https://doi.org/10.51408/issi2025_006

Bridging Classification Systems: The Potentialities of Artificial Intelligence in Developing Concordance Tables for Science, Technology, and Policy

Guendalina Capece[1], Cinzia Daraio[2], Flavia Di Costa[3]

[1] *guendalina.capece@unimercatorum.it*
Department of Engineering and Sciences, Universitas Mercatorum, Rome (Italy)

[2] *daraio@diag.uniroma1.it*
DIAG Sapienza University of Rome, Via Ariosto, 25, 00185 Rome (Italy)

[3] *flavia.dicosta@unimercatorum.it*
Department of Humanities and Social Sciences, Universitas Mercatorum, Rome (Italy)

Abstract

This paper explores the challenges and methodologies involved in aligning diverse subject classification systems through the development of concordance tables. It investigates prior efforts, identifies successful implementations, and evaluates employed methods. Using a multi-method approach, the research combines a literature review with Artificial Intelligence (AI)-enhanced content analysis in Scopus to identify trends and gaps in existing studies. The findings highlight the potential of AI-driven methodologies to improve automation and reliability in creating concordance tables while identifying areas for future research. The study emphasizes the importance and the limits of using AI for integrating classification systems, supporting knowledge organization, and facilitating science and innovation policy decision-making.

Introduction and research questions

As the pursuit of interdisciplinary research frequently encounters diverse and complex systems of knowledge classification, the challenge of aligning these disparate systems becomes increasingly significant. This paper delves into the intricate task of harmonizing various subject classification frameworks by developing concordance tables. By examining prior efforts and successful implementations, while also evaluating the methods employed, this study offers a comprehensive review using a multi-method approach.

Patent data, rich in technological details, have been crucial in showcasing the technological composition of industries (Griliches, 1990). The classification systems used by patent authorities provide high-resolution and hierarchical structures, essential for systematically linking technologies and industries for research purposes (Lafond and Kim, 2019). Historically, patent data have measured technological changes within industries through citation-weighted patent counts. While most changes are incremental and hard to detect without considerable technological shocks, advancements in data collection, natural language processing, and network analysis have introduced new indices to capture gradual technological shifts within industries (Kelly et al., 2018).

Regarding the importance of subject classification systems in informetrics, the key topics include:

1) *Scientific, literature-based*: Systems like Scopus and Web of Science (WoS) fall into this category, often referred to as paper classification.
2) *Technical, patent-based*: The International Patent Classification (IPC) is an example here, based on prior art classification.
3) *Industry sector-based*: This category organizes subjects according to various industrial sectors.

Connections can be established between the different types of classification. For instance: citations in patents to scientific literature may create a link between a patent classification and paper classifications; industries as funders of research papers may establish a link between industrial sector and paper classifications. Industries as assignees of patents may create a link between industrial sector and patent classifications. This study focuses on subject classification systems and aims to systematically analyse which attempts have been made to develop concordance tables between different subject classifications; which concordance tables have actually been created; which methods were used to create these and how successful these methods were, in terms of the degree of validity of the proposed concordance.

Studies on the science-technology-industry interface are confronted with the need to create concordance tables between technology (patent) and industry subject classifications (Schmoch et al., 2003; Schmoch, 2008; Lybbert & Zolas, 2014; Dorner, & Harhoff, 2018; Neuhäusler, Frietsch & Kroll, 2019; Goldschlag, Lybbert, & Zolas, 2020).

Goldschlag, Lybbert & Zolas (2020) applied a probabilistic linkage methodology, pioneered by Lybbert and Zolas (2014), to create concordances between USPC and CPC technology codes and various industry and product classifications, including the International Standard Industrial Classification (ISIC), the North American Industrial Classification System (NAICS), the Standard International Trade Classification (SITC), and the Harmonized System (HS) product codes. Utilizing these concordances, the analysis examined how technology-industry relationships evolved over time by allowing the set of contributing patents to vary. Findings revealed that the link between technologies and industries showed remarkable persistence, with a recent increase in the rate of change after decades of decline. Additionally, the research provided suggestive evidence demonstrating the economic relevance of the measure of technological change. Changes in the industry-technology composition were correlated with shifts in occupational composition, aligning with existing literature on the labor market effects of new technologies.

Neuhäusler, Frietsch & Kroll (2019) enhanced the probabilistic concordance between industry sectors and technology fields, building on prior work (Neuhäusler et al., 2017) by reallocating patents to industry sectors and expanding the database. The analysis further extended to a concordance between scientific disciplines and technology fields. The paper provided valuable insights into the nexus between technological and scientific outputs and economic sectors, building on previous research by Frietsch et al. (2017) and Neuhäusler et al. (2017). This study employed probabilistic concordances at the micro level, linking patents to industry sectors and

publications to technology fields. This method aggregated patents and publications into matrices of patent shares per technology field and sector, and publication shares per discipline and technology field.
Subject classification systems are essential for organizing and accessing knowledge across scientific, technological, and economic domains. However, the coexistence of multiple classification systems often creates challenges in ensuring consistency and interoperability. This paper systematically examines the development and implementation of concordance tables designed to align these diverse systems. Specifically, we address the following research questions: (1) What attempts have been made to develop concordance tables between various subject classifications? (2) Which concordance tables have been successfully created? (3) What methods have been employed in their development? (4) How effective are these methods in terms of the validity, accuracy, and utility of the proposed concordance? (5) How artificial intelligence (AI) could help in developing, maintaining and updating concordance tables?
The paper is organized as follows. The next section outlines the history of concordance tables up to its latest development. The subsequent section describes how Scopus AI can be useful in completing a selected review. The following section presents the potential usefulness of AI in developing, maintaining and updating concordance tables and the last section concludes the paper by highlighting its policy relevance.

