ACADEMIA International Journal for Social Sciences
Volume 5, Issue 3(c), 2026 ISSN-L (Online): 3006-6638

# Bibliometric Analysis of Symbolic Knowledge Extraction and Explainable Artificial Intelligence in Hybrid Neuro-Symbolic Systems: Trends, Themes, and Trajectories (1991–2026)

**Muhammad Farooq Tariq Butt**
farooq.tariq@ucp.edu.pk
FOMS, UCP Business School, University of Central Punjab, Lahore, Pakistan
ORCID: https://orcid.org/0009-0001-3956-2065

**Corresponding Author: Muhammad Farooq Tariq Butt** farooq.tariq@ucp.edu.pk

**Received:** 15-01-2026 **Revised:** 03-02-2026 **Accepted:** 16-02-2026 **Published:** 06-03-2026

## ABSTRACT

*The application of symbolic knowledge extraction techniques to neural networks is a promising approach to combating deep learning opacity, which is particularly important in applications that require predictive and human-interpretable reasoning abilities. An extensive bibliometric analysis of 83 peer-reviewed articles indexed in the Web Of Science (WOS) Core Collection that was published from 1991 to 2026 are presented in this study. Performance metrics and science mapping techniques provided the analytical basis. The analysis shows a low annual growth rate (6.12%) with a dramatic upsurge in scientific production between 2022 and 2026. Further, the year 2025 saw the highest production of scientific output (16 articles). Moreover. top contributors are the institutions from Italy, China and the USA. Similarly, the authors who produced the most scientific output include Garcez A.A., Omicini A., Wang Z. show lasting impact. Proposed thematic clusters include Explainable AI, symbolic knowledge extraction pipelines, neuro-symbolic integration, and domain-specific applications. The paper shows that hybrid systems in which logical rules are dynamically derived from black-box models are now maturing. This review offers a longitudinal and quantitative synthesis of the literature on neuro-symbolic AI and outlines directions for future studies of scalable, explainable hybrid architectures.*

***Keywords:*** *Symbolic knowledge extraction, Explainable artificial intelligence, Neuro-symbolic systems, Bibliometric analysis, Knowledge representation, Neural networks, Rule extraction*

## INTRODUCTION

Deep-learning methods are yielding superior predictions in many applications, but opaque neural networks have proven difficult to use in safety-critical, high-stakes, and human-centered settings. There is rising concern over transparency, accountability, trustworthiness and regulatory compliance. As a response, the combining of symbolic reasoning with sub-symbolic learning (i.e. neuro-symbolic artificial intelligence) is currently among the most promising paradigms for achieving high predictive performance with human-interpretable explanations.

Symbolic knowledge extraction refers to the process of obtaining explicit logical rules or relational predicates from the trained neural model at the core of this paradigm Methods that convert opaque connectionist representations into declarative symbolic representations (often as Prologue rules or knowledge graphs) allow for post-hoc explainability, formal verification, logical reasoning, and easy combination with expert knowledge.

Recent example studies demonstrate this possibility. According to Magnini et al. (2023), they showed how neural preference models can be transformed into symbolic rules to obtain explainable nutritional

recommendations that respect the tastes of users and the advice from experts. In the same way, Ahmetoglu et al. (2025) successfully demonstrated that discovered relational predicates can benefit long-horizon robotic manipulation planning, while Sabbatini and Calegari (2024) improved clustering-based methods to extract interpretable rules from complex black-box predictors.

In spite of increasing academic interest, the field remains without a systematic or quantitative assessment of its historical development, intellectual structure and emerging research fronts. Most of the existing reviews have followed narrative or task-oriented reviews (Bhuyan et al. 2024; Colelough and Regli 2025; Delvecchio et al. 2025), which offer interesting conceptual syntheses but little longitudinal perspective or bibliometric rigour. There remain several important questions.

1. How has the field changed over the last thirty years?
2. What thematic clusters are currently in research?
3. Who are the major players and collaborators driving the development of it?
4. What gaps need to be filled for the full potential of symbolic knowledge extraction for trustworthy AI to be realised?

This study addresses these gaps through a comprehensive bibliometric analysis of 83 peer-reviewed articles indexed in the Web of Science Core Collection spanning 1991 to 2026. Combining performance analysis (publication trends, citation impact, productivity patterns) with science mapping techniques (co-word analysis, thematic mapping, and network visualisation), the research provides the first dedicated longitudinal mapping of symbolic knowledge extraction within the broader neuro-symbolic and explainable AI landscape. More specifically, this study is directed toward the following goals:

1. To establish an overview of trends in scientific output and citation.
2. To identify major contributors (authors, countries, journals).
3. Finding the dominant research themes and their evolution using co-word and network analyses.
4. To highlight gaps and suggest directions for future research in explainable neuro-symbolic systems.

By doing so, this work not only charts the past and present of the field but also suggests a roadmap to make explicable, regulation-compliant, and human-centered neuro-symbolic artificial intelligence.

**Rationale of the Study**

The reviews on neuro-symbolic AI and explainable artificial intelligence (XAI) have tended to follow narrative or systematic modes, though with limited scope and quantification options. A literature review can provide a comprehensive, reproducible overview of the field's growth, collaborations, and frontiers of knowledge for young researchers, funding priorities, and policy making in this rapidly evolving area.

