Computers, Materials & Continua

Tech Science Press

Doi:10.32604/cmc.2025.065149

REVIEW

Research Trends and Networks in Self-Explaining Autonomous Systems: A Bibliometric Study

Oscar Peña-Cáceres[1,2,*], Elvis Garay-Silupu[3], Darwin Aguilar-Chuquizuta[4] and Henry Silva-Marchan[4]

[1]Department of Informatics, Universitat de València, Burjassot, 46100, Spain
[2]Professional School of Industrial Engineering, Universidad Tecnológica del Perú, Lima, 15001, Peru
[3]Department of Informatics, Hospital Especializado San Juan de Dios, Piura, 20001, Peru
[4]Department of Mathematics, Statistics and Informatics, Universidad Nacional de Tumbes, Tumbes, 24000, Peru

*Corresponding Author: Oscar Peña-Cáceres. Email: osjmarpe@alumni.uv.es

Received: 05 March 2025; Accepted: 14 May 2025; Published: 03 July 2025

ABSTRACT: Self-Explaining Autonomous Systems (SEAS) have emerged as a strategic frontier within Artificial Intelligence (AI), responding to growing demands for transparency and interpretability in autonomous decision-making. This study presents a comprehensive bibliometric analysis of SEAS research published between 2020 and February 2025, drawing upon 1380 documents indexed in Scopus. The analysis applies co-citation mapping, keyword co-occurrence, and author collaboration networks using VOSviewer, MASHA, and Python to examine scientific production, intellectual structure, and global collaboration patterns. The results indicate a sustained annual growth rate of 41.38%, with an h-index of 57 and an average of 21.97 citations per document. A normalized citation rate was computed to address temporal bias, enabling balanced evaluation across publication cohorts. Thematic analysis reveals four consolidated research fronts: interpretability in machine learning, explainability in deep neural networks, transparency in generative models, and optimization strategies in autonomous control. Author co-citation analysis identifies four distinct research communities, and keyword evolution shows growing interdisciplinary links with medicine, cybersecurity, and industrial automation. The United States leads in scientific output and citation impact at the geographical level, while countries like India and China show high productivity with varied influence. However, international collaboration remains limited at 7.39%, reflecting a fragmented research landscape. As discussed in this study, SEAS research is expanding rapidly yet remains epistemologically dispersed, with uneven integration of ethical and human-centered perspectives. This work offers a structured and data-driven perspective on SEAS development, highlights key contributors and thematic trends, and outlines critical directions for advancing responsible and transparent autonomous systems.

KEYWORDS: Self-explaining; autonomous systems; explainable AI; machine learning; deep learning; artificial intelligence

1 Introduction

Autonomous systems have evolved rapidly in recent years due to the integration of Artificial Intelligence (AI), Machine Learning (ML), and Explainable Artificial Intelligence (XAI) [1]. These advances have enabled the emergence of self-explanatory autonomous systems, which are designed to enhance transparency and reliability through real-time dynamic explanations that support decision-making processes [2–4]. As their deployment expands into critical domains such as healthcare, transportation, and industrial automation [5],

there is a growing need to examine the scientific progression and research trends surrounding these systems to foster innovation and ensure their effective adoption [6–8].

In this context, a central challenge lies in how autonomous systems communicate and justify their decisions to end users. SEAS aims to meet this challenge by offering user-centered, real-time explanations that enhance interpretability and support meaningful interaction between humans and machines. Beyond their technical capabilities, SEAS is becoming increasingly relevant in light of evolving regulatory frameworks, such as the EU AI Act, which stresses the importance of transparency, auditability, and ethical accountability in algorithmic decision-making [9].

Despite growing academic and industrial interest in SEAS, the research landscape remains highly fragmented. As an inherently interdisciplinary domain that draws from interpretable machine learning, causal inference, human-computer interaction, and knowledge representation, SEAS research is scattered across multiple fields. While meaningful contributions have emerged in application areas such as medicine, cybersecurity, and autonomous transport [10–12], there is still no unified perspective on the field's evolution, the leading actors and institutions, or the emerging thematic priorities. This fragmentation limits the ability to comprehensively map the intellectual development of SEAS and hinders the identification of research gaps.