Methods

We employ a multi-method approach combining a comprehensive literature review with AI-enhanced content analysis in Scopus. Scopus AI is used to identify patterns, trends, and gaps in existing studies.
An analytical overview of the key characteristics highlights the sophisticated functionalities of the system as listed below:

1) Enhanced robustness through advanced content analysis: By employing Scopus AI, our multi-method approach becomes significantly more robust. The tool's sophisticated algorithms provide detailed insights by analyzing the vast volumes of documents available in the Scopus database. This analysis aids researchers in positioning their own work effectively within the current academic landscape.
2) Identification of patterns and trends: Scopus AI can recognize and highlight recurring patterns and emerging trends within academic publications, offering researchers a thorough understanding of the latest developments and gaps in their field. This facilitates the identification of research opportunities and helps in crafting more relevant and impactful studies.
3) Comprehensive literature review: By combining traditional literature review methods with AI-enhanced analysis, researchers can streamline their review process. Scopus AI quickly processes and categorizes data, ensuring a more exhaustive and precise literature review.

4) Data-driven insights: Scopus AI provides researchers with data-driven insights and analytics, helping them to make informed decisions about their research direction and focus areas. This data can be pivotal in identifying under-researched topics or confirming the significance of ongoing studies.
5) Efficiency and accuracy: The use of AI in content analysis significantly reduces the time and effort required to sift through massive amounts of literature while increasing accuracy. Researchers can rely on Scopus AI to update them with the most relevant and recent publications in their domain.

Results

State of the art on concordance tables

The requirements of a classification system, regardless of the specific application area (Fettke and Loos, 2003) are listed below:

- *Completeness*: the specific application domain should be completely covered by the classification scheme.
- *Precision*: a classification scheme must describe models at different levels of detail. The precision of a classification can be increased by defining new classes narrower.
- *Consistency*: the classification scheme must be free from contradictions.
- *Extensibility*: a classification system should be extensible so that they can be adopted in the future. It has extensibility if the new classification characteristics remain stable even after the addition or removal of some classes from the scheme.
- *User-friendliness*: classification scheme should be clearly understood.
- *Economic efficiency*: different type of costs for development and implementation of classification system.

The aim of classification is ordering entities into groups or classes based on their similarity: that means, from a statistical point of view, trying to minimize within-group variance and maximizing the between-group variance. Consequently, it seeks to realize groups that are as different (non-overlapping) possible with the maximum degree of similarity within each group. The basic rule of classification is set up classes that are both exhaustive and mutually exclusive. Typology is another term for a classification: it is multi-dimensional and conceptual. Taxonomy is a term similar to Typology, used as synonym, although it ought to be preferably used for classification of empirical entity (Bayley, 1994).
Listed in the

Table are pros and cons of classification schemes.

Table 1. Classification schemes - pros and cons.

PROS	*CONS*
It is a descriptive tool	Classification is descriptive, pre-explanatory or non-explanatory
Reduction of complexity that allows to synthesize a large amount of data in a smaller number of Types (taxa) significant	Static classification
Identification of the similarities and/or identification of differences in a complementary manner	Difficulty to choose the size and finding cases for classification
Submission of an exhaustive list of dimensions	The logic of the classes because typologies are criticized as dependent on the logic of classes rather than the use of continuous date as in the modern statistical techniques.
Comparison of Types	Although the types are often purely descriptive, however, they serve for the study of relationships and also for the specification of hypotheses concerning these relationships

1. Patent Classification

Internationally, the classification system is the International Patent Classification (IPC) which is updated periodically. The IPC was established in 1971 by the Strasbourg Agreement to provide and ensure a harmonized, hierarchical system for classifying the technology contained in patents and utility models. The current version of the IPC (2022) divides technology into eight sections (A-H) with approximately 75,000 subdivisions. According to the last version of IPC guide (2022):

“The Classification, being a means for obtaining an internationally uniform classification of patent documents has as its primary purpose the establishment of an effective search tool for the retrieval of patent documents by intellectual property offices and other users, in order to establish the novelty and evaluate the inventive step or non-obviousness (including the assessment of technical advance and useful results or utility) of technical disclosures in patent applications”.

The Classification, furthermore, has the important purposes of serving as:

a) an instrument for the orderly arrangement of patent documents in order to facilitate access to the technological and legal information contained therein;

b) a basis for selective dissemination of information to all users of patent information;
c) a basis for investigating the state of the art in given fields of technology;
d) a basis for the preparation of industrial property statistics which in turn permit the assessment of technological development in various areas".

"The IPC is a hierarchical classification system. The contents of lower hierarchical levels are subdivisions of the contents of the higher hierarchical levels to which the lower levels are subordinated. The Classification separates the whole body of technical knowledge using the hierarchical levels, i.e., section, class, subclass, group and subgroup, in descending order of hierarchy".

The patenting system has a classification problem. The current classification system is based on technological and functional principles. The classification scheme is built from a technical point of view: an invention is normally classified according to its function or intrinsic nature.

According to IPC Guide (2022) page 22:

"As an application-oriented reference usually points from a function-oriented place to an application-oriented place, so an informative reference usually points from an application-oriented place to a function-oriented place".

"When it is unclear whether to classify a technical subject in a function-oriented place or in an application-oriented place, the following should be observed:

a) If a particular application is mentioned, but not specifically disclosed or fully identified, classification is made in the function-oriented place, if available. This is likely to be the case when several applications are broadly stated.
b) If the essential technical characteristics of the subject relate both to the intrinsic nature or function of a thing and to its particular use, or its special adaptation to or incorporation into a larger system, classification is made in both the function-oriented place and the application-oriented place, if available.
c) If guidance indicated in subparagraphs (a) and (b), above, cannot be used, classification is made in both the function-oriented place and the relevant application-oriented places".

2. Industry sectors Classification (IPC-industry concordances)

Industry classification or industry taxonomy organizes companies into industrial groupings based on similar production processes, similar products, or similar behavior in financial markets. A wide variety of taxonomies is in use, sponsored by different organizations and based on different criteria (Table 2).

Table 2. Industry classifications
Industry classification. Retrieved 08:29, January 21, 2025, from https://en.wikipedia.org/w/index.php?title=Industry_classification&oldid=1220947550.