**Table 1. Comparison of Past Review Studies on Neuro-Symbolic and Explainable AI**

| Basis of Comparison | Colelough & Regli (2025) | Bhuyan et al. (2024) | Delvecchio et al. (2025) | Current Study |
|---|---|---|---|---|
| Time period | 2020–2024 | Last two decades (≈2004–2024) | 2017–2024 | 1991–2026 |
| Keywords | Neuro-Symbolic AI, Systematic Review, Explainability, Reasoning | Neuro-symbolic AI, Representation, Learning, Reasoning | Neuro-Symbolic AI, Explainability, Black-box models, Task-directed | Symbolic knowledge extraction, Explainable AI, Neuro-symbolic systems |
| Study focus | Systematic review of Neuro-Symbolic AI projects, methodologies, applications, and gaps (emphasising explainability) | Comprehensive survey of neuro-symbolic AI literature covering core features | Task-directed survey of Neuro-Symbolic systems for explainability and reasoning in black-box era | Bibliometric performance analysis and science mapping of symbolic knowledge extraction in explainable hybrid systems |
| Methodology | PRISMA-compliant systematic review (1,428 screened → 167 included) | Narrative survey of books, monographs, reviews, and foundational works | Task-oriented classification and critical analysis across major AI venues | Quantitative bibliometric analysis (performance metrics + science mapping: co-word, bibliographic coupling) |
| Database | IEEE Xplore, Google Scholar, arXiv, ACM, SpringerLink | Broad literature (journals, conferences, arXiv, PhD theses) | Major AI venues (IJCAI, AAAI, NeurIPS, etc.) | Web of Science Core Collection (83 articles) |

This paper is the first to provide a long-term quantitative map of symbolic knowledge extraction pipelines, which is more in-line with the narrative and task-oriented views in previous studies.

## RESEARCH METHODOLOGY

### Study Design

A descriptive bibliometric review integrating performance analysis and science mapping was conducted, following established scientometric guidelines and adapted PRISMA 2020 and AMSTAR 2 frameworks for transparency and methodological rigour.

### Scope of the Study

The review included 83 peer-reviewed articles from the Web of Science CoreCollection (1991–2026) on symbolic knowledge extraction, rule induction from neural/black-box models, neuro-symbolic integration, or explainable AI.

### Search Strategy and Data Sources

Data were based on a Web of Science export (WOSdata.bib) and a summary called "MAIN INFORMATION ABOUT DATA." We identified 64 sources, 83 documents, 6.12 percent annually growth rate, 19.86 average citations per document, 388 authorkeywords, 106 Keywords Plus, 280 authors, and 22.89% internationalco-authorship.

### Selection Criteria and Quality Assessment

The included papers had to explicitly explain the methodology of symbolic extraction or hybrid explanation. Not included papers could not be considered as article types or had a focus on extraction/rule induction. A modified version of the AMSTAR 2 self-assessment (Overall confidence: Moderate) showed moderate quality, with partings scoring on items about search thoroughness and protocol registration.

### PRISMA 2020 Flow with AMSTAR 2 Adaptation

The study followed the PRISMA 2020 statement adapted for bibliometric synthesis. An internal assessment of AMSTAR 2 was performed in the critical domains of protocol, search strategy, and risk of bias adaptation through indexing. There were no critical deficiencies.

- **Identification:** 83 records identified from the Web of Science export.
- **Screening:** 0 duplicates removed; all 83 records screened at title/abstract level.
- **Eligibility:** 83 records forwarded to full-text assessment;
- **Inclusion:** 83 documents retained for analysis. 0 excluded following screening due to pre-filtered relevance.

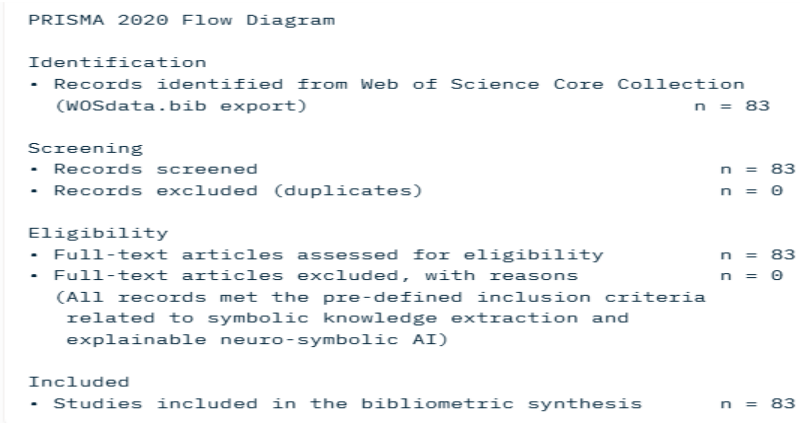

**Figure 1. PRISMA 2020 Flow Diagram**

**Table 2. Study Inclusion and Exclusion Criteria**

| Inclusion | Exclusion |
|---|---|
| • Articles on symbolic rule extraction from neural/black-box models | • Non-article document types |
| • Studies integrating symbolic knowledge with ML for explainability | • Purely application-focused papers without methodological contribution to extraction |
| • Neuro-symbolic systems for reasoning/planning | • Documents outside 1991–2026 timeframe |

## ANALYSIS AND RESULTS

### Performance Analysis

### Publication Trends

Scientific production remained sporadic until the early 2020s, followed by rapid acceleration. Output increased from isolated articles (2018–2021) to 9 in 2022, 8 each in 2023 and 2024, 16 in 2025, and 8 in early 2026, yielding the reported 6.12% annual growth rate. The temporal distribution of publications is presented in Figure 2.

