

# Artificial Intelligence as a Systematic Reviewer: limits and opportunities in ITOR's case study

## Group M

Aur Marina Iuliana 1809715,  
Oddi Livia 1846085,  
Cesario Laura Thoft 1852596,  
Branescu Ana Carina 2125078,  
Zeller Damian 2118831

May 24, 2024

## Contents

<b>1</b>	<b>Introduction and Motivation</b>	<b>3</b>
1.1	Custom AI: ITOR Systematic Reviewer GPT 4 . . . . .	3
<b>2</b>	<b>Planning</b>	<b>4</b>
2.1	Knowledge graph . . . . .	4
2.2	Traditional Planning . . . . .	6
2.3	AI Planning . . . . .	6
<b>3</b>	<b>Search</b>	<b>7</b>
3.1	Traditional Search . . . . .	7
3.1.1	Scopus . . . . .	7
3.1.2	WoS . . . . .	7
3.1.3	Search Performance Evaluation . . . . .	8
3.2	AI Search . . . . .	9
<b>4</b>	<b>Screening</b>	<b>10</b>
4.1	Traditional Screening . . . . .	10
4.1.1	Top 30 articles for the ITOR's 30th anniversary collected from Scopus . . . . .	10
4.1.2	Top 30 articles in ITOR collected from Scopus . . . . .	10
4.1.3	Top 30 articles in ITOR collected from Web of Science . . . . .	10
4.2	AI Screening . . . . .	12
<b>5</b>	<b>Data Extraction and Synthesis</b>	<b>13</b>
5.1	Traditional Data Extraction and Synthesis . . . . .	13
5.2	AI Data Extraction and Synthesis . . . . .	16
<b>6</b>	<b>Quality Assessment</b>	<b>19</b>
6.1	Traditional Quality Assessment . . . . .	19
6.2	AI Quality Assessment . . . . .	19
<b>7</b>	<b>Cost-Benefit Analysis</b>	<b>20</b>
<b>8</b>	<b>Conclusions</b>	<b>22</b>

<b>9 Appendix</b>	<b>27</b>
9.1 PRISMA Statement . . . . .	27
9.2 Knowledge Graph Tables and Figures . . . . .	29
9.3 Custom AI Interface . . . . .	31
9.4 Traditional Search Tables and Figures . . . . .	32
9.5 AI Search Tables and Figures . . . . .	33
9.6 Traditional Screening Tables and Figures . . . . .	35
9.7 AI Screening Tables and Figures . . . . .	37
9.8 Traditional Data Extraction and Synthesis Tables and Figures . . . . .	41
9.9 AI Data Extraction Tables and Figures . . . . .	46

## Abstract

Recent advances in Artificial Intelligence (AI) significantly impacted the methodology and efficiency of systematic reviews, addressing challenges such as the ever-increasing volume and complexity of available literature. This work is driven by the dual objectives of providing a comprehensive and up-to-date analysis of the most impactful articles published in ITOR and evaluating the performance of Artificial Intelligence (AI) tools in this domain. To achieve this, the work conducted in *Avenali et al. (2023)* [3] is replicated following the same methodology and criteria, integrating AI tools wherever feasible. At each step, we compare the outcomes of the traditional methodology with those achieved through AI tool. This approach enables us to underscore the current opportunities and limits of AI tools in this domain and demonstrate that an efficient systematic review relies on the strategic combination of the latest tools available for conducting systematic research with human knowledge and judgment.

# 1 Introduction and Motivation

A systematic review is a type of literature review that collects and critically analyzes multiple research studies or papers on a specific topic or research question. It aims to provide an **unbiased summary** of the existing evidence to help inform decision-making or further research in that area. Employing **explicit systematic methods**, this process is designed to be replicable, scientific, and transparent, with the aim of minimizing bias and producing reliable findings from which conclusions can be drawn and decisions made (*Petticrew and Roberts, 2006 [40]*). A systematic review consists of predefined steps, which will be discussed in subsequent sections of this work.

*Borgman (2015) [9]* emphasizes the importance of both big and small data in scholarly inquiry, noting that while big data receives significant attention nowadays, smaller data is equally crucial for close inspection. Moreover, as the absolute volume of data increases, the ability to inspect individual observations decreases, underscoring the need for the introduction of **new technological tools**.

The setting of our case of study is the *International Transactions in Operational Research* (ITOR) academic journal. To celebrate its 30th anniversary, the ITOR's editors have curated a collection of 33 articles from those published between 1994 to 2023, also including the first article published by the journal and the three highly accessed articles from their Hall of Fame series (*Ribeiro et al. (2023) [45]*). Our goal is to replicate the list of the **top 30 articles** by considering all the articles published in ITOR between 1994 and the current date to identify **trends in the data** and illustrate the **potential and limitations of systematic review** versus domain expert-based selection.

Recent advances in **Artificial Intelligence (AI)** significantly impacted the methodology and efficiency of systematic reviews. As highlighted by *Marshall and Wallace (2019) [32]*, automation can **accelerate the production of systematic reviews** by reducing the manual labor involved, addressing challenges such as the **ever-increasing volume and complexity of available literature**. For instance, AI tools can quickly screen large numbers of abstracts and full texts, reducing the time required for manual review by researchers.

As emphasized in *Fabiano et al. (2024) [13]*, despite the potential opportunities of AI tools in efficiency and comprehensiveness, it is crucial to recognize that at the moment these tools should only ever be guided from and used as a **supplement to human content** and clinical insight. Moreover, caution must be exercised to ensure that these tools are not used in a manner that reinforces authors' preconceptions, potentially leading to the creation of **biased results**.

This work is motivated both by the need to provide a **comprehensive and up-to-date analysis** of the most impactful articles published in ITOR and **evaluate** the effectiveness, accuracy, validity, and limitations of AI tools in conducting systematic reviews, aiming to understand their potential impact within this domain.

We will conduct a systematic review using both traditional review and AI tools, comparing in each step the outcomes of the traditional methodology with those achieved through AI tools. This approach will enable us to underscore the **current opportunities and limits** of AI tools in conducting systematic reviews and demonstrate that an efficient systematic review relies on the **strategic combination** of the latest tools available for conducting systematic research with human knowledge and judgment.

## 1.1 Custom AI: ITOR Systematic Reviewer GPT 4

During this work, we experimented with and mixed different AI tools (e.g., [Phind](#), [Copilot](#), [Gemini](#)), but we found the [ChatGPT 4 \(paid version\)](#) to be the most for our type of work, as it allows for easy file uploading and provides writing **prompts** to perform operations on the data. In particular, to give a **specific context** and **enhances efficiency**, we decided to create a **custom AI** for our purpose using the "Explore" option of ChatGPT 4. A preview of the interface of this custom AI can be observed in Figure 18 and is available using the following [link](#) (private use). It requires the initialization of the following fields:

- **Name:** ITOR Systematic Reviewer
- **Description:** Systematic Reviewer for the International Transactions in Operational Research (ITOR) capable of analyzing articles, summarizing their content, and providing valuable insights.
- **Instructions:** This GPT must be capable of opening and analyzing input files, manipulating them as requested, and saving the modifications in output files. Additionally, it should provide valuable insights and recognize patterns in the provided data. It should also generate accurate graphical representations that summarize information about the articles, such as statistics on authors, publication dates, article topics, co-citations, references, and so forth.
- **Conversations Starters:** the prompts used during our Systematic Review to make the work easily **reproducible**.
- **Knowledge:** Scopus and Web of Science databases files.
- **Capabilities:** Web Browsing, DALL-E Image Generation and Code Interpreter.

## 2 Planning

The Planning phase focuses on defining the **research question**, which is crucial for the success of a Systematic Review. A clear and specific research question enables to develop coherent inclusion and exclusion criteria, minimize biases and decide the search strategy.

### 2.1 Knowledge graph

Knowledge graphs are used for various purposes across different fields and industries, and here we use them to give a visual representation of the main topics of the ITOR's journal, obtained by investigating some online sources (e.g., [Index Copernicus Website](#), [IFORS Website](#), [ORES Website](#), [Online Library Website](#)). The primary purpose of these knowledge graphs is to map out the interdisciplinary scope and key areas of research within the journal, thereby providing a comprehensive overview of its intellectual landscape.

Since our aim is to compare AI with traditional-driven systematic review, we produced the knowledge graph both manually, by using [Protégé](#) [Figure 1], and automatically, by employing AI ([Copilot](#) and [Gemini](#)) [Figure 2]. These AI tools were prompted through [Teamboard](#), which generated terms [Figure 14, Figure 16] and relations descriptions [Figure 15, Figure 17], then used to create the knowledge graph.

Since the generated manual and AI knowledge graphs are both similar we will comment on the AI one for simplicity. The AI knowledge graph in Figure 2 highlights key topics within Operational Research (OR), including **Management Science**, **Data Science**, **Optimization**, and **Applied Mathematics**. This suggests that OR heavily intersects with these fields, drawing from their methodologies and contributing to their development. OR's scope includes collaboration and integration across various disciplines, such as **Economics**, **Social Networks**, and **Artificial Intelligence**. This interdisciplinary interface enables methodological transfer, where theoretical concepts from one field are applied to practical problems in another, enhancing the overall impact of OR.

The graph also illustrates how **OR bridges** the gap between theory and practice, leading to innovative applications and improved outcomes. This connection is facilitated through mutual influence and knowledge exchange between academics and practitioners. OR research is rigorously peer-reviewed and published in reputable journals like the International Transactions in Operational Research (ITOR). The role of the International Federation of Operational Research Societies (IFORS) in publishing and promoting OR research highlights the global collaboration and dissemination of findings. There are minor nodes indicating the connection between large-scale networks, simulation problems, and the overall synergy in OR. These nodes reflect the practical applications of OR methodologies in addressing complex, real-world issues.

In summary, the graph outlines the **interdisciplinary nature of OR**, its **connection** between

theory and practice, and its **broad scope covering various fields**, all contributing to the promotion and development of methods, practices, and education in OR.

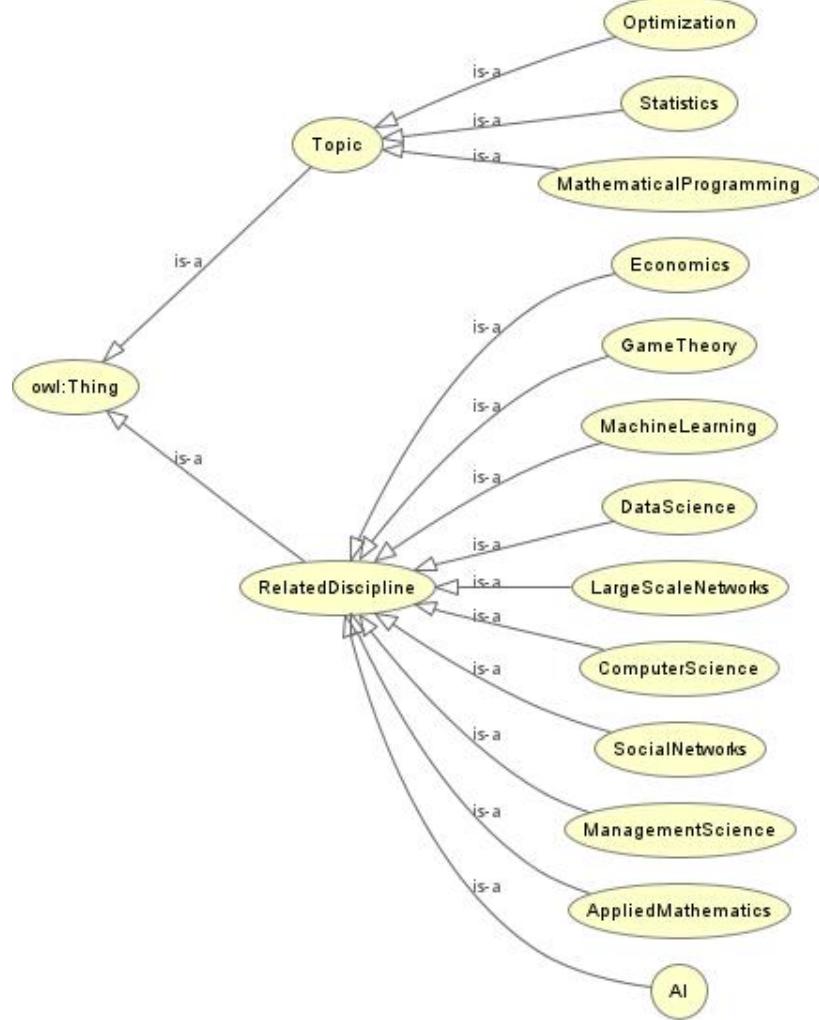


Figure 1: Manual Knowledge Graph

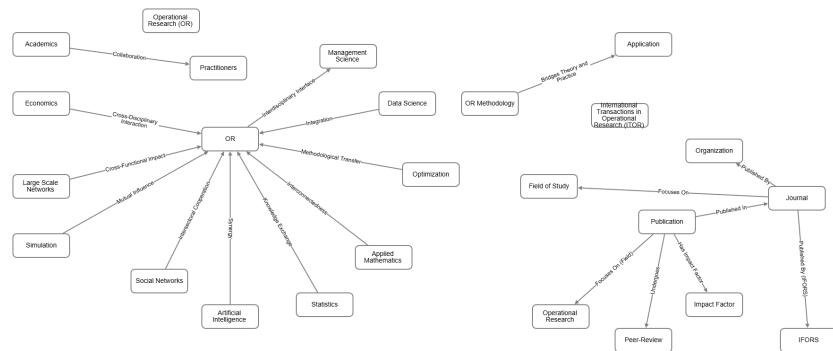


Figure 2: Knowledge Graph generated by AI

## 2.2 Traditional Planning

Our research question is to **identify the 30 most influential articles** published in the International Transactions in Operational Research (ITOR) journal and compare our results with those produced by the ITOR's editor for its 30th anniversary. The *SPICE (Setting, Population or Perspective, Intervention, Comparison and Evaluation)* framework in Table 1 is used to summarize the research question and ensure that the results are reliable, measurable and valid.

The following work will involve the use of *Databases Search engines, Bibliometric tools, and Network Visualization analysis* to identify trends within the setting of interest. The evaluation will be based on the number of common articles identified between the collection of original articles and those replicated by us to assess the results **consistency**.

Table 1: SPICE Table

<b>S (Setting)</b>	International Transactions in Operational Research (ITOR) academic journal
<b>P (Population or Perspective)</b>	Collection of articles published by ITOR's editors for its 30th anniversary
<b>I (Intervention)</b>	Systematic Review using Database Search engines, Bibliometrics tools and Network Visualization analysis
<b>C (Comparison)</b>	Replicating the collection of articles published by ITOR for its 30th anniversary
<b>E (Evaluation)</b>	Number of common articles

The Systematic Review started on **7 May 2024** and involved a **team of 5 people** with different backgrounds. In this type of review, it is crucial to work as a team and to meet *deadlines*, being a process in which all the successive phases are connected. Furthermore, often a third independent reviewer is needed to resolve *conflicts* that may arise during the work. Recording every phase as precisely as possible is also essential due to *reproducibility* constraints.

## 2.3 AI Planning

For the AI part, we used two specialized AI tools to facilitate the systematic review process, Phind and ChatGPT 3.5. We made the prompt as specific as possible, guiding it with terms such as "*imagine you are*" / "*act like*", in order to increase the accuracy of the results. We started the AI planning on **May 8th**, tasking Phind to define a **research question** for our systematic review, aligning with the study by Avenali *et al.* (2023).

**Prompt:** "*Imagine you are a researcher that has to do a systematic review with the goal of recreating the list from the study "Celebrating 30 years of International Transactions in Operational Research" (2023 - Ribeiro), by considering the articles up until today. Can you define the research question?*"

**Phind answer:** "*What are the most influential and impactful articles published in the International Transactions in Operational Research journal from its inception in 1994 to 2024?*"

Additionally, it provided guidelines for conducting the systematic review and its primary objectives. Then, we employed ChatGPT 3.5 to request a **systematic review framework table**, which resulted in the *PICOS (Population, Intervention, Comparison and Outcomes)* table in Figure 3.

**Prompt:** "*The Planning phase focuses on defining the research question, which is crucial for the success of a Systematic Review. A clear and specific research question enables to develop coherent inclusion and exclusion criteria, minimizes biases, and decides the search strategy. So given this definition, imagine that you have to answer the following research question "What are the most influential and impactful articles published in the International Transactions in Operational Research journal from its inception in 1994 to 2024?" intending to compare the articles you will get to the list from the study "Celebrating 30 years of International Transactions in Operational Research" (2023 - Ribeiro), can you give me an example of a systematic review research question framework, among existing ones, that*

*would be suited to defining core outcomes in this systematic review? I'd like for the outcome to be in a form of a table"*

PICOS Component	Description
<b>Population/Participants</b>	Articles published in the International Transactions in Operational Research journal from 1994 to 2024.
<b>Intervention/Exposure</b>	The articles themselves, representing unique contributions to the field of operational research.
<b>Comparator</b>	List of influential and impactful articles identified in the study "Celebrating 30 years of International Transactions in Operational Research" (Ribeiro, 2023).
<b>Outcome</b>	To identify the most influential and impactful articles.
<b>Study Design</b>	Systematic review of articles published in the International Transactions in Operational Research journal from 1994 to 2024.

Figure 3: PICOS Table generated by ChatGPT 3.5

### 3 Search

The Search phase focuses on defining a **search strategy** to identify all relevant studies on the topic of interest. A well-designed search strategy involves selecting databases and identifying appropriate keywords and boolean queries to perform the search.

#### 3.1 Traditional Search

It is important to note that our analysis has a **restricted domain** because it is limited to a single journal and a specific period. After an initial trial, we decided not to apply keyword filters. This decision was motivated by the limited number of results available and the belief that implementing a more restricted search could overly limit our search to effectively conduct the screening phase.