ABBREVIATION	*FULL NAME*	*SPONSOR*	*CRITERION/ UNIT*	*NODE COUNT BY LEVEL*	*ISSUED*
ANZSIC	Australian and New Zealand Standard Industrial Classification	Governments of Australia and New Zealand			1993, 2006
BICS	Bloomberg Industry Classification Standard[2]	Bloomberg L.P.		10/.../2294	
GICS	Global Industry Classification Standard	Standard & Poor's, MSCI	market/ company	2-8 digits 11/24/69/158	1999–present (2018)
HSICS	Hang Seng Industry Classification System[3]	Hang Seng Indexes Company	Revenue source	11/31/89	
IBBICS	Industry Building Blocks[4]	Industry Building Blocks	Market line of business	19/130/550/3000/20200	2002
ICB	Industry Classification Benchmark	FTSE	market/ company	11/20/45/173	2005–present (2019)
ISIC	International Standard Industrial Classification of All Economic Activities	United Nations Statistics Division	production/ establishment	4 digits 21/88/238/419	1948–present (Rev. 4, 2008)
MGECS	Morningstar Global Equity Classificati	Morningstar, Inc.	Securities behavior	3/14/69/148	

ABBREVIATION	FULL NAME	SPONSOR	CRITERION/ UNIT	NODE COUNT BY LEVEL	ISSUED
	on System[5]				
NACE	Statistical Classification of Economic Activities in the European Community	European Union	production/ establishment	6 digits	1970, 1990, 2006, 2023
NAICS	North American Industry Classification System	Governments of the United States, Canada, and Mexico	production/ establishment	6 digits 17/99/313/724/1175 (/19745)[1]	1997, 2002, 2012, 2017, 2022
RBICS	FactSet Revere Business Industry Classification System	FactSet, acquired in 2013[6]	line of business	11000	
SIC	Standard Industrial Classification	Government of the United States	production/ establishment	4 digits 1004 categories	1937–1987 (superseded by NAICS, but still used in some applications)
SNI	Swedish Standard Industrial Classification	Government of Sweden			
TRBC	The Refinitiv Business Classification	Refinitiv	market/ company	10 digits 13/33/62/154/898[7]	2004, 2008, 2012, 2020[8]

ABBREVIATION	*FULL NAME*	*SPONSOR*	*CRITERION/ UNIT*	*NODE COUNT BY LEVEL*	*ISSUED*
UKSIC	United Kingdom Standard Industrial Classification of Economic Activities	Government of the United Kingdom			1948–present (2007)
UNSPSC	United Nations Standard Products and Services Code	United Nations	Product	8 digits (optional 9th) (four levels)	1998–present

The patent classification (IPC) and the industrial classification are not directly comparable. Three are the criteria for assigning an invention to an industry:

1) origin based: patents are assigned to the industrial sector of origin (the main economic sector of inventing / applicant company) (industry of origin);
2) producer based;
3) user based: patents are assigned to the sector where it is in use (the main industry to which belongs the product incorporating the invention) (industry of destination or industry of use).

There are three levels at which patents can be linked to economic activity:

1) macro-level (country): for study rate of innovation, country's innovative capacity effects of patent harmonization;
2) meso level (industry): for studying relationship between patenting and economic activity through time, space and technological classes;
3) micro-level (firm): patenting as part of firm-level strategies.

At meso level the link between patent and industry is based on concordance tables.

3. Concordances

Over the past decades, several notable concordance tables have been developed to map different classification systems (e.g., patent classifications to industry or product codes). Rather than describing each approach in detail within the text, we summarize the main characteristics of these concordances in Table 3. The table highlights the year, classification systems, methodology (e.g., probabilistic vs. direct mappings), and principal contributions of each notable concordance effort.
This consolidated view underscores the diverse methodological approaches—from manual mapping of IPC subclasses to industrial codes (Schmoch et al., 2003) to

algorithmic linkages using textual descriptions (Lybbert & Zolas, 2014)—and reveals how each method addresses particular research needs. It also illustrates how concordances have gradually become more probabilistic and data-driven, reflecting broader trends in AI and big data analytics.

By presenting these concordances side by side, we provide readers with a straightforward means to compare their strengths, limitations, and contexts of application (e.g., macro-level policy analysis vs. micro-level firm strategy). We refer to specific details of each study only when needed for interpreting our results, thus avoiding repetitive textual descriptions in the main body.

Table 3. Comparative Overview of Key Concordance Tables.

Concordance	Year	Classification Systems Mapped	Mapping Method	Key Contribution/Notes
Yale Technology Concordance (YTC) (Evenson et al., 1991)	1991	Patent (Canadian) → Industry (IOO & IUO)	Probabilistic/Manual	Early effort linking patents to industry of origin/use
DG Concordance (Schmoch et al., 2003)	2003	IPC → ISIC (44 sectors)	Manual mapping	Widely used in patent statistics; basis for Eurostat
ALP (Lybbert & Zolas, 2014)	2014	IPC → ISIC/SITC	Algorithmic/probabilistic	Introduces textual keywords & Bayesian weighting
Inventor-Establishment (Dorner & Harhoff) (2018)	2018	Patent (EPO) → Industry (NACE)	Micro-level matching	Leverages inventor-level data for higher precision
Others (OECD, MERIT, etc.)	Various	Patent → Industry/Product classifications	Mixed methods	Show incremental improvements & expansions

Notes:

- *IPC: International Patent Classification; ISIC: International Standard Industrial Classification; SITC: Standard International Trade Classification; NACE: Statistical Classification of Economic Activities in the European Community.*
- *IOO/IUO: Industry of Origin/Industry of Use.*
- *The approaches vary in granularity (e.g., macro-level vs. micro-level) and complexity (manual vs. algorithmic).*

The few studies whose goal was to design a concordance between industry sectors and technology classifications are listed in the following table (Table 4).

Table 4. Articles with concordance tables.

Concordance scheme	*Content*
OTAF Concordance (1974)	A computerized method has been developed by the OTAF at the U.S. Patent and Trademark Office to establish links based on a concordance of detailed patent classification codes and industry codes.
Evenson, R. E., Putnam, J. & Kortum, S. (1991)	Concordance table based on the industry classification made by the Canadian Intellectual Property Office that assigned both an industry of origin code (IOO) and an industry use code (IUO) to Canadian patents.
Kortum, S., & Putnam, J. (1997)	YCT developed within a probabilistic framework, linking potential industries. A statistical model predicts industry standard errors.
Verspagen, B., Morgastel, T. v., Slabbers, M. (1994)	MERIT Concordance matches IPC subclasses to 22 industrial classes based on a mix of 2- and 3-digit ISIC codes.
Johnson, D. K. (2002)	The OECD Technology Concordance (OTC), similar to the Yale Technology Concordance, serves as an instrument for converting IPC-based patent data into patent counts categorized by economic sector.
Schmoch, U., Laville, F., Patel, P. & Frietsch, R. (2003)	The "DG Concordance" aims to align IPC subclasses with ISIC industry classifications, assigning 625 IPC subclasses to 44 manufacturing sectors, each associated with one or more ISIC codes.
Lybbert, T. J., & Zolas, N. J. (2014)	The "Algorithmic Links with Probabilities" (ALP) approach constructs concordances between the IPC system and industry classification systems like SITC and ISIC. It uses keywords from industry descriptions and a probabilistic framework to match data, providing meso-level mappings that complement macro- and firm-level mappings.