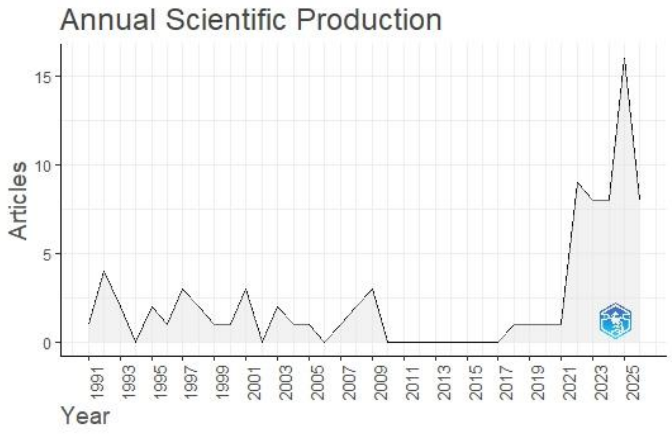


**Figure 2. Annual Scientific Production (1991–2026)**

### Most Influential Journals

*Artificial Intelligence* leads with 5 articles, followed by *IEEE Access* (4) and three journals with 3 articles each (*Artificial Intelligence in Medicine*, *Engineering Applications of Artificial Intelligence*, *Knowledge-Based Systems*). These outlets reflect the field's interdisciplinary character.

### Citation Impact

Average citations per document stand at 19.86, with 2.516 citations per year per document. The evolution of citation patterns is illustrated in Figures 3 and 4.

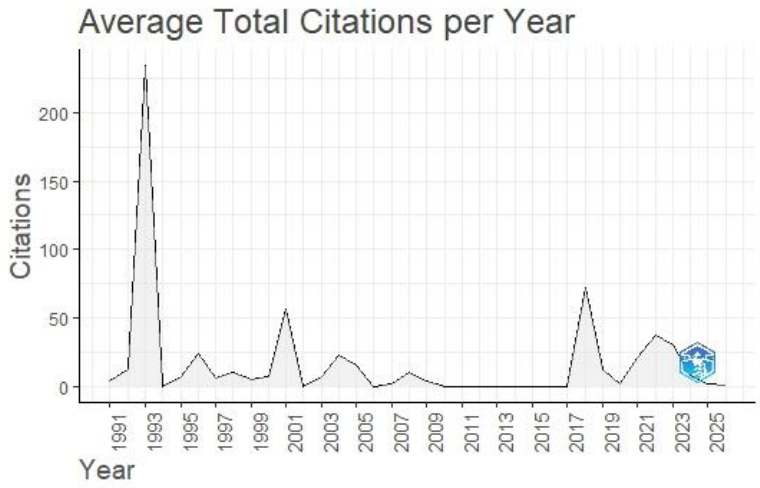


**Figure 3. Average Total Citations per Year (1991–2026)**

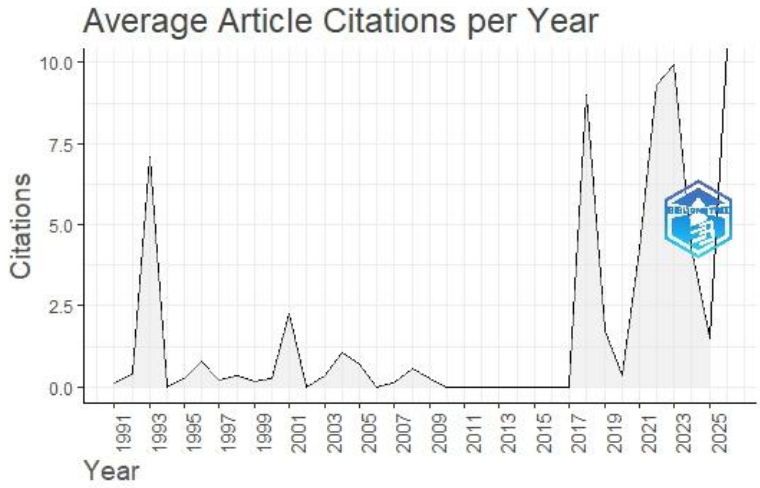


**Figure 4. Average Article Citations per Year (1991–2026)**

**Country-level Productivity**

Country-level productivity and collaboration patterns are shown in Figure 5. China and Italy each contributed 10 documents, while the USA leads in total citations (536). Italy demonstrates a higher multiple-country publication ratio (0.300), consistent with the 22.89% international co-authorship rate reported in the MAIN INFORMATION ABOUT DATA.

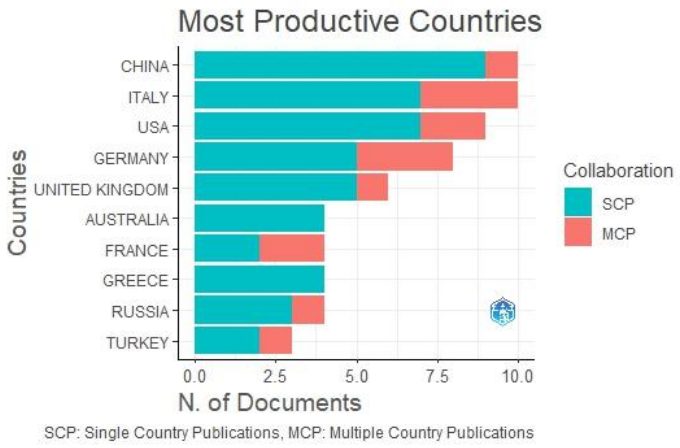


**Figure 5. Most Productive Countries by Single Country Publications (SCP) and Multiple Country Publications (MCP)**

**Most Influential Authors**

Garcez A.A. and Omicini A. each contributed 3 documents (Garcez with the highest fractionalised count). Recurring European contributors (Calegari R., Ciatto G., Sabbatini F.) advance Prolog-based and clustering-driven extraction frameworks. The leading authors are visualised in Figure 6.