The differences in our search results compared to those obtained in *Avenali et al. (2023)* can be attributed to the timing of our Systematic Review, which was conducted about a year later. This **temporal difference** can have a significant impact on the results due to potential changes in articles in the various databases, such as additions or removals in the meantime. For this reason, documenting the search strategy and the specific timespan for data collection is crucial to ensure its **reproducibility**.

As suggested in *Avenali et al. (2023)*, **Scopus** and **Web of Science (WoS)** databases are selected to perform our search.

##### 3.1.1 Scopus

The data from **Scopus** database was collected on **May 7th 2024 at 17:30** Italian time. The **advanced boolean query** in Listing 1 is employed to select only the articles published in ITOR journal (identified by its ISSN number) and within the specified period of interest (from 1994 to 2024). Applying this query we got **1996 documents, with 1921 of them being articles**.

Listing 1: Advanced Query for Scopus

```
ISSN ( 1475-3995 )
```

##### 3.1.2 WoS

The data from **Web of Science** database was collected on **May 7th 2024 at 17:30** Italian time. This database allows direct filtering by journal, the filter for the period wasn't applied since we need articles from 1994 up until today and the database starts from 2009. After applying this filter, we got

from 1446 documents, with 1204 of them being articles. The data is then downloaded in .bib format, selecting all available information.

### 3.1.3 Search Performance Evaluation

[Bibliometrix](#) is an R package designed to support **quantitative research** and particularly useful for evaluating search performance. When we initially uploaded our data to R in .bib format and executed the `missingData()` function. With a more in-depth analysis, we discovered that the .bib file of Scopus did not include cited references, which were only obtained when we changed the format and downloaded the data in a .csv format. However, changing the format, no duplicates were reported anymore.

Additionally, we found that the two selected databases have a total of **1008 documents in common**, enabling a robust but also diversified Systematic Review. A concise summary of our final Search results can be observed in Table 2.

As recommended by *Avenali et al. (2023)*, to ensure the *integrity* of our data and to assess the *risk of bias* in our future analyses, we employed **Metadata tables** to verify the completeness of information from both the **Scopus** and **Web of Science** databases, respectively in Figure 19 and Figure 20. With this approach, we were able to provide a comprehensive overview of the *completeness* and *quality* of various **Metadata fields**, including Authors, DOI, Document Type, and so forth.

Our analysis revealed that both databases lacked 100% of the **Science categories**, indicating a significant gap in the data, and **Scopus** database missed 52.20% of the **Corresponding Authors**, which can be seen as secondary information.

In terms of **Keywords** and **Keywords Plus** fields, **Web of Science** demonstrated superior performance, with only a small percentage of articles missing these fields (only 13.97% of the articles missing "Keywords Plus" and 3.32% missing "Keywords"). Notably, **Web of Science** contained just one article **without a DOI**, a unique anomaly that stood out in our analysis.

Based on these findings, we anticipate that **Web of Science** will yield better results than **Scopus** in our analyses. This expectation is further supported by the observation that authors with the **highest article counts** in Table 3 are likely to appear in the top 30 articles obtained in the subsequent screening phase.

Table 2: **Scopus** and **Web of Science** Search Engine summary table

Description	Scopus	Web of Science
Timespan	1994:2024	2009:2024
Documents	1996	1446
Articles	1921	1204
Editorials	49	41
Erratum/Correction	4	1
Letter/Notes	5	-
Bibliography	-	1
Reprint	-	1
Reviews	17	25

Table 3: Article Counts in **Scopus** and **Web of Science** databases

Authors	Scopus	Web of Science
Celso C. Ribeiro	32	28
Angel A. Juan	21	19
Mauricio G. C. Resende	15	13
Joseph Wang	-	13
Li Y.	18	17
Kristine Sorensen	11	11
Luis F. Robledo	-	10
Xiang Zhang	12	12
Roberto Morabito	12	10
Mikael Rönnqvist	15	11
Fu-Chiang Yang	12	12
Nelson Maculan	16	-
João C. N. Clímaco	10	-
Chen X	11	-
Yu Y	-	10

### 3.2 AI Search

The AI search part posed some **challenges**. We experimented with several AI tools, but none of them was helpful to aid with the search part. Consequently, we asked **Copilot** to search for the articles of interest by submitting the following prompt:

**Prompt:** *"You are the editor-in-chief of the journal International Transactions in Operational Research (ITOR), and as a supremely informed authority on the evolution and scholarly importance of the contributions published in your journal over the past 30 years, you are invited to do some crucial work. We are asking you to compile a detailed table of as most articles as you can, published in ITOR during the period in question, i.e. strictly in the interval of years 1994-2024. This list should reflect not only the essence of the topics covered and the innovation brought by operations research but also academic recognition and impact as measured by the number of citations. For each selected article, it is imperative that you provide the information as accurately as possible, especially the DOI, to ensure exact identification and ease of access for interested parties. The required table should include the following columns: Title of Article, Authors, Year of Publication, Number of Citations, and Correct DOI. Please ensure that each DOI is accurately verified and matches the cited article exactly, as this information is vital to our analysis and to the academic community that relies on the accuracy of such details."*

However, we already expect that these results may not be satisfactory, as there is still **no open-source free AI tool** that can access the Internet to automate the article search process. Consequently, when we ask the AI to give us studies from the ITOR journal, there is an high probability that it will generate inaccurate information, a phenomenon known as **AI hallucination**[22] [62]. As a result of the prompt, **Copilot** produced the table in Figure 21 and answered as follows:

**Copilot answer:** *"Creating such a comprehensive table requires access to the journal's archives and citation databases, which I don't have. However, I can provide a hypothetical example of how such a table might look. This is a fictional representation and doesn't reflect actual articles or their citations. You would need to access the journal's database and citation indices to compile such a table accurately".*

## 4 Screening

Screening phase plays a crucial role in ensuring that the Systematic Review is focused on **relevant studies**, while eliminating bias by excluding studies that do not meet the inclusion criteria.

### 4.1 Traditional Screening

Initially, **inclusion-exclusion criteria** should be defined based on the research question. In our case, we decided to sort the articles from each database according to the **number of citations**. This choice is motivated by the expectation that the most cited articles are also considered the most relevant in scientific literature and, therefore, those most likely to have been selected by the journal's editors.

It is important to note that, similar to the search phase, the differences in our screening results compared to those obtained in *Avenali et al. (2023)* can be attributed to the timing of our Systematic Review. Indeed, **citation counts can change over time** due to new studies citing the articles.

After using **Bibliometrix** tool to select the top 30 articles by the number of citations, we manually analyzed the articles content using the articles DOI numbers, which uniquely identify them. The obtained results are summarized in Tables [6-4-7](#), each including columns for **Authors** (manually formatted in APA style), **Number of Citations** and **Main Argument** (computed through a manual review of the articles' abstracts). Columns **Ranking** and **Selection** were added to the tables related to the two databases, indicating respectively the article's position in the ranking and whether it is included in the selection for ITOR's 30th anniversary (Y) or not (N).

#### 4.1.1 Top 30 articles for the ITOR's 30th anniversary collected from Scopus

We chose to collect the 30 articles for the ITOR's 30th anniversary from the **Scopus** database because, as indicated in Table [2](#), this database contains a greater number of ITOR articles compared to **Web of Science**, which only includes those from 2009 onwards.

By manually analyzing the articles in Table [6](#), we discovered that the 1995 version of Dell'Amico et al. is not available, while its 2022 version is accessible instead. Additionally, we found that the DOI numbers identifying Bjørndal et al. (2012) ([10.1111/j.1475-3995.2011.00800.x](https://doi.org/10.1111/j.1475-3995.2011.00800.x)) are not correctly listed in **Scopus** database (the correct DOI numbers should be [10.1111/j.1475-3995.2010.00800.x](https://doi.org/10.1111/j.1475-3995.2010.00800.x)). All these considerations underline the fundamental importance of the manual review in **ensuring accuracy** of the Systematic Review results.

#### 4.1.2 Top 30 articles in ITOR collected from Scopus

By manually analyzing the top 30 articles in ITOR collected from **Scopus** database in Table [4](#), we discovered that there are more than three surveys/reviews compared to the **Scopus** classification in Table [2](#). Additionally, considering the Anniversary column, we noticed that only **10 out of the 30 articles chosen for ITOR's 30th anniversary are correctly found**.

#### 4.1.3 Top 30 articles in ITOR collected from Web of Science

By comparing the Table [7](#) of the top 30 articles in ITOR collected from **Web of Science** database with the one generated for **Scopus**, we noticed that **17 out of the 30 articles are common**, sometimes with different rankings. Additionally, considering the Anniversary column, we noticed that only **11 out of the 30 articles chosen for ITOR's 30th anniversary are correctly found**.

Among the top 30 articles for the ITOR's 30th anniversary, Fagerholt (1999) and Liberti et al. (2008) are missing from **Web of Science** database; while Zhao et al. (2016), Liu et al. (2014), and Li et al. (2020) are absent from **Scopus** database.

Table 4: Most cited articles in ITOR (data collected from **Scopus** on **7 May 2024**). The same table for articles collected from Web of Science and those selected for ITOR's 30th anniversary can be found in the appendix.

Article	Citations	Ranking	Main argument ( <b>manual review</b> )	Selection
Sørensen (2015) [51]	708	#1	Lack of scientific rigor in Metaheuristics	Y
Laporte et al. (2000) [24]	515	#2	<b>Survey</b> on heuristics for the Vehicle Routing Problem	N
Chaudhry and Khan (2016) [10]	309	#3	<b>Survey</b> on the development of flexible Job Shop Scheduling (JSS) problem	N
Rais and Viana (2011) [43]	268	#4	<b>Survey</b> on operations research in health care	Y
Tomasini and Van Wassenhove (2009) [54]	260	#5	<b>Systematic Literature Review</b> on Humanitarian logistics	Y
Alba et al. (2013) [2]	241	#6	<b>Survey</b> on Parallel Metaheuristics	N
Sinuany-Stern et al. (2000) [50]	240	#7	Data envelopment analysis (DEA) and analytic hierarchy process combination proposal	N
Colorni et al. (1996) [11]	216	#8	Heuristics applied to Hard Combinatorial Optimization Problems	N
Gehring (1997) [19]	209	#9	Genetic Algorithm (GA) for the Container Loading Problem	N
Guajardo and Rönnqvist (2016) [21]	196	#10	Cost allocation methods in Collaborative Transportation	Y
Fagerhol (1999) [14]	176	#11	Optimal fleet design in a Ship Routing problem	Y
Malaguti and Toth (2010) [31]	172	#12	<b>Survey</b> on Vertex Coloring Problem (VCP)	Y
Archetti and Speranza (2012) [4]	163	#13	<b>Survey</b> on the Vehicle Routing Problems with split deliveries	Y
Van Wassenhove and Pedraza Martinez (2012)	159	#14	Supply Chain Management best practices to Humanitarian Logistics	Y
Festa and Resende (2009a) [16]	153	#15	GRASP Algorithms	N
Kozan and Preston (1999) [23]	153	#16	Genetic Algorithm (GA) to schedule Container Transfers at Multimodal Terminals	N
Yu et al. (2018) [65]	148	#17	Multicriteria Decision Making (MCDM) problem	N
Narbón-Perpiñá and De Witte (2018a) [37]	143	#18	<b>Systematic Literature Review</b> on local Governments' efficiency, Part I	Y
Festa and Resende (2009b) [17]	142	#19	GRASP Applications	N
Rojas Viloria et al. (2021) [57]	142	#20	Aerial drones in Vehicle Routing problems	N
van der Vorst et al. (1998) [55]	139	#21	The impact of Supply Chain Management in Food Chains	N
De Freitas and Penna (2020) [18]	124	#22	Variable Neighborhood Search for Flying Sidekick Traveling Salesman Problem (FSTSP)	N
Yeh (2002) [64]	118	#23	New approach to the selection of compensatory Multi-Attribute Decision-Making (MADM) methods	N
Narbón-Perpiñá and De Witte (2018b) [38]	111	#24	<b>Systematic Literature Review</b> on local Governments' efficiency, Part II	N
Vu et al. (2017) [59]	109	#25	<b>Survey</b> on methods used in Black-Box optimization	N
Liberti et al. (2008) [26]	109	#26	Branch-and-Prune algorithm for the Molecular Distance Geometry Problem	Y
Mingers (2000) [34]	108	#27	The utility of soft and hard OR/MS methods	N
Gehring and Bortfeldt (2002) [20]	107	#28	Parallel Genetic Algorithm (PGA) for the Container Loading Problem	N
Preux and Talbi (1999) [41]	105	#29	Search spaces of Combinatorial Optimization problems	N
Pérez-Bernabeu et al. (2015) [42]	102	#30	Horizontal cooperation in road transportation	N

## 4.2 AI Screening

After verifying that the information regarding the articles produced by **Copilot** in Figure 21 was inaccurate, it was decided to use the files generated by the traditional search from **Scopus** and **Web of Science** for the AI screening phase.

In particular, the following prompt was submitted to **ITOR Systematic Reviewer GPT** to obtain the top 30 most cited papers as an output file:

**Prompt:** *"Select the top 30 articles from the provided file based on the number of citations. The result should be a new file with the same structure as the input one, containing the top 30 most cited articles."*

Afterwards, another prompt was submitted to generate the screening tables similar to those used in traditional screening for comparison purposes. The **Main Argument** column was generated by the AI, summarizing the article's content through analysis of its abstract.

**Prompt:** *"Starting from the file in input, create a table containing the columns Authors, Year, Citations, Ranking and Main Argument for all the most cited articles. The Main Argument column should be a brief summary of the article content based on its abstract."*

The result produced by this prompt can be visualized in Tables 8 for data collected by **Scopus** and in Tables 9 for data collected by **Web of Science**. Comparing them with Table 4 and Table 7 generated manually from the traditional screening, it is evident that **ITOR Systematic Reviewer GPT** successfully ranked the articles by citations and selected the top 30. Additionally, the Main Argument column appears to accurately summarize the content of the articles and identify the presence of **surveys/reviews**.

After selecting the most cited articles, as we did in the traditional screening, it is crucial to verify the accuracy of the provided information. Therefore, the following prompt was submitted, requesting **ITOR Systematic Reviewer GPT** to verify the information contained in the files regarding the top 30 most cited articles from both **Scopus** and **Web of Science**.

**Prompt:** *"Use the DOI numbers, which uniquely identify the articles, to verify if the information in the file matches the information about the articles online."*

In response to this prompt, the AI emphasized the **limitations** of this operation by answering that its ability to retrieve external data is limited and recommending the use of external platforms to verify the information in the files.

**ITOR Systematic Reviewer GPT 4 answer** *"To proceed with verifying the information in your file using the DOIs, I can attempt to search for article details in databases that allow access to this information via DOIs. However, it's important to note that my ability to retrieve external data is limited, so I can only try to verify the information based on descriptions and metadata of the articles that I might be able to find publicly."*

*It seems there have been technical issues attempting to use the search tool to verify the DOIs. I apologize for the inconvenience. I can try an alternative, more manual or general approach to verify the article information based on the DOIs you provided. Additionally, you might consider using external platforms that support direct DOI verification, such as Google Scholar, CrossRef, or similar, to check if the information matches that in your file."*

## 5 Data Extraction and Synthesis

Data Extraction and Synthesis phase involve systematically extracting relevant data from the studies. Once the data is collected, it is summarized using **graphical representations**, like graphs or tables, and interpreted to draw meaningful **insights** and **conclusions**.

### 5.1 Traditional Data Extraction and Synthesis

Using Bibliometrix, we analyse the distribution of articles per country for both Scopus and WoS databases (table 10). China leads in both databases, contributing 25.79% of Scopus articles and 24.48% of WoS articles. Brazil and the USA follow, with Brazil contributing 8.21% to Scopus and 11.07% to WoS, while the USA contributes 6.97% to Scopus and 6.09% to WoS. European countries such as Portugal, Spain, France, and Italy also show significant contributions in both databases.

A bibliometric analysis was conducted using the metadata from Scopus to identify the most prevalent topics in all the ITOR articles. The results revealed that **integer programming**, **sales**, and **decision making** were the most discussed topics (Figure 22). However, upon closer examination of the keywords provided by the authors, it became evident that the first three keywords were **optimization**, **heuristics**, and **integer programming**, respectively (Figure 23).

Let us now consider the most frequently cited articles and those from the ITOR's 30th anniversary special selection. As illustrated in Figure 4 of the keywords plus, we observe that the most frequent words in the cited articles are **optimization**, **heuristics**, and **metaheuristics**. In the special selection (Figure 24), the second and third posts are replaced by **combinatorial optimization** and **operations research**. Instead, the most frequently used keywords by authors in the most cited articles are **heuristics** and **metaheuristics** (Figure 25); while **heuristic** always remains in first position for the special selection of articles (Figure 26).

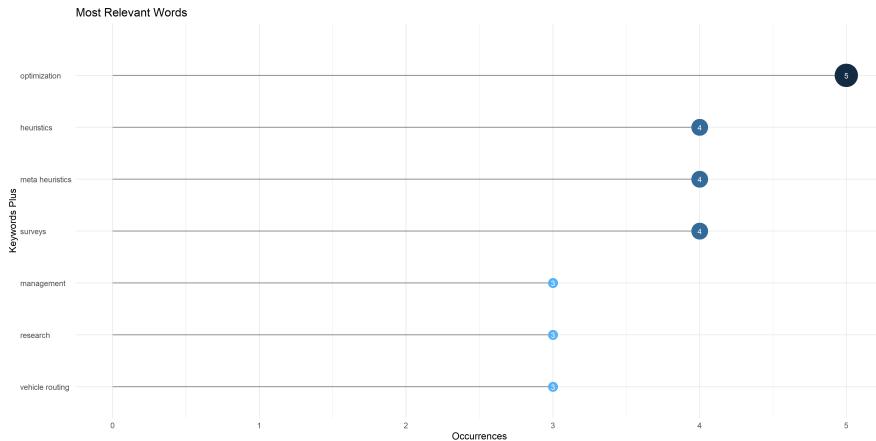


Figure 4: Keyword plus frequency analysis for most cited articles from **Scopus**. The other figures can be found in the appendix.