Concordance scheme	*Content*
van Looy, B., Vereyen, C., & Schmoch, U. (2014)	The concordance update, addressing the limitations of the 2003 Schmoch et al. version, widely utilized by Eurostat for patent statistics, reviewed 44 technology definitions, assigned IPC 4-digit codes, and incorporated the NACE 2 classification.
Dorner, M., & Harhoff, D. (2018)	Concordance tables, address the preference for linking industries to their knowledge and technological opportunities ("industry of origin") using linked inventor-establishment data for Germany to generate accurate industry of origin information for patents.
Neuhäusler, P., Frietsch, R., & Kroll, H. (2019)	Concordance tables integrate micro-level data from patent applicants and authors, aggregated at sector and technology field levels, utilizing sources like NACE Rev. 2 and the 35 WIPO fields, and applying probabilistic methods to generate comprehensive concordances.
Goldschlag, N., Lybbert, T. J., & Zolas, N. J. (2019)	Concordance tables, using a probabilistic linkage methodology, were created to map USPC and CPC technology codes to various industry and product classifications, including ISIC, NAICS, SITC, and HS.

An application of Scopus AI

The following steps outline the procedure for effectively utilizing Scopus AI in research, ensuring a comprehensive and data-driven understanding of the field (Table 5):

Table 5. Scopus AI procedure.

Step	*Description*
1. Accessing Scopus AI	Start by logging into the Scopus account, ensuring that the institution has access to the Scopus AI features.
2. Formulating Queries	Natural language is used to type questions or statements into the Scopus AI search box. No need for complex search strings.
3. AI-Enhanced Search	Sophisticated algorithms analyze the vast volume of documents, identifying patterns, trends, and knowledge gaps.
4. Reviewing Topic Summary	Scopus AI synthesizes abstracts from relevant documents to generate a Topic Summary, offering an overarching view of the subject.
5. Exploring Expanded Summary	Detailed information is provided through the Expanded Summary, offering comprehensive insights and references to supporting documents.
6. Utilizing Concept Maps	Keywords from research abstracts are used to generate an interactive Concept Map, illustrating the relationship with various subtopics.
7. Identifying Key Papers and Researchers	Scopus AI highlights influential papers and top researchers linked to the query, helping identify critical publications and leading experts.
8. Data-Driven Insights	Insights and analytics provided by Scopus AI are utilized to make informed decisions about research direction, identifying under-researched areas.
9. Saving and Exporting Results	After reviewing, results can be saved and exported for further analysis and reference. Summaries and insights can be easily shared with the research team.

To test the potentialities of the Scopus-AI tool, bibliographical research was conducted as described below.
A sentence containing the keywords (concordance table, subject classification, science classification, patent classification, industry classification) from Daraio, Di Costa & Moed (2014) was entered into the search box.
Input in search box
"How can concordance tables facilitate the alignment of different subject classification, science classification, patent classification, and industry classification systems"
The Scopus -AI hidden procedure that generates the results is the following:

1) **Creating a plan to answer your query**
2) **Performing natural language search:**
How do concordance tables facilitate the alignment of different classification systems in subjects, sciences, patents, and industries?

3) **Performing keyword search:**
("concordance" OR "alignment" OR "correspondence" OR "mapping") AND ("tables" OR "charts" OR "lists" OR "diagrams") AND ("classification" OR "categorization" OR "taxonomy" OR "system") AND ("systems" OR "frameworks" OR "models" OR "structures")
4) **Generating summary**

Scopus-AI creates a structured summary in the following main sections, accompanied by the respective references:

1) How Concordance Tables Facilitate Alignment of Different Classification Systems
2) Linking Different Classifications
3) Policy Planning and Technology Targeting
4) Empirical Innovation Studies
5) Advantages and Applications
6) Challenges and Considerations
7) Conclusion

Scopus AI utilizes keywords extracted from research abstracts to generate interactive Concept Maps. This tool offers a comprehensive overview of the topic landscape, illustrating connections with various research areas, including those that may extend beyond the user's traditional scope of inquiry. Figure 1 displays the results of the example under consideration.

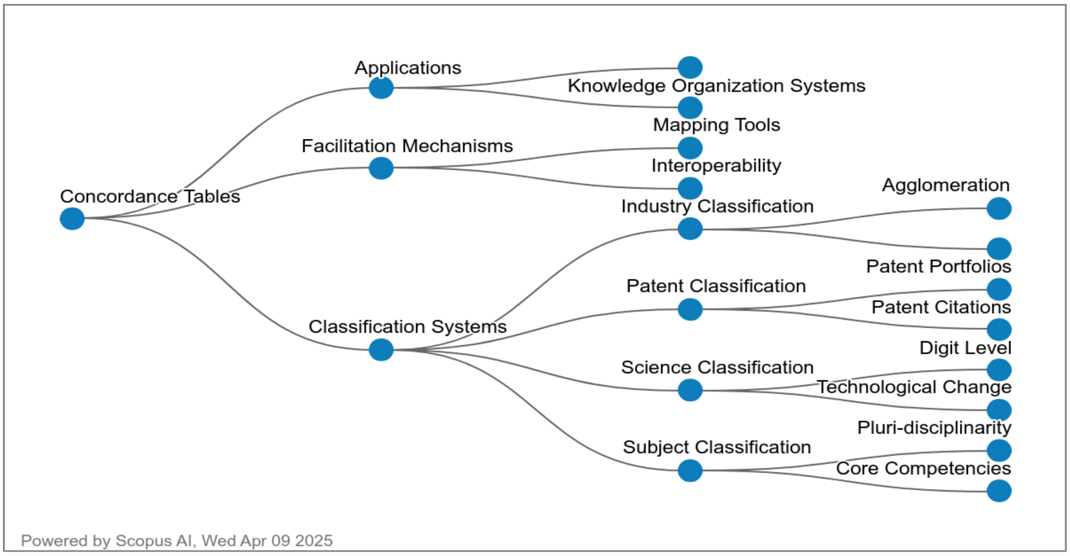

Figure 1. Concept Map generated by Scopus AI on Apr 09 2025.

