Article

Not peer-reviewed version

ChatGPT as an Artificial Intelligence Tool to Provide an Analytical Boost to Text Mining: An Application to TUCKER Models for Multiway Tables

Roberto Cascante-Yarlequé [*] , Purificación Galindo-Villardón , Fabricio Guevara-Viejó

Posted Date: 25 March 2026

doi: 10.20944/preprints202507.1477.v2

Keywords: tucker decomposition; canonical biplot; ChatGPT; neural networks; scientific literature review

doi:10.20944/preprints202507.1477.v2

Article

ChatGPT as an Artificial Intelligence Tool to Provide an Analytical Boost to Text Mining: An Application to TUCKER Models for Multiway Tables

Roberto Cascante-Yarlequé [1,2,*], Purificación Galindo-Villardón [1,2,3] and Fabricio Guevara-Viejó [2]

[1] Department of Statistics, University of Salamanca, 37008 Salamanca, Spain
[2] Centro de Estudios Estadísticos, Universidad Estatal de Milagro (UNEMI), Ecuador
[3] Centro de Estudios e Investigaciones Estadísticas (CEIE), Escuela Superior Politécnica del Litoral (ESPOL), Ecuador
* Corresponding: rcascant@usal.es

Abstract

In this comprehensive study, we meticulously investigated multidimensional data analysis techniques, particularly focusing on Tucker decomposition methods, spanning the period from 2000 to 2025. Our primary objective was to discern trends, advancements, and applications of these techniques across various domains of knowledge and how they have evolved over time. An extensive corpus of 288 scientific articles related to tensor decompositions, Tucker models and applications was previously reviewed. Multivariate methods such as text mining using IraMuteq software and MANOVA-Biplot were employed to visualize identified data patterns, and the analytical capability of ChatGPT artificial intelligence was assessed to provide contextual insights and add another layer of information to the research. Our conclusions underscore the importance of blending traditional statistical approaches with natural language processing prowess to achieve a profound understanding of the data. This analysis offers a comprehensive perspective on the evolution and application of multidimensional data analysis techniques, with a special emphasis on the enduring relevance of Tucker techniques in this new millennium.

Keywords: tucker decomposition; canonical biplot; ChatGPT; neural networks; scientific literature review

---

1. Introduction

Systematic reviews are an essential tool in scientific research, allowing the collection, analysis, and synthesis of existing information in a specific field [1]. In recent years, advances in technology and artificial intelligence have introduced new approaches to perform this task more efficiently and accurately. Two innovative approaches proposed in this study are Biplot-based Text Mining and ChatGPT artificial intelligence, which offer different perspectives on extracting scientific information.

Text Mining refers to the process of using natural language processing and data mining techniques to analyze large volumes of text and extract relevant information [2]. In a systematic review, this information can be extracted from the abstracts of scientific articles published in high-impact journals. Additionally, Biplot techniques are presented as an ideal complement to enrich text analysis, as they effectively visualize the relationships between keywords and groups of documents. This facilitates the interpretation of patterns, relationships, and trends in textual data, allowing for a deeper understanding of a particular domain. Traditionally, this has been done with Correspondence Analysis (CA) by Benzecri [3], which penalizes the most frequently used words in the text corpus that may be very common; however, they may have little semantic or informative relevance.

On the other hand, ChatGPT is a form of artificial intelligence based on generative language models, which has the ability to understand and generate text in natural language [4]. With its ability for contextual understanding and generating coherent responses, ChatGPT can be used for tasks such as classification, summarization, and analysis of scientific information.

In this article, we present our proposal on the publications related to Tucker models in this new millennium. Tucker models, named after the eminent mathematician Ledyard R. Tucker, are highly useful tools today as they are multiway data models that play a crucial role in the decomposition and understanding of complex multidimensional data [5].

Despite the growing importance and widespread application of Tucker models, a comprehensive understanding of how research in this field has evolved over the past two decades remains elusive. Existing literature reviews on multiway analysis tend to be either narrowly focused on specific applications or purely methodological, lacking a global perspective on disciplinary trends, emerging topics, and shifts in research focus over time. Furthermore, traditional approaches to synthesizing such a diverse body of literature are labor-intensive and may inadvertently overlook subtle but significant patterns buried within large collections of abstracts. This gap in our understanding limits the ability of new researchers to quickly grasp the state of the field and prevents the identification of cross-disciplinary fertilization and emerging application domains. Therefore, there is a clear need for a systematic, data-driven approach that can objectively map the intellectual landscape of Tucker model research and track its evolution.

Conducting a systematic literature review on these models is an extremely complex task, given the wide range of applications they have in the scientific world. Numerous re- searchers are dedicated to exploring and leveraging the findings provided by these models, reflecting their importance and relevance in contemporary research.

To address this gap, the present study proposes and demonstrates a hybrid methodological framework that integrates traditional text mining and multivariate statistical analysis with the contextual reasoning capabilities of large language models. Specifically, we combine lexical analysis using IraMuteq software with Canonical Biplot (MANOVA-Biplot) techniques to obtain a statistically robust visualization of research trends. We then augment these findings with a thematic analysis performed by ChatGPT, which provides a nuanced, context-aware interpretation of the identified patterns. Our primary objective is twofold: first, to conduct a systematic literature review of Tucker models from 2000 to 2025, identifying key research themes, their evolution over time, and emerging application areas; and second, to critically evaluate the synergies and complementarities between traditional statistical approaches and AI-powered analysis in the context of scientific literature review. The remainder of this paper is organized as follows: Section 2 provides the mathematical foundations of Tucker models. Section 3 describes the materials and methods, including the PRISMA-guided data collection process, the text mining workflow, the Canonical Biplot analysis, and the integration of ChatGPT. Section 4 presents the results from both analytical approaches. Section 5 discusses the findings, compares the two methodologies, and addresses the study's limitations. Finally, Section 6 offers conclusions and directions for future research.

2. Tucker Models

Tucker models are a generalization of singular value decomposition (SVD) specifically designed for tensors. In this context, a tensor can be understood as a multidimensional data structure that generalizes the concept of matrices to more than two dimensions. While a matrix has rows and columns, a tensor can have multiple dimensions or modes.

Within three-way data analysis, the **STATIS** techniques (*Structuration des Tableaux à Trois Indices de la Statistique*) [6] are noteworthy. These techniques treat three-way data as tables or data matrices to obtain a two-dimensional array, thus capturing only the stable structure of the original data [7]. In contrast, Tucker models [5] analyze the entire data cube, constructing a simplified model, thereby capturing the dynamic part of the data. This ability of Tucker models to examine the complete data cube results in much more detailed and content-rich information. These models allow for the representation of a three-dimensional tensor (or higher-order tensor) as a combination of a core tensor and a set of mode matrices that capture the underlying dynamic structure of the data [8]. This hierarchical and multidimensional structure provides a more compact and meaningful representation of the data, facilitating the identification of patterns and relationships in complex datasets [9].

In the ***Tucker-3*** model, the three-way data tensor $\underline{\mathbf{X}} = (x_{ijk})$, where $i \in \{1,2,\ldots,I\}$, $j \in \{1,2,\ldots,J\}$, $k \in \{1,2,\ldots,K\}$, which we assume, without loss of generality, consists of I *individuals*, J *variables*, and K *occasions (or time points)*, can be approximated by the tensor $\underline{\widehat{\mathbf{X}}} = (\hat{x}_{ijk})$ using the equation

$$\underline{\mathbf{X}} = \underline{\widehat{\mathbf{X}}} + \underline{\mathbf{E}}, \qquad (1)$$

where $\underline{\mathbf{E}} = (e_{ijk})$ is a tensor of *residuals*.

To construct the estimated tensor $\underline{\widehat{\mathbf{X}}}$, the Tucker-3 decomposition seeks to find three orthogonal loading matrices $A = (a_{ip})$, $B = (b_{jq})$, and $C = (c_{kr})$; where $p \in \{1,2,\ldots,P\}$, $q \in \{1,2,\ldots,Q\}$, and $r \in \{1,2,\ldots,R\}$. These matrices A, B, and C correspond to the weights for *individuals*, *variables*, and *occasions* respectively, and the decomposition also involves a *core tensor* $\underline{\mathbf{G}} = (g_{pqr})$, also known as *core matrix* [5].

Thus, the ijk-th element of the tensor $\underline{\mathbf{X}}$ in equation (1) can be expressed as

$$x_{ijk} = (\sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} g_{pqr} a_{ip} b_{jq} c_{kr}) + e_{ijk}. \qquad (2)$$

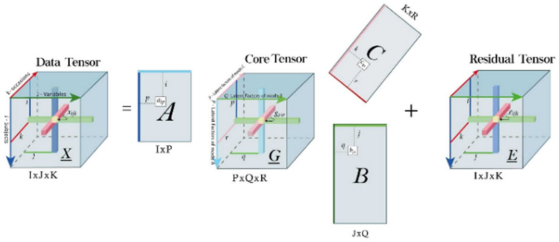

Figure 1. Graphical representation of the Tucker-3 model decomposition.

The *core tensor* $\underline{\mathbf{G}}$, as shown in Figure 1, is the central element of the Tucker-3 model; it represents the interaction between the three *modes* of the original tensor $\underline{\mathbf{X}}$ [5,8]. It resembles a three-dimensional matrix and contains the amount of variance explained by each combination of components. This tensor $\underline{\mathbf{G}}$ has dimensions $P \times Q \times R$, where P, Q, and R ($P < I$, $Q < J$, $R < K$) are the number of components or latent factors required for each mode respectively, as a result of the dimensionality reduction of the tensor $\underline{\mathbf{X}}$. In other words, the *core tensor* $\underline{\mathbf{G}}$ is considered a reduced version of the data tensor $\underline{\mathbf{X}}$ [10]. Moreover, the selection of the number of components of the *core tensor* is still a subject of study today, with various algorithms developed in this regard.

One of the first algorithms developed to estimate the number of components, by Kroonenberg and De Leeuw in 1980 [11], is the algorithm known as *TUCKALS3*. This algorithm, through alternating least squares fitting, aims to minimize the function

$$g(A,B,C,G) = \| \mathbf{X} - A\mathbf{G}(C \otimes B)^T \|^2, \qquad (3)$$

where $\|\cdot\|$ denotes the *Euclidean* norm, and the matrices (also called unfolded tensors) $\mathbf{X}$ and $\mathbf{G}$ are obtained by concatenating the K frontal slices of the tensors $\underline{\mathbf{X}}$ and $\underline{\mathbf{G}}$, respectively.

Later, Timmerman and Kiers in 2000 [12] proposed the well-known algorithm ***DIFFIT***, which determines an optimal proportion of the difference in fit or sum of squares explained by the solutions of these models known as *Three-Mode Principal Components Analysis* (***3MPCA***).

The modes are essential for defining how data is organized and related in the three-dimensional tensor [13]. The three mode matrices are used to describe how dimensions are grouped within each mode and, together, determine the structure of the *core tensor* [14]. The

Tucker-3 model can be represented using matrix equations for each of its three modes as follows:

$$Mode\ A: X_A = AG_A(C \otimes B)^T + E_A, \qquad (4)$$

$$Mode\ B: X_B = BG_B(C \otimes A)^T + E_B, \qquad (5)$$

$$Mode\ C: X_C = CG_C(B \otimes A)^T + E_C, \qquad (6)$$

In equations (4), (5), and (6), the operator $\otimes$ is the *Kronecker* product [5]. Taking mode A as an example, the matrices X_A, G_A, and E_A are two-way arrays of order $(I \times JK)$, $(P \times QR)$, and $(I \times JK)$, respectively, and they are the *matricized* versions of the tensors $\underline{\mathbf{X}}$, $\underline{\mathbf{G}}$, and $\underline{\mathbf{E}}$.

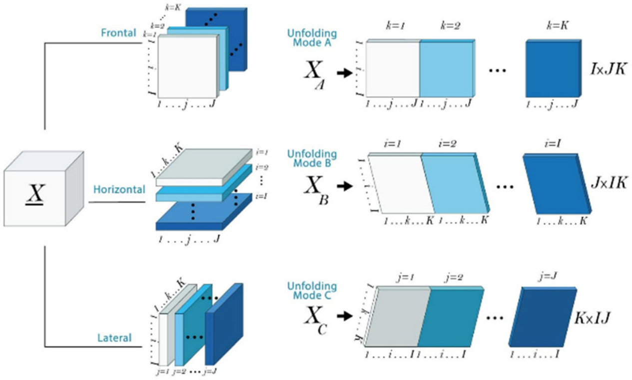

Figure 2. Unfolding of the data tensor $\underline{\mathbf{X}}$ in its different modes.

As shown in Figure 2, the matricization or unfolding of the tensor $\underline{\mathbf{X}}$ is obtained by concatenating the different slices, whether they are frontal for mode A, horizontal for mode B, or lateral for mode C, in order to obtain their two-way matrix versions.

The ***Tucker-2*** models are specific variants of the Tucker-3 model [15]. In a Tucker-2 model, out of the three available modes, two are reduced, resulting in three possible configurations of Tucker-2 models; these are ***Tucker-2-AB***, ***Tucker-2-AC***, and ***Tucker-2-BC***. Similarly, Kroonenberg and De Leeuw developed the ***TUCKALS2*** algorithm to find optimal solutions with alternating least squares. For example, in the Tucker-2-AB model, mode A is reduced to P components ($P < I$) and mode B is reduced to Q components ($Q < J$). Mode C, on the other hand, is not reduced and remains with its K original slices. This same principle applies to the other modes.

From another perspective, the ***Tucker-1*** models are also classified as specific instances of the Tucker-3 model. In a Tucker-1 model, out of the three available modes, only one is reduced. In total, three variants of Tucker-1 models are distinguished: ***Tucker-1-A***, which involves a component analysis on two-dimensional tables for the frontal slice supermatrix; ***Tucker-1-B***, which involves a component analysis on two-dimensional tables for the horizontal slice supermatrix; and ***Tucker-1-C***, which corresponds to a component analysis on two-dimensional tables for the vertical slice supermatrix. The ***TUCKALS1*** algorithm is used for these three Tucker-1 models. For example, in the Tucker-1-A model, mode A is reduced to P components ($P < I$). Modes B and C are not reduced and remain with their J and K original slices, respectively.

The challenges encountered in interpreting the results of Tucker models have driven the development of new approaches based on simpler hypotheses. In 1970, Carroll and Chang [14], as well as Harshman [16], independently proposed a model to decompose three-way tables with fully interrelated modes more directly. While the former authors called it ***CANDECOMP*** (*Canonical Decomposition*), Harshman named it ***PARAFAC*** (*Parallel Factor Analysis*).