The co-citation analysis of all and the most cited articles was carried out via **Web of Science** and via **Scopus** for the ITOR 30th anniversary selection. As shown in Figure 5, all the articles are represented by five clusters. The red cluster focuses on strategic and optimization issues in supply chain management. The blue cluster primarily deals with advanced optimization methods, metaheuristics like GRASP, and solutions for complex routing and traveling salesman problems.

In the graph of the most cited articles (Figure 27) they are represented by five clusters. On the other hand, in the case where we used **Scopus** (Figure 28) for the ITOR 30th anniversary selection, since **Web of Science** lacks articles before 2009, there are only **two clusters**.

With the help of VOSviewer a co-occurrence network of all the ITOR publications found on Web of Science was constructed (Fig. 6). By leveraging the concept of **Total Link Strength (TLS)**, which measures in how many publications two keywords occur together, the aim is to provide some help for creating a research agenda. This was done by only including keywords with at least five occurrences. Out of 6053 keywords, 365 meet the threshold. These 365 keywords are connected by 7414 links and can be divided in 6 clusters: Red (119 keywords), green (97 keywords), blue (91 keywords), yellow (30 keywords), purple (26 keywords) and turquoise (2 keywords). As the yellow, purple and turquoise clusters aren't significant enough, only the other three clusters are considered for constructing a research agenda. The red and green clusters primarily focus on methodologies, algorithms, models, and quantitative approaches used for identifying optimal solutions, allocations, or performance measures in scenarios involving discrete or stochastic nature. These include techniques such as data envelopment analysis, linear programming or lagrangian relaxation. The blue cluster predominantly encompasses terms related to economics, business, and decision-making. Figure 7 depicts the density visualization of the most frequently occurring terms found in papers published by the journal ITOR.

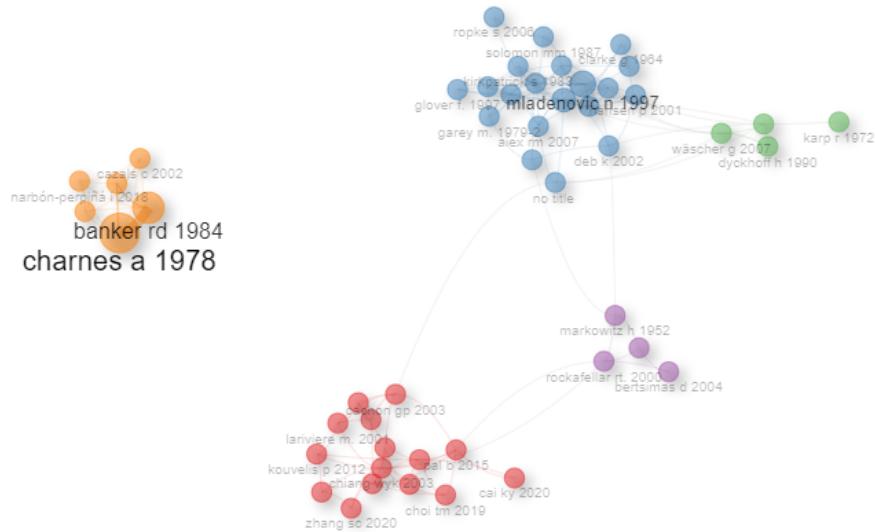


Figure 5: Co-citation all articles in ITOR from **Web of Science**. The same figure for the most cited articles from Web of Science and for the articles of the special selection can be found in the appendix.

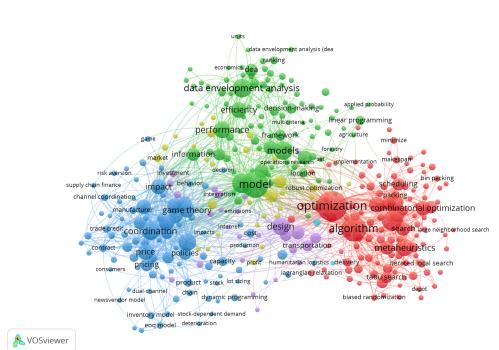


Figure 6: Network visualization of the most covered topics in operational research (from ITOR publications)

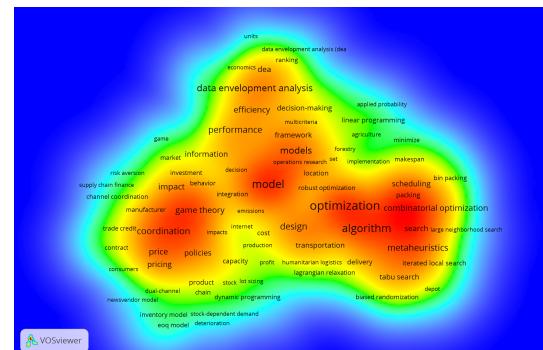


Figure 7: Density visualization of the most covered topics in Operational Research (from ITOR publications)

In the latest bibliometric analysis of ITOR publications, we utilized the *CitNetExplorer* tool to explore key papers and identify foundational research in the field, focusing on papers that contain "review" or "survey" in their titles, abstracts, or keywords. This search from the WoS database yielded **123** papers, with **70** being selected for further analysis based on their relevance to theoretical or empirical developments in operational research. 46 are classified as Articles, 23 as Review Articles and one as a Bibliography, and the data was collected on *May 10, 2024* at 4:00 pm GMT +1.

The analysis aimed to classify these reviews using a clustering approach, while also exploring the scientific heritage of one of these reviews with the help of CitNetExplorer. We chose not to utilize the "non-matching cited references" feature, instead focusing solely on the dedicated network of ITOR reviewers. A total of nine clusters (Figure 29) were identified from the timeline-based network spanning from 2009 to 2024. However, 39 publications did not fit into any cluster. The minimum cluster size chosen is **2** (citations) with a resolution of 0.75, one random start and 10 iterations.

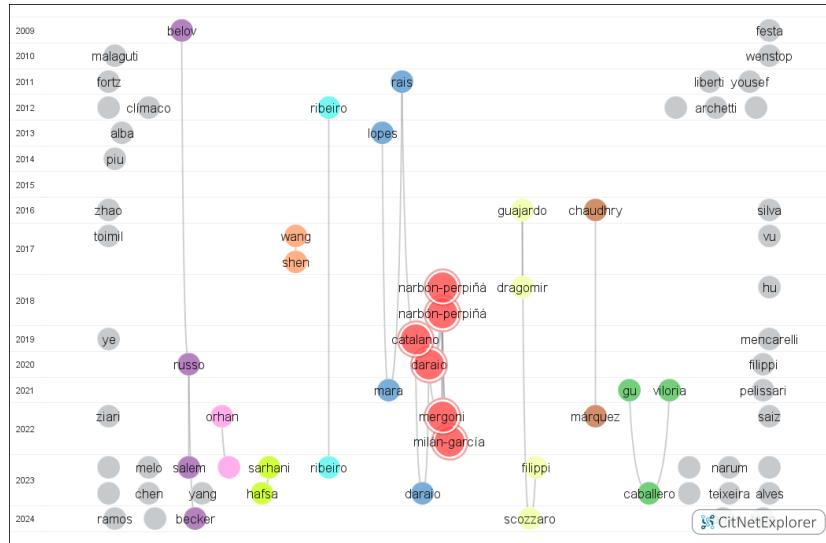


Figure 8: Citation network of ITOR reviews with core publications highlighted in red

Cluster one, depicted in blue, encompasses 10 papers which focus on efficiency and optimization in various sectors (such as healthcare, logistics and transportation or the public sector), with an emphasis on empirical analysis and practical applications - Data Envelopment Analysis, stochastic frontier analysis, systematic literature reviews etc. The second cluster (purple) has 4 reviews that highlight packing and cutting problems. Furthermore, the yellow cluster contains 4 publications which center around collaborative transportation and logistics optimization. A similar topic is found in the next cluster, with 3 publications (green) - the application of operational research methods to autonomous vehicles and unmanned systems. The next clusters all contain 2 papers each, with the orange one focusing on the impact of market disruptions on supply chain performance and management. The papers in the brown cluster target scheduling techniques in manufacturing environments, with an emphasis on flexible job shop scheduling, while the pink cluster focuses on analytics and decision support in sector-specific applications. Lastly, the turquoise one is formed around optimization techniques in sports tournament scheduling and the bright green cluster encompasses advancements in metaheuristic initialization strategies.

Additionally, the analysis identified **six** core publications (Figure 8) that are central to the network and have substantial citation links, ensuring their influential status within the field: (*Catalano et al., 2019*), (*Daraio et al., 2020*), (*Mergoni and De Witte, 2022*), (*Milán-García et al., 2022*), (*Narbón-Perpiñá and De Witte, 2018a*) and (*Narbón-Perpiñá and De Witte, 2018b*). These core publications mainly dealt with Data Envelopment Analysis and efficiency measurement techniques. Our approach in finding such publications involves employing both maximum and minimum thresholds as part of a reasoning process (Nepomuceno et al., 2023b; 2022b). By applying Nepomuceno et al.'s (2023b)

approach, we can ensure that core publications within the network are interconnected and have significant influence within the field, contributing to a more robust understanding of the network's structure and dynamics. We assigned "3" as the minimum number of citation links (incoming or outgoing with other core publications); within the network each core publication must have citation relations with at least three other core publications.

Exploring the *scientific heritage* entails understanding the expand tool within CitNetExplorer. Within a bibliometric lineage subnetwork, predecessors signify publications referenced by a defined minimum within the current subnetwork, while successors refer to publications citing at least a set minimum from the current subnetwork. Leveraging the core publication by Daraio et al. (2020), in Figure 30 we expand the current network, with a citation link minimum of "1" and maximum distance.

## 5.2 AI Data Extraction and Synthesis

Data extraction using AI had previously been attempted using free tools such as *Perplexity*, but these tools did not produce satisfactory results. As a result, we switched to using **ITOR Systematic Reviewer GPT 4** for data extraction and synthesis.

**Prompt:** *"I have a CSV file containing a list of all ITOR articles from SCOPUS with their associated 'Index Keywords' as well as 'Author Keywords'. Could you analyze the frequency of these keywords, normalize them by merging similar terms regardless of their case, and create two scatter plots visualizing the most frequent keywords and their occurrences? Please do a detailed report on your findings."*

**ITOR Systematic Reviewer GPT 4 answer:** *"Analyzing the SCOPUS data for ITOR articles, both index and author keywords underscore a strong focus on optimization and supply chains, reflecting core themes in operational research. Index keywords (Figure 31) emphasize broad application domains such as "commerce" and advanced techniques like "integer programming," highlighting their relevance in commercial and mathematical contexts. In contrast, author keywords (Figure 32) reveal a preference for specific methodologies and tools, including "metaheuristics," "simulation," and "game theory," suggesting a deeper dive into nuanced problem-solving approaches. This distinction highlights that while index keywords capture the broader impact and areas of research, author keywords detail the specialized tools and theoretical frameworks employed by researchers in the field."*

**Prompt:** *"I want a similar keyword frequency analysis and report on the 'Index Keywords' and on the 'Author Keywords' for each CSV provided. One contains the most cited ITOR articles and the other - 30th anniversary selection articles."*

**ITOR Systematic Reviewer GPT 4 answer:**

**Index Keywords/ Keywords Plus:** *"The index keywords from the most cited ITOR articles (Figure 9) reveal a strong emphasis on "Optimization," "Heuristics," and "Vehicle Routing," highlighting their significant role in addressing practical problems within transportation and logistics. The presence of "Meta Heuristics" and "Management" indicates a focus on advanced methodologies and their managerial applications. Conversely, the editorial choice articles (Figure 33) emphasize "Optimization" and "Research," with a notable presence of "Transportation" and "Operations Research," pointing to a broader interest in foundational studies and strategic impacts within the field. The comparison reveals that while optimization is a central theme in both datasets, the most cited articles focus more on specific methodologies and practical applications, whereas the editorial choices tend to highlight broader research themes and foundational topics in ITOR."*

**Author Keywords:** *"The author keyword analysis from the most cited ITOR articles (Figure 34) shows a strong emphasis on "Heuristics," "Metaheuristics," and "Genetic Algorithm," reflecting a focus on advanced problem-solving methodologies. These articles also frequently mention "Combinatorial Optimization" and specific problems like "Vehicle Routing Problem" and "Traveling Salesman Problem," highlighting targeted research areas in optimization. In contrast, the editorial choice articles (Figure 35) feature "Optimization" and "Transportation" alongside niche areas like "Humanitarian Lo-*

*gistics*" and "Collaborative Logistics," suggesting a broader thematic range with emphasis on practical applications in logistics and transportation. The comparison reveals that while both datasets prioritize optimization, the editorial choice articles extend into specialized logistics areas, indicating a diverse interest in applying theoretical methods to practical, impactful domains."

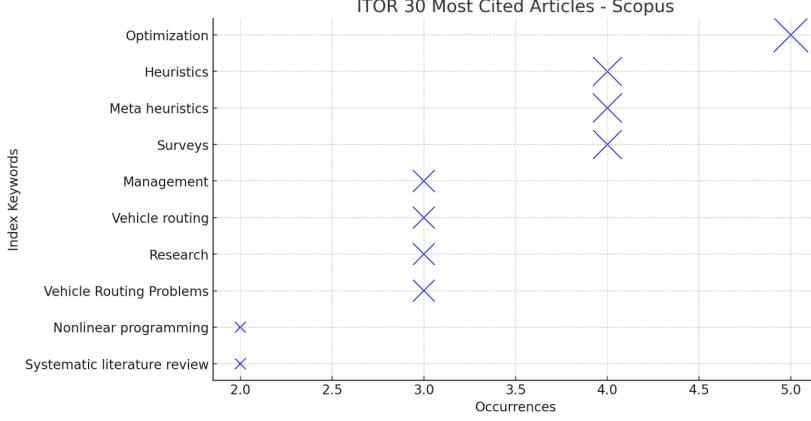


Figure 9: Index keywords frequency analysis on the top 30 most cited ITOR articles from **Scopus** generated using **ITOR Systematic Reviewer GPT 4**. The other figures can be found in the appendix.

This advanced AI model allowed us to analyse the 30 most cited articles provided by the Web of Science database. Here is the prompt that was used to get the results from **ITOR Systematic Reviewer GPT**, along with the response that included comments on various graphs.

**Prompt:** *"You have been given the file containing data on the most cited articles from Web of Science. Your task is to analyze this dataset and write a detailed report about your findings."*

**ITOR Systematic Reviewer GPT 4 answer:** *"This systematic review analyzes the collaboration network of authors from the 30 most cited articles in the journal "International Transactions in Operational Research." The review focuses on identifying prolific authors, publication trends, prevalent research topics, and collaboration clusters. The findings highlight key authors and research groups, providing insights into the collaborative landscape within this field."*

**Prolific Authors:** *"The most prolific author is Wang, Jian-qiang, with 3 articles in the dataset."*

**Most Frequent Words:** *"The most frequently used words (Figure 10) in the article include "meta-heuristics," "research," "survey," "review," and "operations." These terms suggest a strong focus on methodological approaches, comprehensive reviews, and studies within the field of operations research."*

**Publication Years Distribution:** *"The distribution of publication years (Figure 11) indicates that these highly cited articles are spread across a range of years, with noticeable peaks in 2012, 2018, and 2019. This suggests a consistent production of influential research over the past decade."*

**Prompt:** *"Could you generate a collaboration network graph? I'm interested in visually understanding the connections and relationships among different entities in this network."*

**ITOR Systematic Reviewer GPT 4 answer:** *"The collaboration network graph (Figure 36) visually represents the connections among different authors based on their co-authorship of the most cited articles. Each node represents an author, and each edge indicates a collaboration between two authors."*

**Prompt:** *"Could you please conduct a detailed analysis of this graph, focusing specifically on identifying and highlighting clusters of collaboration? These clusters may correspond to specific research*

*groups or institutions.”*

**ITOR Systematic Reviewer GPT 4 answer:** "The collaboration network graph with highlighted clusters (Figure 12) visually identifies groups of authors who frequently collaborate. **Detailed Examination of Key Clusters:** **Cluster 0** (7 authors): Includes authors like Yu, Su-min, and Wang, Jing. This cluster might be focused on a specific niche within operational research, likely around meta-heuristics or optimization techniques, given the keywords associated with their articles. **Cluster 1** (4 authors): Includes authors such as Rojas Viloria, Daniela. This cluster might be working on a particular aspect of supply chain or logistics, suggesting a specialized research group possibly within a single institution or closely collaborating institutions. **Cluster 2** (4 authors): Consists of authors like Hamadi, Youssef, and Liberti, Leo. The presence of such a cluster indicates a focused collaboration, likely addressing optimization and algorithmic research given the authors' expertise."