Benefits and challenges

Concordance tables offer numerous benefits, guiding policymakers in understanding the gaps between science, technology, and industry, aiding in targeted policy planning and technology development (Wong, & Fung, 2017). They facilitate the utilization of often-underutilized patent documents and technical information, enabling the visualization and analysis of relationships among technologies, which supports more informed decision-making (Pasek, 2021; Leydesdorff, 2008). By highlighting connections between different fields, concordance tables foster innovation by identifying opportunities for cross-disciplinary research and development, helping track the evolution of technologies and industries, and providing a roadmap for future innovation (Lee, 2018; Wong, & Fung, 2017).

However, several challenges accompany concordance tables. The alignment of different classification systems like IPC and CPC, each with distinct logic and granularity, is complex and often leads to misalignment and oversight of emerging technological trends (Lobo, & Strumsky, 2019; Alisova, 2013). Aggregating bibliographic data from diverse sources poses technical difficulties, requiring high-quality mappings to resolve defects such as missing or incorrect relations (Pfeffer, 2016; Ivanova & Lambrix, 2013). Automated methods, while less resource-intensive, may not achieve the same level of accuracy (Pfeffer, 2016). Despite these challenges, concordance tables remain a valuable tool at the intersection of technology and industry. See Table 6 for a summary of the main challenges and benefits of concordance tables.

Table 6. Challenges and benefits of concordance tables to align classification systems.

Challenges	*Benefits*
Complexity and diversity of systems (Lobo, & Strumsky, 2019; Alisova, 2013; Lee, 2018)	Policy and planning support (Wong, & Fung, 2017)
Data integration and quality issues (Pfeffer, 2016; Ivanova & Lambrix, 2013)	Enhanced data utilization (Pasek, 2021; Leydesdorff, 2008)
Resource intensity (Pfeffer, 2016)	Support for innovation (Lee, 2018; Wong, & Fung, 2017)

Potentialities and limits of AI usage for concordance tables

AI-driven techniques provide solutions to several key challenges in mapping classification systems, including scalability, semantic ambiguity, and the need for dynamic updates. By combining deep learning for semantic understanding, clustering for pattern detection, and predictive modeling for adaptability, AI introduces a powerful set of tools to automate, refine, and expedite the concordance process. Moreover, the hybrid integration of AI with human expertise ensures that the benefits of automation are paired with contextual precision, resulting in robust and accurate mapping frameworks.

Artificial intelligence—specifically, natural language processing (NLP) and machine learning (ML)—be utilized to *automate*, *refine*, *accelerate* and *validate* the mapping process.

Natural language processing (NLP) can analyze and interpret textual descriptions, category names, and associated metadata in classification systems. Specific applications include i) *Semantic Analysis* in which NLP algorithms extract the meaning of terms and phrases from classification systems, identifying synonyms, hierarchical relationships, and contextual overlaps; ii) *Entity Recognition* in which NLP can identify and tag key concepts, entities, and terms from textual data, enabling precise alignment between systems; iii) *Text Clustering*: based on NLP-powered clustering that groups similar terms across classifications, revealing patterns of equivalence or correspondence.

Machine learning algorithms can improve the precision of concordance mapping through i) *Supervised Learning* through which ML models trained on labeled datasets can learn to map terms from one classification system to another, generalizing their knowledge to new, unseen classifications; ii) *Unsupervised Learning* based on techniques like clustering or topic modeling can identify hidden relationships in datasets without requiring pre-labeled data, making them ideal for exploratory concordance creation; iii) *Contextual Embeddings* based on advanced ML methods like transformer-based models that can embed terms and categories in high-dimensional spaces, enabling similarity detection based on context.

AI technologies significantly reduce the time and effort required for mapping by automating repetitive and computationally intensive tasks. Large datasets spanning multiple classification systems can be processed simultaneously, scaling concordance efforts beyond manual capabilities. AI systems can automatically update mappings as classification systems evolve or new data becomes available.

Finally, AI can enhance the reliability and validity of the mappings by error detection, and identifying inconsistencies or ambiguities in the concordance through anomaly detection algorithms.

AI-driven techniques—such as deep learning for semantic analysis, clustering for unsupervised categorization, and predictive modeling for trend analysis— could be incorporated to overcome the limitations of traditional tools with *hybrid methodologies*, based on the combination of manual expertise with automated AI processing. AI systems are powerful but not infallible. They are particularly proficient at handling large-scale, repetitive tasks, while human experts excel at nuanced judgment and contextual understanding. By combining AI-driven techniques with manual expertise, the limitations of both approaches can be mitigated.

Additionally, AI enables dynamic and adaptive concordance tables that evolve with new data inputs, reflecting real-time changes in classifications.

AI tools offer transformative opportunities for policymakers to make better use of concordance tables by enhancing their accessibility, adaptability, and utility in decision-making processes. By incorporating AI-driven techniques, concordance tables can be transformed from static tools into dynamic, interactive systems that provide real-time updates, visual analytics, and predictive insights.

AI can monitor data continuously updated, such as scientific publications and patents, to ensure concordance tables remain current and reflect the latest developments. This allows policymakers to make decisions based on the most up-to-date information, particularly in rapidly evolving fields. Furthermore, AI-powered visual analytics tools, such as dashboards and network graphs, can present complex concordance relationships in an intuitive and actionable format. For example, policymakers could use these tools to identify overlaps or gaps in innovation funding across sectors or to explore regional trends in research output.