This lack of consolidation represents a critical gap in the literature. In the absence of a structured, data-driven view of SEAS research, scholars and practitioners risk overlooking influential contributions, failing to spot emerging trends, and missing opportunities for innovation. Thus, we believe a bibliometric approach offers a rigorous framework to address this problem by systematically mapping scientific output, intellectual influence, and collaborative networks. Furthermore, it facilitates the identification of under-explored areas and promotes interdisciplinary integration, especially with fields such as the ethics of artificial intelligence and human-computer interaction [13–15]. Based on these needs, the following section presents the motivation for conducting a bibliometric analysis and examines the main gaps not addressed by previous studies.

1.1 Motivation of the Bibliometric Analysis and Gaps Identified in Previous Studies

To better understand and address the fragmentation identified in the previous section, we conducted a focused review of representative studies related to autonomous systems and explainability. Table 1 provides a comparative analysis of these works, outlining their thematic focus, methodological scope, and our study's contribution in addressing the reported limitations. While these studies have advanced the field in different ways, several limitations remain evident. Most focus on specific application areas, such as autonomous vehicles, unmanned aerial vehicles [16], or maritime systems, while often neglecting explainability as a unifying research dimension across disciplines.

Table 1: Comparative analysis of studies related to autonomous systems and explainability

Ref.	Year	Focus	Limitations	Contribution in response to limitations
[17]	2025	Conceptual classification of explainability methods in autonomous vehicles, organized by explanatory task, type of information, and communication strategy.	The study does not include bibliometric analysis and remains limited to a conceptual taxonomy within the domain of vehicles.	Applies bibliometric techniques to analyze explainability across various types of autonomous systems, incorporating trend evolution and network structures.

(Continued)

Table 1 (continued)

Ref.	Year	Focus	Limitations	Contribution in response to limitations
[18]	2023	Bibliometric mapping of research on autonomous vessels using publication and authorship indicators from Scopus.	The scope is limited to maritime applications and does not incorporate explainability as a research dimension.	Introduces explainability as a variable within the bibliometric mapping of autonomous systems beyond the maritime context.
[19]	2023	Bibliometric mapping of social acceptance in autonomous vehicles, emphasizing collaboration networks and topic evolution.	The study is limited to social acceptance and does not explore explainability or communication between humans and autonomous systems.	Explainability is integrated as a thematic axis in the bibliometric mapping of autonomous systems, evaluating its relationship with the priority given to human-system interaction.
[20]	2022	Bibliometric analysis of autonomous vehicles in mixed traffic, identifying thematic clusters, prolific authors, and key publication sources.	The study focuses solely on traffic-related scenarios and omits aspects related to explainability and human-centered design.	Expands bibliometric analysis by including explainability and cognitive interaction as relevant dimensions in autonomous systems research.
[21]	2019	Scientometric and bibliometric analysis of research trends in autonomous vehicles using CiteSpace and Web of Science data.	The analysis is restricted to vehicular applications and does not consider interdisciplinary components such as explainability or interaction models.	We examine explainability within an interdisciplinary bibliometric framework that includes multiple categories of autonomous systems.

This observation highlights the absence of an integrated, data-driven perspective capable of mapping the intellectual landscape of SEAS research. Hence, several questions arise. How has the dimension of explainability been introduced into autonomous systems? Which countries, publishers, and journals drive the field of study? Which authors and publications exert the greatest influence in shaping the field? And to what extent do patterns of co-citation and keyword co-occurrence reflect the emergence of a cohesive scientific community around SEAS?

The reviewed literature also reveals that few studies adopt a quantitative data-driven approach that allows for an interdisciplinary analysis of the dimension of explainability in autonomous systems. For example, although conceptual analyses such as those by Tekkesinoglu et al. [17] explore key theoretical foundations, they lack an empirical mapping of the scientific development of the field. Similarly, bibliometric studies such as those of Chaal et al. [18] and Ho et al. [19] offer valuable insights within specific domains but do not address explainability as a central focus of research. This fragmentation highlights the need for a broader, more integrative approach.

In response, the present study applies bibliometric techniques to analyze SEAS-related research's evolution, structure, and dynamics. Starting from the identified gaps, the temporal progression of key topics is investigated, the impact of journals and publishers is assessed, influential authors and emerging areas are identified, and keyword co-occurrence and co-citation networks are mapped to better understand the consolidation of SEAS as a scientific domain. These objectives are detailed in the following section.