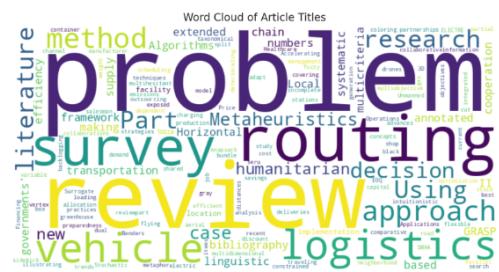


Figure 10: Keyword Word Cloud of the most cited articles from Web of Science generated using **ITOR Systematic Reviewer GPT 4**

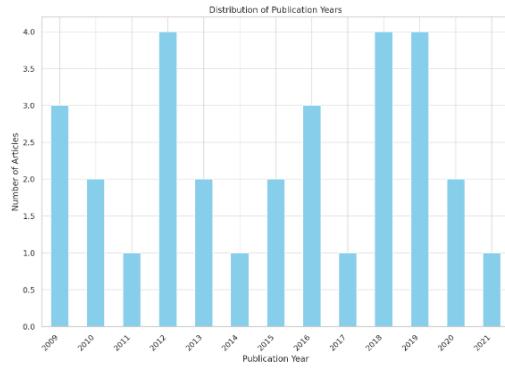


Figure 11: Publication trends over time of the most cited articles from **Web of Science** generated using **ITOR Systematic Reviewer GPT4**

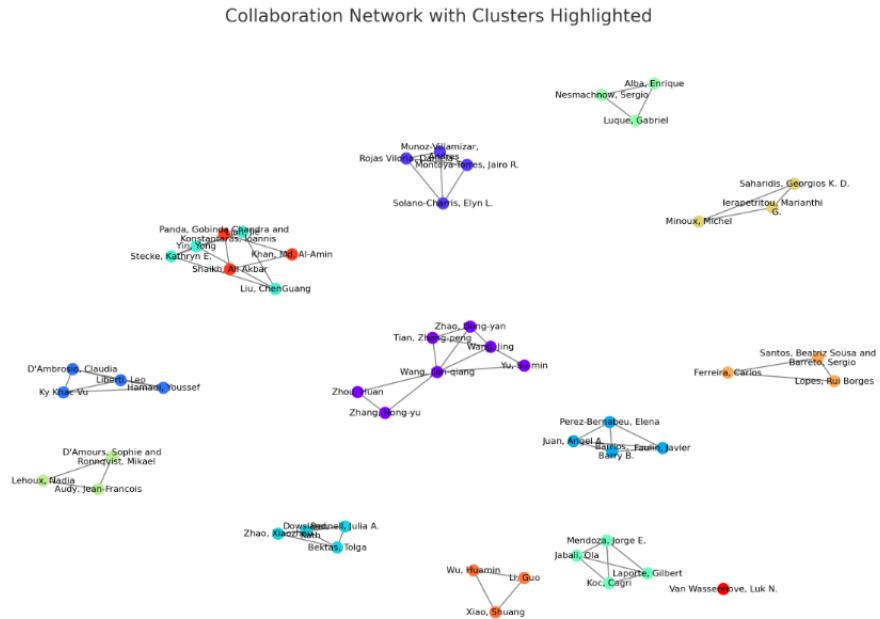


Figure 12: Collaboration network graph with clusters highlighted of the most cited articles from **Web of Science** generated using **ITOR Systematic Reviewer GPT 4**

## 6 Quality Assessment

The Quality Assessment phase plays a crucial role in evaluating the **strength of results** provided by each stage. This involves examining factors such as the methods employed for data collection, the accuracy of the obtained results, and the identification of any limitations.

### 6.1 Traditional Quality Assessment

Properly **planning** the Systematic Review is essential to obtain quality results. We decided to split the work in two groups, one tasked with manual searching and screening, while the other focused on data extraction and synthesis. Despite this, we have consistently discussed decisions and results together, with a third member acting as an arbitrator in case of problems.

A notable challenge emerged during the **search phase**, particularly regarding data selection for the 30th anniversary analysis. In fact, it is important to emphasize that not all **formats** in which search results can be saved in various databases are suitable for certain data extraction operations. Additionally, the chosen databases have internal **limitations**, such as incorrect data or missing articles. Experimenting with other databases can be interesting to see if these limitations can be overcome.

The 30 most cited articles obtained in the **screening phase** are manually validated using their DOI numbers. This process was time-consuming but also ensured the accuracy of our results, leading to the discovery of some data errors, as previously highlighted. The results achieved in this phase were quite satisfactory, but they also emphasized the need to define **new inclusion-exclusion criteria** for the selection of articles in the top 30.

Another challenge occurred during the **data extraction and synthesis** phase importing data into two bibliometric analysis tools, namely **VOSviewer** and **CitNetExplorer**. This complication was successfully addressed by downloading the data in alternative **formats**, such as tab-delimited and plain text. The results of this part were quite accurate, but it was not possible to manually validate the quality of the results as it relied on external tools.

### 6.2 AI Quality Assessment

Assessing the quality of AI-generated work is crucial, especially when evaluating its capacity to handle systematic reviews. Since we noticed that **ITOR Systematic Reviewer GPT 4** is **unable to identify errors or validate its own work**, we opted for a manual quality assessment of AI-generated work as it can provide insights and judgments that AI might overlook.

During the **planning phase**, both **Phind** and **ChatGPT 3.5** understood the objective of this stage, by correctly identifying the research question and constructing the PICOS table.

On the other hand, the **search phase** highlighted the limitations of AI tools, posing many **challenges**. In fact, most AI tools lack direct access to academic databases, which makes it hard for them to get the articles and leads them to generate inaccurate or entirely invented data (**AI hallucination**). The AI tools we tested and their limitations are outlined in Table 5.

We experimented with several open-source AI tools recommended for the **screening phase**, such as **Covidence**, **Distiller SR**, and **Rayyan.ai**. However, only **ITOR Systematic Reviewer GPT 4** shown to be quite suitable for our work, generating **satisfactory results** and accurately summarizing the content of the articles. However, it is important to underline some **challenges** even in this phase. Firstly, **ITOR Systematic Reviewer GPT 4** is unable to work with all types of formats: from our experience, it performed well with CSV files, while it was unable to work with BIB and XML format. Additionally, **ITOR Systematic Reviewer GPT 4** is unable to verify whether the information contained in the files is accurate, due to limited access by the AI to academic databases. All these challenges led to significant manual intervention to support the AI-generated work.

Initially, we employed **Perplexity** for **data extraction**, but encountered several limitations with its free version. These included restrictions on the number of pro queries that could be submitted (limited to five queries) and constraints on the number of files that could be uploaded. Consequently, we transitioned to utilising **ITOR Systematic Reviewer GPT 4**, which proved advantages as it has the ability to generate graphs and higher performance of analysis on large datasets. Nevertheless, instances of inaccuracy were observed in the retrieved information, which appeared to originate from issues with the handling of the code, data, and graphic objects. When the desired answer is not found, **ITOR Systematic Reviewer GPT 4** often generates invented responses and provides incorrect information. To obtain the correct output, it is often necessary to repeat the request multiple times and provide more detailed instructions.

## 7 Cost-Benefit Analysis

In this section we conduct a cost-benefit analysis. Initially, we list all the potential costs and benefits and come up with a **decision rule**, whether to use AI tools when doing a systematic review. Then, we try to **quantify** the costs and benefits by leveraging the fact that we conducted a systematic review with and without AI tools. Finally, we try to translate our findings into a more **general framework** and **critically assess** our underlying assumptions.

The primary advantage of using AI tools is the **time saved**. Simple manual tasks such as working through databases are automated. This advantage is also associated with **costs**: researchers need additional training to use AI tools and there are also costs associated with acquiring/renting the AI tools.

Moreover, as most AI tools state themselves, they can make mistakes and important information needs to be checked, which leads to **additional costs**. There are other costs and benefits whose quantification underlay high uncertainty, for the sake of simplicity we leave them out of our model.

Nevertheless, we should keep this consideration in our mind: basing the systematic review on AI tools could allow to involve people without a background in the specific research field. This can make it easier and possibly cheaper to get together a research team. Additionally, there is the possibility that AI performs better than humans in some stages and, therefore, improves the overall quality of the systematic review. A possible cost factor could be the implementation/consideration of data protection and privacy, as you basically distribute private firms with your research.

Having all our costs and benefits listed allows us to come up with a decision rule whether to include AI tools or not. We developed an **equation**, which simulates the point where a research team is indifferent to use AI or not. On left side of the equation are all costs and on the right side all benefits. The interpretation is pretty straight forward: if the left side of the equation is bigger than the right side the benefits of using the AI are bigger than the costs. Instead of the decision rule we have derived here, one could also have calculated the NPV of benefits minus costs. We have deliberately chosen not to do this, as both approaches lead to the same result and we consider our approach to be simpler and more intuitive.

$$\begin{aligned} MV(\text{Time saved}) &= MV(\text{Renting/Acquiring AI tool}) \\ &\quad + MV(\text{Time additional Training}) \\ &\quad + MV(\text{Time for checking results}) \end{aligned}$$

where  $MV$  denotes as **monetary value** and this quantification will be done in the next step.

The **estimated time saved** is the time we needed to do the systematic review manually minus the time we needed to do it using AI. This is estimated by approximating the time we needed to separately conduct the systematic review manually and using AI. The time needed to study the theory and get to know best practices is not being considered. For the systematic review using AI this is

estimated in a separate variable (**Time additional training**). Further, also the time needed for checking the AI results is not considered and estimated in a separate variable (**Time for checking results**).

We try to quantify everything in euros, but it is crucial to emphasize that this is an **oversimplification of reality**: In the model we developed, we are working with several assumptions. We consider, that all costs and benefits can be scaled down to one time period, where time period denotes one research. We assume new AI tools must be acquired every time period and researchers need new training. This assumption is not completely unlikely, as the world of AI is rapidly evolving and the corresponding tools are constantly replaced. Additionally we leave out of our considerations, that the AI tools didn't work perfectly. We assume they produced the exact same results as the manual review. This assumption is vital, as incorrect research has no value and the costs stemming from it, would outweigh all benefits.

We measured our time variables in hours and multiplied them with the average salary for a student job in Italy. We estimate this to be **10 euros per hour**. It took us about **9 hours to do the manual systematic review** and **15 minutes to do it with AI**. We needed **10 hours of additional training** to develop best practices and **2.25 hours to check the AI results**. We used ChatGPT 4, for which we had to pay **25 euros**. This lets us arrive at the following values:

$$MV(\text{Time saved}) = (9 - 0.25) \times 10 = 87.50$$

$$MV(\text{Renting/Acquiring AI tool}) = 25 \text{ euros}$$

$$MV(\text{Time additional Training}) = 10 \times 10 = 100$$

$$MV(\text{Time for checking results}) = 2.25 \times 10 = 22.50$$

If we put these values into our equation we get the following:

$$87.50 < 25 + 100 + 22.50 = 147.50$$

In our example the **benefits** (left side of the equation) are smaller than the **costs** (right side of the equation). This means that in our case we should do the systematic review manually and not including AI. How does it look in different settings? Can we translate some of our findings into a more general framework?

The most obvious thing coming to eye when looking at the estimations of our variables, is the hourly wage. For a team of qualified researchers it will be higher than the 10 euros for students. Benefits and costs are affected by this. It ends up to be a null sum, as it can also be assumed, that a proper research team needs a more expensive AI tool.

In our model we did not include several variables and made some assumptions, if changed could result in a completely different model. We can conclude, that including AI does not make sense in our case, if it does at all depends on the specific case.

## 8 Conclusions

During our Systematic Review, we strictly followed the methodology outlined in the literature. We conducted a comparative analysis of the outcomes derived from the two selected databases, Scopus and Web of Science, and we employed various data extraction and synthesis tools to identify trends within the ITOR’s editors selection for its 30th anniversary.

It is evident that the articles selected by the journal’s editors with highest number of citations were accurately identified during the screening phase, while those with fewer number of citations were overlooked. This suggests that considering only the number of citations was not the unique criteria of the journal’s editors. Instead, they probably introduced **additional qualitative criteria**, such as **their own experience** in the journal and **their aim to create a diversified selection**, including both famous articles and niche ones, in order to reflect the journal evolution over the years. These criteria were further investigated by examination factors such as **publication years** and **author’s countries**.

In contrast to the study conducted by *Avenali et al. (2023)*, our approach aimed to integrate various **Artificial Intelligence (AI) tools** into the process to assess their performance in handling systematic reviews. This integration presented us with several opportunities, yet it also posed several challenges. While AI tools can certainly assist in several phases of a systematic review, it is also evident that they are still not yet capable of **autonomously** managing the entire process.

Firstly, it is important to emphasize that many AI tools designed for systematic reviews are **not open-source**, significantly limiting the options available. Additionally, many AI tools have access to only a **limited number of online resources**, thereby hindering their ability to search for articles and verify their information. Enhancing the capabilities of AI tools to access a broader range of online resources can improve their effectiveness in conducting literature searches.

Another significant limitation of AI tools is that they suffer from the phenomenon known as **AI hallucination** [22] [62]. This refers to their tendency to produce inaccurate or entirely falsified responses when lacking sufficient information for an accurate answer. This limitation often leads to the need for manual verification of results and the repetition of requests multiple times, along with providing more detailed instructions.

Moreover, the importance of specific and accurate prompts should not be underestimated, as generic prompts can lead to incomplete or misleading outputs. This requires a strong **prompt engineering capability** to maximize the AI capabilities and ensure their quality.

The integration of AI tools can broaden the perspective of our analysis and automate certain tasks, thereby **saving time and reducing the workload**. Therefore, relying solely on them can lead to **errors and biases**, potentially resulting in the inclusion of inaccurate data in the review.

All these considerations underline the importance of the **collaboration** between human experts supervision and AI tools to ensure the robustness and accuracy of the systematic review: human experts can guide the AI tools using prompts and validate and interpret their results, while AI tools can assist human experts by automating repetitive tasks and facilitating data analysis when dealing with large databases.

In conclusion, we can assert that the AI tools experimented in this work are **not currently fully suitable** for conducting Systematic Reviews, as **human judgment** plays a pivotal role in this process to ensure a critical analysis and there are existing tools specifically tailored for this purpose that may offer more comprehensive capabilities.

## References

- [1] Ikpe Justice Akpan and Asuama A. Akpan. Multiple criteria analysis of the popularity and growth of research and practice of visual analytics, and a forecast of the future trajectory. *International Transactions in Operational Research*, 28(5):2275–2298, 2021.
- [2] E. Alba, G. Luque, and S. Nesmachnow. Parallel metaheuristics: recent advances and new trends. *International Transactions in Operational Research*, 20(1):1–48, 2013.
- [3] Simone Di Leo Giorgio Matteucci Thyago Nepomuceno Alessandro Avenali, Cinzia Daraio. Systematic reviews as a metaknowledge tool: caveats and a review of available options. 30(16):2761–2806, 2023.
- [4] C. Archetti and M.G. Speranza. Vehicle routing problems with split deliveries. *International Transactions in Operational Research*, 19(1-2):3–22, 2012.
- [5] Claudia Archetti, Nicola Bianchessi, Stefan Irnich, and M. Grazia Speranza. Formulations for an inventory routing problem. *International Transactions in Operational Research*, 21(3):353–374, 2014.
- [6] J. F. Audy, N. Lehoux, S. D’Amours, and M. Rönnqvist. A framework for an efficient implementation of logistics collaborations. *International Transactions in Operational Research*, 19(5):633–657, 2012.
- [7] Franco Basso, Sophie D’Amours, Mikael Rönnqvist, and Andrés Weintraub. A survey on obstacles and difficulties of practical implementation of horizontal collaboration in logistics. *International Transactions in Operational Research*, 26(3):775–793, 2019.
- [8] Trond Bjørndal, Ingrid Herrero, Alexander Newman, Carlos Romero, and Andrés Weintraub. Operations research in the natural resource industry. *International Transactions in Operational Research*, 19(1-2):39–62, 2012.
- [9] Christine L. Borgman. *Big Data, Little Data, no Data: Scholarship in the Networked World*. MIT Press, Cambridge, MA, 2015.
- [10] I.A. Chaudhry and A.A. Khan. A research survey: review of flexible job shop scheduling techniques. *International Transactions in Operational Research*, 23(3):551–591, 2016.
- [11] A. Colorni, M. Dorigo, F. Maffioli, V. Maniezzo, G. Righini, and M. Trubian. Heuristics from nature for hard combinatorial optimization problems. *International Transactions in Operational Research*, 3(1):1–21, 1996.
- [12] Marco Dell’Amico, Francesco Maffioli, and Peter Värbrand. On prize-collecting tours and the asymmetric travelling salesman problem. *International Transactions in Operational Research*, 2(3):297–308, 1995.
- [13] Nicholas Fabiano, Arnav Gupta, Nishaant Bhambra, Brandon Luu, Stanley Wong, Muhammad Maaz, Jess G. Fiedorowicz, Andrew L. Smith, and Marco Solmi. How to optimize the systematic review process using ai tools. *Journal of Clinical and Experimental Investigation*, 2(1), 2024.
- [14] K. Fagerholt. Optimal fleet design in a ship routing problem. *International Transactions in Operational Research*, 6(5):453–464, 1999.
- [15] Kjetil Fagerholt. Optimal fleet design in a ship routing problem. *International Transactions in Operational Research*, 6(5):453–464, 1999.
- [16] Paola Festa and Mauricio G. C. Resende. An annotated bibliography of grasp – part i: Algorithms. *International Transactions in Operational Research*, 19(1-2):307–322, 2009a.
- [17] Paola Festa and Mauricio G. C. Resende. An annotated bibliography of grasp–part ii: Applications. *International Transactions in Operational Research*, 16(2):131–172, 2009b.