Another critical capability of AI is its ability to provide predictive insights and scenario modeling. By simulating the potential outcomes of classification alignments, AI tools can help policymakers anticipate the effects of their decisions on various sectors, such as predicting the impact of aligning academic and industry classifications on workforce development or innovation growth. Moreover, these tools can be tailored to provide policy-specific recommendations, allowing policymakers to explore how concordance relationships affect their goals and constraints.

Despite these advantages, the use of AI in this domain is not without *significant limitations*. A major challenge lies in the dependence on the *quality* and *representativeness* of the training datasets. *Incomplete* or *biased* data can lead to inaccuracies in concordance mappings, perpetuating existing discrepancies rather than resolving them. Additionally, many AI models operate as opaque *black boxes*, where their processes and outputs are not easily interpretable. This *lack of transparency* can undermine trust among stakeholders and impede the reproducibility of results, a critical aspect of scientific rigor.

AI tools also face difficulties in capturing domain-specific distinctions, particularly in highly specialized or interdisciplinary fields. While AI excels at automating repetitive tasks, its ability to make contextually informed decisions is limited, *necessitating continued human supervision*. Furthermore, the dynamic nature of classification systems, while well-suited to AI's adaptive capabilities, introduces challenges in maintaining long-term consistency. Frequent updates to concordance tables can lead to *fragmentation* or misalignment of historical data.

The *ethical implications* of AI use are another pressing concern. *Bias* in AI models, if unchecked, can exacerbate existing misuses, and the use of proprietary or sensitive data raises questions about *privacy* and *intellectual property*. Infrastructure and expertise requirements present additional barriers, as deploying and maintaining sophisticated AI systems often demands significant computational resources and technical skills. These constraints can limit accessibility for smaller organizations or underfunded research initiatives, creating disparities in who can leverage these tools effectively. Moreover, overreliance on AI risks neglecting the critical evaluative role that human judgment plays in ensuring accuracy and relevance.

To address these challenges, the research community and policymakers must prioritize efforts to mitigate biases in training datasets and promote the development of transparent, interpretable AI models. Collaborative frameworks that bring together AI developers, domain experts, and decision-makers are essential to ensure that AI-driven concordance tools produce balanced and meaningful outcomes. The creation of resource-efficient and cost-effective solutions is equally important to expand accessibility across diverse institutions. Ethical oversight and accountability mechanisms should be established to monitor AI usage, safeguard data privacy, and foster trust.

While the limitations of AI tools are significant, they do not compensate the transformative potential these technologies hold for enhancing the efficiency, precision, and adaptability of concordance tables. By addressing these issues thoughtfully, AI can serve as a powerful catalyst for aligning classification systems and advancing innovation policies in an increasingly interconnected world.

AI-driven techniques have greatly enhanced the creation and maintenance of concordance tables by automating key tasks such as large-scale text analysis, semantic clustering, and predictive modelling. Through natural language processing (NLP) and machine learning (ML) methods—including deep learning and transformer-based embeddings—AI can reduce manual effort, scale mapping efforts across large, diverse datasets, and dynamically update concordances in response to newly available information. In doing so, policymakers and researchers gain real-time insights, enabling more informed decisions about resource allocation, innovation funding, and strategic planning.

Despite its advantages, the effectiveness of AI depends heavily on high-quality, representative training data and transparent, interpretable models. Biased or incomplete datasets can reinforce existing discrepancies, while black-box approaches make it difficult to validate or reproduce results. AI also struggles with domain-specific nuances, requiring ongoing human supervision and expert input. Moreover, adopting sophisticated AI systems often demands significant computational resources, specialized expertise, and robust data governance—factors that may restrict access for smaller organizations. Ethical and legal considerations, such as bias, data privacy, and intellectual property, further complicate large-scale adoption.

Addressing these challenges calls for collaborative frameworks among AI developers, domain experts, and policymakers, alongside efforts to develop resource-efficient, explainable AI solutions. With proper oversight and strategies to mitigate biases, AI tools can serve as powerful catalysts for enhancing the efficiency, precision, and adaptability of concordance tables, promoting more effective alignment of classification systems in an increasingly interconnected world.

AI Techniques for Developing and Updating Concordance Tables

Recent advancements in Artificial Intelligence (AI)—particularly in Natural Language Processing (NLP) and Machine Learning (ML)—offer powerful tools to address the complexities of aligning diverse classification systems (e.g., patent, industry, and scientific taxonomies). In this context, AI can**:**

1. *Automate Large-Scale Text Analysis*. NLP methods such as named-entity recognition and text clustering enable the systematic extraction and grouping of relevant terms or codes from extensive document corpora (patents, scientific articles, etc.). These methods can detect semantic overlaps, synonyms, or

hierarchical relationships that inform how different classification systems interrelate.
2. *Improve Accuracy and Reduce Redundancies***.** By applying supervised learning (e.g., Random Forest, SVM, or neural networks) to labeled training sets, AI algorithms learn to associate the descriptive content of documents with specific industry or patent classes. This reduces time-consuming manual mapping and can facilitate ongoing updates as new data emerge.
3. *Identify Ambiguities and Cross-Disciplinary Links***.** NLP-driven topic modeling and clustering (e.g., LDA, DBSCAN) can discover hidden patterns and overlapping categories. This is particularly useful when dealing with interdisciplinary fields, where traditional classification schemes may lack clarity or granularity.
4. *Enable Dynamic, Scalable Concordance Tables***.** Machine learning approaches can update mappings in near real-time, reflecting evolving research frontiers or emerging technologies. Hybrid "human-in-the-loop" workflows, in which experts validate uncertain assignments, further enhance reliability and transparency.

Practical Examples

Works such as Lybbert and Zolas (2014) employ algorithmic text-matching to link International Patent Classification (IPC) codes with economic and industry classifications (e.g., ISIC, SITC). Similarly, Dorner and Harhoff (2018) leverage inventor-establishment data to refine the accuracy of patent-to-industry correspondences. Although in the previously cited approaches there are not AI techniques applied, these methods illustrate how AI-driven techniques can increase both the speed and precision of concordance-building efforts, helping policymakers and scholars navigate constantly evolving classification systems.

Addressing Ethical and Infrastructural Challenges

Despite these advantages, AI-based approaches also raise important ethical and infrastructural considerations (Table 7).

Table 7. AI-based approaches ethical and infrastructural considerations.