The ***CANDECOMP/PARAFAC*** **(CP)** model is a constrained version of the Tucker-3 model and is defined as

$$X_C = C\mathbf{H}(B \otimes A)^T + E_C, \qquad (7)$$

where the matrix $\mathbf{H}$ is the two-dimensional version of order $R \times R^2$ of the three-dimensional *superidentity* tensor $\underline{\mathbf{H}}$; that is, $\underline{\mathbf{H}} = (h_{pqr})$, with $h_{pqr} = 1$ when $p = q = r$, and $h_{pqr} = 0$ otherwise.

The model is uniquely defined under certain (weak) conditions, which are often met in practice. That is, the estimates of A, B, and C are unique up to an arbitrary scaling of the columns in two of the component matrices and an arbitrary simultaneous permutation of the columns in the component matrices.

These Tucker models have found applications in a wide variety of fields, from spectroscopy [17–27], agriculture [28–31], data mining [32], and neuroimaging [33,34]. Their versatility and ability to handle high-dimensional data have driven their adoption in scientific research and have contributed to a better understanding of complex phenomena [35]. Therefore, it is crucial to document and advance the development of Tucker models due to their fundamental role in contemporary and future scientific research in the new millennium.

In this article, we conduct a systematic literature review using Biplot-based Text Mining and ChatGPT 5.2 for scientific information extraction, focusing on a specific case: the search and analysis of the literature from the year 2000 to 2025 related to Tucker models. For this purpose, searches were conducted in the Scopus and Web of Science databases, obtaining a set of relevant articles. These articles were processed using a Text Mining tool, the IraMuteq software, generating a lexical table and subsequently performing a Canonical Biplot analysis. In turn, ChatGPT 5.2 was used to perform a similar classification, evaluating similarities and differences with the Biplot-based Text Mining approach, thereby adding an additional layer of information to the systematic literature review.

3. Materials and Methods

3.1. Materials

A comprehensive search for articles in high-impact journals was conducted using a broad set of keywords related to Tucker models and their equivalent formulations. The search terms included specific Tucker model variants (*Tucker1, Tucker-1, Tucker2, Tucker-2, Tucker3, Tucker-3, Tucker decomposition, Tucker model*), equivalent mathematical formulations (*Higher-Order Singular Value Decomposition, HOSVD, Multilinear Singular Value Decomposition, MLSVD, three-mode principal component analysis, 3MPCA, Higher-order orthogonal iteration, HOOI*), and closely related concepts (*tensor decomposition, multiway analysis, three-way data analysis, CANDECOMP/PARAFAC, PARAFAC2, tensor train, core tensor, multilinear algebra, multilinear rank*). The search was performed from January 1, 2000 to December 31, 2025 in the Scopus and Web of Science (WoS) databases, specifically.

In the search conducted in Scopus, a total of 351 publications were obtained, while 328 publications were found in WoS. In total, an initial database of 679 publications was obtained, which was subjected to a screening process. For this screening, the principles established in the ***PRISMA*** 2020 statement (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*), which defines guidelines for the systematic review and meta-analysis of studies [36], were applied. These guidelines allowed for the evaluation of the relevance and methodological quality of the articles, ensuring the inclusion of studies that met rigorous standards.

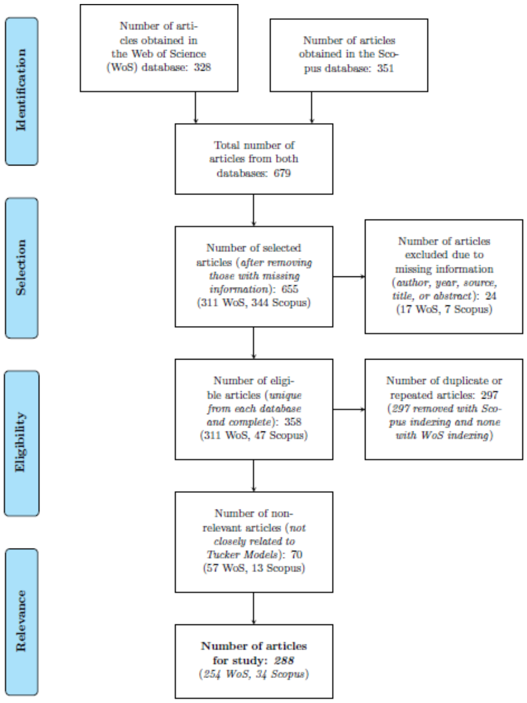

Figure 3. Flowchart of the sample selection of articles related to Tucker Models from 2000 to 2025 according to the PRISMA 2020 statement.

During the inclusion and exclusion process of articles, a series of stages were carried out to ensure the relevance of the set of articles for the study. Initially, publications with missing information in critical fields such as author, year, source, title, or abstract were discarded. These fields are considered essential for the proper analysis and understanding of the articles. As a result of this filter, 24 articles were removed (17 from WoS and 7 from Scopus), leaving a total of 655 articles, of which 311 are indexed in WoS and 344 in Scopus.

Subsequently, duplicate articles were removed; that is, articles that were found in both databases. If an article is indexed in both databases, we removed those indexed in Scopus and retained those published in WoS as a priority. We identified a total of 297 duplicate articles, which were removed from Scopus, resulting in a new total database consisting of 358 articles, of which 311 are indexed in WoS and 47 in Scopus; this database no longer contains duplicate articles.

Finally, the relevance of each article to the study, focused on Tucker models, was evaluated. At this stage, the contribution of the research and its close relation to Tucker models were considered of great importance. For this purpose, a new column was created in the MsExcel document containing the data, allowing logical values of 1 to be assigned to relevant articles and 0 to those that were not. After this evaluation, 70 articles were discarded (57 from WoS and 13 from Scopus), resulting in a final set of 288 articles, of which 254 are indexed in WoS and 34 in Scopus, ready for Text Mining analysis using IraMuteq software. This screening process is summarized in Figure 3.

The resulting database, composed of 288 publications, was used as the definitive set to conduct the study with Text Mining tools, Canonical Biplot, and ChatGPT. This analysis will allow us to evaluate the efficiency and accuracy of each technique in extracting specific scientific information about Tucker models, as well as to identify potential advantages and disadvantages of each approach.

3.2. Research Methods

3.2.1. Text Mining with IraMuteq

The statistical analysis of textual data, commonly referred to as SATD, is a methodology that combines text analysis techniques with statistical techniques to extract meaningful information and reveal patterns and trends in large volumes of textual data [37]. The primary objective of this analysis is to explore and understand the structure and content of the texts, as well as to identify relationships between words and their context [38].

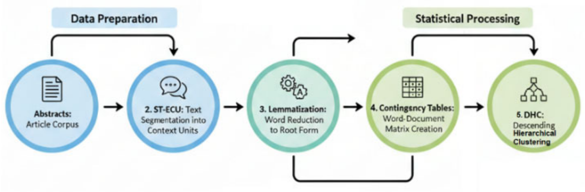

Figure 4. Flowchart of the process for creating lexical tables with IraMuteq software.

In this study, SATD was used as part of the Text Mining approach to analyze the database of 288 publications related to Tucker models from the new millennium, specifically from the year 2000 to 2025. SATD was conducted using the IraMuteq software (Interface for R for Multidimensional Analysis of Texts and Questionnaires), a free software developed by Pierre Ratinaud at the LERASS Laboratory of the University of Toulouse [39]. IraMuteq software allows for multidimensional analyses of texts of various natures, including official documents, web pages, news, and open-ended survey responses [39,40], and it enables lexical analysis and the generation of a lexical frequency table of keywords used in article abstracts [41].

The analysis with IraMuteq software is based on correspondence analysis methods and hierarchical classifications, with a focus on descending hierarchical clustering [41]. This procedure, shown in Figure 4, involves reducing context units to more manageable text segments (ECUs), followed by lemmatization to reduce words to their canonical forms [39]. Then, context units (CU) are created that meet certain criteria for active and supplementary forms [39,41].

Descending hierarchical classification is performed in stages, using correspondence factor analysis and optimization of inter-class inertias to form clusters/classes of ECUs [42]. This analysis allows for the identification of textual structures in the text corpus [43,44].

Each article in the corpus is accompanied by its corresponding abstract. These abstracts provide a concise overview of the contents and context of the full article. Abstracts are a rich source of textual information that allows the software to analyze lexical and semantic patterns based on the keywords used and their relationships [45].

It is essential that the selected corpus is focused on a specific and coherent topic, in this case, Tucker models. The thematic homogeneity ensures that the lexical and semantic analyses conducted by the IraMuteq software are relevant and meaningful in the research context [46].

Once the corpus is prepared, the next step is the segmentation of the text into Elementary Context Units (ECUs). These ECUs are short text fragments, usually two or three lines, representing indivisible units of analysis [47]. Segmentation into ECUs allows the IraMuteq software to analyze and process each fragment independently, facilitating the identification of specific patterns and features in the text [48]. Before proceeding with lexical analysis, IraMuteq software performs an automatic lemmatization stage of the words [39]. Lemmatization involves reducing each word to its canonical or root lexical form [49]. This way, different forms of the same word are grouped under a single form, simplifying the analysis and ensuring greater consistency in the results obtained [50].

Once the words have been lemmatized, the IraMuteq software proceeds to create Context Units (CU) [50]. Each CU is a set of ECUs that share a certain minimum number of "active forms," which are generally verbs, nouns, adverbs, and adjectives. These active forms are relevant for lexical and semantic analysis as they contain key information about the content of the text [51].

From the CUs, the IraMuteq software constructs frequency tables, which are tables showing the frequency of occurrence of each active form in each CU [52]. These tables are crucial for correspondence analysis and identifying patterns in textual data. The IraMuteq software performs a

Correspondence Factor Analysis (CFA) on these contingency tables to detect relationships and similarities between the active forms and the CUs [3].

The product we will use from the IraMuteq software analysis is a lexical table that shows the relationships between active forms (lemmatized words) and CUs [53]. In this table, each row represents a CU, and each column represents an active form. The cells of the table contain information about the frequency of occurrence of each active form in each CU [42]. The lexical table is a visual representation of the lexical and semantic relationships in the analyzed corpus, allowing for the identification of patterns and trends in textual data [51].

3.2.2. Multivariate Analysis of the Lexical Table with Canonical Biplot

The lexical table obtained from the IraMuteq software is a cross-frequency table that has traditionally been used for Benzecri's *Correspondence Factor Analysis* (CFA). CFA is a multivariate technique that examines dependency relationships between categorical variables and generates profiles based on relative frequencies. In this method, the most frequent words are penalized using a weighted chi-square distance to measure relationships between categories in a contingency table. These weights are essential for calculating the weighted chi-square distance and are an integral component of the method.

However, CFA has limitations in describing the group structure of individuals and assessing significant differences between them. In these cases, Canonical Analysis is used, particularly Canonical Biplot or MANOVA Biplot, which provides a graphical representation associated with MANOVA or Discriminant Analysis. This approach offers a more comprehensive view of the group structure and the statistical significance of the differences between them [54].

Amaro et al. [54] justify the use of Canonical Biplot instead of CFA in the analysis of data with group structure. They note that, from a statistical standpoint, CFA does not consider the group structure of individuals and is not suitable for studying the significance of differences between groups. In contrast, MANOVA can be used to study the significance of differences between groups, and Canonical Analysis is used to study the dimensionality of the alternative hypothesis when the null hypothesis is rejected in MANOVA [55].

The Canonical Biplot is an extension of Canonical Analysis that combines information about group means and variables from a data matrix [54]. It allows for a joint representation of group means and variables in a two-dimensional space, where group means are projected onto the canonical variates, which are linear combinations of observed variables that best explain the differences between groups [56]. The variables are positioned on the plot in directions that best predict the actual mean values for each variable [57].

The matrix approach to the MANOVA model for n variables is defined as follows:

$$\mathbf{X} = \mathbf{AB} + \mathbf{R}. \qquad (8)$$

In this formulation, $\mathbf{X}$ represents a matrix of dimensions $m \times n$, which stores the observations. The matrix $\mathbf{A}$, of rank t, refers to the design matrix, which is set by the researcher, while $\mathbf{B}$ corresponds to the matrix of unknown parameters. Lastly, $\mathbf{R}$, a matrix of size $m \times n$, stores the model residuals or random deviations [57].

The MANOVA-Biplot, once the model is fitted, presents the null hypothesis, which is fundamental for testing linear hypotheses such as

$$H_0: \mathbf{MBN} = 0, \qquad (9)$$

where $\mathbf{M}$ and $\mathbf{N}$ are full-rank matrices, and their rank is $p \leq t$ [58].

The selection of the matrices $\mathbf{M}$ and $\mathbf{N}$ can vary, allowing for the construction of Biplots for different study hypotheses. Typically, $\mathbf{M}$ and $\mathbf{N}$ include a set of coefficients for making contrasts.

In the Canonical Biplot, similarities between groups are interpreted based on Mahalanobis distances; the variables responsible for differences between groups are projected onto the group directions, and correlations between variables are represented as angles between vectors on the plot [59]. Acute angles indicate positive relationships, obtuse angles indicate negative relationships, and right angles are interpreted as independence [60].

The interpretation of the canonical axes, which are linear combinations of variables derived in Canonical Component Analysis (CCA), is typically done through the correlations between the observed variables and the canonical axes, although these relationships are not maximized by the technique [54]. Instead, the use of the quality of representation of group means is proposed as a measure of result accuracy. The quality of data representation is obtained through the Quality Representation Index (QRI) [61]. This index measures the percentage of variability of group means represented by the reduced-dimension solution and corresponds to the squared cosine of the angle between the point/vector in multidimensional space and its projection in the low-dimensional solution [62]. A QRI is calculated for both group means and variables [63]. In the Canonical Biplot, it is possible to draw a confidence circle around the projection of the means for each variable. These constructed circles allow us to visually identify significant differences between groups in terms of the variables being analyzed [54].

In this study, once lexical tables were obtained from text analysis using IraMuteq software, a multivariate analysis was performed using the Canonical Biplot or MANOVA Biplot method. The main objective of this analysis was to group the articles based on their publication periods and explore trends and changes over time in relation to Tucker models.

To carry out the Canonical Biplot, the articles were divided into four groups based on six-year publication periods: *group 1 (2000-2005)*, *group 2 (2006-2011)*, *group 3 (2012-2018)*, and *group 4 (2019-2025)*. Each group represented a specific period during which Tucker models were studied, allowing for the analysis of the evolution of these models over time.

In the Canonical Biplot, the lexical tables obtained through the IraMuteq software were used to represent the groups of articles and the variables (keywords) in a two-dimensional space. The group means (represented by the articles belonging to each previously defined period) were projected onto the canonical axes, which are linear combinations of the lexical variables that best explain the differences between groups. Thus, each group was positioned in the Canonical Biplot space according to its specific lexical characteristics.