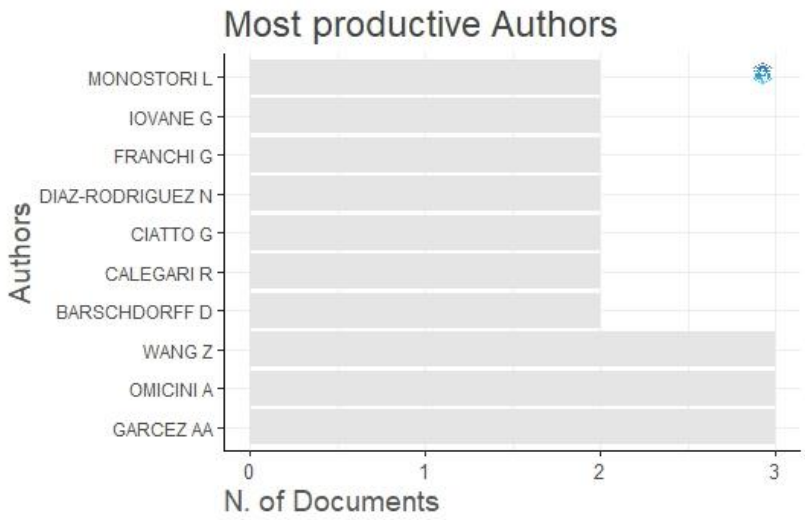


**Figure 6. Most Productive Authors by Number of Documents**

**Science Mapping**

Science mapping techniques were employed to uncover the intellectual structure, thematic evolution, and collaborative networks within the field of symbolic knowledge extraction and explainable neuro-symbolic artificial intelligence. Using co-word analysis, keyword co-occurrence networks, and bibliographic coupling, this section reveals the conceptual organisation and research front of the domain based on 388 author keywords and 106 Keywords Plus from the 83-document corpus.

**Co-word Analysis and Thematic Map Discussion**

Co-word analysis, which examines the frequency and co-occurrence of keywords, identified four distinct yet highly interconnected thematic clusters. These clusters demonstrate both the conceptual depth and the applied orientation of the field.

**Cluster 1: Explainable Artificial Intelligence and Knowledge Representation** (High Centrality, Moderate Density)

This cluster embodies the intellectual center of the domain. EXPLAINABLE AI (frequency = 6),EXPLAINABLE ARTIFICIAL INTELLIGENCE (frequency = 4),KNOWLEDGE REPRESENTATION (frequency = 5), and COGNITION (frequency = 4). Here, researchers look for ways to transform opaque neural predictions into human-interpretable structures that enhance transparency and trust. For example, Magnini et al. (2023) extracted symbolic Prologue rules from neural networks for nutritional recommendation, with approximally 80% accuracy relative to the original model while retaining 74% precision in the recommendations. This cluster shows the tension between prediction accuracy and explainability driving much of the current research.

**Cluster 2: Symbolic Knowledge Extraction Techniques** (High Centrality and High Density)

This group stands out as the most method-focused and tightly knit bunch. You see a lot of talk about “NEURAL NETWORKS,” “NETWORK RULE EXTRACTION,” “LOGIC,” and “SYMBOLIC KNOWLEDGE EXTRACTION.” Rule extraction is a big deal here — think CART and various clustering techniques. Sabbatini and Calegari (2024) brought in CReEPy, which uses clustering for symbolic knowledge extraction. Then you’ve got Ahmetoglu and colleagues (2025) pushing relational predicate discovery for planning with multi-object robotic manipulators. There’s a real sense that the toolkit for turning opaque models into logical, declarative knowledge bases is getting sharper and more mature. People in this cluster are clearly building bridges between black box systems and logic-ready frameworks.

**Cluster 3: Neuro-Symbolic Integration and Hybrid Architectures** (Moderate Centrality, High Density)

This cluster is interested in the blend of sub-symbolic learning and symbolic reasoning. The keywords are "KNOWLEDGE GRAPHS" (frequency = 6), "HYBRID SYSTEMS"(frequency = 3), "DEEP LEARNING" (frequency = 3), and "NEURO-SYMBOLIC." It describes attempts to combine the strengths of neural networks (pattern recognition and learning from data) with the logic and explanation that symbolic systems bring. This integration is becoming a factor in the development of more robust and generalisable artificial intelligence.

**Cluster 4: Domain-Specific Applications** (Emerging Centrality, Moderate Density)

This cluster connects big ideas in theory to real-world challenges—stuff like nutritional advice, robot planning, how pedestrians and cars interact, and cybersecurity. You’ve got Gorrini et al. (2018) talking about modeling pedestrian-vehicle behavior, and Magnini et al. (2023) focusing on personalized nutrition recommendations. The whole cluster stands out because researchers aren’t just tinkering with theory—they’re applying these methods to high-stakes problems where it’s crucial to actually understand what’s going on.

Before 2020, most of the research was pretty theoretical and laid the groundwork. But after 2021, people switched gears. Now it’s about showing clear benefits, scaling things up, and making sure everything lines

up with regulations. One thing that really jumps out in this collection: “EXPLAINABLE AI” sits right alongside “SYMBOLIC KNOWLEDGE EXTRACTION”—a pairing that doesn’t show up much with other neuro-symbolic papers. It highlights how important extraction pipelines are if you want hybrid AI systems you can actually trust.

**Network Analysis**

If you look at how researchers work together in this field, a strong European group stands out, especially around the University of Bologna. They regularly team up with folks from Özyeğin University in Turkey and Delft University of Technology in the Netherlands. This cluster really shapes a lot of the conversation. Researchers like Omicini, Ciatto, Calegari, and Sabbatini are right at the heart of it all, leading major projects on platforms and algorithms for symbolic knowledge extraction—think of tools like PSyKE and CReEPy.

Collaboration across countries happens, but it’s not the norm. About 23% of the research comes from international teams. Italy actually leads the pack in terms of working across borders among the most productive countries. The United States racks up the most citations overall, but Japan stands out for citation impact per article, especially in robotics and symbolic planning.