Challenges	*Description*
Bias in Training Data	Algorithmic decisions can inadvertently reflect biases in the underlying datasets, especially if certain industries, countries, or languages are underrepresented. Periodic audits and balanced data sampling can help reduce these distortions
Data Privacy and Confidentiality	Large-scale text analysis often involves sensitive corporate data, personal details (e.g., inventor information), or confidential product descriptions. Employing robust data governance strategies—such as anonymization protocols and secure

	storage—ensures compliance with legal and ethical standards
Interpretability and Accountability	Many advanced AI models (e.g., deep neural networks) operate as “black boxes,” complicating the explanation of how specific concordances are generated. Solutions include explainable AI frameworks and transparent reporting of model decisions.
Infrastructure and Accessibility	Training and deploying AI models can require significant computational resources and specialized expertise. Smaller research groups or institutions may lack the necessary hardware, software, or funding to implement advanced methods, potentially widening the gap in data capabilities across organizations
Dynamic Maintenance Over Time	Because classification systems evolve, concordance tables must be continuously updated. AI can facilitate automated or semi-automated revision, but this introduces ongoing costs in software maintenance, model retraining, and data curation

By actively managing these *ethical and infrastructural challenges*, researchers and policymakers can maximize the benefits of AI-driven concordance mapping—greater speed, scalability, and accuracy—while ensuring fair, secure, and transparent processes.

Conclusions and further development

This study emphasizes the central role of concordance tables in harmonizing diverse classification systems across scientific, technological, and industrial domains. Through a detailed exploration of historical developments, methodological advancements, and the integration of Artificial Intelligence (AI), our research sheds light on the opportunities and challenges inherent in aligning these systems. Concordance tables are indispensable tools for fostering interoperability, facilitating knowledge organization, and supporting evidence-based decision-making in science and innovation policy.

The integration of AI into the creation and maintenance of concordance tables marks a significant step forward. AI-driven tools such as Scopus AI demonstrate transformative potential in this context, enhancing automation, precision, and scalability. For instance, Scopus AI's ability to analyze vast datasets, identify patterns, and generate concept maps provides researchers with a comprehensive understanding of classification relationships. By summarizing information from diverse sources, the tool reveals knowledge gaps and highlights emerging trends, enabling the development of concordance tables that remain relevant in an ever-evolving landscape. These capabilities were evident in the AI-generated concept maps and summaries, which revealed the intricate connections between subject classifications, patent classifications, and industry frameworks.

Despite these advances, the application of AI in this domain faces notable limitations. AI tools depend heavily on the quality, diversity, and neutrality of training datasets. Inadequate or biased data can lead to inaccuracies in mappings, undermining the reliability of concordance tables. Additionally, the *black-box* nature of many AI models poses challenges for interpretability and transparency, complicating efforts to validate and trust their outputs. Domain-specific distinctions and the dynamic evolution of classification systems further complicate the process, as these require a combination of automated processing and expert judgment to address.

The ethical and infrastructural considerations associated with AI tools also warrant attention. Biases in AI models, if unchecked, can exacerbate systemic biases, while issues related to data privacy and intellectual property remain pressing concerns. The computational resources and expertise required to implement sophisticated AI systems often limit their accessibility to well-funded organizations, creating disparities across the research landscape.

To harness the full potential of AI while addressing its limitations, future research must focus on several key areas. First, ensuring transparency and fairness in AI methodologies is essential. This involves developing explainable AI models and employing diverse training datasets to mitigate biases. Second, collaborative efforts between AI developers, domain experts, and policymakers are necessary to balance the computational power of AI with the contextual precision of human oversight. Third, the creation of resource-efficient AI tools can enhance accessibility, enabling broader participation in the development of concordance tables.

The Scopus AI tool offers an indication into the future of AI-powered research, demonstrating how interactive visualizations and real-time data analysis can support decision-making. By aligning patent classifications with academic research and industry frameworks, these tools provide actionable insights into the innovation ecosystem. For example, identifying gaps between research activity and patent filings could inform targeted funding strategies or reveal emerging technologies requiring early support.

Building on these insights, we propose *five* recommendations to guide both future research and policy initiatives:

i) *Promote Open, Interoperable Datasets by* establishing standardized metadata protocols so that patent, scientific, and industry data can be more easily integrated and by encouraging data sharing across institutions and countries through open-access repositories, enabling the creation of more accurate and universally applicable concordance tables.

ii) Develop Transparent and Fair AI Tools by prioritizing explainable AI approaches that allow for auditing and improving algorithmic decisions and by adopting bias-mitigation strategies, including balanced sampling and periodic model audits, to ensure underrepresented fields or regions are adequately reflected.

iii) *Enhance Collaborative Frameworks* by fostering partnerships between domain experts, AI developers, and policymakers to combine technical expertise with contextual knowledge and by encouraging cross-sectoral working groups to refine AI methodologies and evaluate their impact on policy decisions.

iv) *Create Policy Incentives for Dynamic Concordances* by integrating human-in-the-loop governance in official guidelines, ensuring experts validate AI outputs for sensitive sectors and by supporting sustainable funding models to maintain and update concordance tables, reflecting changes in classification systems and emerging technologies.

v) *Strengthen Ethical and Legal Frameworks* by implementing data privacy regulations that protect sensitive information while allowing large-scale text analysis and by enforcing accountability mechanisms (e.g., impact assessments, review boards) for teams employing AI in classifying or mapping potentially sensitive data.

By adopting these recommendations, researchers and policymakers can collaboratively move toward more effective, equitable, and innovative concordance-building efforts. In conclusion, while AI introduces significant advancements in the development and maintenance of concordance tables, its successful implementation requires a careful balance between automation and human expertise. Addressing the ethical, technical, and infrastructural challenges associated with AI is crucial for realizing its full potential. By adopting hybrid approaches and fostering collaborative frameworks, concordance tables can evolve into dynamic tools that not only align classification systems but also drive innovation and policy development in a rapidly changing knowledge-based world.

Acknowledgments

This work builds upon and expands the research initiated in Daraio, Di Costa, and Moed (2014) and is dedicated to the memory of Henk Moed.

References

Alisova, N. V. (2013). Biomedical engineering in international patent classification. Biomedical Engineering, 47(3), 164-168.