The keywords (variables) were also placed on the plot based on their relationships with the groups. The words responsible for the most significant differences between groups were projected in the directions of the corresponding groups, allowing the identification of which terms were most associated with each specific period.

Additionally, the Canonical Biplot facilitated the interpretation of similarities and differences between groups and variables. Mahalanobis distances were used to assess dissimilarities between the groups on the plot, while the angles between the keyword vectors allowed for the analysis of correlations among them.

The application of the Canonical Biplot in this study provided a clear visualization of the group structure of the articles based on their publication periods and revealed patterns and trends related to Tucker models over time. It also provided a rich and detailed graphical representation that helped better understand the relationships between the keywords and groups, facilitating a deeper interpretation of the results obtained.

3.2.3. ChatGPT for Classification

ChatGPT, developed by OpenAI, is a powerful natural language processing system powered by artificial intelligence (AI) [64]. It belongs to the GPT (Generative Pre-trained Transformer) series, which has stood out for training massive language models on large amounts of text to learn linguistic patterns and complex grammatical structures [65].

The latest known free version, GPT-5.2, also called ChatGPT, represents a milestone in the evolution of AI [66]. This model has been able to generate coherent and relevant text thanks to its architecture based on neural networks known as Transformers, which allows it to understand and generate text based on the provided context [65]. The GPT-5.2 model, like its predecessors, is characterized by its "pre-training" approach [64,67]. This means that the model is trained on large amounts of text without a specific task in mind. During the training process, the model learns linguistic patterns, semantic relationships, and general encyclopedic knowledge [68].

The Transformer architecture is notable for its self-attention mechanism and its encoder-decoder layers [69]. The self-attention mechanism allows the model to focus on relevant parts of the input text by assigning weights to different words based on their relevance to understanding the context [70]. Additionally, GPT-5.2 has a massive structure consisting of 175 billion parameters [71]. This makes it one of the largest and most powerful language models ever created, allowing it to generate highly coherent and contextually relevant text [72].

The pre-training process of ChatGPT involves presenting the model with a large amount of text collected from the web. During this process, the model tries to predict the next word or token in a given sequence, allowing it to learn the probability of occurrence of different words based on the context [73].

Once the model has been pre-trained, it can be fine-tuned for specific tasks by providing labeled examples [74]. This allows developers to tailor the model for specific tasks, such as answering questions, translating languages, generating creative text, and, as in this work, adding an additional layer of relevant information to the Systematic Literature Review [75].

It is important to note that while GPT-5.2 is a marvel of artificial intelligence, it still has limitations and challenges. For example, it can generate plausible but inaccurate responses and may be sensitive to biases present in the training data [76]. These issues are the subject of ongoing research and development to improve the model's accuracy and fairness; in fact, at the time of this research, there are newer versions available for purchase [77].

The use of ChatGPT is diverse and ranges from virtual assistants to chatbots and text generation for various applications [78]. Users can interact with it through text commands (prompts), and the model will respond in a contextually relevant manner [78]. One of the key concepts in using ChatGPT is "prompts", or text commands used to guide text generation [77]. Prompts are instructions or questions provided to the model to direct its output based on the specific task to be performed [65]. Proper use of prompts is essential to obtaining accurate and relevant responses [78].

Optimizing the use of prompts involves carefully choosing words and grammatical structures to clearly communicate the desired task or question [71]. It may also require adding specific constraints or indications to avoid inappropriate or irrelevant responses. Iteration and experimentation with different prompts are common practices to refine the desired text generation [71].

According to [65], although language models like ChatGPT can provide valuable information and complementary analysis, it is important to keep in mind that in certain cases, the generated information may not be completely reliable. This is due to several reasons:

1. *Training data:* Language models like ChatGPT learn from large amounts of textual data available online. If these data contain incorrect or biased information, the model may generate responses that reflect those limitations.
2. *Pattern-based generation:* Language models do not understand context in the same way humans do. They often generate responses based on patterns present in the training data, even if the response is not coherent or accurate.
3. *Unverified information:* Language models cannot verify the accuracy of the information they generate. They may produce statements that sound convincing but are not supported by reliable evidence.
4. *Bias and neutrality:* Models can capture biases present in the training data and replicate them in the generated responses. This can lead to responses that reinforce biases or stereotypes.

To mitigate these issues, it is crucial to properly train and fine-tune language models before using them for specific tasks [76]. This involves providing relevant and correct examples and guiding the model toward generating accurate and reliable responses [77]. It is also essential to cross-check the results generated by the AI tool with information from reliable sources and experts in the field [78]. When using artificial intelligence as a support tool, it is necessary to exercise critical judgment

doi:10.20944/preprints202507.1477.v2

and verify the consistency and accuracy of the generated information to ensure the integrity of analyses and results [74].

It is important to highlight that the use of language models like ChatGPT for text analysis complements traditional techniques such as those used with IraMuteq software and allows for more robust and detailed results, especially in large datasets and in research that requires a more sophisticated and automated approach. However, it is also important to note that these language models should be used with caution and that the results obtained should be validated through additional analyses and expert review in the field of study.

4. Results

4.1. Data Preparation for Text Mining

To generate the data matrix used in the analysis, certain methodological adjustments were made. First, articles were obtained from the prestigious academic databases Scopus and Web of Science. These articles were selected focusing on the fundamental fields for our study, including the author's name, the year of publication, the source of publication, the title of the article, and, most importantly, the abstract. It is important to note that the text mining analysis will be based on the latter field, as it provides a concise but informative summary of the article's content.

Following the methodological guidelines established by Caballero *et al*. [79], a lexical table was constructed using a specific approach. A plain text file was created that includes a unique identifier for each article, accompanied by its respective abstract. It is crucial to mention that, in this context, it is pertinent to perform additional cleaning of the abstracts. Occasionally, some journals add copyright clauses at the end of the abstracts that are not the focus of the text mining analysis. Therefore, it is recommended to exclude this content to ensure the integrity and homogeneity of the analysis.

Additionally, it has been observed that the IraMuteq software tends to omit the first abstract during the analysis. To address this issue, it is suggested to duplicate the first abstract in the plain text file. This measure ensures that the analysis with IraMuteq software is conducted completely and accurately, without losing valuable information that may be contained in the first abstract of the selected articles.

4.2. Results of Text Mining

Once the textual corpus under study was loaded into the IraMuteq software, we present the results obtained, starting with the similarity analysis, based on graph theory. A graph, in this context, represents a set of vertices (which correspond to words or forms) and edges (which represent the relationships between them). The purpose of this analysis is to investigate the proximity and relationships between the elements of a set, optimizing the simplification of the number of connections until a connected acyclic graph is obtained. This is characterized by being a closed path in which no vertex is repeated, except for the first one, which acts as both the start and end of the path, as illustrated in Figure 5.

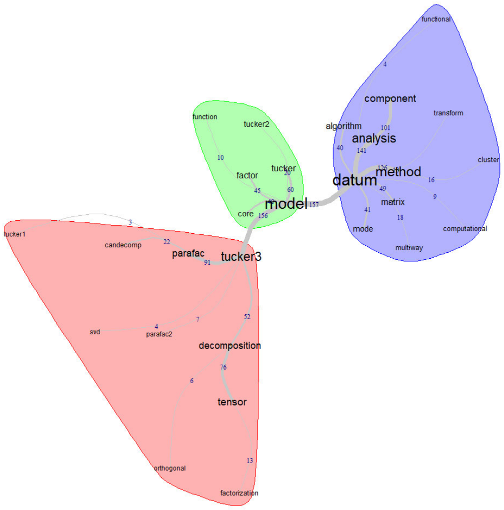

Figure 5. Similarity graph of the textual corpus obtained by IraMuteq software.

As can be seen, to obtain the graph shown, some adjustments were made in the IraMuteq software, considering words that, in addition to having high frequencies, are connected to the topic in question, the Tucker models. In the similarity graph, three clusters are observed, represented by different colors. Within these clusters, there are larger words (called the maximum tree of each cluster that interconnects them: “model”, “datum”, and “tucker3”). At the same time, within each cluster, these higher-frequency words are connected to other words that had strong connections in the textual corpus.

In the central cluster, we have the term “model”, which acts as the central or most prominent node. Within this cluster, some terms are strongly connected, such as: “factor”, “tucker”, “tucker2”, “function”, and “core”; this suggests the presence of factorial analysis and mathematical functions for Tucker models, particularly Tucker-2, where the core tensor, also known as the core matrix, is involved.

In the left cluster, we identified the term “tucker3”, which acts as the maximum tree and is strongly associated with other words such as: “parafac”, “parafac2”, “decom- position”, “sdv”, “tensor”, “tucker1”, “candecomp”, “orthogonal”, and “factorization”. For example, “parafac” and “parafac2” refer to variants or extensions of the PARAFAC model, while “decomposition” indicates tensor decomposition. On the other hand, “candecomp” refers to CANDECOMP/PARAFAC decomposition and also indicates the orthogonality property associated with certain decomposition methods.

The third cluster, located on the right, has “datum” as the maximum tree, which is connected to a series of keywords such as: “method”, “analysis”, “component”, “transform”, “functional”, “cluster”, “multiway”, “computational”, “matrix”, “mode”, and “algorithm”. This structure suggests a variety of related concepts and terms. For example, it indicates different approaches or methods for data analysis, which are related to decomposing data into simpler components and data transformation for various analysis purposes, including cluster analysis.

4.3. Data Preparation for MANOVA-Biplot

One of the primary functions of IraMuteq software lies in its ability to generate a lexical table or matrix called *tableafmc*, which can later be subjected to Correspondence Factor Analysis (CFA) as proposed by Benzecri [3], a statistical technique used to analyze contingency tables to explore relationships between the categories of two categorical variables. In the context of textual analysis, this table shown in Table 1 incorporates the identifiers assigned to each article as variables and the words identified from the analyzed corpus as rows.

Table 1. Partial result of Correspondence Factor Analysis from IraMuteq software.

	***ID_1**	***ID_10**	**...**	***ID_99**
similarity	0	0	...	0
dynamic	0	0	...	0
calculate	0	0	...	0
risk	0	0	...	0
solution	0	0	...	0
batch	6	0	...	0
approximation	0	0	...	0
tucker3	2	0	...	0
…	…	…	…	…
validation	0	0	…	0

The lexical table is a fundamental representation that allows us to analyze and understand the frequency of keywords in the corpus of articles related to Tucker models.

Through this table, the distribution of words and their relevance in each publication period can be visualized.

Constructing the lexical table involved transforming the corpus of articles into a matrix where the rows represent the articles and the columns represent the unique words present in the corpus. For each cell in the matrix, a numerical value is assigned indicating the frequency of the word's appearance in the corresponding article. This process provides a quantitative view of the most relevant words in the dataset.

Using the IraMuteq software, we were able to analyze this lexical table and generate a visual representation of the distribution of keywords across different publication periods. Lexical analysis techniques, such as *Correspondence Factor Analysis*, allowed us to identify patterns and trends in keyword frequency over time.

Following the recommendations of Caballero *et al*. [79], a characterization factor was applied using the formula:

$$f_{ij}^{*} = \frac{f_{ij}}{\sqrt{max\, f_i}\sqrt{max\, f_j}} \qquad (10)$$

where f_{ij} represents the frequency of a word located in row i and column j in the *tableafmc* matrix. This transformation aims to simplify the matrix and enhance subsequent analysis.

To perform this characterization, the integrated development environment *RStudio* [80] was used, where the tableafmc matrix was initially transposed and the characterization formula applied. The result was saved in a new file called data, which was stored in Excel format. To further enrich the dataset, a column indicating the publication year corresponding to each article in the original database was added. This step can be easily performed using the filters available in the software, as shown in part in Table 2.

Table 2. Simplified characterized frequency table.

ID	**similarity**	**dynamic**	**calculate**	**…**	**validation**
*ID_1	0	0	0	...	0

*ID_10	0	0	0	...	0
*ID_100	0	0.169	0	...	0
*ID_101	0	0	0	...	0
*ID_102	0	0	0	...	0
…	…	…	…	...	…
*ID_99	0	0	0	...	0

Furthermore, for conducting a group study following the MANOVA-Biplot approach, an additional column numbered from 1 to 4 was added to the data file, indicating the group to which each article belongs. This classification is based on a date table specifying the publication time intervals, where each number represents a different period, as shown in Table 3.

4. 4 Results of the MANOVA-Biplot

Once the groups corresponding to the publication periods of the articles were defined, in our case, we have 4 groups, each spanning 6 and 7 years according to Table 3, chosen to maintain homogeneity concerning the number of years analyzed.

Table 3. Correspondence between groups and article publication periods .

Group	**Publication Period**
1	January 1, 2000 - December 31, 2005
2	January 1, 2006 - December 31, 2011
3	January 1, 2012 - December 31, 2018
4	January 1, 2019 - December 31, 2025

We then proceeded to load the database called data into the *MultBiplot* software [81] version *18.0312*, a tool developed in the Department of Statistics at the University of Salamanca, widely used in multivariate data analysis. In this context, the software allows for the import of data from Excel and the definition of the group variable as nominal, along with the specification of the 4 categories corresponding to that variable. Subsequently, we performed the one-way MANOVA-Biplot analysis.

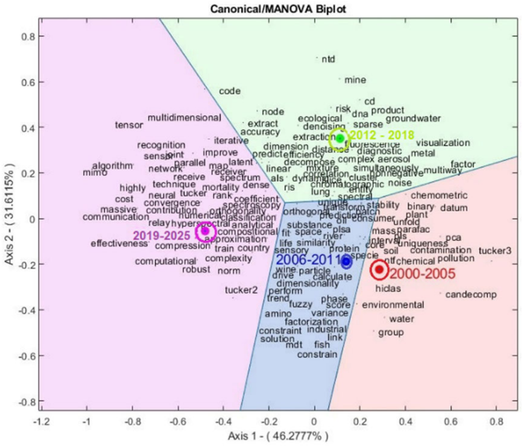

Figure 6. One-way MANOVA-Biplot obtained from the characterized lexical table.

After running the analysis in the software, we made various modifications to adjust and improve data visualization, thus obtaining a more accurate representation, as shown in Figure 6. In this regard, advanced visualization techniques were implemented, including Voronoi regions and confidence circles. These regions, named after the Russian mathematician Georgy Voronoi, are used to delimit areas on a plane based on proximity to a given set of points, allowing for better interpretation and understanding of the groups identified in the analysis [82]. On the other hand, the confidence circles shown highlight the significant differences between the means since they do not intersect.