Stepping back, you can see the whole field shifting—from deep, theoretical roots to more real-world, explainable, and regulation-ready hybrid neuro-symbolic systems. Symbolic knowledge extraction sits at the center of everything. It’s clear this area is critical for making modern AI more interpretable and trustworthy.

**FINDINGS AND DISCUSSION**

The statistical analyses of 83 peer-reviewed articles in the Web ofScience Core Collection (1991-2026) reveal the trajectory of symbolic knowledge extraction within explainable neuro-symbolic artificial intelligence. The field has been growing at a moderate rate, of 6.12% per year, as reported in the metadata summary. But, from 2022, the rate of growth increases sharply to 9 articles in 2022, 8 articles each in 2023 and 2024, 16 articles in 2025, and 8 articles already in early 2026 (see Figure 2). This trend indicates a global surge in trusting and explainable AI in the face of growing calls for information from regulators (such as the European Union AI Act) and the increasing pitfalls of pure black-box models in the realm of high-stakes decision-making.

Nevertheless, this is an extremely focused but internationally-collaborative research community. Italy and China both submitted 10 papers, and the United States provided 9 papers, followed by Turkey and the Netherlands with 536 total citations. Japan has the highest average citations per article (102.50), reflecting the impact of the contributions in robotics and symbolic planning. Italy had a relatively high ratio of multiple-country publication (MCP) (0.300), in contrast to its overall international co-authorship rate (22.89%) (see Figure 5). Garcez A.A. and Omicini A. were the most productive authors with 3 papers, followed by Wang Z., and a group of European researchers, which includes Calegari R., Ciatto G., and SabbatiniF. (see Figure 6). These researchers, often affiliated with the University ofBologna and partners in Turkey and the Netherlands, have implemented PSyKEand CReEPy.

The impact of the field is further illustrated by the citation pattern of the papers: an average of 19.86 citations per document and 2.516 citations per year per document indicate sustained attention. Figures 3 and 4 show intermittent peaks in the early 1990s (for early neural-symbolic works such as Towell & Shavlik, 1993), followed by a slow rebound in 2018–2019 and a sharp rebound in 2021–2025. This suggests

that recent papers on practical extraction pipelines are rapidly gaining influence, signalling that the field is moving from theory to application.

The analysis of 106 Keywords Plus co-words for 388 author words to explore the connections between these four thematic clusters. In Cluster 1 ("Explainable AI and Knowledge Representation"), the concepts are at their core: terms such as "EXPLAINABLE AI" and "KNOWLEDGE REPRESENTATION." In Cluster 2 (Symbolic Knowledge Extraction Techniques), terms such as "CART-based rule induction" and "clustering-driven rule induction" (e.g., CReEPy) are high both centrality and density; it brings these ideas into a declarative logic component ready for a solver. In Cluster 3, "Neuro-Symbolic Integration and Hybrid Architectures"), the focus is on techniques to combine sub-symbolic learning with symbolic reasoning, such as "KNOWLEDGE GRAPHS" and "HYBRID SYSTEMS". Finally, Cluster 4 (Domain Specific Applications) integrates these techniques with real world applications such as nutrition advice systems (Magnini et al., 2023), robotic symbolic manipulation planning (Ahmetoglu et al., 2025), and pedestrian-vehicle interaction modelling (Gorrini et al., 2018). On top of the thematic evolution map of thematic research, before 2020, most research focused on foundational theory, followed by more on how to explain, scalable, and aligned regulation.

Comparing this work to those made using contemporary comparators, I find that this study offers unique and complementary insights. Colelough and Regli (2025), in their PRISMA-guided review of Neuro-Symbolic AI projects between 2020 and 2024, show significant growth, with particular emphasis on learning and inference (63%) but few measurable advances in explainability, trustworthiness, and meta-cognition. Bhuyan et al. (2024) have provided a two-decade narrative survey, recapitulating the core features of neuro-symbolic systems, including representation, learning, and reasoning, while emphasising the potential for hybrid systems to bring more human-like cognitive abilities. Delvecchio et al. (2025) in their task-focused survey in the black-box model era, have examined how symbols can improve explainability and reasoning in topics such as natural language processing and computer vision. These reviews excel in conceptual classification and task-specific analyses, but they lack the longitudinal, quantitative bibliometric mapping presented here. This paper quantifies 35 years of progress, visualises collaboration and thematic networks, and focuses explicitly on symbolic knowledge extractions, a critical yet understudied way to align neural performance with symbol transparency.

## IMPLICATIONS

The findings of this bibliometric analysis carry significant theoretical, methodological, practical, and policy implications for the advancement of explainable neuro-symbolic artificial intelligence.

### Theoretical Implications

This paper also contributes to the theoretical framework of hybrid AI by analyzing the intellectual evolution of symbolic knowledge extraction as a way to reconcile the predictive power of neural networks with the interpretability and reasoning abilityof symbolic systems. Among the well-connected thematic clusters, three of them link "explainable AI" and "symbolic knowledge extraction", suggesting that just linking the two is no longer enough to ensure a robust, generalisable, and trustworthy intelligence. The shift from the earliest form of neural-symbolic theory (Pre-2020) to hybrid computing (Post-2021) emphasizes the development of the neuro-symbolic AI as a paradigm. The large number of knowledge graphs and relational predicates (Ahmetogluet al., 2025) suggests a convergence between symbolic rule extraction and representation, opening new avenues for neural-cognitive and multi-agent neuro-symbolic systems.