1.2 Contributions

Building on the gaps identified in previous research, this study provides a bibliometric analysis of the scientific landscape related to SEAS. Using data extracted from the Scopus database, we apply keyword co-occurrence mapping, co-citation analysis, and performance analysis to examine the intellectual structure, publication trends, and conceptual evolution of the field. Bibliometric methods provide a rigorous framework to capture scientific dynamics, identify influential contributions, and delineate the structural underpinnings of an emerging research area [22–24]. Considering the rapid growth and multidisciplinary nature of SEAS, such a systematic investigation is timely and necessary. Accordingly, the specific objectives of this study are as follows:

- To identify the most frequent keywords through co-occurrence analysis, examine their temporal evolution, and explore their association with the countries leading SEAS research.
- To evaluate the scientific output by journals and publishers and assess their impact through citation-based metrics.
- To analyze the historical development of the field by identifying influential publications, prolific authors, and emerging research areas.
- To examine the geographical distribution of the literature and map the structure of author co-citation networks, with the aim of understanding their role in the consolidation of SEAS as a scientific domain.

The insights derived from this study aim to guide researchers and practitioners interested in advancing the development and implementation of SEAS. By uncovering historical and contemporary research patterns, stakeholders will be better equipped to identify critical gaps and promising directions for future inquiry [25]. Ultimately, this analysis seeks to ensure that SEAS evolves in alignment with both societal priorities and industrial needs.

This article is structured as follows. Section 2 describes the methodology used for data collection and bibliometric analysis. Section 3 presents the results, including citation patterns, co-authorship networks, and thematic groupings. Section 4 provides a critical interpretation of the results, emphasizing the main technological developments and their implications. In the end, Section 5 outlines the theoretical and practical implications of the study and proposes directions for future research.

2 Research Methodology

This section outlines the methodology employed to conduct the bibliometric analysis of scientific production within the field of study. The process commenced with data collection from the Scopus database, selected for its extensive indexing of peer-reviewed scientific literature and its capacity to provide robust research impact indicators [26]. A detailed justification for the selection of this database is presented in Section 2.2.

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) approach was used to structure the literature review and analysis, allowing for a rigorous process in identifying, selecting, and evaluating scientific papers. Based on the search strategy, a filtering process was carried out that

included eliminating out-of-context papers, reviewing abstracts and titles, and selecting relevant studies for bibliometric analysis [27]. Fig. 1 presents the article selection process flowchart based on PRISMA.