Table 4 details how the main dimensions capture variability in the data. The first dimension or first canonical variable has an eigenvalue of 20.49 and an explained variance of 46.28%, meaning it explains 46.28% of the variability between groups. This dominant factor captures nearly half of the variability between groups, while the second component explains 31.61%; in other words, the first two canonical variables explain 77.89% of the total variability, suggesting a deep understanding of the data structure. Additionally, the p-values associated with the canonical variables are nearly zero, indicating high statistical significance and confirming the robustness of the model. This detailed evaluation provides a comprehensive view of the underlying relationships between the two sets of variables, thus supporting the validity and reliability of the analysis.

Table 4. Eigenvalues and Variability between groups explained and cumulative.

Dimension	Eigenvalue	% Expl.	Cumm.	p-val.
1	20.49	46.28	46.28	0
2	16.94	31.61	77.89	0
3	14.17	22.11	100	0

The results in Table 5, *Quality of representation of group means*, show how the different year groups are represented on the axes of the canonical biplot analysis. Each group corresponds to a specific period (e.g., “2000-2005”, “2006-2011”, etc.), and the values in the “Axis 1” and “Axis 2” columns represent the coordinates of that group on the two main axes of the Biplot; additionally, we multiplied their original values by 1000 and added an accumulated quality column to facilitate interpretation.

Table 5. Quality of representation of group means x1000 .

Group	Period	Axis 1	Axis 2	Cum

1	2000-2005	340	209	549
2	2006-2011	153	282	435
3	2012-2018	93	907	1000
4	2019-2025	975	13	988

Groups 3 and 4, i.e., the years 2012-2018 and 2019-2025, present the highest qualities of representation on the first two factorial axes (1000 and 988, respectively), followed by group 1 (549) and finally group 2 (435) with relatively low quality compared to the other groups. This suggests considering that this group will be better represented on axes 2 and 3. However, in general terms, we have a very good quality representation of the canonical groups on the first two axes.

The graph of the Canonical Biplot analysis is presented, visualizing the relationships between groups corresponding to different year periods. Each group is represented in the Biplot by a set of coordinates on the main axes. This analysis allows us to observe the trends and patterns in the evolution of keywords over time concerning Tucker models. Focusing on the result of the MANOVA-Biplot analysis obtained in Figure 6, four regions representing the publication periods of the articles are identified: “2000-2005”, “2006-2011”, “2012-2018”, and “2019-2025”. A notable feature is the significant increase in the number of keywords as we move forward in time. This suggests an expansion in the scope and focus of research, with diversification in the thematic areas addressed concerning Tucker models.

It is also observed that there are many similarities between the groups of articles from the periods 2000-2005 and 2006-2011 (groups 1 and 2) since they are very close on the graph; on the other hand, groups 3 and 4 are far from the others, suggesting that the topics covered in those periods are significantly different but within the scope of Tucker models.

In the period “2000-2005”, the inclusion of the terms “hiclas” and “binary” suggests a focus on data organization and hierarchy, which corresponds to hierarchical classes of binary data through Tucker models. This indicates the exploration of complex structures and relationships between variables. The appearance of “Candecomp” refers to CANDECOMP (Canonical Decomposition); in addition, terms such as “Tucker3”, “PCA”, and “PARAFAC”, which are fundamental techniques in Tucker models, are included. Their inclusion in this period indicates continued interest in researching and developing these techniques in the early years of this millennium, although they are models developed in the last century.

The inclusion of terms such as “pollution”, “chemical”, “chemometric”, “water”, and “contamination” suggests a focus on applying Tucker models in contexts related to the assessment and modeling of pollution and water quality. This indicates an interest in analyzing environmental and health-related data during that period.

In group 2, period “2006-2011”, the repeated presence of terms such as “dimensionality” and “transform” indicates a concentration on generating and transforming data using Tucker models during this period. These terms suggest a focus on the manipulation and conversion of data into useful and analyzable formats.

The presence of the term “fuzzy” indicates a focus on studying Tucker models with imprecise or fuzzy data. Taken together, these findings suggest an emphasis on innovation and advancement in tensor analysis during this specific period, highlighting the importance of exploring and understanding the complexities of Tucker models to address challenges in various application areas.

During the period 2012-2018, the published articles present a varied focus covering different aspects of the research. There is an interest in updating methods and techniques, such as “sparse”. Additionally, the term “cd” suggests an interest in studying Tucker models for compositional data.

There is also an evident focus on data visualization (“visualization”) and dimension estimation (“dimension”), which suggests an emphasis on understanding and visually representing information; these terms also suggest the study of the number of components or dimensions to retain for each mode in Tucker models. The presence of terms like “dna”, “efficiency”, “diagnostic”, "fluorescence", "mine", "ecological", and "denoising" indicates a diversity of application areas, ranging from environmental research to molecular biology. These terms reveal multidisciplinary and multifaceted research during this period, covering everything from the refinement of analytical techniques to the application of tensor analysis in various fields of study.

Finally, in the group corresponding to the period "2019-2025", there is a concentration of keywords such as "robust", "highly", "algorithm", "multidimensional", "massive", "network", "numerical", "mimo", and "communication". These keywords reflect a technically and conceptually advanced landscape concerning Tucker models in the most recent period. There is a focus on areas such as advanced algorithms, image processing, computational efficiency, pattern recognition, and neural network applications in communication, among others.

The knowledge development line in Tucker models has evolved from their fundamental development to their application in complex and dynamic problems across various fields. Over the years, a deeper understanding of Tucker models and their integration with other analytical methods has been achieved, leading to significant advances in multidimensional data analysis and understanding patterns and trends over time. Up to this point in the study, we have conducted a traditional, subjective, and superficial interpretation of the scientific contributions of the new millennium within the context of Tucker models. In this sense, it is essential to add an additional layer of information using artificial intelligence.

4. 5 How Can AI Enhance the Results of Text Mining?

Artificial intelligence, specifically ChatGPT-5.2, enhances text mining on article abstracts related to Tucker models by offering a deep interpretation of the context, generating additional summaries that highlight key points, identifying emerging trends and patterns, providing practical examples of application, and analyzing fundamental opinions. This facilitates a more comprehensive understanding of research in this field and guides future research areas with greater clarity and precision.

To maximize the effectiveness of ChatGPT-5.2 in text mining for Tucker models, it is crucial to design well-structured prompts that ensure accuracy, coherence, and relevance in the generated responses. The way a prompt is formulated significantly influences the quality of insights derived from the AI model. Table 6 presents key strategies for optimizing prompt design, including specificity, step-by-step structuring, the use of examples and constraints, iterative refinement, and bias control. Implementing these strategies enhances the precision of text analysis, reduces irrelevant information, and allows for a more structured extraction of meaningful patterns from the literature.

Table 6. Strategies for Prompt Design in ChatGPT-5.2 and their impact .

Strategy	**Description**	**Impact on results**
Specificity and Context	Provide clear details about the task and objective of the query to reduce ambiguities.	Improves accuracy and relevance of the response.
Step-by-Step Structuring	Divide the query into sequential steps to better	Produces more organized and coherent responses.

	guide content generation.	
Use of Examples and Constraints	Include specific examples and limit the scope of the response to avoid irrelevant information.	Reduces generic responses and improves information relevance.
Iterative Refinement	Reformulate prompts based on the initial output to improve ac- curacy and depth.	Adjusts the generated content to be more precise and aligned with the objectives.
Bias Control and Cross-Validation	Evaluate responses from different perspectives and verify information with external sources.	Minimizes errors, biases, and ensures more reliable information.

To conduct this study, we accessed the OpenAI website (*www.chat.openai.com*) and used the ChatGPT version 5.2 platform. It is recommended that users first create an account on the platform to establish the proper use of artificial intelligence. Then, we entered the initial prompt while maintaining the structure shown in Table 7.

Table 7. Initial Prompt for article abstract analysis using ChatGPT-5.2.

Prompt	**Description**
Prompt 1	I will provide a set of research article abstracts. Your task is to analyze them and extract key insights related to Tucker models.
Prompt 2	I will provide you with a table containing information from 20 articles. The first column of this table is the year of publication, the

	second column contains the authors, and the third column contains the corresponding abstract of each article. In chronological order, identify and summarize the main findings and present the information grouped by year.

With this set of instructions to ChatGPT, capturing the main findings of each article is ensured. As shown in Figure 7, ChatGPT is initially prepared to receive the information from the corresponding articles for each group. Specifying the format of the provided data is important to subsequently create a timeline to capture the main findings detected. Additionally, due to the processing limitations of this free version of ChatGPT-5.2, the dataset to be entered was divided into groups of no more than 20 articles per entry until the total was reached.

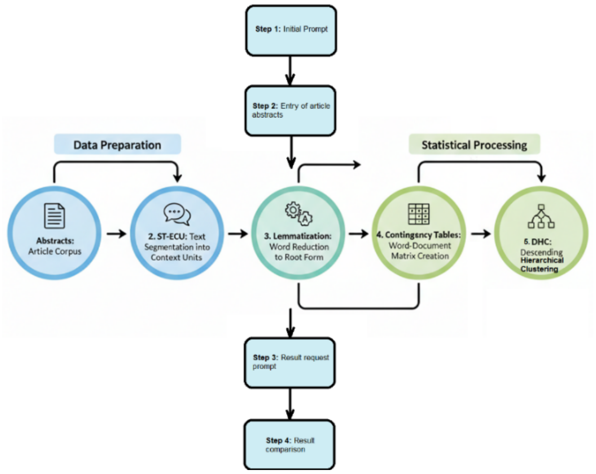

Figure 7. Process of textual corpus analysis with ChatGPT-5.2.

Once the abstracts were entered, the artificial intelligence identifies the main findings. These results are then compared with the information captured in the MANOVA-Biplot analysis. This multi-layered validation ensures consistency, highlights key patterns, and identifies potential discrepancies between AI-generated insights and traditional statistical approaches. This process is crucial because ChatGPT can generate erroneous results, potentially biased by the user's prior information, query region, language, and other factors.

4. 6 ChatGPT-5.2 results for textual analysis of articles related to Tucker Models in the period 2000-2025.

Following the prompt design strategies outlined in Section 4.5, the initial prompts shown in Table 7 were entered into the ChatGPT-5.2 platform. Subsequently, the abstracts of the 288 articles were introduced in chronological order, grouped in batches of no more than 20 articles per session to ensure processing accuracy. The model then generated year-by-year summaries of the main findings related to Tucker models.

Table 8 presents a representative sample of the outputs obtained for the initial years of the analysis (2000-2002). The complete set of results, covering the entire 2000-2025 period, was generated following the same procedure.

Table 8. Sample of ChatGPT-5.2 outputs for article abstracts analysis (2000-2002) .

Year	Authors	Main findings
2000	Louwerse & Smilde	Extended the theory of batch-process MSPC control charts and developed improved charts based on Unfold-PCA, PARAFAC, and Tucker3. They showed that performance depends on the fault type and the chosen model/chart.
2000	Timmerman & Kiers	Proposed DIFFIT to select the optimal number of components in trilinear PCA (3MPCA). In simulations, DIFFIT more accurately identified the correct number of components than competing methods.
2001	Gurden et al.	Presented a “grey” model combining known and unknown information about a chemical system using Tucker3, illustrated with batch chemical reaction data.
2001	Bro et al.	Highlighted fundamental differences between PARAFAC and Tucker3 and proposed a modification of the multilinear partial least squares regression model.

2001	Simeonov et al.	Applied multivariate statistical methods to drinking water data to characterize water quality across different systems (multiway/multivariate analytical context).
2001	De Juan & Tauler	Compared three-dimensional chemical data resolution methods and suggested Tucker3 and MCR-ALS are suitable for non-trilinear data.
2001	Estienne et al.	Used Tucker3 to extract useful structure from complex EEG data and to identify irrelevant sources of variation.
2002	Gemperline et al.	Applied Tucker3 to multidimensional fish-spawning catch data to identify migratory patterns.
2002	Amaya & Pacheco	Described the Tucker3 method for three-way data analysis and demonstrated its practical application.
2002	Lopes & Menezes	Compared PARAFAC and Tucker3 for industrial process control, showing that model choice affects monitoring performance and interpretability.

4. 7 *Timeline of the Development of Tucker Models in the New Millennium*

Overall, the period from 2000 to 2025 has witnessed significant progress in the field of tensor decomposition and multidimensional data analysis. Researchers have achieved advances in theory, algorithms, and practical applications, leading to a greater understanding and use of Tucker models across various disciplines. Figure 8 presents the main findings related to Tucker models in the new millennium, resulting from Text Mining based on Biplot and ChatGPT-5.2.

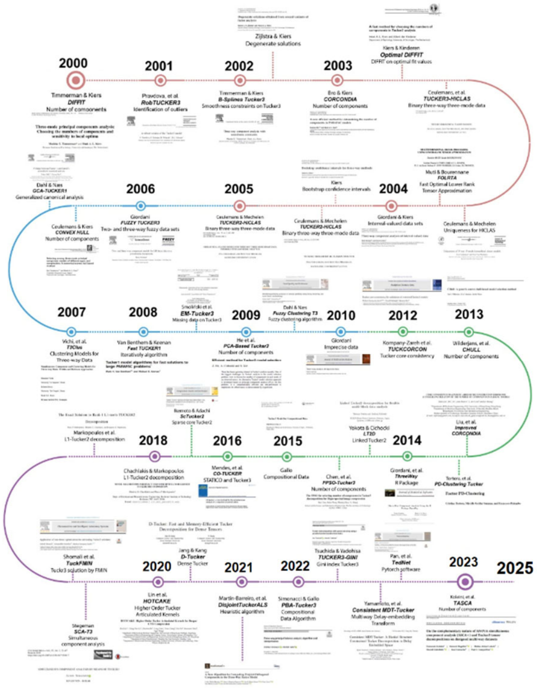

Figure 8. Some prominent research on Tucker models from 2000 to 2025.

At the beginning of this new millennium, Timmerman and Kiers in 2000 [12] introduced a new method, called **DIFFIT**, that determines the optimal number of components to use in the Tucker-3 model, a challenge that persists to this day. The effectiveness of DIFFIT is compared with two methods based on two-way PCAs through a simulation study, concluding that DIFFIT significantly outperforms the other methods in indicating the correct number of components. Furthermore, the sensitivity of the TUCKALS3 algorithm in estimating the 3MPCA is examined, finding that, when the number of com- ponents is correctly specified, it avoids local optima. However, the occurrence of local optima increases with the discrepancy between the underlying components and those estimated by TUCKALS3, with fewer local optima occurring in runs initiated rationally compared to random starts. The following year, Pravdova, et al. [81] presented a robust method, called **RobTUCKER3**, for identifying outliers in the Tucker-3 model, highlighting its effectiveness in both simulated and real data.