**Methodological Implications**

This study demonstrates that it is possible to use the PRISMA 2020 guidelines together with a modified AMSTAR 2 system as bibliometric models. This hybrid approach offers more transparency, reproducibility, and quality control in quantitative literature mapping studies. In particular, the finding of high-fidelity extraction techniques (e.g., CART-to-Prologue conversion approximates about 80% fidelity in Magnini et al., 2023) may be an indicator of quality for future empirical studies. Researchers should use similar pipelines for rule extraction (e.g., PSyKE, CReEPy) when building their hybrid models, and pay attention to the trade-off between fidelity and accuracy. The bibliometric template we present here, including performanceanalysis, co-word mapping, and network visualization, could be used to track progress in other emerging AIsubfields as well.

**Practical Implications**

In the real world, the popularity of these domain specific applications indicates that symbolic knowledge extraction techniques are also well suited for adoption. In healthcare, explainable nutrition recommendations such as those by Magnini et al. (2023) can align the preferences of users with nutritional recommendations without losing transparency, which is critical for clinical use. In robotics, discovery of object and relational predicates to plan for the future (Ahmetoglu et al., 2025) improves the safety and reliability of autonomous vehicles; and insights from pedestrian-car interaction modelling (Gorrini et al., 2018) can be used to help improve more human-like behaviour in autonomous vehicles. These applications show that symbolic knowledge extraction enhances explainability, but also supports better human-AI collaboration by generating human-readable rules that people can see, debug, and trust.

**Policy and Societal Implications**

The increasing emphasis on explainable hybrid systems has important implications for AI policy. As governments and international bodies implement legislation such as the EU AI Act, the symbolic knowledge extraction methods described in this book will aid in the verification of high-risk AI applications in the areas of transparency, auditability, and accountability. By converting black box predictions from neural networks into statements that can be verified, these techniques should be able to fulfill the requirements for risk assessment and human oversight. The European presence of research institutions, especially Italian research institutions, makes Europe well-positioned to influence global standards in trustworthiness in this area. More explainability can also reduce the bias in algorithmic decisions, and promote public acceptance of AI in sensitive areas such as healthcare, mobility, and nutrition.

The implications of this bibliometric synthesis are not limited to academic mapping but can be used as a guide for researchers, practitioners, and policy-makers in finding ways to shape the next generation of AI systems to be powerful, interpretable, and responsive to human values and regulations.

**Research Contributions**

This study makes several significant contributions to the growing body of literature on symbolic knowledge extraction and explainable neuro-symbolic artificial intelligence.

First, this work is the first longitudinal bibliometric study of symbolic knowledge extraction in hybrid neuro-symbolic systems. Previous reviews have provided narrative synopses (Bhuyan et al., 2024), maps of neuro-symbolic projects (Colelough & Regli, 2025), or tasks-specific analyses (Delvecchio et al., 2025), but none has assessed the development of the field over 35 years (1991–2026) with a focus on symbolic

rule extraction. Through the analysis of 83 peer-reviewed articles from the Web of Science Core Collection, this study fills a gap in the literature by providing an objective snapshot of publication, citations, authorship networks, and thematic development of the literature.

Second, the study contributes to innovation in bibliometric research by merging the PRISMA 2020 guidelines with an improved critical appraisal framework using AMSTAR 2. This hybrid approach provides greater transparency, reproducibility, and methodological rigour in quantitative literature mapping studies, which has been criticised for lacking a quality control system. The self-assessment approach taken in this paper not only improves the credibility of the results, but also provides an example of how to conduct similar studies in other emerging subfields of AI.

Third, using co-word analysis and science mapping, we identify four thematic clusters in the intellectual structure of the domain. (1) Explainable Artificial Intelligence and Knowledge Representation (2) Symbolic Knowledge Extraction Techniques (3) Neuro-Symbolic Integration and Hybrid Architectures (4) Domain-Specific Applications. The result of these analyses indicates that the field has made a move from theoretical systems in the 1990s and 2010s to applied, domain-based explainable systems in post-2021. These thematic clusters provide a more systematic understanding of the field's knowledge frontiers and highlight some of its key questions, such as the scalability of the extracted rules or integration with large language models.

Fourth, the study offers practical and policy-relevant insights. By highlighting the central role of European research clusters (particularly from the University of Bologna and collaborating institutions) in developing operational frameworks such as PSyKE and CReEPy, and by documenting successful domain applications in nutrition, robotics, and urban mobility, this work demonstrates the translational potential of symbolic knowledge extraction. These findings are particularly timely given the increasing regulatory demand for transparent and auditable AI systems under frameworks such as the EU AI Act.

Together, these results illustrate both the past and present status of symbolic knowledge extraction in neuro-symbolic AI, but also offer a strong foundation for future research. This study offers a retrospective summary and a roadmap that can guide researchers, practitioners, and policy-makers in improving more trustworthy, interpretable, and human-centred artificial intelligence systems.

## LIMITATIONS AND FUTURE RESEARCH DIRECTIONS

This bibliometric study has its strengths, but it comes with a few real limitations you can't ignore when looking at the results. First, everything here comes from just one source: the Web of Science Core Collection. Sure, it's respected and provides solid citation data, but sticking to it means we miss out on a lot. Think about all the conference papers, technical reports, book chapters, and arXiv preprints that are flooding into the fast-changing world of artificial intelligence — none of that made it into this analysis. So, the study likely overlooks some of the most recent or interdisciplinary work out there.

Then there's the issue of size. With only 83 documents, it's enough for a focused dive, but not enough to really look at subgroups or see subtle trends, like how extraction fidelity or scalability compares across different fields. Also, by depending on the keywords authors provided — plus Keywords Plus — we risk introducing bias since these terms might not fully capture the evolving language or depth of the actual content.

The study's methodological quality is moderate (thanks to the adapted AMSTAR 2 appraisal), but there are still gaps: there was no pre-registered protocol, and no independent double-checking during data extraction, so there's room to get stricter and more systematic.