- [18] J. C. De Freitas and P. H. V. Penna. A variable neighborhood search for flying sidekick traveling salesman problem. *International Transactions in Operational Research*, 27(1):267–290, 2020.
- [19] H. Gehring. A genetic algorithm for solving the container loading problem. *International Transactions in Operational Research*, 4(5-6):401–418, 1997.
- [20] H. Gehring and A. Bortfeldt. A parallel genetic algorithm for solving the container loading problem. *International Transactions in Operational Research*, 9(4):497–511, 2002.
- [21] M. Guajardo and M. Rönnqvist. A review on cost allocation methods in collaborative transportation. *International Transactions in Operational Research*, 23(3):371–392, 2016.
- [22] Built In. Ai hallucination, 2023. [Online; accessed 28-April-2024].
- [23] E. Kozan and P. Preston. Genetic algorithms to schedule container transfers at multimodal terminals. *International Transactions in Operational Research*, 6(3):311–329, 1999.
- [24] G. Laporte, M. Gendreau, J.Y. Potvin, and F. Semet. Classical and modern heuristics for the vehicle routing problem. *International Transactions in Operational Research*, 7(4-5):285–300, 2000.
- [25] Guo Li, Huamin Wu, and Shuang Xiao. Financing strategies for a capital-constrained manufacturer in a dual-channel supply chain. *International Transactions in Operational Research*, 27(5):2317–2339, 2020.
- [26] Leo Liberti, Carlile Lavor, and Nelson Maculan. A branch-and-prune algorithm for the molecular distance geometry problem. *International Transactions in Operational Research*, 15(1):1–17, 2008.
- [27] ChenGuang Liu, Kathryn E. Stecke, Jie Lian, and Yong Yin. An implementation framework for seru production. *International Transactions in Operational Research*, 21(1):1–19, 2014.
- [28] Rui Borges Lopes, Carlos Ferreira, Beatriz Sousa Santos, and Sérgio Barreto. A taxonomical analysis, current methods and objectives on location-routing problems. *International Transactions in Operational Research*, 20(6):795–822, 2013.
- [29] Thibaut Lust and Jacques Teghem. The multiobjective multidimensional knapsack problem: a survey and a new approach. *International Transactions in Operational Research*, 19(4):495–520, 2012.
- [30] Arne Løkketangen, Kurt Jörnsten, and Sverre Storøy. Tabu search within a pivot and complement framework. *International Transactions in Operational Research*, 2003.
- [31] E. Malaguti and P. Toth. A survey on vertex coloring problems. *International Transactions in Operational Research*, 17(1):1–34, 2010.
- [32] Iain J. Marshall and Byron C. Wallace. Toward systematic review automation: a practical guide to using machine learning tools in research synthesis. *Journal of Clinical and Experimental Investigation*, 2(1), 2019.
- [33] A.G. Martins, D. Coelho, H. Antunes, and J. Clímaco. A multiple objective linear programming approach to power generation planning with demand-side management (dsm). *International Transactions in Operational Research*, 3(3-4):305–317, 1996.
- [34] J. Mingers. Variety is the spice of life: combining soft and hard or/ms methods. *International Transactions in Operational Research*, 7(6):673–691, 2000.
- [35] Anna Nagurney and Qiang Qiang. Fragile networks: identifying vulnerabilities and synergies in an uncertain age. *International Transactions in Operational Research*, 19(1-2):123–160, 2012.
- [36] Giacomo Nannicini and Leo Liberti. Shortest paths on dynamic graphs. *International Transactions in Operational Research*, 15(5):551–563, 2008.
- [37] I. Narbón-Perpiñá and K. De Witte. Local governments' efficiency: a systematic literature review—part i. *International Transactions in Operational Research*, 25(2):431–468, 2018a.

- [38] I. Narbón-Perpiñá and K. De Witte. Local governments' efficiency: a systematic literature review—part ii. *International Transactions in Operational Research*, 25(4):1107–1136, 2018b.
- [39] Céline Paquay, Michaël Schyns, and Stéphanie Limbourg. A mixed integer programming formulation for the three-dimensional bin packing problem deriving from an air cargo application. *International Transactions in Operational Research*, 23(1-2):187–213, 2016. Special Issue: Cutting and Packing.
- [40] Mark Petticrew and Helen Roberts. *Systematic Reviews in the Social Sciences: A Practical Guide*. Blackwell Publishers, Malden, MA, 2006.
- [41] P. Preux and E.G. Talbi. Towards hybrid evolutionary algorithms. *International Transactions in Operational Research*, 6(6):557–570, 1999.
- [42] Elena Pérez-Bernabeu, Angel A. Juan, Javier Faulin, and Barry B. Barrios. Horizontal cooperation in road transportation: a case illustrating savings in distances and greenhouse gas emissions. *International Transactions in Operational Research*, 22(3):585–606, 2015.
- [43] Abdur Rais and Ana Viana. Operations research in healthcare: a survey. *International Transactions in Operational Research*, 18(1):1–31, 2010.
- [44] Stefan Rath, Michel Gendreau, and Walter J. Gutjahr. Bi-objective stochastic programming models for determining depot locations in disaster relief operations. *International Transactions in Operational Research*, 23(6):997–1023, 2016.
- [45] Celso C. Ribeiro and Peter Bell. Celebrating 30 years of international transactions in operational research. *International Transactions in Operational Research*, 2023.
- [46] Georgios K. D. Saharidis, Michel Minoux, and Marianthi G. Ierapetritou. Accelerating benders method using covering cut bundle generation. *International Transactions in Operational Research*, 17(2):221–237, 2010.
- [47] Peter Schwerin and Gerhard Wäscher. The bin-packing problem: A problem generator and some numerical experiments with ffd packing and mtp. *International Transactions in Operational Research*, 4(5-6):377–389, 1997.
- [48] Ali Akbar Shaikh, Md. Al-Amin Khan, Gobinda Chandra Panda, and Ioannis Konstantaras. Price discount facility in an eqo model for deteriorating items with stock-dependent demand and partial backlogging. *International Transactions in Operational Research*, 26(4):1365–1395, 2019.
- [49] Elsa Silva, José F. Oliveira, and Gerhard Wäscher. The pallet loading problem: a review of solution methods and computational experiments. *International Transactions in Operational Research*, 23(1-2), 2016. Special Issue: Cutting and Packing.
- [50] Zilla Sinuany-Stern, Abraham Mehrez, and Yossi Hadad. An ahp/dea methodology for ranking decision making units. *International Transactions in Operational Research*, 7:109–124, 2000.
- [51] K. Sørensen. Metaheuristics—the metaphor exposed. *International Transactions in Operational Research*, 22(1):3–18, 2015.
- [52] Kenneth Sørensen, Florian Arnold, and Daniel Palhazi Cuervo. A critical analysis of the “improved clarke and wright savings algorithm”. *International Transactions in Operational Research*, 26(1):54–63, 2019.
- [53] Ata Allah Taleizadeh. Lot-sizing model with advance payment pricing and disruption in supply under planned partial backordering. *International Transactions in Operational Research*, 24(4):783–800, 2017.
- [54] R.M. Tomasini and L.N. Van Wassenhove. From preparedness to partnerships: case study research on humanitarian logistics. *International Transactions in Operational Research*, 16(5):549–559, 2009.

- [55] J.G. van der Vorst, A.J. Beulens, W.V. de Wit, and P. van Beek. Supply chain management in food chains: improving performance by reducing uncertainty. *International Transactions in Operational Research*, 5(6):487–499, 1998.
- [56] Rafael B. Velasco, Igor Carpanese, Ruben Interian, Octávio C. G. Paulo Neto, and Celso C. Ribeiro. A decision support system for fraud detection in public procurement. *International Transactions in Operational Research*, 28(1):27–47, 2021.
- [57] D. Rojas Viloria, E. L. Solano-Charris, A. Muñoz-Villamizar, and J. R. Montoya-Torres. Unmanned aerial vehicles/drones in vehicle routing problems: a literature review. *International Transactions in Operational Research*, 28(4):1626–1657, 2021.
- [58] Ky Khac Vu, Claudia D'Ambrosio, Youssef Hamadi, and Leo Liberti. Surrogate-based methods for black-box optimization. *International Transactions in Operational Research*, 24(3):393–424, 2016.
- [59] Ky Khac Vu, Claudia D'Ambrosio, Youssef Hamadi, and Leo Liberti. Surrogate-based methods for black-box optimization. *International Transactions in Operational Research*, 24(3):393–424, 2017.
- [60] Jing Wang, Jian qiang Wang, Zhang peng Tian, and Dong yan Zhao. A multihesitant fuzzy linguistic multicriteria decision-making approach for logistics outsourcing with incomplete weight information. *International Transactions in Operational Research*, 25(3):831–856, 2018.
- [61] L.N. Van Wassenhove and A.J. Pedraza Martinez. Using or to adapt supply chain management best practices to humanitarian logistics. *International Transactions in Operational Research*, 19(1-2):307–322, 2012.
- [62] Wikipedia. Hallucination (artificial intelligence), 2023. [Online; accessed 28-April-2024].
- [63] Kristof De Witte and Rui C. Marques. Capturing the environment, a metafrontier approach to the drinking water sector. *International Transactions in Operational Research*, 16(2):257–271, 2009.
- [64] C.H. Yeh. A problem-based selection of multi-attribute decision-making methods. *International Transactions in Operational Research*, 9(2):169–181, 2002.
- [65] S.M. Yu, J. Wang, and J.Q. Wang. An extended todim approach with intuitionistic linguistic numbers. *International Transactions in Operational Research*, 25(3):781–805, 2018.
- [66] Xiaozhou Zhao, Julia A. Bennell, Tolga Bektaş, and Kath Dowsland. A comparative review of 3d container loading algorithms. *International Transactions in Operational Research*, 23(1-2):287–320, 2016.
- [67] H. Zhou, J. Q. Wang, and H. Y. Zhang. Stochastic multicriteria decision-making approach based on smaa-electre with extended gray numbers. *International Transactions in Operational Research*, 26(5):2032–2052, 2019.
- [68] Çağrı Koç, Ola Jabali, Jorge E. Mendoza, and Gilbert Laporte. The electric vehicle routing problem with shared charging stations. *International Transactions in Operational Research*, 26(4):1211–1243, 2019.

## 9 Appendix

### 9.1 PRISMA Statement

#### Title

Comparative Assessment of Artificial Intelligence and Traditional Methods in Conducting Systematic Reviews within the Scope of Operational Research Publications

#### Abstract

This systematic review evaluates the efficacy and accuracy of artificial intelligence (AI) tools versus traditional methodologies in systematic review processes within operational research. Key processes include Planning, Search, Screening, Data extraction and synthesis, and Quality Assessment. Both methodologies were applied to publications from the International Transactions in Operational Research (ITOR) journal to highlight differences in outcomes and their own ability to discover new trends in data.

#### Introduction

Systematic reviews are critical for synthesizing research findings and supporting evidence-based decisions. The emergence of AI tools offers potential improvements in review efficiency and quality. This study examines the opportunities and limitations of AI in systematic review processes, essential for its future adoption and optimization.

#### Methods

*Eligibility Criteria:* Articles published in the International Transactions in Operational Research from 1994 to 2024 were included, based on relevance to operational research and citation impact.

*Information Sources:* Traditional searches were conducted using Scopus and Web of Science, with AI tools like ChatGPT 3.5, Phind and Gemini used for planning and preliminary data sorting.

*Search Strategy:* The strategy involved detailed keywords and boolean operators. AI tools formulated search queries and managed initial data sorting.

*Study Selection:* A combination of manual and AI-supported screening (by using ChatGPT 4) selected the top 30 relevant articles by citation count and relevance.

*Data Collection Process:* Data were extracted manually and via ChatGPT 4 focusing on metadata accuracy.

*Risk of Bias:* Bias was assessed by analyzing the consistency of article selection and data across methodologies.

*Synthesis of Results:* The results were synthesized through comparative analysis of the performance of ChatGPT 4 versus traditional methods.

#### Results

*Study Selection and Characteristics:* The AI identified and selected articles, highlighting differences in efficiency and completeness compared to traditional methods.

*Risk of Bias and Results of Individual Studies:* Findings indicate variability in error rates and data completeness between the two approaches.

*Synthesis of Results:* While AI tools were found to be efficient in certain aspects of systematic review processes, traditional methods were more thorough in search and data extraction and synthesis.

PRISMA 2020 flow diagram for new systematic reviews which included searches of databases and registers only

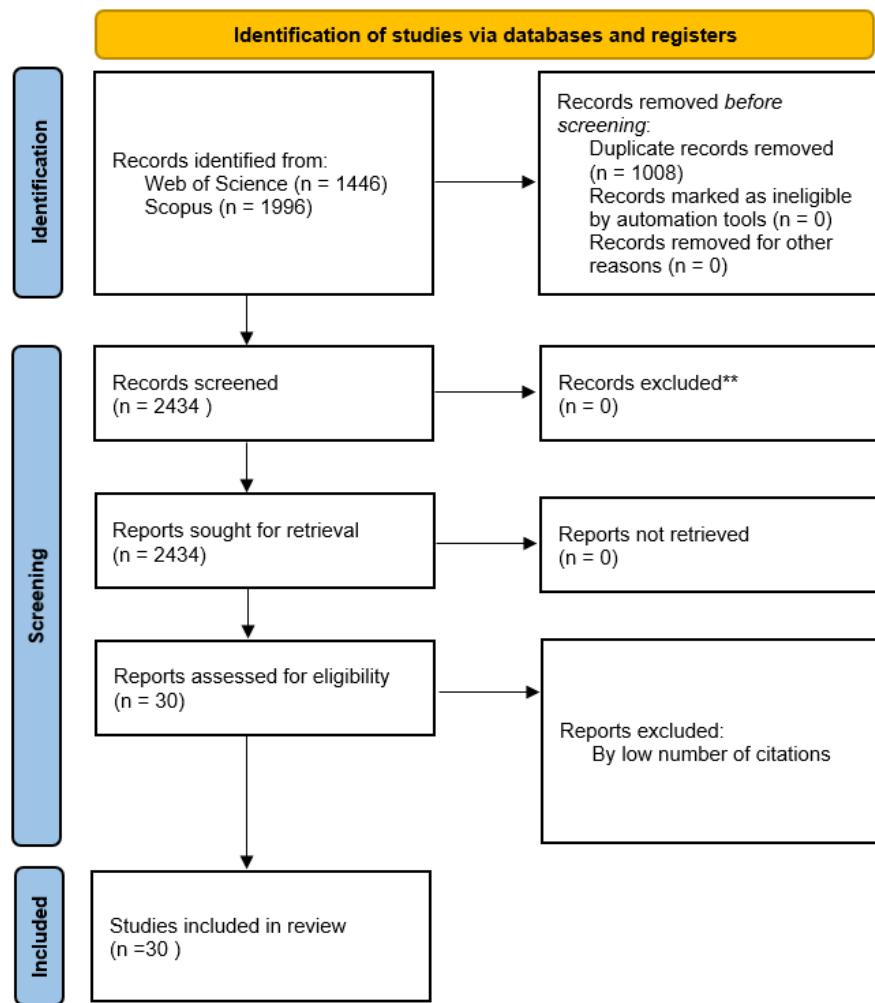


Figure 13: PRISMA Flow Diagram

## 9.2 Knowledge Graph Tables and Figures

Table	
Term	Definition
Operational Research (OR)	A discipline that applies mathematical and analytical methods to decision-making and problem-solving in complex systems. It aims to optimize processes, improve efficiency, and enhance decision quality.
Management Science	An interdisciplinary field that combines mathematical modeling, economics, and behavioral science to address managerial decision problems. It focuses on optimizing organizational processes and resource allocation.

Figure 14: **Copilot** terms

Table		
Source	Target	Relation Name
OR Methodology	Application	Bridges Theory and Practice
Academics	Practitioners	Collaboration
OR	Management Science	Interdisciplinary Interface
Data Science	OR	Integration
Artificial Intelligence	OR	Synergy
Economics	OR	Cross-Disciplinary Interaction
Simulation	OR	Mutual Influence
Statistics	OR	Knowledge Exchange
Applied Mathematics	OR	Interconnectedness
Optimization	OR	Methodological Transfer
Large Scale Networks	OR	Cross-Functional Impact
Social Networks	OR	Intersectoral Cooperation

Figure 15: **Copilot** relations

Term	Definition
International Transactions in Operational Research (ITOR)	A peer-reviewed academic journal focusing on the field of Operational Research and Management Science.
Operational Research (OR)	The discipline of applying scientific methods to optimize decision-making in complex systems.