In 2002, there were two important contributions. On one hand, Timmerman and Kiers [83] proposed a three-dimensional component analysis with smoothness constraints, which improved the estimation of model parameters in functional data. In this research, smoothness conditions were posed using the family of B-Splines functions. On the other hand, [84] investigated degenerate solutions, highlighting the importance of unique com- ponents in generating stable and interpretable solutions.

Methodological development was also highlighted by Bro and Kiers in 2003 [85], who introduced a new efficient method, the core consistency diagnostic called **CORCON- DIA**, to determine the appropriate number of components in multi-dimensional models such as PARAFAC. This tool proved effective in selecting appropriate models, although it was noted that the theoretical understanding was not yet complete. This diagnostic is based on evaluating the “appropriation” of the structural model concerning the data and the estimated parameters of gradually increased models, and it has proven to be an effective tool in determining the number of components. Continuing with the challenge of the number of components, Kiers and Kinderen, 2003 [86] proposed a quick method for choosing the number of components in Tucker-3 analysis, the improved DIFFIT. Their alternative procedure is based on a single, fast analysis of the three-dimensional dataset, providing comparable effectiveness results to the original method proposed by Timmerman and Kiers in 2000.

Regarding binary data, Ceulemans, et al., 2003 [87] proposed a new model for 3-way 3-mode, called **Tucker3-HICLAS**. Along the same lines, Ceulemans and Mechelen, 2003 [88] investigated and demonstrated two uniqueness theorems for hierarchical class models in N-way N-mode data, that is, for multi-way binary data. Additionally, Muti and Bourennane, 2003 [89] presented a new method for the approximation of low-rank tensors, known as **FOLRTA**, highlighting its efficiency in noise reduction in applications such as color image processing and seismic signals. These studies underline the diversity of approaches to addressing problems related to multi-way models and the importance of efficient methods for determining the number of components in multidimensional data analysis.

Regarding variants of three-dimensional component analysis, Giordani and Kiers, 2004 [90] extended principal component analysis to three-dimensional data with interval values, applying methods such as Tucker-3 and CANDECOMP/PARAFAC. Simultaneously, Kiers, 2004 [91] introduced a bootstrap procedure to provide confidence intervals for all output parameters in three-dimensional analysis, enabling the evaluation of the stability of the obtained solutions.

Ceulemans focused several of his studies on binary data; thus, in 2004 [92] introduced a new hierarchical class model called **Tucker2-HICLAS**, and the following year [93] they presented **Tucker1-HICLAS**, thereby addressing binary data for the various modes. These models include a hierarchical classification of the elements of each mode, highlighting a distinctive feature that retains the differences between the association pat- terns of the elements in each of the modes.

Giordani, 2006 [94] addressed the problem of data reduction for fuzzy data, introducing several component models to handle two-way and three-way fuzzy datasets. For Tucker models, this includes the **Fuzzy Tucker3** model. The two-way models are based on classical principal component analysis (PCA), while the three-way models are based on three-dimensional generalizations of PCA, such as Tucker-3 and CANDECOMP/PARAFAC. These models leverage possibilistic regression, treating component models for fuzzy data as regression analyses between a set of observed fuzzy variables (response variables) and a set of unobserved crisp variables (explanatory variables). On the other hand, Ceulemans and Kiers, 2006 [95] addressed the selection of three-mode principal component models of different types and complexities. They proposed a numerical method based on a convex hull, known as **CONVEX HULL**, to select the appropriate model and complexity for a specific dataset, showing that this method works almost perfectly in most cases, except for Tucker3 data arrays with at least one small mode and a relatively large amount of error.

Regarding canonical analysis techniques, Dahl and Næs, 2006 [96] presented the **GCA-Tucker1** model, which analyzes relationships between and within multiple data matrices, unifying Carroll's Generalized Canonical Analysis (GCA) methods with the Tucker1 method for principal component analysis of multiple matrices. Meanwhile, Vichi et al., 2007 [97] implemented clustering techniques for three-way data, presenting the **T3Clus** model.

In the development of algorithms that accelerate computational performance for large volumes of three-way data, Van Benthem and Keenan, 2008 [98] described the **Fast- TUCKER1** method for performing trilinear analysis on large datasets using a modification of the PARAFAC-ALS algorithm based on the Tucker1 model, allowing for a solution 60 times faster for massive PARAFAC data problems that do not fit in the computer's main memory, resulting in a significant improvement in performance. Meanwhile, Smolinski et al., 2008 [99] proposed using the expectation-maximization (EM) algorithm on the Tucker-3 model, called **EM-Tucker3**, to handle missing data in chemometric analyses. Additionally, He et al., 2009 [100] presented an efficient approach for model se- lection in Tucker-3 analysis based on principal component analysis (PCA), **PCA-Based Tucker3**, demonstrating computational efficiency and effectiveness in selecting the number of components for

doi:10.20944/preprints202507.1477.v2

each mode. Meanwhile, Dahl and Næs, 2009 [101] introduced a method, **Fuzzy Clustering T3**, for identifying atypical evaluators in descriptive sensory analysis based on fuzzy clustering and multiple block methodology.

Giordani, 2010 [102] proposed an approach for the analysis of three-way imprecise data, focusing on transforming these data into fuzzy sets for further analysis. This method involves using generalized Tucker-3 and CANDECOMP/PARAFAC models to examine the underlying structure of the resulting fuzzy sets. Additionally, the statistical validity of the identified structure is evaluated using bootstrapping techniques, providing a robust tool for summarizing samples of uncertain data. On the other hand, Kompany-Zareh et al., 2012 [103] extended the core consistency diagnostic to validate restricted Tucker3 models, now with the **TUCKCORCON** model, offering a tool to assess the adequacy of the constraints applied to Tucker-3 analysis.

In 2013, methods continued to be developed for determining the number of components for each mode. Wilderjans et al., 2013 [104] proposed the model called **CHULL**, which addresses the model selection problem researchers face when analyzing data. This problem includes determining the optimal number of components or factors in principal component analysis (PCA) or factor analysis, as well as identifying the most important predictors in regression analysis. Meanwhile, Liu et al., 2013 [105] proposed a method to improve accuracy in determining the number of components in PARAFAC models, using a core consistency diagnostic approach with error reconstruction called Improved **CORCONDIA**. Tortora et al., 2013 [106] extended probabilistic distance clustering to the context of factorial methods, **PD-Clustering Tucker**, combining Tucker-3 decom- position with PD clustering for greater stability in clustering complex data.

Giordani et al., 2014 [107] presented the R package, named **ThreeWay**, highlighting its main features and functions, such as T3 and CP, which implement the Tucker-3 and Candecomp/Parafac methods, respectively, for analyzing three-dimensional arrays. Yokota and Cichocki, 2014 [108] proposed the linked Tucker-2 decomposition model (**Linked Tucker2 - LT2D**) for flexible analysis of multi-block data, demonstrating its efficacy and convergence properties. On the other hand, Chen et al., 2014 [109] proposed the **FPSO-Tucker3** method to select the number of components in Tucker-3 decomposition for hyperspectral image compression, using a particle swarm optimization- based approach (FPSO).

Gallo, 2015 [110] described the compositional data Tucker-3 analysis, highlighting the importance of considering the specific characteristics of this type of data. Mendes et al., 2017 [111] proposed the **CO-TUCKER** method for the simultaneous analysis of a sequence of paired tables, offering a new perspective for understanding structure-function relationships in ecological communities. Additionally, Ikemoto and Adachi [112] developed a procedure to find an intermediate model between Tucker-2 and Parafac for analyzing three-dimensional data. In the proposed model, called "**Sparse Core Tucker2**" (**ScTucker2**), the Tucker-2 loss function is minimized subject to a specified number of zero core elements, whose locations are unknown; the optimal locations of the zero elements and the values of the non-zero parameters are simultaneously estimated. An alternating least squares algorithm is presented for ScTucker2 along with a procedure to select an appropriate number of zero elements. Additionally, Tortora et al., 2016

[113] proposed the **FPDC-Tucker3** model, which consists of clustering factors using probabilistic distances.

Chachlakis and Markopoulos, in 2018, proposed an approximate algorithm for the **L1-TUCKER2** decomposition of three-dimensional tensors, highlighting its resistance to data corruption compared to conventional methods [114]. Additionally, they developed exact and efficient algorithms for this decomposition, demonstrating their robustness against outliers and outperforming counterparts based on L2 and L1 norms [115]. On the other hand, Shomali et al. applied nonlinear optimization techniques to estimate Tucker-3 solutions by minimizing the objective function (**TuckFMIN**), while Stegeman introduced a model for the simultaneous analysis of components using Tucker-3, demonstrating its effectiveness across various datasets [116,117]. Additionally, Lin and collaborators proposed **HOTCAKE** for compressing convolutional neural networks, while Jang and

doi:10.20944/preprints202507.1477.v2

Kang presented **D-Tucker**, a fast and efficient method for the decomposition of dense tensors, noted for its speed and memory efficiency [118,119].

Martin-Barreiro and collaborators, in 2021, proposed a heuristic algorithm for computing disjoint orthogonal components in the three-way Tucker model, known as **DisjointTuckerALS**, thereby facilitating the analysis of three-dimensional data and the interpretation of results [120]. Meanwhile, Simonacci and Gallo, in 2022, presented **PBA-Tucker3** as a three-way principal component analysis using Tucker-3 decomposition, with applications in the composition of academic recruitment fields by macro-region and gender/role in Italy [121]. Tsuchida and Yadohisa proposed a Tucker-3 decomposition with a sparse core using a penalty function based on the Gini index, known as **TUCKER3-GINI**, allowing for a simpler interpretation of the results [122]. Additionally, Yamamoto and collaborators presented **Consistent MDT-Tucker**, a low dimensional tensor decomposition method that preserves the Hankel structure in data transformed by Multiway Delay-embedding Transform [123]. Pan and collaborators developed **TedNet**, a **PyTorch** tool for tensor decomposition networks, implementing five types of tensor decomposition in traditional deep neural network layers, facilitating the construction of various neural network models [124]. Finally, Koleini and collaborators, in 2023, demonstrated the complementary nature of simultaneous analysis of variance components (**ASCA+**) and Tucker-3 tensor decompositions in designed datasets, showing how ASCA+ can identify statistically sufficient and significant Tucker-3 models. This model is called **TASCA**, thereby simplifying the visualization and interpretation of the data [125].

Methodological advancements in 2024 included Han et al.'s **sparsity-enhanced Tucker-2** model for fMRI data, which improved schizophrenia classification by 7% [126]. Pagani et al. [127] extended DD-SIMCA to multiway data using Tucker-3 for one-class classification of adulterated food samples. Frutos-Bernal et al. [87] proposed **Tucker3-PCovR** for simultaneous dimension reduction and prediction with biplot interpretation. Kirch et al. combined Tucker3 with joint plots and Procrustes analysis for multi-response genotype evaluation [128].

Zhang et al. (2025) developed **TDSA-SuperPoint**, a novel scene matching navigation method using Tucker-2 decomposition to reduce channel dimensions in neural network weights, achieving 58% fewer parameters, 18% higher efficiency, and 50.45% better accuracy compared to traditional SuperPoint [129]. Graham et al. [130] introduced a constrained Tucker-1 tensor decomposition for multivariate geochronology data to trace sediment sources, validated on detrital zircon data from the Andes. Lestari et al. [131] investigated Tucker3 tensor decomposition for the standardized residual hypermatrix in three-way correspondence analysis, presenting mathematical properties and algorithms for constructing correspondence plots. Schenone et al. [132] applied a sparse Tucker3 clustering framework to analyze quality of life across 174 cities (2016-2024), identifying distinct urban clusters based on environmental, economic, and social indicators. Frutos-Bernal et al. [8] proposed T3-PCA, a global model for simultaneous analysis of coupled three-way and two-way data using Tucker3 and PCA, demonstrating superior performance over sequential strategies in simulation studies and anxiety data applications.

Saylor et al. [133] applied non-negative Tucker-1 decomposition to multivariate detrital zircon data from till samples in British Columbia, successfully identifying two endmembers corresponding to non-oxidized (low Cu-ore potential) and oxidized-hydrous (high Cu-ore potential) igneous sources, with Source 2 proportions decreasing with distance from ore bodies. Martyna et al. [134] developed a hybrid likelihood ratio model coupled with Tucker3 decomposition for forensic comparison of weathered diesel fuel samples in fire investigations, where Tucker3 decomposed GC-MS data into GC, MS, and sample modes, enabling correct source identification regardless of weathering state. Ginige et al. [135] proposed novel joint channel estimation and prediction strategies for beyond diagonal RIS-assisted MIMO systems using Tucker2 decomposition with bilinear alternating least squares, achieving 98% reduction in pilot overhead and robust performance under channel aging.

Mao et al. [136] introduced a tensor-based approach for joint hybrid beamforming and artificial noise design in secure mmWave MU-MIMO-OFDM systems, formulating a Tucker2 decomposition subproblem for analog beamforming and employing HOSVD-based statistical null-space algorithm for artificial noise precoding. Gu et al. [137] presented TensorTrack, a video object tracking method combining L1-norm tensor decomposition for background subtraction and Tucker2 decomposition for integrating appearance and motion patterns, achieving 15.8% improvement in AUC and ten-fold speed increase over deep learning methods on OTB100 benchmark. Cascante-Yarlequé et al. [138] proposed a methodological framework for three-way analysis of cost of living and quality of life indices using dynamic HJ-Biplot and sparse Tucker3/Parafac models, applied to ten American countries across six years (2019-2024), revealing relationships between economic variables and quality of life indicators.

5. Discussion

In this study, we employed a novel hybrid methodology combining statistical techniques (MANOVA-Biplot) with advanced natural language processing (ChatGPT-5.2) to analyze 288 scientific articles on Tucker models published between 2000 and 2025. This approach provided a multidimensional perspective on the evolution and application of tensor decomposition techniques across diverse scientific domains.

Integration of Methods. The combination of text mining, Canonical Biplot, and AI-powered analysis proved particularly valuable for understanding the intellectual structure of Tucker model research. The similarity graph revealed three distinct lexical clusters centered on "model," "tucker3," and "datum," reflecting the fundamental tripartite structure of the field: methodological developments, specific model variants, and data-centric applications. The MANOVA-Biplot enhanced this view by incorporating temporal dynamics, showing clear lexical shifts across the four periods. Groups 1 and 2 clustered closely, indicating thematic continuity in early research focused on foundational methods. In contrast, groups 3 and 4 diverged significantly, reflecting diversification into new application domains and methodological specialization in recent years. The first two canonical dimensions explained 77.89% of the variability between groups, confirming the robustness of the temporal differentiation observed in the data.

