking if there is information regarding periodic reviews of the papers.	Is there a mechanism in place for periodic reviews of the standardization process?	**1**	**1**	**1**	**1**
			How are revisions made when new data or methodologies become available?				
Stakeholder Engagement	Engage with stakeholders, such as users of the indicators, data providers, and subject matter experts, to gather feedback and ensure the standardized indicators meet their needs.	Checking if there are any stakeholders that contributed to writing the reports.	Who are the key stakeholders for these indicators?	**1**	**1**	**1**	**1**
			How often are stakeholders engaged in the standardization process?				
Ethical Considerations	Ensure data privacy, security and other ethical are addressed during the standardization process.	Assessing if ethical and data privacy issues are addressed.	What measures are in place to ensure data privacy?	**1**	**1**	**1**	**1**
			How are other ethical considerations, like data security, addressed in the standardization process?				