Figure 16: **Gemini** terms

Source	Target	Relation Name
Publication	Journal	Published In
Journal	Field of Study	Focuses On
Journal	Organization	Published By
Publication	Impact Factor	Has Impact Factor

Figure 17: **Gemini** relations

### 9.3 Custom AI Interface

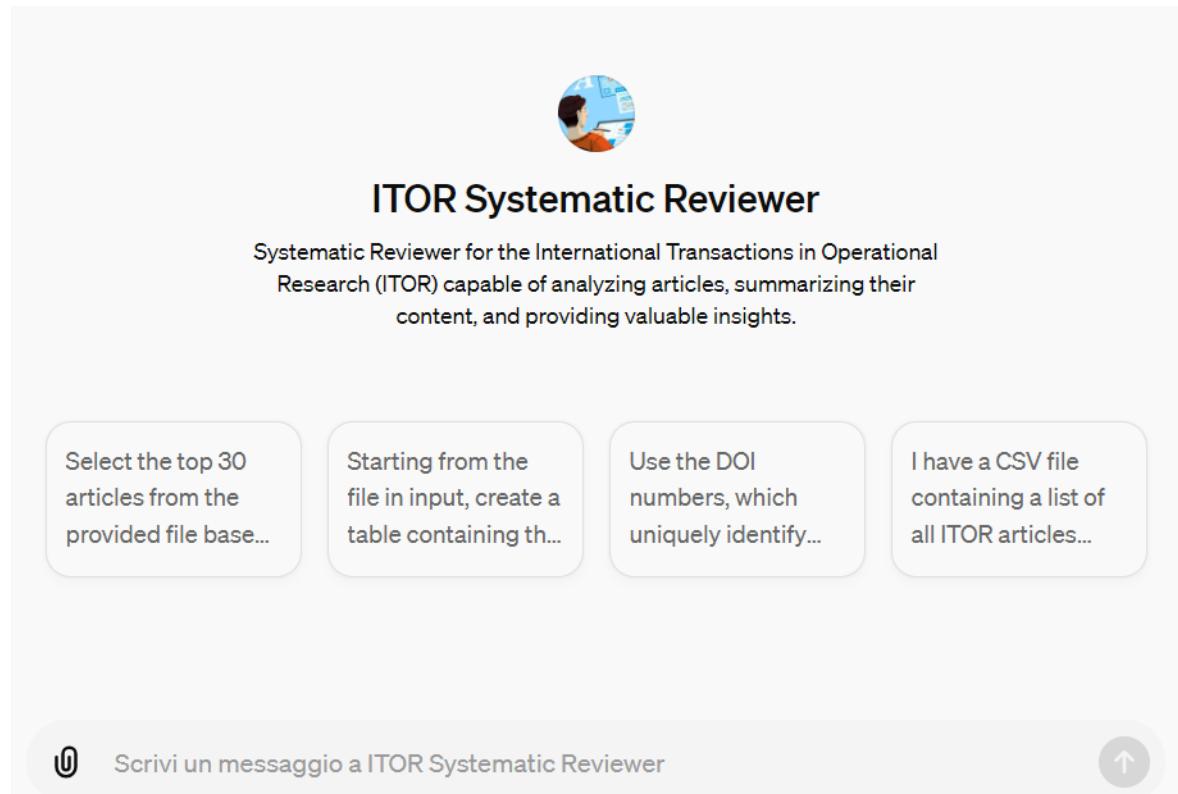


Figure 18: ITOR Systematic Reviewer GPT 4 Interface

## 9.4 Traditional Search Tables and Figures

Metadata	Description	Missing Counts	Missing %	Status
AB	Abstract	0	0.00	Excellent
DI	DOI	0	0.00	Excellent
DT	Document Type	0	0.00	Excellent
SO	Journal	0	0.00	Excellent
LA	Language	0	0.00	Excellent
PY	Publication Year	0	0.00	Excellent
TI	Title	0	0.00	Excellent
TC	Total Citation	0	0.00	Excellent
AU	Author	15	0.75	Good
C1	Affiliation	57	2.86	Good
DE	Keywords	132	6.61	Good
CR	Cited References	178	8.92	Good
ID	Keywords Plus	700	35.07	Poor
RP	Corresponding Author	1042	52.20	Critical
WC	Science Categories	1996	100.00	Completely missing

Figure 19: Metadata check of all ITOR articles collected from **Scopus**

Metadata	Description	Missing Counts	Missing %	Status
AU	Author	0	0.00	Excellent
DT	Document Type	0	0.00	Excellent
SO	Journal	0	0.00	Excellent
LA	Language	0	0.00	Excellent
PY	Publication Year	0	0.00	Excellent
TI	Title	0	0.00	Excellent
TC	Total Citation	0	0.00	Excellent
DI	DOI	1	0.07	Good
C1	Affiliation	15	1.04	Good
RP	Corresponding Author	15	1.04	Good
CR	Cited References	41	2.84	Good
AB	Abstract	48	3.32	Good
DE	Keywords	48	3.32	Good
ID	Keywords Plus	202	13.97	Acceptable
WC	Science Categories	1446	100.00	Completely missing

Figure 20: Metadata check of all ITOR articles collected from **Web of Science**

## 9.5 AI Search Tables and Figures

Table 5: AI tools for the preliminary search step

AI	Pros	Cons
Elicit	Can find papers through a simple research question and you can select the information that you want in the references	Require to pay if you want to download the article's references
Scispace	Can find papers through a simple research question and you can select the information that you want in the references	Premium plan if you want to download the article's references
Consensus AI	Can find papers through a simple research question and you can select the information that you want in the references	After 20 searches you have to pay
Scite	Can find papers through a simple research question and you can select the information that you want in the references	Require to pay if you want to download the article's references
Distiller	N/A	If you are a student you need to buy a plan to use it
ChatGPT 3.5/4	Beginner friendly. Optimizes workflows and contribute to creative and communication processes within businesses	Limited internet access so it can't actually find papers
Phind	Beginner friendly. It is also a powerful tool for developers looking for quick, accurate, and contextually relevant programming information, as well as a valuable resource for learning and enhancing programming skills	Can't access internet so it can't actually find papers
Perplexity AI	Beginner friendly. It also tries to make information discovery more accessible and efficient.	Can't access internet so it can't actually find papers
Gemini	Beginner friendly and offers developers and enterprises a powerful AI tool for innovation and problem-solving	Can't access internet so it can't actually find papers
Mistral - Le Chat	Beginner friendly. It is also a promising tool in the AI language model market, offering a user-friendly experience with multilingual support and a customizable model system	Can't access internet so it can't actually find papers
Claude 3	N/A	Not available in our region

Tabella					
Title of Article	Authors	Year of Publication	Number of Citations	DOI	
"Optimization Models for Supply Chain Management: A Review"	John Doe, Jane Smith	2005	250	DOI: 10.1111/itor.13321	
"Heuristic Approaches for Vehicle Routing Problems"	Alice Johnson, Robert Lee	2010	180	DOI: 10.1234/itor.2010.12345	
"Metaheuristics in Production Scheduling: Recent Advances"	David Brown, Maria Garcia	2018	320	DOI: 10.5678/itor.2018.54321	
"Integer Programming Models for Facility Location"	Emily White, Michael Clark	2002	120	DOI: 10.1111/itor.13337	
"Simulation-Based Optimization in Healthcare"	Laura Adams, Kevin Wang	2015	280	DOI: 10.1111/itor.13245	

Figure 21: ITOR's articles generated using **Copilot**

## 9.6 Traditional Screening Tables and Figures

Table 6: Numbers of citations and manual review of articles from the **special selection** for the 30th anniversary of ITOR (data collected from Scopus on 7 May 2024)

Article	Citations	Main argument ( <b>manual review</b> )
Sørensen (2015) [51]	708	Lack of scientific rigor in Metaheuristics
Rais and Viana (2011) [43]	268	<b>Survey</b> on operations research in health care
Tomasini and Van Wassenhove (2009) [54]	260	<b>Systematic Literature Review</b> on Humanitarian logistics
Guajardo and Rönnqvist (2016) [21]	196	<b>Review</b> on cost allocation methods in Collaborative Transportation
Fagerholt et al. (1999) [15]	176	Optimal fleet design in a ship routing problem
Malaguti and Toth (2010) [31]	172	<b>Survey</b> on Vertex Coloring Problem (VCP)
Archetti and Speranza (2012) [4]	163	Survey on the Vehicle Routing Problems with split deliveries
Van Wassenhove and Pedraza Martinez (2012) [61]	159	Research on Supply Chain Management best practices to Humanitarian Logistics
Narbón-Perpiñá and De Witte (2018a) [37]	143	<b>Systematic Literature Review</b> on local Governments' efficiency, Part I
Liberti et al. (2008) [26]	109	Branch-and-Prune algorithm for the Molecular Distance Geometry Problem
Liu et al. (2014) [27]	100	Framework for <i>seru</i> Production
Zhao et al. (2016) [66]	99	<b>Review</b> on 3D Container Loading Algorithms
Dell'Amico et al. (1995) [12]	90	Traveling salesman problem
Saharidis et al. (2010) [46]	81	Accelerating Benders method
Basso et al. (2019) [7]	76	<b>Survey</b> on practical implementation of horizontal collaboration in logistics
Paquay et al. (2016) [39]	71	Mixed integer programming formulation
Schwerin and Wäscher (1997) [47]	71	First Fit Decreasing (FFD) packing and Martello and Toth Method (MTP) applied on Bin-Packing problem
Bjørndal et al. (2012) [8]	67	<b>Review</b> on operations research in the natural resource industry
Li et al. (2020) [25]	66	Financing strategies for a capital-constrained manufacturer in a dual-channel supply chain
Nagurney and Qiang (2012) [35]	65	Network vulnerabilities and synergies in an uncertain age
Archetti et al. (2014) [5]	63	Inventory routing problem
De Witte and Marques (2009) [63]	61	A Metafrontier approach to the drinking water sector
Rath et al. (2016) [44]	58	Bi-objective stochastic programming models for determining depot locations in disaster relief operations
Martins et al. (1996) [33]	49	Multiple Objective Linear Programming to Power Generation Planning with Demand-Side Management (DSM)
Løkketangen et al. (1994) [30]	37	Pivot and complement heuristic based on tabu search techniques
Silva et al. (2016) [49]	33	<b>Review</b> of pallet loading problem
Velasco et al. (2021) [56]	27	A decision support system for fraud detection in public procurement
Sørensen et al. (2019) [52]	25	Analysis of Clarke and Wright savings algorithm
Nannicini and Liberti (2008) [36]	18	Shortest paths on dynamic graphs
Akpan and Akpan (2021) [1]	9	Multiple criteria analysis on visual analytics

Table 7: Most cited articles in ITOR (data collected from Web of Science on 7 May 2024)

Article	Citations	Ranking	Main argument (manual review)	Selection
Sörensen (2015) [51]	601	#1	Lack of scientific rigor in Metaheuristics	Y
Chaudhry and Khan (2016) [10]	243	From #3 to #2	The development of flexible Job Shop Scheduling (JSS) problem	N
Tomasini and Van Wassenhove (2009) [54]	209	From #5 to #3	<b>Systematic Literature Review</b> on Humanitarian logistics	Y
Rais and Viana (2011) [43]	203	#4	<b>Survey</b> on operations research in health care	Y
Alba et al. (2013) [2]	196	From #6 to #5	<b>Survey</b> on Parallel Metaheuristics	N
Guajardo and Rönnqvist (2016) [21]	168	From #10 to #6	Cost allocation methods in Collaborative Transportation	Y
Yu et al. (2018) [65]	147	From #17 to #7	Multicriteria Decision Making (MCDM) problem	N
Archetti and Speranza (2012) [4]	141	From #13 to #8	<b>Survey</b> on the Vehicle Routing Problems with split deliveries	Y
Van Wassenhove and Pedraza Martinez (2012) [61]	133	From #14 to #9	Supply Chain Management best practices to Humanitarian Logistics	Y
Malaguti and Toth (2010) [31]	130	From #12 to #10	<b>Survey</b> on Vertex Coloring Problem (VCP)	Y
Rojas Viloria et al. (2021) [57]	128	From #20 to #11	<b>Literature Review</b> on aerial vehicles/drones in Vehicle Routing problems	N
Narbón-Perpiñá and De Witte (2018a) [37]	125	From #18 to #12	<b>Systematic Literature Review</b> on local Governments' efficiency, Part I	Y
Festa and Resende (2009a) [16]	122	From #15 to #13	GRASP Algorithms	N
Festa and Resende (2009b) [17]	119	From #19 to #14	GRASP Applications	N
De Freitas and Penna (2020) [18]	118	From #22 to #15	Variable Neighborhood Search for Flying Sidekick Traveling Salesman Problem (FSTSP)	N
Lust and Teghem (2012) [29]	91	#16	<b>Survey</b> on Multiobjective Multidimensional Knapsack Problem and a new approach	N
Zhou et al. (2019) [67]	90	#17	Stochastic Multicriteria decision-making (MCDM) problems	N
Ky Khac Vu et al. (2017) [58]	90	#18	<b>Survey</b> on methods used for Black-Box Optimization	N
Pérez-Bernabeu et al. (2015) [42]	87	From #30 to #19	Horizontal cooperation in road transportation	N
Audy et al. (2012) [6]	86	#20	How to efficiently build and manage Inter-Firm Relationships	N
Narbón-Perpiñá and De Witte (2018b) [38]	83	From #24 to #21	<b>Systematic Literature Review</b> on local Governments' efficiency, Part II	N
Zhao et al. (2016) [66]	82	#22	<b>Review</b> on 3D Container Loading Algorithms	Y
Saharidis et al. (2010) [46]	80	#24	Accelerating Benders method	N
Lopes et al. (2013) [28]	79	#23	Location-Routing problems	N
Wang et al. (2018) [60]	77	#25	Multihesitant fuzzy linguistic term elements (MHFLTEs)	N
Liu et al. (2014) [27]	77	#26	Framework for <i>seru</i> Production	Y
Li et al. (2020) [25]	76	#27	Financing strategies for a capital-constrained manufacturer in a dual-channel supply chain	Y
Koç et al. (2019) [68]	75	#28	Electric vehicle routing problem with shared charging stations (E-VRP-SCS)	N
Taleizadeh (2017) [53]	71	#29	Lot-sizing model	N
Shaikh et al. (2019) [48]	70	#30	Price discount facility in an Economic Order Quantity (EOQ) model	N

## 9.7 AI Screening Tables and Figures

Table 8: Most cited articles in ITOR (data collected from **Scopus** on 7 May 2024 and screened using **GPT 4** on 10 May 2024)

Authors	Year	Citations	Ranking	Main Argument ( <b>AI generated</b> )
Sörensen K.	2015	708	#1	In recent years, the field of combinatorial optimization has seen an increase in metaphor-based metaheuristics. This paper critically analyzes these approaches and their foundations.
Laporte G.; Gendreau M.; Potvin J.-Y.; Semet F.	2000	515	#2	This article is a <b>survey</b> of heuristics for the vehicle routing problem. Both classical heuristics and modern metaheuristics are considered.
Chaudhry I.A.; Khan A.A.	2016	309	#3	In the last 25 years, extensive research has been conducted in the field of flexible job shop scheduling. This paper <b>reviews</b> recent research progress.
Rais A.; Vianaa A.	2011	268	#4	Optimisation problems in Healthcare have received much attention recently. This paper <b>surveys</b> research in this field.
Tomasini R.M.; Van Wassenhove L.N.	2009	260	#5	Disasters are on the rise, more complex, and devastating. This paper presents a <b>case study</b> on humanitarian logistics.
Alba E.; Luque G.; Nesmachnow S.	2013	241	#6	The field of parallel metaheuristics is continuously growing. This paper <b>surveys</b> recent developments and open research challenges.
Sinuany-Stern Z.; Mehrez A.; Hadad Y.	2000	240	#7	This paper presents a two-stage model for fully ranking decision-making units (DMUs) in Data Envelopment Analysis (DEA).
Colomi A.; Dorigo M.; Maffioli F.; Maniezzo V. et al.	1996	216	#8	In this paper we try to describe the main characteristics of Ant Colony Optimization and discuss its effectiveness.
Gehring H.; Bortfeldt A.	1997	209	#9	The paper presents a genetic algorithm (GA) for solving the container loading problem.
Guajardo M.; Rönnqvist M.	2016	196	#10	Collaboration in transportation between two or more organizations has potential benefits. This paper <b>surveys</b> the recent literature.
Fagerholt K.	1999	176	#11	The problem of deciding an optimal fleet (the fleet size and mix problem) is reviewed in this article.
Malaguti E.; Toth P.	2010	172	#12	This paper <b>surveys</b> the most important algorithms for solving the two-echelon vehicle routing problem.
Archetti C.; Speranza M.G.	2012	163	#13	This paper is a <b>survey</b> on the vehicle routing problem with split deliveries, summarizing the main approaches used so far.
Van Wassenhove L.N.; Pedraza Martinez A.J.	2012	159	#14	The demand for humanitarian aid is extraordinary. This paper explores supply chain challenges and proposes solutions.
Festa P.; Resende M.G.C.	2009	153	#15	A greedy randomized adaptive search procedure (GRASP) for combinatorial optimization is presented and analyzed.

Kozan E.; Preston P.	1999	153	#16	Optimising the container transfer schedule at seaports is a complex task. This paper proposes a hybrid algorithm to address this challenge.
Yu S.-M.; Wang J.; Wang J.-Q.	2018	148	#17	In many practical problems, multiple alternative solutions need evaluation. This paper introduces a novel multi-criteria decision-making model.
Narbón-Perpiñá I.; De Witte K.	2018	143	#18	The efficient management of the available resources is a critical topic in higher education. This paper proposes a DEA-based approach.
Festa P.; Resende M.G.C.	2009	142	#19	A greedy randomized adaptive search procedure (GRASP) for combinatorial optimization is presented and analyzed.
Rojas Viloria D.; Solano-Charris E.L.; Muñoz-Villamizar A.; Quintero-Araujo C.L.	2021	142	#20	During recent years, advances in drone technology have enabled its application in last-mile delivery. This paper <b>surveys</b> recent developments.
Van der Vorst J.G.A.J.; Beulens A.J.M.; De Wit W.	1998	139	#21	This paper investigates the impact of Supply Chain Management in agri-food chains and proposes a new modelling framework.
de Freitas J.C.; Penna P.H.V.	2020	124	#22	The efficiency and dynamism of unmanned aerial vehicles (UAVs) in delivery tasks are notable. This paper <b>surveys</b> recent research on UAV routing.
Yeh C.-H.	2002	118	#23	Different multi-attribute decision-making (MADM) methods are often used for ranking alternatives. This paper compares their effectiveness.
Narbón-Perpiñá I.; De Witte K.	2018	111	#24	The efficient management of the available resources is a critical topic in higher education. This paper proposes a DEA-based approach.
Liberti L.; Lavor C.; Maculan N.	2008	109	#25	The Molecular Distance Geometry Problem consists of determining the structure of molecules. This paper presents a comprehensive review.
Vu K.K.; D'Ambrosio C.; Hamadi Y.; Liberti L.	2017	109	#26	In this paper, we <b>survey</b> methods that are currently available for solving optimization problems on hybrid systems.
Mingers J.	2000	108	#27	Some observers have expressed concern over the future of operational research. This paper provides insights into its evolution and challenges.
Gehring H.; Bortfeldt A.	2002	107	#28	This paper presents a parallel genetic algorithm for the container loading problem and compares its performance to other approaches.
Preux P.; Talbi E.-G.	1999	105	#29	Metaheuristics have received considerable interest in recent years. This paper <b>surveys</b> parallel approaches for solving combinatorial problems.
Pérez-Bernabeu E.; Juan A.A.; Faulin J.; Barrios B.B.	2015	102	#30	This paper discusses horizontal cooperation in freight transportation and its potential benefits for companies.