Temporal Evolution of Research Themes. The lexical analysis revealed a clear evolutionary pattern that aligns with the historical development of tensor decomposition methods. The 2000-2005 period focused on fundamental techniques such as Tucker3, PARAFAC, and Candecomp, along with early applications in chemometrics and environmental science, reflecting the consolidation of multiway analysis as a recognized methodological framework. The 2006-2011 period introduced methodological variants including fuzzy models, dimensionality reduction techniques, and the first applications to sensory analysis and process monitoring. The 2012-2018 period showed marked diversification into molecular biology, ecology, medical diagnostics, and image processing, demonstrating the cross-disciplinary fertilization of Tucker methodologies. Finally, the 2019-2025 period revealed maturation of the field, with emphasis on massive data, network analysis, MIMO communications, robust algorithms, and deep learning integration. This trajectory confirms that Tucker models have evolved from specialized chemometric tools to versatile techniques addressing contemporary challenges in telecommunications, neuroimaging, and artificial intelligence.

Synergy Between Statistical and AI-Based Approaches. The integration of ChatGPT-5.2 added a crucial contextual layer to the statistical analysis, addressing a fundamental limitation of purely lexical methods. While the Biplot identified keyword associations and group structures with statistical rigor, it could not discern the specific contributions, authors, or application contexts underlying those patterns. ChatGPT bridged this gap by extracting narrative information from abstracts, transforming abstract lexical clusters into coherent research trajectories. For instance, the Biplot's identification of "mimo" and "communication" as characteristic terms for 2019-2025 was substantiated by ChatGPT's extraction of specific contributions on RIS-assisted MIMO systems and secure beamforming. Similarly, the appearance of "sparse" and "visualization" in 2012-2018 was contextualized by contributions on sparse Tucker variants and factor clustering methods. This synergy between quantitative pattern recognition and qualitative content extraction represents the primary methodological contribution of this work, demonstrating that AI-assisted literature review can transcend the limitations of both purely manual and purely automated approaches.

Methodological Considerations for AI-Assisted Review. The integration of large language models into systematic review workflows requires careful methodological design. In this study, we implemented a structured prompt engineering strategy that emphasized specificity, step-by-step structuring, and iterative refinement. The initial prompts established the task context and output format, while subsequent interactions refined the extraction quality. Processing abstracts in batches of no more than 20 articles per session ensured that the model maintained focus and coherence across the 288-article corpus. The sample output shown in Table 8 demonstrates the level of detail achievable with this approach. However, manual verification of a subset of extractions confirmed that AI-generated summaries, while highly accurate for main findings, occasionally omitted nuanced methodological details or misinterpreted domain-specific terminology. This underscores that AI should be viewed as an augmentative tool rather than a replacement for human expertise in literature synthesis.

Limitations and Future Directions. Despite the expanded keyword set including HOSVD, MLSVD, tensor train, and related formulations, some relevant publications may remain unidentified due to terminology variations not captured in our search strategy. The final corpus of 288 articles, while comprehensive, may underrepresent contributions from conferences, preprints, and non-English publications. Additionally, analysis based solely on abstracts, while efficient for large-scale review, may miss methodological nuances present only in full manuscripts. The reliance on ChatGPT-5.2, while state-of-the-art at the time of analysis, carries inherent risks of hallucination or bias, which we mitigated through cross-validation with Biplot results and manual verification of representative samples. Future studies should consider integrating multiple AI models for triangulation, expanding search parameters to include additional databases and document types, and developing automated validation protocols to further enhance reproducibility. The methodological framework presented here, however, provides a robust foundation for such extensions, offering a replicable template for AI-assisted systematic review across scientific disciplines.

6. Conclusions

This study conducted a systematic literature review of Tucker models from 2000 to 2025 using a hybrid methodology that combined text mining, Canonical Biplot analysis, and ChatGPT-5.2. Analysis of 288 articles from Scopus and Web of Science yielded several key conclusions.

First, research on Tucker models has evolved significantly over the past quarter-century. The field has progressed from foundational work on Tucker3, PARAFAC, and Candecomp in the early 2000s, through methodological diversification including fuzzy models and sparse variants in the 2006-2011 period, to broad application across biological, environmental, and diagnostic domains in 2012-2018, and finally to contemporary focus on massive data, telecommunications, and deep learning integration in 2019-2025. This trajectory confirms the enduring relevance and remarkable adaptability of Tucker decomposition techniques, which continue to find new applications as data complexity increases across scientific disciplines.

Second, the comparison between Canonical Biplot and ChatGPT-5.2 demonstrated the complementarity of traditional statistical methods and AI-powered analysis. The Biplot provided statistically robust visualization of keyword associations and group structures, revealing latent thematic relationships across time periods with rigorous mathematical foundations. ChatGPT contributed contextual depth by extracting specific contributions, author information, and application contexts, transforming abstract lexical patterns into narratively coherent research trajectories. Their combination yielded insights unavailable from either method alone, highlighting the value of methodological pluralism in data analysis.

Third, the methodological framework proposed here offers a replicable template for AI-assisted systematic literature review. The structured prompt design, batch processing strategy, and cross-validation with statistical techniques address common concerns about AI reliability and reproducibility. This approach is particularly valuable for rapidly evolving fields where manual synthesis of large literature corpora is impractical, and where the volume of publications exceeds the capacity of traditional review methods.

Fourth, this study underscores the importance of human oversight in AI-assisted research. While ChatGPT-5.2 demonstrated high accuracy in extracting main findings from abstracts, manual verification revealed occasional omissions of nuanced details and domain-specific terminology. AI should therefore be viewed as an augmentative tool that enhances human capabilities rather than a replacement for expert judgment in literature synthesis.

Finally, the hybrid methodology presented here has broader implications for scientific knowledge synthesis. As AI capabilities continue to advance, the integration of quantitative and qualitative techniques, statistical rigor and contextual understanding, human expertise and machine assistance will become increasingly essential for navigating the growing complexity of scientific literature. The framework developed in this study provides a foundation for such integrative approaches, with potential applications extending beyond tensor decomposition to any research domain requiring systematic analysis of large document collections.

Acknowledgement: The authors arc grateful to xxxxx at the ICAR-CentralInstitutc of Agricultural Enginccring, Bhopal, India for the assistanccin data analysis.
Funding: This work does not funding support.

Ethical Statement: This research is based exclusively on the analysis of published scientific literature and secondary textual data extracted for methodological purposes. No human or animal subjects were involved, and no personal, clinical, behavioral, or identifiable data were collected or analyzed. Consequently, institutional review board approval and informed consent were not required.

Conflicts of Interest :The authors declare that they have no conflicts of interest to this work.

References

1. Brignardello-Petersen, R., Santesso, N., & Guyatt, G. H. (2025). Systematic reviews of the literature: an introduction to current methods. *American Journal of Epidemiology, 194*(2), 536–542. https://doi.org/10.1093/aje/kwae232
2. Taha, K., Yoo, P. D., Yeun, C., Homouz, D., & Taha, A. (2024). A comprehensive survey of text classification techniques and their research applications: Observational and experimental insights. *Computer Science Review, 53*, 100664. https://doi.org/10.1016/j.cosrev.2024.100664
3. Beh, E. J. (2024). Correspondence analysis using the Cressie-Read family of divergence statistics. *International Statistical Review, 92*(1), 17-42. https://doi.org/10.1111/insr.12541
4. Min, B., Ross, H., Sulem, E., Veyseh, A. P. B., Nguyen, T. H., Sainz, O., Agirre, E., Heinz, I., & Roth, D. (2023). Recent advances in natural language processing via large pre-trained language models: A survey. *ACM Computing Surveys, 56*(2), Article 30. https://doi.org/10.1145/3605943
5. Tucker, L. R. (1966). *Some mathematical notes on three-mode factor analysis. Psychometrika, 31*(3), 279-311. https://doi.org/10.1007/BF02289464
6. Yu, J. C., Crotty, M. F., Gautam, N. K. S., van den Berg, R. A., Schmidtke, L. M., & Bastian, S. E. P. (2024). Sparse multiple factor analysis, sparse STATIS, and sparse multiple factor analysis for contingency tables. *Journal of Chemometrics, 38*(5), e3443. https://doi.org/10.1002/cem.3443
7. Abdesselam, R. (2024). Topological analysis of multiple tables. *Journal of AppliedMath, 2*(1), Article 424. https://doi.org/10.59400/jam.v2i1.424
8. Frutos-Bernal, E., Ceulemans, E., Galindo-Villardón, P., & Wilderjans, T. F. (2025). Data fusion by T3-PCA: A global model for the simultaneous analysis of coupled three-way and two-way real-valued data. *British Journal of Mathematical and Statistical Psychology, 78*(2), 672-709. https://doi.org/10.1111/bmsp.12372

9. Tokcan, N., Sofi, S. S., Pham, V. T., Prévost, C., Kharbech, S., Magnier, B., Nguyen, T. P., Zniyed, Y., & De Lathauwer, L. (2026). Tensor decompositions for signal processing: Theory, advances, and applications. *Signal Processing, 238*, 110191. https://doi.org/10.1016/j.sigpro.2025.110191
10. Yu, H., Larsen, K. G., & Christiansen, O. (2025). Optimization methods for tensor decomposition: A comparison of new algorithms for fitting the CP (CANDECOMP/PARAFAC) model. *Chemometrics and Intelligent Laboratory Systems, 257*, 105290. https://doi.org/10.1016/j.chemolab.2024.105290
11. Bilius, L.-B., Pentiuc, S.-G., & Vatavu, R.-D. (2023). TIGER: A Tucker-based instrument for gesture recognition with inertial sensors. *Pattern Recognition Letters, 165*, 1–8. https://doi.org/10.1016/j.patrec.2022.11.028
12. Timmerman, M. E., & Kiers, H. A. L. (2000). Three-mode principal components analysis: Choosing the numbers of components and sensitivity to local optima. *British Journal of Mathematical and Statistical Psychology, 53*(1), 1-16. https://doi.org/10.1348/000711000159132
13. An, J., Wilson, D. I., Deed, R. C., Kilmartin, P. A., Young, B. R., & Yu, W. (2023). The importance of outlier rejection and significant explanatory variable selection for Pinot noir wine soft sensor development. *Current Research in Food Science, 6*, 100514. https://doi.org/10.1016/j.crfs.2023.100514
14. Zhang, Q., Dong, Y., Yuan, Q., Song, M., & Yu, H. (2023). Combined deep priors with low-rank tensor factorization for hyperspectral image restoration. *IEEE Geoscience and Remote Sensing Letters, 20*, 5500105. https://doi.org/10.1109/LGRS.2023.3236341
15. Mishra, P., Roger, J.-M., Jouan-Rimbaud-Bouveresse, D., Biancolillo, A., Marini, F., Nordon, A., & Rutledge, D. N. (2021). Recent trends in multi-block data analysis in chemometrics for multi-source data integration. *TrAC Trends in Analytical Chemistry, 137*, 116206. https://doi.org/10.1016/j.trac.2021.116206
16. Phougat, M., Sahni, N. S., & Choudhury, D. (2023). Multiway analysis reveals hydrophobicity as the sole determinant of dynamic peptide-acetonitrile-water association behavior. *Journal of Physical Chemistry B, 127*(23), 5256–5267. https://doi.org/10.1021/acs.jpcb.3c02642
17. Ramos, M., Ascencio, J. G., Hinojosa, M. V., Vera, F., Ruiz, O., Jimenez-Feijoo, M. I., & Galindo, P. (2021). Multivariate statistical process control methods for batch production: A review focused on applications. *Production & Manufacturing Research, 9*(1), 33-55. https://doi.org/10.1080/21693277.2020.1871441
18. Guo, L., Yu, H., Li, Y., Zhang, C., & Kharbach, M. (2023). Tensor methods in data analysis of chromatography/mass spectroscopy-based plant metabolomics. *Plant Methods, 19*, 130. https://doi.org/10.1186/s13007-023-01105-y
19. Yu, H., Guo, L., Kharbach, M., & Han, W. (2021). Multi-way analysis coupled with near-infrared spectroscopy in food industry: Models and applications. *Foods, 10*(4), 802. https://doi.org/10.3390/foods10040802
20. Guembe-García, M., Magnaghi, L. R., Franceschi, G. E., Bova, A., & Biesuz, R. (2026). Multi-way data analysis nowadays: Taking advanced chemometric tools to everyday analytical chemistry applications. *Chemosensors, 14*(2), 37. https://doi.org/10.3390/chemosensors14020037
21. Ranaweera, R. K. R., Gilpin, M., Boulton, R., & Thevathasan, T. (2021). A review of wine authentication using spectroscopic approaches in combination with chemometrics. *Molecules, 26*(14), 4334. https://doi.org/10.3390/molecules26144334
22. Peris-Díaz, M. D., & Krężel, A. (2021). A guide to good practice in chemometric methods for vibrational spectroscopy, electrochemistry, and hyphenated mass spectrometry. *TrAC Trends in Analytical Chemistry, 135*, 116157. https://doi.org/10.1016/j.trac.2020.116157
23. Li, Z. Y., Li, X. K., Lin, Y., Feng, N., Zhang, X.-Z., Li, Q.-L., & Li, B. Q. (2022). A comparative study of three chemometrics methods combined with excitation–emission matrix fluorescence for quantification of the bioactive compounds aesculin and aesculetin in Cortex Fraxini. *Frontiers in Chemistry, 10*, 984010. https://doi.org/10.3389/fchem.2022.984010
24. Dinç, E., Ertekin, Z. C., & Büker, E. (2024). Two-way and three-way resolutions of fluorescence excitation-emission dataset for the co-estimation of two pharmaceuticals in a binary mixture. *Chemometrics and Intelligent Laboratory Systems, 245*, 104873. https://doi.org/10.1016/j.chemolab.2023.104873