Bailey, K. D. (1994). *Typologies and taxonomies: An introduction to classification techniques*. Sage Publications, Inc.

Daraio, C., Di Costa, F., & Moed, H. F. (2014). Towards Concordance Tables of Different Subject Classification Systems. A literature review with policy implications. STI 2014 Leiden, 132-135.

Dorner, M., & Harhoff, D. (2018). A novel technology-industry concordance table based on linked inventor-establishment data. *Research Policy*, 47(4), 768-781.

Evenson, R. E., Putnam, J. & Kortum, S. (1991). Estimating patent counts by industry using the Yale-Canada concordance. Final Report to the National Science Foundation.

Fettke, P., & Loos, P. (2003). Classification of reference models: a methodology and its application. *Information systems and e-business management*, *1*, 35-53.

Frietsch, R., Kladroba, A., Markianidou, P., Neuhäusler, P., Peter, V., Ravet, J., ... & Schneider, J. (2017). Final Report on the Collection of Patents and Business Indicators by Economic Sector: Societal Grand Challenges and Key Enabling Technologies Collection and Analysis of Private R&D Investment and Patent Data in Different Sectors, Thematic Areas and Societal Challenges. Luxembourg: Publications Office of the European Union, Luxembourg, Accessed 27 January 2025.

Goldschlag, N., Lybbert, T. J., & Zolas, N. J. (2020). Tracking the technological composition of industries with algorithmic patent concordances. *Economics of Innovation and New Technology*, *29*(6), 582-602.

Griliches, Z. (1990). Patent Statistics as Economic Indicators: A Survey. *Journal of Economic Literature* 28 (4): 1661–1707.

World Intellectual Property Organization. (2023). Guide to the International Patent Classification (IPC 2022 edition). World Intellectual Property Organization. Available at: https://www.wipo.int/edocs/pubdocs/en/wipo-guide-ipc-2022-en-guide-to-the-international-patent-classification-2022.pdf

Ivanova, V., & Lambrix, P. (2013). A unified approach for aligning taxonomies and debugging taxonomies and their alignments. In The Semantic Web: Semantics and Big Data: 10th International Conference, ESWC 2013, Montpellier, France, May 26-30, 2013. Proceedings 10 (pp. 1-15). Springer Berlin Heidelberg.

Johnson, D. K. (2002). The OECD Technology Concordance (OTC): Patents by industry of manufacture and sector of use.

Kelly, B., Papanikolaou, D., Seru, A., & Taddy, M. (2021). Measuring technological innovation over the long run. *American Economic Review: Insights*, *3*(3), 303-320.

Kortum, S., & Putnam, J. (1997). Assigning patents to industries: tests of the Yale Technology Concordance. Economic Systems Research, 9(2), 161-176.

Lafond, F., & Kim, D. (2019). Long-run dynamics of the US patent classification system. *Journal of Evolutionary Economics*, *29*(2), 631-664.

Lee, H. (2018, October). Research on the impact of technology taxonomy for the tracking of technology convergence. In *2018 International Conference on Information and Communication Technology Convergence (ICTC)* (pp. 1452-1456). IEEE.

Leydesdorff, L. (2008). Patent classifications as indicators of intellectual organization. *Journal of the American Society for Information Science and Technology*, *59*(10), 1582-1597.

Lobo, J., & Strumsky, D. (2019). Sources of inventive novelty: two patent classification schemas, same story. *Scientometrics*, *120*(1), 19-37.

Lybbert, T. J., & Zolas, N. J. (2014). Getting patents and economic data to speak to each other: An 'algorithmic links with probabilities' approach for joint analyses of patenting and economic activity. *Research Policy*, 43(3), 530-542.

Neuhäusler, P., Frietsch, R., & Kroll, H. (2019). Probabilistic concordance schemes for the re-assignment of patents to economic sectors and scientific publications to technology fields (No. 60). Fraunhofer ISI Discussion Papers-Innovation Systems and Policy Analysis.

Pasek, J. E. (2021). Trends in bioengineering patents granted. *Biomedical Sciences Instrumentation*, 57(2), 61-73.

Neuhäusler, P., Frietsch, R., Mund, C., & Eckl, V. (2017). Identifying the technology profiles of R&D performing firms—a matching of R&D and patent data. *International Journal of Innovation and Technology Management*, *14*(01), 1740003. Trends in bioengineering patents granted. Biomedical Sciences Instrumentation, 57(2), 61-73

Office of Technology Assessment and Forecast. (1974). OTAF Concordance between U.S. Patent Classification and Standard Industrial Classification Systems.

Pfeffer M. (2016). Automatic creation of mappings between classification systems for bibliographic data. In Proceedings of the 2016 International Conference on Dublin Core and Metadata Applications (DCMI'16). Dublin Core Metadata Initiative, 75–84.

Putnam, J., & Evenson, R. E. (1994). Inter-sectoral technology flows: Estimates from a patent concordance with an application to Italy. Mimeograph, Yale University, New Haven, CT.

Schmoch, U. (2008). Concept of a Technology Classification for Country Comparisons. Final Report to the World Intellectual Property Office (WIPO), Karlsruhe: Fraunhofer ISI.

Schmoch, U., Laville, F., Patel, P., & Frietsch, R. (2003). Linking technology areas to industrial sectors. Final Report to the European Commission, DG Research.

van Looy, B., Vereyen, C., & Schmoch, U. (2014). Patent statistics: Concordance IPC V8–NACE rev. 2. Eurostat, Euopean Commission.

Verspagen, B., Moergastel, T. V., & Slabbers, M. (1994). MERIT concordance table: IPC-ISIC (rev. 2).

Verspagen, B. (1997). Measuring intersectoral technology spillovers: estimates from the European and US patent office databases. *Economic Systems Research*, 9(1), 47-65.

Wikipedia contributors. (2024, April 26). Industry classification. In Wikipedia, The Free Encyclopedia. Retrieved 08:29, January 21, 2025, from https://en.wikipedia.org/w/index.php?title=Industry_classification&oldid=1220947550

Wong, C. Y., & Fung, H. N. (2017). Science-technology-industry correlative indicators for policy targeting on emerging technologies: exploring the core competencies and promising industries of aspirant economies. *Scientometrics*, 111, 841-867.