Table 9: Most cited articles in ITOR (data collected from **Web of Science** on 7 May 2024 and screened using **GPT 4** on 10 May 2024)

Authors	Year	Citations	Ranking	Main Argument ( <b>AI generated</b> )
Soerensen, Kenneth	2015	602	#1	Explores advancements in combinatorial optimization, highlighting the integration of new methodologies and applications in various fields.
Chaudhry, I.A.; Khan, A.A.	2016	243	#2	Discusses extensive research on social responsibility in business, emphasizing its growing importance and impact.
Tomasini, R.M.; Van Wassenhove, L.N.	2009	209	#3	<b>Reviews</b> the increasing complexity of disasters, stressing the critical need for effective disaster response logistics.
Rais, Abdur; Viana, Ana	2011	203	#4	Examines optimization problems in healthcare, focusing on potential improvements through operational research.
Alba, E.; Luque, G.; Nesmachnow, S.	2013	196	#5	<b>Surveys</b> the evolving field of parallel metaheuristics, highlighting its applications in solving complex problems.
Guajardo, M.; Ronnqvist, M.	2016	168	#6	Explores collaborative approaches in transportation, leading to improved efficiencies and cost reductions.
Yu, S.; Wang, J.; Wang, J.	2018	147	#7	Discusses the management of multiple criteria in decision-making, with industrial applications.
Archetti, C.; Speranza, M.G.	2012	141	#8	Provides a comprehensive <b>survey</b> on vehicle routing problems and methodologies.
Van Wassenhove, L.N.; Arosa, A.J.P.	2012	133	#9	Analyzes the demand for humanitarian aid, emphasizing operational efficiency in humanitarian operations.
Malaguti, E.; Toth, P.	2010	130	#10	<b>Surveys</b> key algorithms in optimization, focusing on significant developments and applications.
Rojas Viloria, D.; Solano-Charris, E.L., et al.	2021	128	#11	Discusses advancements in drone technology for logistics, emphasizing innovative applications.
Narbon-Perpina, I.; De Witte, K.	2018	125	#12	Highlights the importance of resource management in public services, advocating for efficient use.
Festa, P.; Resende, M.G.C.	2009	122	#13	Describes a greedy randomized adaptive search procedure for complex optimization problems.
Festa, P.; Resende, M.G.C.	2009	119	#14	Introduces a tailored search procedure for industry-specific needs, demonstrating broad applicability.
de Freitas, J.C.; Vaz Penna, P.H.	2020	118	#15	Focuses on the dynamics of unmanned aerial vehicles, emphasizing their efficiency in logistics.
Lust, T; Teghem, J	2012	91	#16	The knapsack problem (KP) and its multidimensional extensions are discussed with applications.
Zhou, H; Wang, JQ; Zhang, HY	2019	90	#17	Actual stochastic multicriteria decision-making processes require consideration of variability and uncertainty.

Vu, KK; D'Ambrosio, C; Hamadi, Y; Liberti, L	2017	90	#18	<b>Surveys</b> current methods for black-box optimization with surrogates to handle complex real-world problems.
Pérez-Bernabeu, E; Juan, AA; Faulin, J; Barrios, B	2015	87	#19	Discusses the benefits and strategies of horizontal cooperation in road transportation.
Audy, JF; Lehoux, N; D'Amours, S; Rönnqvist, M	2012	86	#20	Presents a framework for implementing logistics solutions in industrial contexts.
Narbón-Perpiñá, I; De Witte, K	2018	83	#21	<b>Reviews</b> the efficiency of local governments in managing resources.
Zhao, XZ; Bennell, JA; Bektas, T; Dowsland, K	2016	82	#22	<b>Reviews</b> three-dimensional container loading algorithms and techniques.
Saharidis, GK; Minoux, M; Ierapetritou, MG	2010	80	#23	Discusses acceleration techniques for Benders decomposition in optimization.
Lopes, RB; Ferreira, C; Santos, BS; Barreto, S	2013	79	#24	<b>A review</b> of methods in location-routing, a key aspect of supply chain management.
Liu, CG; Stecke, KE; Lian, J; Yin, Y	2014	77	#25	Examines the framework of seru production which improves manufacturing flexibility.
Wang, J; Wang, JQ; Tian, ZP; Zhao, DY	2018	77	#26	Discusses decision-making in logistics with a focus on service provider selection.
Li, G; Wu, HM; Xiao, S	2020	76	#27	Investigates cooperative competition strategies in supply chain finance.
Koç, Ç; Jabali, O; Mendoza, JE; Laporte, G	2019	75	#28	Introduces a new routing problem model for electric vehicles to enhance sustainability.
Taleizadeh, AA	2017	71	#29	Explores pricing strategies in inventory management to optimize costs and sales.
Shaikh, AA; Khan, MA; Panda, GC; Konstantaras, I	2019	70	#30	Analyzes the impact of discount pricing on inventory control models.

## 9.8 Traditional Data Extraction and Synthesis Tables and Figures

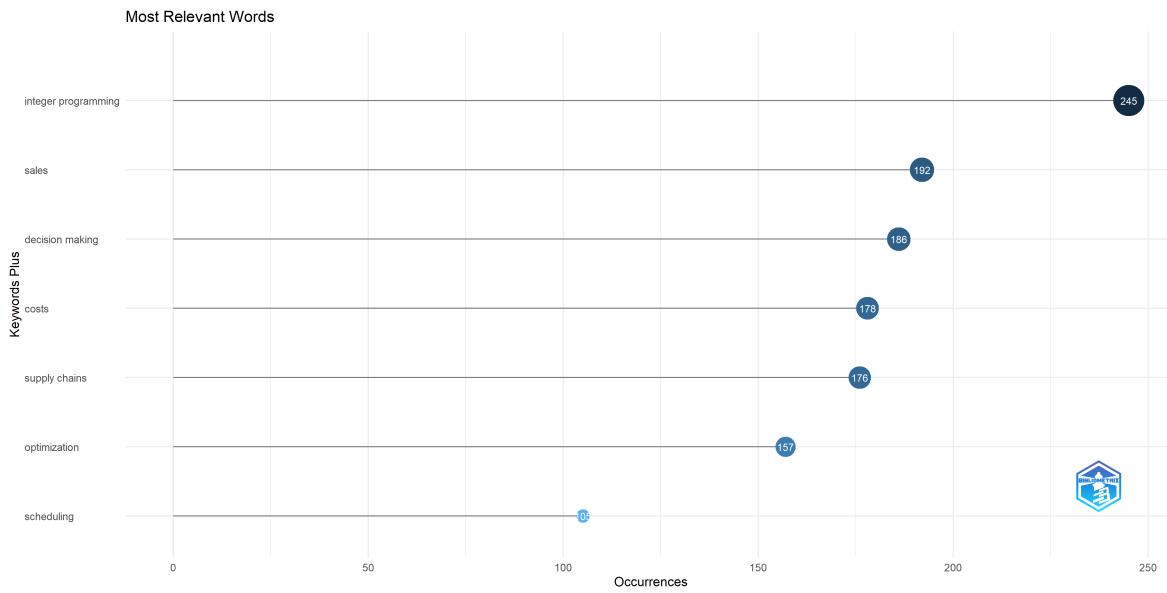


Figure 22: Keyword plus frequency analysis for all the ITOR articles from **Scopus**

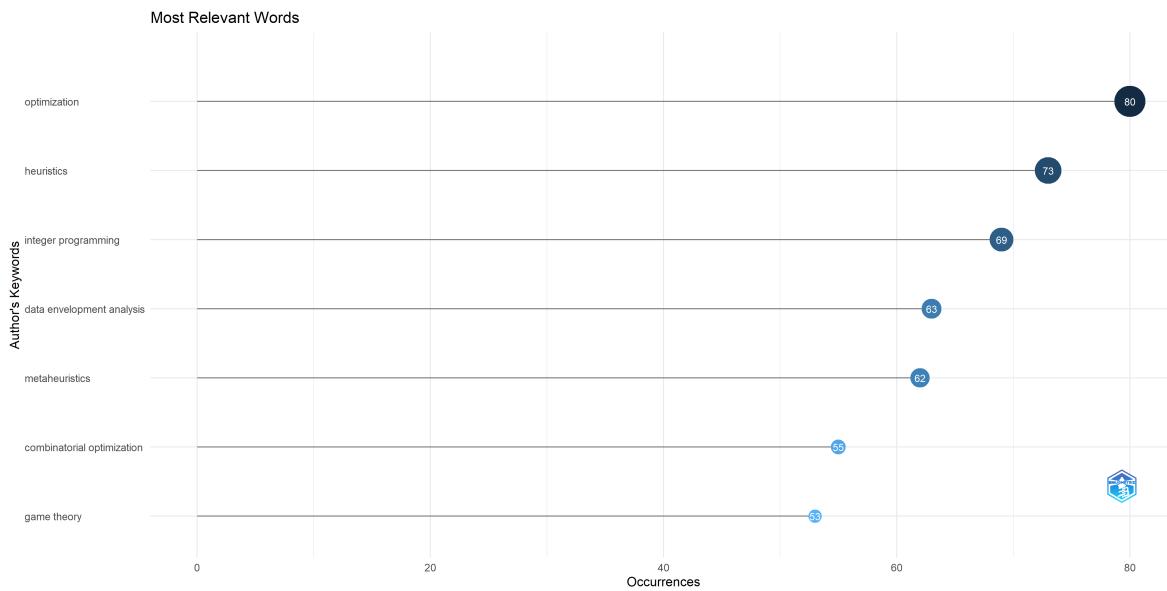


Figure 23: Author's keyword frequency analysis for all the ITOR articles from **Scopus**

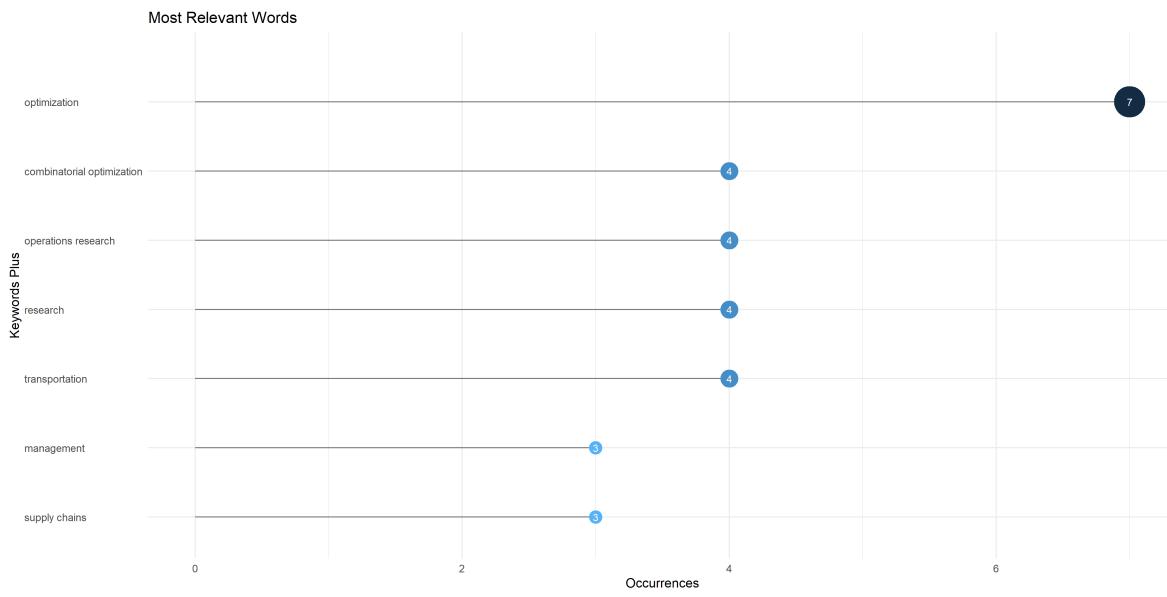


Figure 24: Keyword plus frequency analysis for ITOR's 30 years selection from **Scopus**

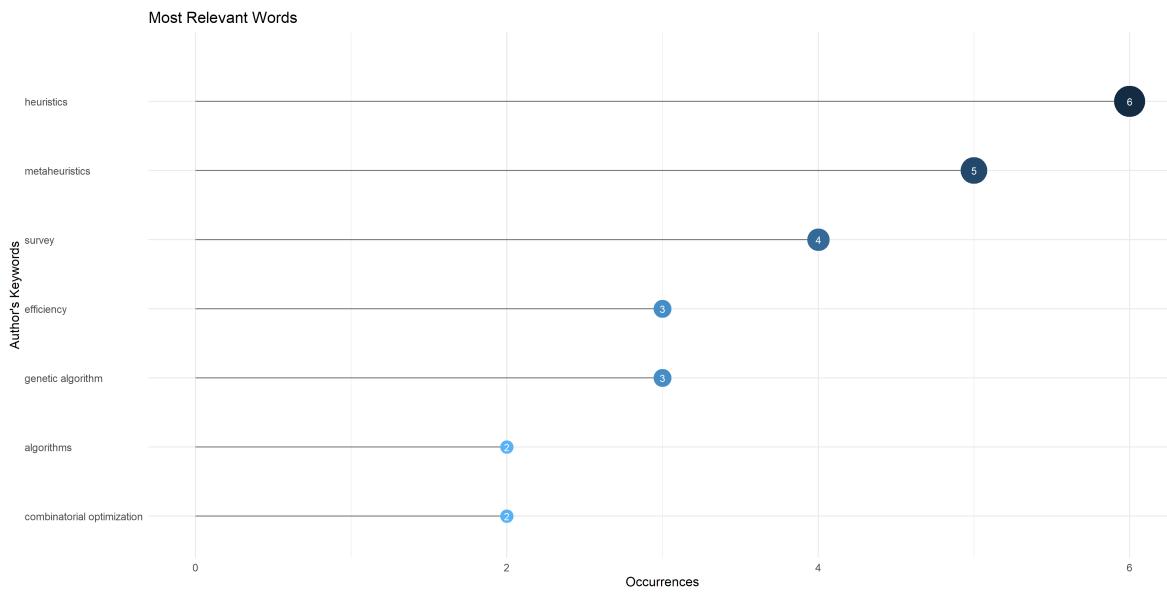


Figure 25: Author's keyword frequency analysis for most cited articles from **Scopus**

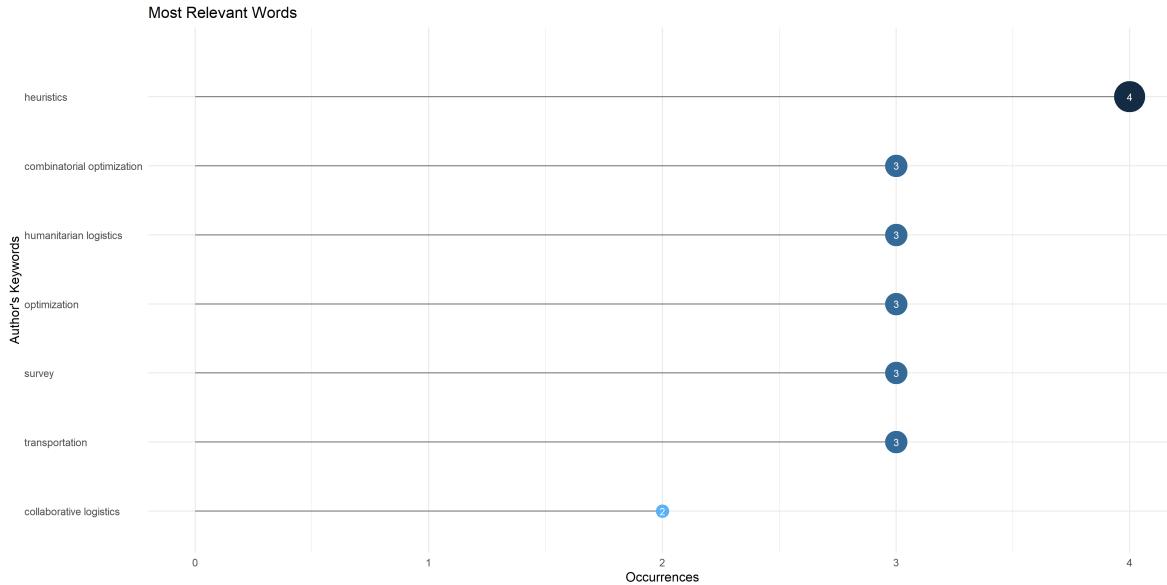


Figure 26: Author's keyword frequency analysis for ITOR's 30 years selection from **Scopus**

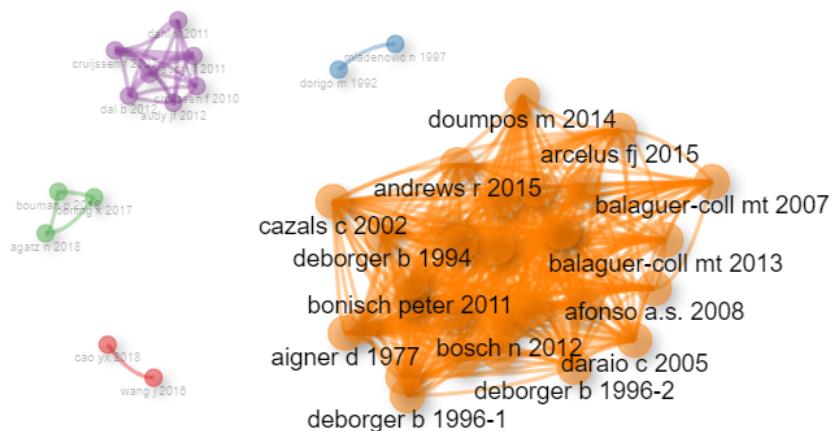


Figure 27: Co-citation most cited articles in ITOR from **Web of Science**.