25. Moro, M. K., dos Santos, F. D., Folli, G. S., Romão, W., & Filgueiras, P. R. (2021). A review of chemometrics models to predict crude oil properties from nuclear magnetic resonance and infrared spectroscopy. *Fuel, 306*, 121283. https://doi.org/10.1016/j.fuel.2021.121283
26. Wang, D., Zhuo, Y., Karfunkle, M., Patil, S. M., Smith, C. J., Keire, D. A., & Chen, K. (2021). NMR spectroscopy for protein higher order structure similarity assessment in formulated drug products. *Molecules, 26*(14), 4251. https://doi.org/10.3390/molecules26144251
27. Raljević, D., Parlov Vuković, J., Smrečki, V., Marinić Pajc, L., Novak, P., Hrenar, T., Jednačak, T., Konjević, L., Pinević, B., & Gašparac, T. (2021). Machine learning approach for predicting crude oil stability based on NMR spectroscopy. *Fuel, 305*, 121561. https://doi.org/10.1016/j.fuel.2021.121561
28. Tiwari, R., Ranjan, N., Chaurasia, M., & Flora, S. J. S. (2025). Hyphenated mass spectroscopic detection of heavy metals in environmental and biological samples: A review. *Journal of Trace Elements and Minerals, 5*, 100273. https://doi.org/10.1016/j.jtemin.2025.100273
29. Rodionova, O. Y., Pomerantsev, A. L., & Oliveri, P. (2024). Chemometrics as an efficient tool for food authentication. *Food Control, 163*, 110511. https://doi.org/10.1016/j.foodcont.2024.110511
30. Sala-Sala, V., Andreu, J. M., Pérez-Gimeno, A., Jordán, M. M., Navarro-Pedreño, J., & Almendro-Candel, M. B. (2025). Spatial and multivariate analysis of groundwater hydrochemistry in the Solana aquifer, SE Spain. *Environments, 12*(9), 323. https://doi.org/10.3390/environments12090323
31. Francis, G. A., Jeyakumar, S., Ray, S., et al. (2025). Emerging roles of FTIR spectroscopy in toxic metal profiling: Innovations for food safety monitoring. *Food Safety and Risk, 12*, 9. https://doi.org/10.1186/s40550-025-00119-9
32. Cohen, J., Bro, R., & Comon, P. (2023). Tensor decompositions: Principles and application to food sciences. *In Source Separation in Physical-Chemical Sensing* (pp. 255–323). https://doi.org/10.1002/9781119137252.ch6
33. Wu, Y., Lu, L., Xu, A., Wang, Y., Li, Z., Yang, Z., Zeng, L., & Li, Q. (2026). Neural networks for epilepsy detection and prediction with EEG signals: A systematic review. *Artificial Intelligence Review, 59*, 30. https://doi.org/10.1007/s10462-025-11441-1
34. Reitsema, A. M., Jeronimus, B. F., van Dijk, M., Ceulemans, E., van Roekel, E., Kuppens, P., & de Jonge, P. (2022). Distinguishing dimensions of emotion dynamics across 12 emotions in adolescents' daily lives. *Emotion, 23*(6), 1549–1561. https://doi.org/10.1037/emo0001173
35. Auddy, A., Xia, D., & Yuan, M. (2025). Tensors in high-dimensional data analysis: Methodological opportunities and theoretical challenges. *Annual Review of Statistics and Its Application, 12*, 527–551. https://doi.org/10.1146/annurev-statistics-112723-034548
36. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71
37. Birkenmaier, L., Lechner, C. M., & Wagner, C. (2024). The search for solid ground in text as data: A systematic review of validation practices and practical recommendations for validation. Communication *Methods and Measures*. https://doi.org/10.1080/19312458.2023.2285765
38. Hankar, M., Kasri, M., & Beni-Hssane, A. (2025). A comprehensive overview of topic modeling: Techniques, applications and challenges. *Neurocomputing, 628*, 129638. https://doi.org/10.1016/j.neucom.2025.129638
39. Kumar, K. (2021). Tucker3 modelling of EEMF spectroscopic data sets using restrained initialization of spectral variables: Fluorimetric analysis of mixtures consisting of bioactive molecules. *Luminescence, 36*(3), 678–689. https://doi.org/10.1002/bio.4052
40. Sousa, Y. S. O. (2021). The use of the Iramuteq software: Fundamentals of lexicometry for qualitative research. *Estudos e Pesquisas em Psicologia, 21*(spe), 1541–1560. https://doi.org/10.12957/epp.2021.64034
41. Silva, S., & Ribeiro, E. A. W. (2021). The IRAMUTEQ software as a methodological tool for qualitative analysis in research in professional and technological education. *Cadernos de Educação Tecnologia e Sociedade, 14*(2), 275–284. https://doi.org/10.14571/brajets.v14.n2.275-284

42. Alhuzali, T., Beh, E. J., & Stojanovski, E. (2024). Using multi-way correspondence analysis to examine the Nobel Prize data in STEM related disciplines. Communications in Statistics: Case Studies, *Data Analysis and Applications, 10*(2), 143–161. https://doi.org/10.1080/23737484.2024.2394421
43. Jouan-Rimbaud Bouveresse, D., Mazerolles, G., Qannari, E. M., Courcoux, P., Lê, S., & Rutledge, D. N. (2024). A synthetic review of some recent extensions of ComDim. *Journal of Chemometrics, 38*(7), e3454. https://doi.org/10.1002/cem.3454
44. Tussupov, J., Kassymova, A., Mukhanova, A., Bissengaliyeva, A., Azhibekova, Z., Yessenova, M., & Abuova, Z. (2025). Analysis of short texts using intelligent clustering methods. *Algorithms, 18*(5), 289. https://doi.org/10.3390/a18050289
45. Pak, A., Ziyaden, A., Saparov, T., Akhmetov, I., & Gelbukh, A. (2024). Word embeddings: A comprehensive survey. *Computación y Sistemas, 28*(4), 2005–2029. https://doi.org/10.13053/cys-28-4-5225
46. Zhang, B., Zhou, Y., & Li, D. (2025). Can human reading validate a topic model? *Sociological Methodology, 55*, 59–90. https://doi.org/10.1177/00811750241265336
47. Leipold, F. M., Kaczmirek, L., & Bosnjak, M. (2025). Detecting respondent burden in online surveys. *Social Science Computer Review*. https://doi.org/10.1177/08944393241247425
48. Panchendrarajan, R., Amaitik, N. M., & Min-Allah, N. (2023). Topic-based influential user detection: A survey. *Applied Intelligence, 53*, 10077–10118. https://doi.org/10.1007/s10489-022-03831-7
49. Incitti, F., Urli, F., & Snidaro, L. (2023). Beyond word embeddings: A survey. *Information Fusion, 89*, 418–436. https://doi.org/10.1016/j.inffus.2022.08.024
50. Jehangir, B., Radhakrishnan, S., & Agarwal, R. (2023). A survey on named entity recognition: Datasets, tools, and methodologies. *Natural Language Processing Journal, 3*, 100017. https://doi.org/10.1016/j.nlp.2023.100017
51. Boselli, R., D'Amico, S., & Nobani, N. (2025). eXplainable AI for word embeddings: A survey. *Cognitive Computation, 17*, 19. https://doi.org/10.1007/s12559-024-10373-2
52. Seow, W. L., Chaturvedi, I., Hogarth, A., Mao, R., & Cambria, E. (2025). A review of named entity recognition: From learning methods to modelling paradigms and tasks. *Artificial Intelligence Review, 58*, 315. https://doi.org/10.1007/s10462-025-11321-8
53. El Zini, J., & Awad, M. (2023). On the explainability of natural language processing deep models. *ACM Computing Surveys, 55*(5), Article 103, 1–31. https://doi.org/10.1145/3529755
54. Sánchez-García, A. B., Zárate-Santana, Z., & Patino-Alonso, C. (2025). A multivariate analysis with MANOVA-Biplot of learning approaches in health science students. *Social Sciences, 14*(7), 403. https://doi.org/10.3390/socsci14070403
55. Ahmed, R. R., Streimikiene, D., Streimikis, J., & Siksnelyte-Butkiene, I. (2024). A comparative analysis of multivariate approaches for data analysis in management sciences. *E+M Ekonomie a Management, 27*(1), 192–210. https://doi.org/10.15240/tul/001/2024-5-001
56. Adachi, K. (2025). Advances in multivariate data analysis. *Behaviormetrika*. https://doi.org/10.1007/s41237-025-00268-3
57. Grané, A., Albarrán, I., & Oliveira, M. (2021). Robust multivariate analysis for mixed-type data: Novel model and algorithm for the analysis and clustering of mixed-type data using a hierarchical approach based on Forward Search. *Sociological Methods & Research, 52*(2), 731–770. https://doi.org/10.1177/0049124119882440
58. Ganey, R., & Gardner-Lubbe, S. (2026). A canonical variate analysis biplot based on the generalized singular value decomposition. *Journal of the Italian Statistical Society*. https://doi.org/10.1007/s10260-025-00831-y
59. Espinel-Obregoso, F., Sánchez-Loor, J., Basurto-Quilligana, R., Peralta-Gamboa, D., & Villacís-Macías, C. (2025). Diversity and application of biplot methods in Ecuadorian research: A systematic literature review. Open *Information Science, 9*(1), 20250015. https://doi.org/10.1515/opis-2025-0015
60. Todorov, V., Simonacci, V., Di Palma, M. A., & Gallo, M. (2025). Robust tools for three-way component analysis of compositional data. *Behaviormetrika*. https://doi.org/10.1007/s41237-025-00276-3
61. Brereton, R. G. (2025). Partial least squares. *Journal of Chemometrics, 39*(12), e70069. https://doi.org/10.1002/cem.70069

62. Grismer, L. L. (2025). Introducing multiple factor analysis (MFA) as a diagnostic taxonomic tool complementing principal component analysis (PCA). *ZooKeys, 1248*, 93–109. https://doi.org/10.3897/zookeys.1248.159516
63. Galvan, D., de Andrade, J. C., Conte-Junior, C. A., Killner, M. H. M., & Bona, E. (2023). DD-ComDim: A data-driven multiblock approach for one-class classifiers. *Chemometrics and Intelligent Laboratory Systems, 233*, 104748. https://doi.org/10.1016/j.chemolab.2022.104748
64. Reusens, M., Stevens, A., Tonglet, J., De Smedt, J., Verbeke, W., vanden Broucke, S., & Baesens, B. (2024). Evaluating text classification: A benchmark study. *Expert Systems with Applications, 254*, 124302. https://doi.org/10.1016/j.eswa.2024.124302
65. Sajun, A. R., Zualkernan, I., & Sankalpa, D. (2024). A historical survey of advances in transformer architectures. *Applied Sciences, 14*(10), 4316. https://doi.org/10.3390/app14104316
66. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33*, 1877–1901. https://doi.org/10.48550/arXiv.2005.14165
67. Islam, S., Elmekki, H., Elsebai, A., Bentahar, J., Drawel, N., Rjoub, G., & Pedrycz, W. (2024). A comprehensive survey on applications of transformers for deep learning tasks. *Expert Systems with Applications, 241*, 122666. https://doi.org/10.1016/j.eswa.2023.122666
68. Raiaan, M. A. K., Mukta, M. S. H., Fatema, K., Fahad, A., Mim, M. M. J., Ahmad, J., Dobaie, A. M., Azad, A. K., & Alsharif, M. H. (2024). A review on large language models: Architectures, applications, taxonomies, open issues and challenges. *IEEE Access, 12*, 26839–26874. https://doi.org/10.1109/ACCESS.2024.3365742
69. Kumar, P., Choudhary, A., Chandran, D., et al. (2024). Large language models (LLMs): Survey, technical frameworks, and future challenges. *Artificial Intelligence Review*. https://doi.org/10.1007/s10462-024-10888-y
70. Trust, P., & Minghim, R. (2024). A study on text classification in the age of large language models. *Machine Learning and Knowledge Extraction, 6*(4), 2688–2721. https://doi.org/10.3390/make6040129
71. Fields, J., Chovanec, K., & Madiraju, P. (2024). A survey of text classification with transformers: How wide, how large, how long, how accurate, how expensive, how safe? *IEEE Access, 12*, 6518–6531. https://doi.org/10.1109/ACCESS.2024.3349952
72. Koç, C., Özyurt, A. S., Alkan, A., Güllü, E., & Döndüren, M. Y. (2025). Survey on latest advances in natural language processing and machine translation methods. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 15(4), e70004. https://doi.org/10.1002/widm.70004
73. Alva Principe, R., & Chiarini, N. (2025). Long document classification in the transformer era: A survey on challenges, advances, and open issues. Wiley Interdisciplinary Reviews: *Data Mining and Knowledge Discovery, 15*(2), e70019. https://doi.org/10.1002/widm.70019
74. Wu, Y., Wan, J., Liu, M., et al. (2025). A survey of text classification based on pre-trained language models. *Neurocomputing, 616*, 128921. https://doi.org/10.1016/j.neucom.2024.128921
75. Țucudean, G., Bucos, M., & Gherman, B. (2024). Natural language processing with transformers: A review. *PeerJ Computer Science, 10*, e2222. https://doi.org/10.7717/peerj-cs.2222
76. Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems, 3*, 121–154. https://doi.org/10.1016/j.iotcps.2023.04.003
77. Cong-Lem, N., Soyoof, A., & Tsering, D. (2024). A systematic review of the limitations and associated opportunities of ChatGPT. *International Journal of Human–Computer Interaction*. https://doi.org/10.1080/10447318.2024.2344142
78. Alawida, M., Alfehaid, A., Ibrahim, S., & Tei, K. (2023). A comprehensive study of ChatGPT: Advancements, applications, prospects, and challenges. *Information, 14*(8), 462. https://doi.org/10.3390/info14080462
79. Caballero-Julia, D., Campillo, P. (2021) Epistemological Considerations of Text Mining:Implications for Systematic Literature Review. *Mathematics, 9*, 1865. https://doi.org/10.3390/math9161865
80. Dege, D., & Brüggemann, P. (2024). Marketing analytics with RStudio: A software review. *Journal of Marketing Analytics, 12*, 465–470. https://doi.org/10.1057/s41270-023-00264-0