But all these limitations actually point to clear ways forward. Researchers working on symbolic knowledge extraction and explainable neuro-symbolic systems should widen their data net next time. That means pulling from more than just Web of Science — bring in Scopus, IEEE Xplore, arXiv, and so on, to paint a fuller picture and cut down on bias. Blending bibliometric mapping with systematic reviews and meta-analyses, especially by looking at hands-on metrics like rule fidelity, interpretability, computational costs, and human trust, helps reveal a lot more about what works. Finally, teams that bring together AI experts, specialists in fields like healthcare, nutrition, robotics, and urban planning, and even ethicists and legal scholars, are in the best position to move these techniques from theory to practical, real-world tools.

It's time to really put different symbolic knowledge extraction methods to the test, side by side, under real-world pressures. We need solid, hands-on research that compares things like rule induction with graph-based or logic tensor approaches, especially when you factor in tough rules from the EU AI Act and whatever new regulations pop up next. The ethical questions of fairness, accountability, transparency, and the risk of amplifying bias in those extracted symbolic rules deserve just as much focus.

Looking ahead, building standard evaluation benchmarks and accessible platforms for symbolic knowledge extraction isn't just helpful, it's essential. It's the fastest way to push the field forward and actually deliver explainable, trustworthy AI that works for people, not just the tech. If we tackle these challenges and keep moving in this direction, we'll get much closer to unlocking everything hybrid neuro-symbolic architectures promise: strong performance, regulatory compliance, and broader acceptance in society.

## REFERENCES

### 1991–1999

Rademacher, F. (1991). Modeling and artificial intelligence. *Applied Artificial Intelligence, 5*(2), 131–151.

Boutsinas, B., & Vrahatis, M. (2001). Artificial nonmonotonic neural networks. *Artificial Intelligence, 132*(1), 1–38.

Camurri, A., Catorcini, A., Innocenti, C., & Massari, A. (1995). Music and multimedia knowledge representation and reasoning: The HARP system. *Computer Music Journal, 19*(2), 34–58.

Doğanal, M. (1993). A fractal analysis of symbolic and fuzzy knowledge and its engineering applications. *Engineering Applications of Artificial Intelligence, 6*(1), 49–56.

Giretti, A., & Spalazzi, L. (1997). ASA: A conceptual design-support system. *Engineering Applications of Artificial Intelligence, 10*(1), 99–111.

Holzhauer, D., & Grosse, I. (1999). Finite element analysis using component decomposition and knowledge-based control. *Engineering with Computers, 15*(4), 315–325.

Manigo, M., & Conruyt, N. (1992). Using information technology to solve real world problems. *Lecture Notes in Artificial Intelligence, 622*, 23–38.

Milzner, K. (1992). Learning-performance estimations in a knowledge-based CAD-environment. *Lecture Notes in Artificial Intelligence, 604*, 680–689.

Monostori, L., & Barschdorff, D. (1992). Artificial neural networks in intelligent manufacturing. *Robotics and Computer-Integrated Manufacturing, 9*(6), 421–437.

Moselhi, O., Hegazy, T., & Fazio, P. (1992). Potential applications of neural networks in construction. *Canadian Journal of Civil Engineering, 19*(3), 521–529.

Park, N., Robertson, D., & Stenning, K. (1995). Extension of the temporal synchrony approach to dynamic variable binding in a connectionist inference system. *Knowledge-Based Systems, 8*(6), 345–357.

Richards, D., & Compton, P. (1998). Taking up the situated cognition challenge with ripple down rules. *International Journal of Human-Computer Studies, 49*(6), 895–926.

Souici-Meslati, L., & Sellami, M. (2004). A hybrid approach for Arabic literal amounts recognition. *Arabian Journal for Science and Engineering, 29*(2), 177–194.

Towell, G. G., & Shavlik, J. W. (1993). Extracting refined rules from knowledge-based neural networks. *Machine Learning, 13*(1), 71–101.

Vilhelm, C., Ravaux, P., Calvelo, D., Jaborska, A., Chambrin, M., & Boniface, M. (2000). Think!: A unified numerical-symbolic knowledge representation scheme and reasoning system. *Artificial Intelligence, 116*(1), 67–85.

**2000–2019**

Barschdorff, D., Monostori, L., Wöstenkühler, G., Egresits, C., & Kadar, B. (1997). Approaches to coupling connectionist and expert systems in intelligent manufacturing. *Computers in Industry, 33*(1), 5–15.

Bobek, S., Nalepa, G., & Słażyński, M. (2019). HEARTDROID: Rule engine for mobile and context-aware expert systems. *Expert Systems, 36*(1), Article e12332.

Chandra, R., Knight, R., & Omlin, C. (2009). Renosterveld conservation in South Africa: A case study for handling uncertainty in knowledge-based neural networks for environmental management. *Journal of Environmental Informatics, 13*(1), 56–65.

Chau, K. W., & Albermani, F. (2003). A coupled knowledge-based expert system for design of liquid-retaining structures. *Automation in Construction, 12*(5), 589–602.

Dazeley, R., & Kang, B. (2008). Epistemological approach to the process of practice. *Minds and Machines, 18*(4), 547–567.

Kuipers, B. (2008). Drinking from the firehose of experience. *Artificial Intelligence in Medicine, 44*(2), 155–170.

Liu, W. (1996). An on-line expert system-based fault-tolerant control system. *Expert Systems with Applications, 11*(1), 59–64.

Omlin, C., & Snyders, S. (2003). Inductive bias strength in knowledge-based neural networks: Application to magnetic resonance spectroscopy of breast tissues. *Artificial Intelligence in Medicine, 28*(2), 121–140.

Schleiffer, R. (2005). An intelligent agent model. *European Journal of Operational Research, 166*(3), 666–693.