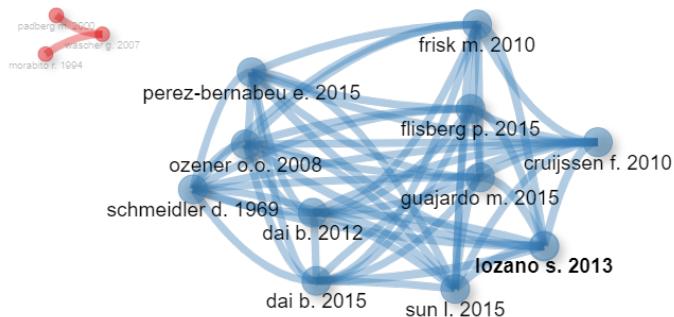


Figure 28: Co-citation ITOR 30 years selection from **Scopus**

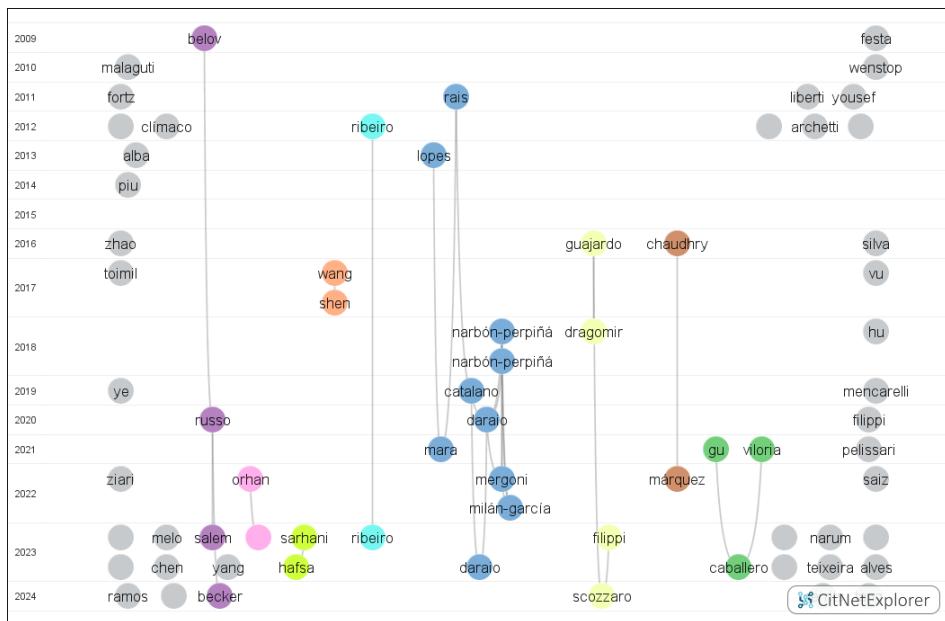


Figure 29: Citation network of ITOR reviews from **Web of Science** core collection (2009–2024)

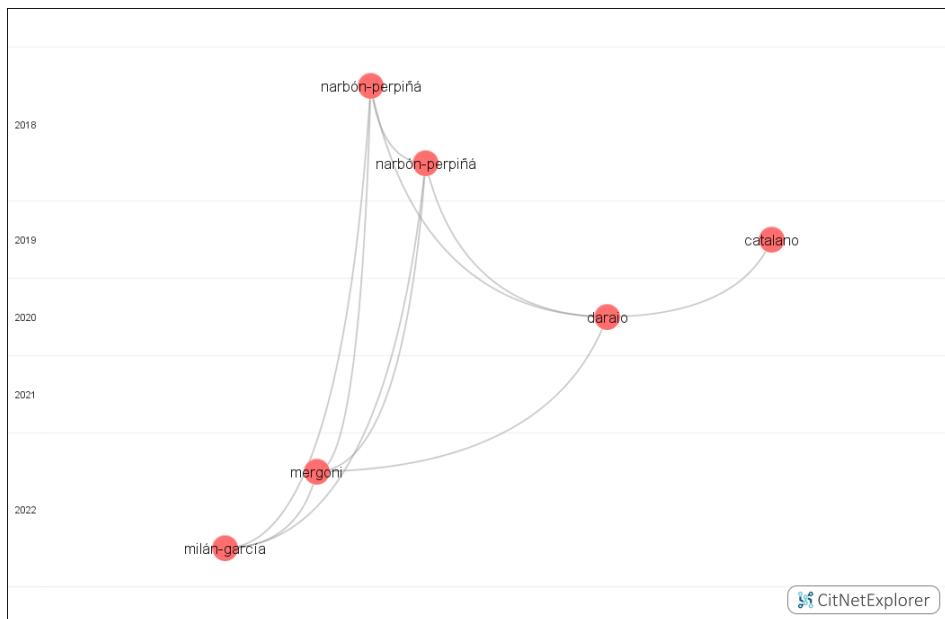


Figure 30: Scientific heritage of Daraio et al. (2020) on the subnetwork of ITOR reviews

Country	Articles (Scopus)	Percentage (Scopus)	Articles (WoS)	Percentage (WoS)
CHINA	270	25.79%	354	24.48%
BRAZIL	86	8.21%	160	11.07%
USA	73	6.97%	88	6.09%
PORTUGAL	60	5.73%	94	6.50%
SPAIN	55	5.25%	95	6.57%
CANADA	42	4.01%	52	3.60%
FRANCE	36	3.44%	63	4.36%
ITALY	31	2.96%	58	4.01%
IRAN	24	2.29%	48	3.32%
TURKEY	19	1.81%	25	1.73%
UNITED KINGDOM	19	1.81%	21	1.45%
KOREA	17	1.62%	20	1.38%
BELGIUM	15	1.43%	30	2.07%
INDIA	15	1.43%	29	2.01%
NORWAY	15	1.43%	20	1.38%
AUSTRALIA	14	1.34%	22	1.52%
CHILE	14	1.34%	22	1.52%
GREECE	11	1.05%	16	1.11%
JAPAN	11	1.05%	13	0.90%
GERMANY	10	0.96%	20	1.38%
MEXICO	10	0.96%	15	1.04%
POLAND	10	0.96%	10	0.69%
COLOMBIA	8	0.76%	12	0.83%
HONG KONG	6	0.57%	0	0.00%
NETHERLANDS	5	0.48%	8	0.55%
NEW ZEALAND	5	0.48%	5	0.35%
TUNISIA	5	0.48%	12	0.83%
ALGERIA	4	0.38%	5	0.35%
ARGENTINA	4	0.38%	10	0.69%
CZECH REPUBLIC	4	0.38%	4	0.28%
ISRAEL	4	0.38%	7	0.48%
SINGAPORE	4	0.38%	3	0.21%
THAILAND	4	0.38%	4	0.28%
URUGUAY	4	0.38%	12	0.83%
IRELAND	3	0.29%	2	0.14%
SERBIA	3	0.29%	8	0.55%
UNITED ARAB EMIRATES	3	0.29%	6	0.41%
AUSTRIA	2	0.19%	6	0.41%
DENMARK	2	0.19%	5	0.35%
FINLAND	2	0.19%	2	0.14%
HUNGARY	2	0.19%	2	0.14%
KUWAIT	2	0.19%	4	0.28%
QATAR	2	0.19%	4	0.28%
SAUDI ARABIA	2	0.19%	3	0.21%
SOUTH AFRICA	2	0.19%	4	0.28%
SWITZERLAND	2	0.19%	4	0.28%
BAHRAIN	1	0.10%	1	0.07%
EGYPT	1	0.10%	1	0.07%
INDONESIA	1	0.10%	1	0.07%
LEBANON	1	0.10%	1	0.07%
MALAYSIA	1	0.10%	0	0.00%
PHILIPPINES	1	0.10%	0	0.00%
SWEDEN	1	0.10%	2	0.14%
UKRAINE	1	0.10%	2	0.14%
BANGLADESH	0	0.00%	1	0.07%
ECUADOR	0	0.00%	1	0.07%
LUXEMBOURG	0	0.00%	1	0.07%
PAKISTAN	0	0.00%	1	0.07%
PARAGUAY	0	0.00%	1	0.07%
<b>Total</b>	<b>1047</b>	<b>100.00%</b>	<b>1446</b>	<b>100.00%</b>

Table 10: Distribution of articles by country in **Scopus** and **Web of Science** databases

## 9.9 AI Data Extraction Tables and Figures

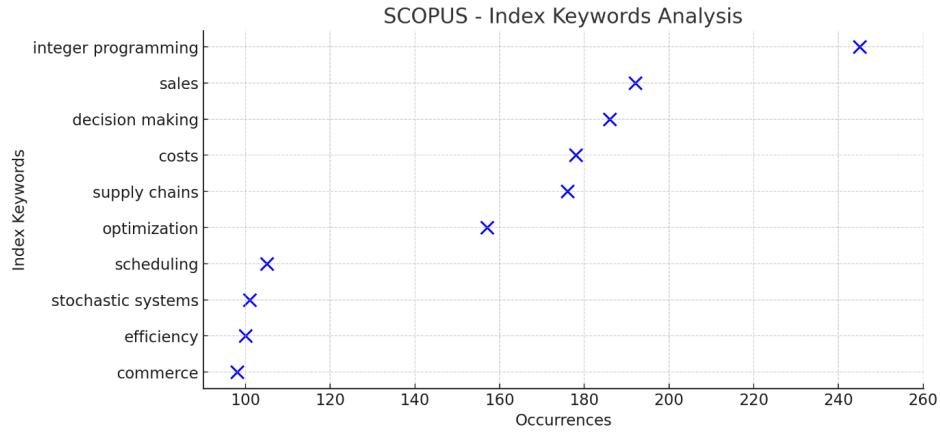


Figure 31: Index keywords frequency analysis on all ITOR articles from **Scopus** generated using **ITOR Systematic Reviewer GPT 4**

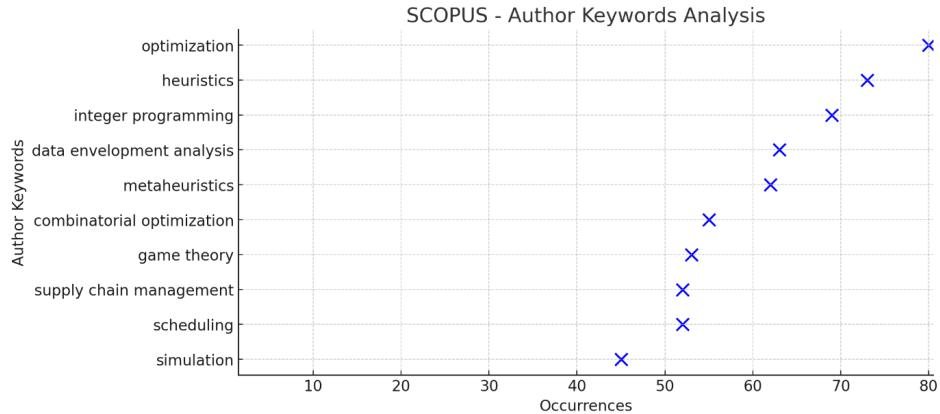


Figure 32: Author keywords frequency analysis on all ITOR articles from **Scopus** generated using **ITOR Systematic Reviewer GPT 4**

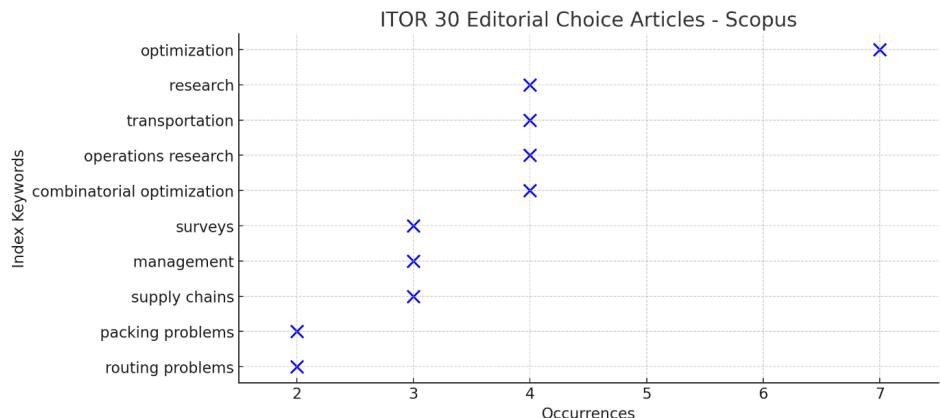


Figure 33: Index keywords frequency analysis on **ITOR's 30 year selection** articles generated using **ITOR Systematic Reviewer GPT 4**

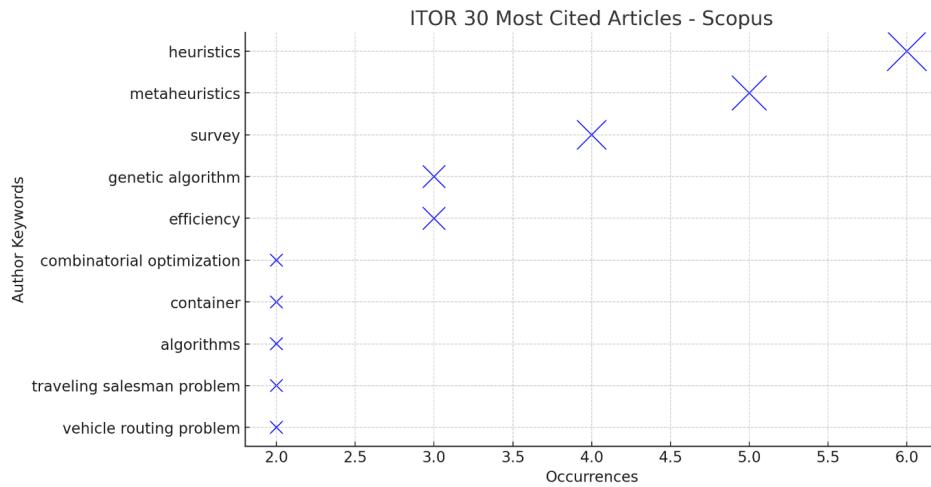


Figure 34: Author keywords frequency analysis on **ITOR's 30 year selection** articles generated using **ITOR Systematic Reviewer GPT 4**

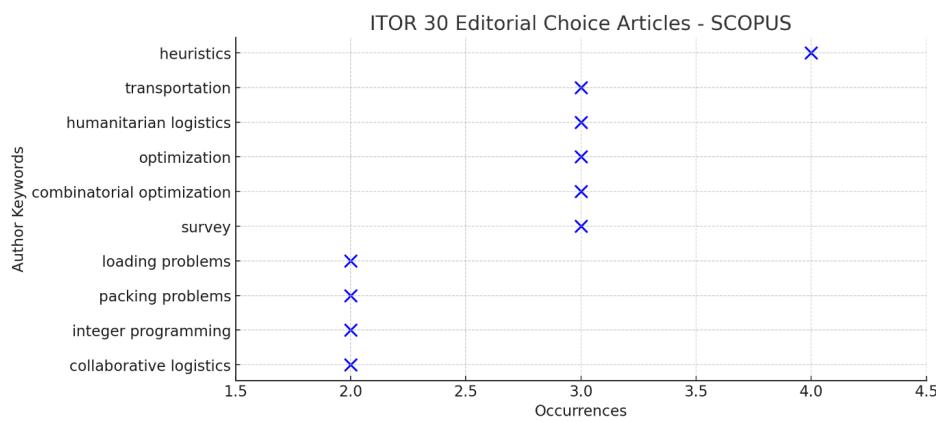


Figure 35: Author keywords frequency analysis on **ITOR's 30 year selection** articles generated using **ITOR Systematic Reviewer GPT 4**

Collaboration Network

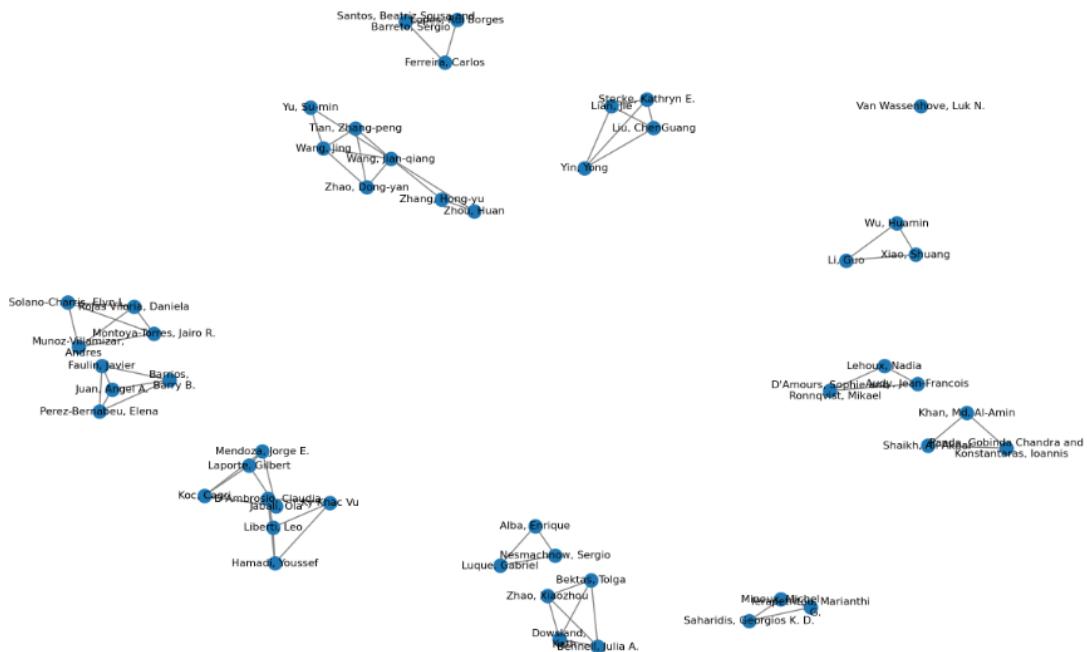


Figure 36: Collaboration network graph of the most cited articles from **Web of Science** generated using **ITOR Systematic Reviewer GPT 4**