81. Graffelman, J. (2025). Biplots for the correlation matrix. *Journal of Computational and Graphical Statistics*. https://doi.org/10.1080/10618600.2025.2469757
82. Nair, V. G., & Bhattacharya, D. (2025). Metric-driven Voronoi diagrams: A comprehensive mathematical framework. *Computation, 13*(9), 212. https://doi.org/10.3390/computation13090212
83. Timmerman, M. E., & Kiers, H. A. L. (2002). Three-way component analysis with smoothness constraints. *Computational Statistics & Data Analysis, 40*(3), 447–470. https://doi.org/10.1016/S0167-9473(02)00059-2
84. Hubert, M., & Hirari, M. (2024). MacroPARAFAC for handling rowwise and cellwise outliers in incomplete multi-way data. *Chemometrics and Intelligent Laboratory Systems, 251*, 105170. https://doi.org/10.1016/j.chemolab.2024.105170
85. Halberg, H. F. F., Bevilacqua, M., & Rinnan, Å. (2024). Resampling as a robust measure of model complexity in PARAFAC models. *Journal of Chemometrics, 38*(12), e3601. https://doi.org/10.1002/cem.3601
86. Frutos-Bernal, E., Vicente-González, L., & Vicente-Villardón, J. L. (2024). Tucker3-PCovR: The Tucker3 principal covariates regression model. *Behavior Research Methods, 56*(4), 3873–3890. https://doi.org/10.3758/s13428-024-02379-3
87. Bottazzi Schenone, M., Iannaccio, T., Mozzetta, I., & Vichi, M. (2026). Three-way data analysis with explainable Tucker3 clustering (XT3Clus). *Applied Stochastic Models in Business and Industry, 42*(1). https://doi.org/10.1002/asmb.70069
88. Ceulemans, E; Van Mechelen, I. (2003). Uniqueness of N-way N-mode hierarchical classes models. *Journal of Mathematical Psychology, 47*(3). https://doi.org/10.1016/S0022-2496(03)00002-6
89. D. Muti and S. Bourennane, (2003). Multidimensional signal processing using lower-rank tensor approximation. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. III-457. https://10.1109/ICASSP.2003.1199510
90. Wang, J., Liu, P., Zhou, W., & Liu, J. (2025). Interval-valued functional principal component analysis method and application under the general distribution. *Journal of Systems Science and Information, 13*(4), 525–549. https://doi.org/10.12012/JSSI-2024-0152
91. Kiers, H.A.L (2004). Bootstrap confidence intervals for three-way methods. *Journal of Chemometrics, 18*, 167-180. https://doi.org/10.1002/cem.841
92. Mitsushita, K., Murakoshi, S., & Koyama, M. (2023). How are various natural disasters cognitively represented? A psychometric study of natural disaster risk perception applying three-mode principal component analysis. *Natural Hazards, 115*, 2523–2545. https://doi.org/10.1007/s11069-022-05708-x
93. Zdunek, R., & Gabor, M. (2022). Nested compression of convolutional neural networks with Tucker-2 decomposition. *In 2022 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–8). IEEE. https://doi.org/10.1109/IJCNN55064.2022.9892959
94. Zhang, L., Wu, Z., Sun, J., Xu, Y., & Wei, Z. (2022). A distributed and parallel method of hyperspectral computational imaging via collaborative Tucker3 tensor decomposition. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 2345–2348). https://doi.org/10.1109/IGARSS46834.2022.9883840
95. Ardah, K., Gherekhloo, S., de Almeida, A. L. F., & Haardt, M. (2022). Double-RIS versus single-RIS aided systems: Tensor-based MIMO channel estimation and design perspectives. *Speech and Signal Processing (ICASSP)* (pp. 5203–5207). https://doi.org/10.1109/ICASSP43922.2022.9746287
96. Lesiuk, M. (2022). Quintic-scaling rank-reduced coupled cluster theory with single and double excitations. *Journal of Chemical Physics, 156*(24), 244107. https://doi.org/10.1063/5.0071916
97. Zhang, J., Liu, M., Xiong, P., Du, H., Zhang, H., Sun, G., Hou, Z., & Liu, X. (2022). Automated localization of myocardial infarction of image-based multilead ECG tensor with Tucker2 decomposition. *IEEE Transactions on Instrumentation and Measurement, 71*, 4002212. https://doi.org/10.1109/TIM.2021.3104394
98. dos Santos, R. F., Paraskevaidi, M., Mann, D. M. A., Allsop, D., Santos, M. C. D., Morais, C. L. M., & Lima, K. M. G. (2022). Alzheimer's disease diagnosis by blood plasma molecular fluorescence spectroscopy (EEM). *Scientific Reports, 12*, 15234. https://doi.org/10.1038/s41598-022-20611-y
99. Gabor, M., & Zdunek, R. (2023). Compressing convolutional neural networks with hierarchical Tucker-2 decomposition. *Applied Soft Computing, 136*, 110081. https://doi.org/10.1016/j.asoc.2022.109856
100. He, Z.; Cichocki, A.; Xie, S. (2009). Efficient method for Tucker3 model selection. *Neurocomputing, 72*, 947-957. https://doi.org/10.1049/el.2009.0635

101. Dahl, T.; Næs, T. (2009). Identifying outlying assessors in sensory profiling using fuzzy clustering and multi-block methodology. *Food Quality and Preference, 20*, 153- 162. https://doi.org/10.1016/j.foodqual.2008.12.001
102. Giordani, P. (2010). Three-way analysis of imprecise data. *Journal of Chemometrics, 24*, 29-38. https://doi.org/10.1016/j.jmva.2009.10.003
103. Kompany-Zareh, M.; Akhlaghi, Y.; Bro, R. (2012). Tucker core consistency for validation of restricted Tucker3 models. *Chemometrics and Intelligent Laboratory Systems, 110*, 46-52. https://doi.org/10.1016/j.aca.2012.02.028
104. Wilderjans, Tom F.; Ceulemans, Eva; Meers, Koen. (2013). CHull: A generic convex-hull- based model selection method. *Behavior Research Methods, 45*, 1-15. https://doi.org/10.3758/s13428-012-0238-5
105. Liu, K.; So, H.C.; Da Costa, J.P.C.L.; Huang, L. (2013). Core consistency diagnostic aided by reconstruction error for accurate enumeration of the number of components in parafac models. *Chemometrics and Intelligent Laboratory Systems, 129*, 12-20. https://10.1109/ICASSP.2013.6638945
106. Tortora, C.; Summa, M.G.; Palumbo, F. (2013). Factor PD-clustering. *Chemometrics and Intelligent Laboratory Systems, 129*, 221-231. https://doi.org/10.1007/978-3-319-00035-0_11
107. Giordani, P.; Kiers, H.A.L.; Del Ferraro, M.A. (2014). Three-way component analysis using the R package ThreeWay. *Journal of Statistical Software, 59*, 1-26. https://doi.org/10.18637/jss.v057.i07
108. Yokota, T.; Cichocki, A. (2014). Linked Tucker2 Decomposition for Flexible Multi-block Data Analysis. *Neurocomputing, 146*, 108-118. https://doi.org/10.1007/978-3-319-12643-2_14
109. Chen, H.; Wang, J.; Zhou, S.; Zhang, Y. (2014). The FPSO for selecting number of components in Tucker3 decomposition for Hyperspectral image compression. *Optik, 125*, 6977-6981. https://doi.org/10.1109/DCC.2014.32
110. Gallo, M. (2015). Tucker3 Model for Compositional Data. *Chemometrics and Intelligent Laboratory Systems, 144*, 77-86. https://doi.org/10.1080/03610926.2013.798664
111. Mendes, S.; Fernandez-Gomez, M. Jose; Marques, Sonia Cotrim; Pardal, Miguel Angelo; Azeiteiro, Ulisses Miranda; Galindo-Villardon, M. Purificacion. (2017). CO-tucker: a new method for the simultaneous analysis of a sequence of paired tables. *Journal of Applied Statistics, 44*(15), 2729-2755. https://doi.org/10.1080/02664763.2016.1261815
112. Ikemoto, H.; Adachi, K. (2016). Sparse Tucker2 analysis of three-way data subject to a constrained number of zero elements in a core array. *Chemometrics and Intelligent Laboratory Systems, 151*, 141-151. https://doi.org/10.1016/j.csda.2015.12.007
113. Tortora, C.; Summa, M.G.; Marino, M.; Palumbo, F. (2016). Factor probabilistic distance clustering (FPDC): a new clustering method. *Statistical Methods & Applications, 25*, 215-238. https://doi.org/10.1007/s11634-015-0219-5
114. Chachlakis, D.G.; Markopoulos, P.P. (2018). Novel Algorithms for Exact and Efficient L1- NORM-BASED Tucker2 Decomposition. *IEEE Transactions on Signal Processing, 66*, 2921-2934. https://doi.org/10.1109/ICASSP.2018.8461839
115. Markopoulos, P.P.; Chachlakis, D.G.; Papalexakis, E.E. (2018). The Exact Solution to Rank-1 L1-Norm TUCKER2 Decomposition. *IEEE Transactions on Signal Processing, 66*, 4712-4725. https://doi.org/10.1109/LSP.2018.2790901
116. Shomali, Z.; Omidikia, N.; Kompany-Zareh, M. (2018). Application of non-linear optimization for estimating Tucker3 solutions. Journal of *Chemometrics, 32*, e3045. https://doi.org/10.1016/j.chemolab.2018.01.006
117. Stegeman, A. (2018). Simultaneous Component Analysis by Means of Tucker3. *Journal of Chemometrics, 32*, e3008. https://doi.org/10.1007/s11336-017-9568-7
118. Lin, R., Ko, C.-Y., He, Z., Chen, C., Cheng, Y., Yu, H., Chesi, G., & Wong, N. (2020). HOTCAKE: Higher order Tucker articulated kernels for deeper CNN compression. IEEE Transactions on Circuits and Systems for Video Technology, 30(2), 362–375. https://doi.org/10.1109/ICSICT49897.2020.9278257
119. Jang, J.-G., & Kang, U. (2020). D-Tucker: Fast and memory-efficient Tucker decomposition for dense tensors. *IEEE Transactions on Signal Processing, 68*, 5271–5284. https://doi.org/10.1109/ICDE48307.2020.00186

120. Martin-Barreiro, C., Ramirez-Figueroa, J. A., Nieto-Librero, A. B., Leiva, V., Martin-Casado, A., & Galindo-Villardón, M. P. (2021). A new algorithm for computing disjoint orthogonal components in the three-way Tucker model. *Mathematics, 35*(10), e3348. https://doi.org/10.3390/math9030203
121. Simonacci, V., & Gallo, M. (2022). Three-way principal balance analysis: Algorithm and interpretation. *Annals of Operations Research, 231*, 104325. https://doi.org/10.1007/s10479-022-04782-5
122. Tsuchida, J., & Yadohisa, H. (2022). Tucker-3 decomposition with sparse core array using a penalty function based on Gini-index. *Japanese Journal of Statistics and Data Science, 484*, 523–536. https://doi.org/10.1007/s42081-022-00179-7
123. Yamamoto, R., Hontani, H., Imakura, A., & Yokota, T. (2022). Consistent MDT-Tucker: A Hankel structure constrained Tucker decomposition in delay embedded space. IEEE *Transactions on Signal Processing, 70*, 2087–2100. https://doi.org/10.23919/APSIPAASC55919.2022.9980035
124. Pan, Y., Wang, M., & Xu, Z. (2022). TedNet: A PyTorch toolkit for tensor decomposition networks. *Neurocomputing, 7*(77), 3668. https://doi.org/10.1016/j.neucom.2021.10.064
125. Koleini, F., Hugelier, S., Lakeh, M. A., Abdollahi, H., Camacho, J., & Gemperline, P. J. (2023). On the complementary nature of ANOVA simultaneous component analysis (ASCA+) and Tucker3 tensor decompositions on designed multi-way datasets. *Journal of Chemometrics, 234*, 104444. https://doi.org/10.1002/cem.3514
126. Han, Y., Lin, Q.-H., Kuang, L.-D., Zhao, B.-H., Gong, X.-F., Cong, F., Wang, Y.-P., & Calhoun, V. D. (2024). A core tensor sparsity enhancement method for solving Tucker-2 model of multi-subject fMRI data. *Biomedical Signal Processing and Control, 408*, 110172. https://doi.org/10.1016/j.bspc.2024.106471
127. Pagani, A. P., Camargo, G., Ibanez, G. A., Olivieri, A. C., Pomerantsev, A. L., & Rodionova, O. Y. (2024). Data-driven version of multiway soft independent modeling of class analogy (N-Way DD-SIMCA): Theory and application. *Analytical Chemistry, 1325*, 343012. https://doi.org/10.1021/acs.analchem.3c05096
128. Kirch, J. L., Spitti, A. M. D. S., Chiorato, A. F., dos Santos Dias, C. T., & de Lima, C. G. (2024). A multi-attribute evaluation of genotype-environment experiments using biplots and joint plots graphics. *Journal of Computational and Graphical Statistics, 116*(4), 1789–1805. https://doi.org/10.1080/10618600.2024.2325445
129. Zhang, H., Miao, C., Lyu, Q., Zhang, L., & Ye, W. (2025). TDSA-SuperPoint: A novel SuperPoint based on Tucker decomposition with self-attention for scene matching navigation. *IEEE Sensor Journal, 61*(2), 2345–2360. https://doi.org/10.1109/JSEN.2024.3516829
130. Graham, N., Richardson, N., Friedlander, M. P., & Saylor, J. (2025). Tracing sedimentary origins in multivariate geochronology via constrained tensor factorization. *Mathematical Geosciences, 57*(3), 412–435. https://doi.org/10.1007/s11004-024-10175-0
131. Lestari, K. E., Yudhanegara, M. R., Nugraha, E. S., & Sylviani, S. (2025). Tucker3 tensor decomposition for the standardized residual hypermatrix on three-way correspondence analysis. *Statistics in Transition, 26*(2), 89–108. https://doi.org/10.22342/jims.v31i2.1491
132. Schenone, M. B., Iannaccio, T., Mozzetta, I., & Vichi, M. (2025). City quality of life by advanced tensor analysis. *Statistics For Innovation, 178*(1), 45–72. https://doi.org/10.1007/978-3-031-96303-2_38
133. Saylor, J. E., Richardson, N., Graham, N., Lee, R. G., & Friedlander, M. P. (2025). Tracking Cu-fertile sediment sources via multivariate petrochronological mixture modeling of detrital zircons. *Journal of Geophysical Research-Earth Surface, 120*(3), 567–589. https://doi.org/10.1029/2025JF008406
134. Martyna, A., Alladio, E., Romagnoli, M., Malaspina, F., & Pazzi, M. (2025). A likelihood ratio model for three-way data coupled with a Tucker3 model. *Chemometrics and Intelligent Laboratory Systems, 368*, 112345. https://doi.org/10.1016/j.chemolab.2025.105464
135. Ginige, N., de Sena, A. S., Mahmood, N. H., Rajatheva, N., & Latva-Aho, M. (2025). Efficient channel prediction for beyond diagonal RIS-assisted MIMO systems with channel aging. *IEEE Transactions on Wireless Communications, 24*(3), 1890–1906. https://doi.org/10.1109/TVT.2025.3556277
136. Mao, D., Li, Z., Li, S., Hao, W., Wang, N., & Xu, W. (2025). Tensor-based joint hybrid beamforming and artificial noise design for secure mmWave MU-MIMO-OFDM communication systems. *IEEE Transactions on Communications, 73*(2), 1123–1140. https://doi.org/10.1109/TIFS.2025.3614447

137. Gu, Y., Zhao, P., Cheng, L., Guo, Y., Wang, H., Ding, W., & Liu, Y. (2025). TensorTrack: Tensor decomposition for video object tracking. *Mathematics, 133*(4), 1023–1045. https://doi.org/10.3390/math13040568
138. Cascante-Yarlequé, R., Martin-Barreiro, C., Cabezas, X., Camacho-Villagomez, F. R., Suarez, C. A., Freire, L., & De Santis, P. R. (2025). Methodological framework for three-way statistical analysis of cost of living and quality of life indices: A case study in the American continent. *Scientific Reports*, *15*, 16585. https://doi.org/10.1038/s41598-025-00672-5