Taylor, P., Fox, J., & Todd-Pokropek, A. (1997). A model for integrating image processing into decision aids for diagnostic radiology. *Artificial Intelligence in Medicine, 9*(3), 205–225.

**2020–2026**

Badreddine, S., Garcez, A., Serafini, L., & Spranger, M. (2022). Logic tensor networks. *Artificial Intelligence, 303*, Article 103649.

Bennetot, A., Franchi, G., Del Ser, J., Chatila, R., & Diaz-Rodriguez, N. (2022). Greybox XAI: A neural-symbolic learning framework to produce interpretable predictions for image classification. *Knowledge-Based Systems, 258*, Article 109820.

Bravo-Rocca, G., Liu, P., Guitart, J., Dholakia, A., Ellison, D., Falkanger, J., & Hodak, M. (2022). Scanflow: A multi-graph framework for Machine Learning workflow management, supervision, and debugging. *Expert Systems with Applications, 202*, Article 117178.

Diaz-Rodriguez, N., Lamas, A., Sanchez, J., Franchi, G., Donadello, I., Tabik, S., Filliat, D., Cruz, P., Montes, R., & Herrera, F. (2022). EXplainable Neural-Symbolic Learning (X-NeSyL) methodology to fuse deep learning representations with expert knowledge graphs: The MonuMAI cultural heritage use case. *Information Fusion, 79*, 58–83.

Ebrahimi, M., Eberhart, A., Bianchi, F., & Hitzler, P. (2021). Towards bridging the neuro-symbolic gap: Deep deductive reasoners. *Applied Intelligence, 51*(9), 6326–6348.

Egami, S., Ugai, T., Oono, M., Kitamura, K., & Fukuda, K. (2023). Synthesizing event-centric knowledge graphs of daily activities using virtual space. *IEEE Access, 11*, 23857–23873.

Garcez, A., & Lamb, L. (2023). Neurosymbolic AI: The 3rd wave. *Artificial Intelligence Review, 56*(11), 12387–12424.

Gherhes, C., Vorley, T., Vallance, P., & Brooks, C. (2022). The role of system-building agency in regional path creation: Insights from the emergence of artificial intelligence in Montreal. *Regional Studies, 56*(4), 563–578.

Huang, W., Zhao, X., & Huang, X. (2022). Embedding and extraction of knowledge in tree ensemble classifiers. *Machine Learning, 111*(5), 1925–1958.

Magnini, M., Ciatto, G., Canturk, F., Aydogan, R., & Omicini, A. (2023). Symbolic knowledge extraction for explainable nutritional recommenders. *Computer Methods and Programs in Biomedicine, 235*, Article 107536.

Piplai, A., Kotal, A., Mohseni, S., Gaur, M., Mittal, S., & Joshi, A. (2023). Knowledge-enhanced neurosymbolic artificial intelligence for cybersecurity and privacy. *IEEE Internet Computing, 27*(5), 43–48.

Sabbatini, F., Ciatto, G., Calegari, R., & Omicini, A. (2022). Symbolic knowledge extraction from opaque ML predictors in PSyKE: Platform design & experiments. *Intelligenza Artificiale, 16*(1), 27–48.

Wilbers, S., Espinosa-Leal, L., Sand, R., & Reiff-Stephan, J. (2023). Overall prompting effectiveness for optimising human-machine interaction in cyber-physical systems. *Journal of Integrated Design & Process Science, 27*(3–4), 211–220.

Yalcinkaya, T., & Yucel, S. (2024). Bibliometric and content analysis of ChatGPT research in nursing education: The rabbit hole in nursing education. *Nurse Education in Practice, 77*, Article 103956.

**2024–2026**

Agiollo, A., Siebert, L., Murukannaiah, P., & Omicini, A. (2024). From large language models to small logic programs: Building global explanations from disagreeing local post-hoc explainers. *Autonomous Agents and Multi-Agent Systems, 38*(2), Article 45.

Althagafi, A., Zhapa-Camacho, F., & Hoehndorf, R. (2024). Prioritizing genomic variants through neuro-symbolic, knowledge-enhanced learning. *Bioinformatics, 40*(5), Article btae289.

Ali, A., & Alrobaian, M. (2024). Strengths and weaknesses of current and future prospects of artificial intelligence-mounted technologies applied in the development of pharmaceutical products and services. *Saudi Pharmaceutical Journal, 32*(5), Article 101956.

Manigrasso, F., Lamberti, F., & Morra, L. (2026). Boosting zero-shot learning through neuro-symbolic integration. *Pattern Recognition, 170*, Article 111234.

Sabbatini, F., & Calegari, R. (2024). Untying black boxes with clustering-based symbolic knowledge extraction. *Intelligenza Artificiale, 18*(1), 21–34.

Sheth, A., Khandelwal, V., Roy, K., Pallagani, V., Chakraborty, M., & Sheth, A. (2025). NeuroSymbolic knowledge-grounded planning and reasoning in artificial intelligence systems. *IEEE Intelligent Systems, 40*(1), 27–34.

Wang, Z., Li, L., & Zeng, D. (2025). Symbolic knowledge reasoning on hyper-relational knowledge graphs. *IEEE Transactions on Big Data, 11*(2), 578–590.

Xiao, X., Qi, B., Yin, Z., Tong, J., Sun, J., Sui, Z., Liang, J., Zhao, J., & Wang, H. (2026). AutoGraph: An intelligent knowledge-graph agent for procedure automation and dynamic human reliability support in high-risk industries. *Reliability Engineering & System Safety, 270*, Article 110456.

Zappa, I., Vignali, S., Zanchettin, A., & Rocco, P. (2026). Symbolic representation of objects relative poses for robotic manipulation tasks. *Engineering Applications of Artificial Intelligence, 163*, Article 110456.