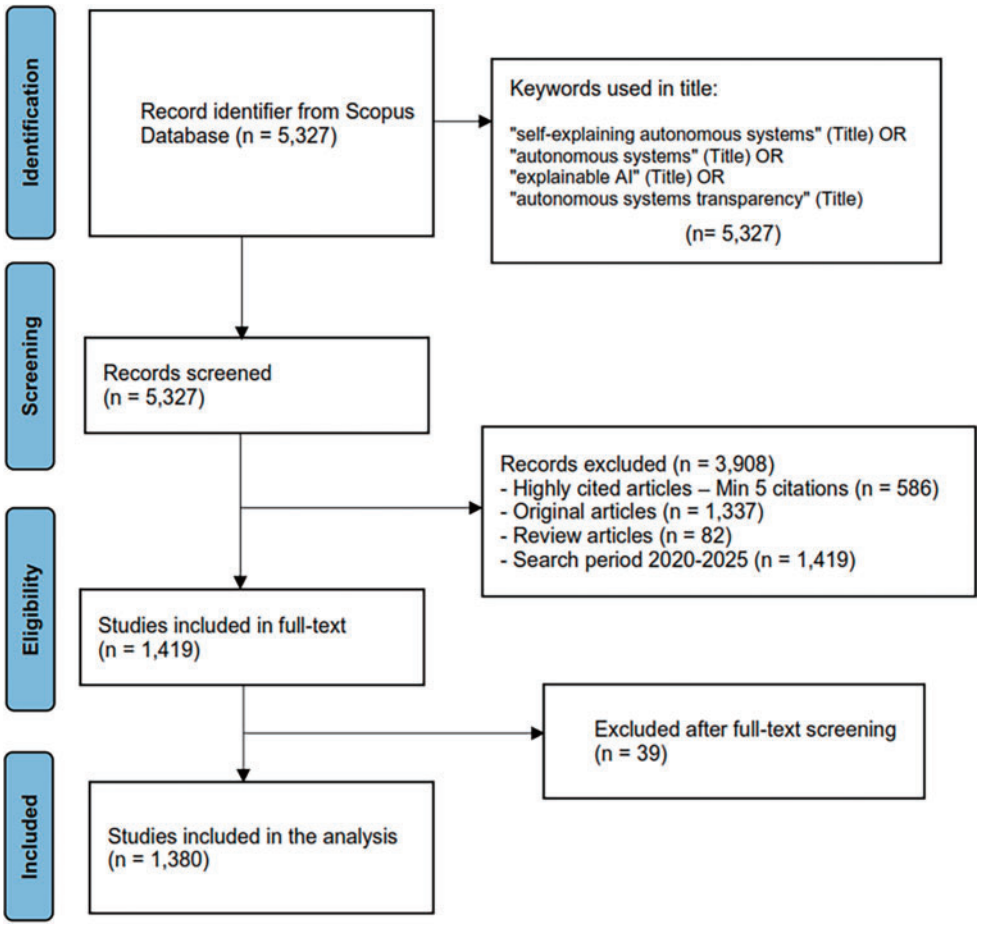

Figure 1: PRISMA flowchart (Reprinted with permission from [28]. © 2021, Page MJ)

2.1 Defining the Research Question

The research question that guides the manuscript is established at this stage. Following the methodological approach proposed by Arksey and Malley [29], the present study aims to analyze existing scientific production on SEAS and assess the evolution of research in this field. Our general research question is: How has research on SEAS evolved in terms of scientific output, thematic trends, bibliometric impact, and global collaborative networks?

2.2 Defining Search Sources

As mentioned above, Scopus was selected as the primary data source for this study due to its extensive coverage of peer-reviewed literature, especially in fields central to SEAS, such as artificial intelligence, robotics, and human-machine interaction. Compared to Web of Science (WoS), Scopus provides broader indexing of conference proceedings and interdisciplinary studies, which is relevant for capturing emerging developments in technology-driven domains [30,31]. While WoS offers a more selective scope, Scopus's broader indexing strategy better aligns with the study's objective of mapping the evolving SEAS research

landscape. Moreover, the bibliometric tools employed in this study, in particular, MASHA (Metrics-Analysis-Science-Hub-Analytics) [32], offer full compatibility with the data formats exported from Scopus, allowing for more efficient, consistent, and secure analysis. However, we recognize the limitation of relying exclusively on Scopus, as this may result in the omission of niche or emerging research not indexed in the database.

2.3 Defining a Search

This section outlines the keywords and their combinations used to retrieve relevant literature. Drawing on insights from prior research, we selected terms closely associated with SEAS to enhance both the accuracy and comprehensiveness of the resulting dataset.

The keyword selection was informed by a preliminary review of influential studies in the fields of XAI and autonomous systems. For instance, Barredo Arrieta et al. [33] and Alonso et al. [34] conducted systematic reviews of XAI, consolidating core terms such as "explainable AI." Similarly, Sharma et al. [35] emphasized the importance of XAI in addressing legal, ethical, and social implications within the research community. In contrast, the term "autonomous systems" is a central concept in this study due to its close relationship with the challenges related to explainability and transparency in intelligent environments.

Based on this review, a search query was formulated in the Scopus database using the following terms in the title field: "`self-explaining autonomous systems`" OR "`autonomous systems`" OR "`explainable AI`" OR "`autonomous systems transparency`". This initial search constituted the first stage in the compilation of relevant literature. In the next subsection, we describe in detail the filtering criteria applied to refine the initial set of results, ensuring that the final selection of publications aligns fully with the research objectives.

2.4 Conducting a Search

The search process was conducted based on the query defined in Section 2.3. To structure the analysis, we followed the PRISMA guidelines, which support comprehensive and transparent reporting of literature selection. These guidelines include three phases: identification, screening, and inclusion. The search and data collection took place on 18 February 2025.

The initial search retrieved a total of 5327 papers. To ensure the relevance of the results, the search was limited to document titles, thereby avoiding the inclusion of articles that only mentioned the keywords superficially. Filters were then applied, reducing the number of records to 3908.

A document type filter was then applied to include only research and review articles, resulting in a final set of 1419 documents. All selected records were accessed and reviewed in full text. The full-text criterion implied the availability of the full content of the article, accessing the sections necessary to verify its thematic focus, such as introduction, development, and conclusions, through institutional subscriptions or open access through Scopus. Each document was evaluated in its entirety in order to confirm its thematic relevance in the field of study. All selected records were in English, which facilitated the analysis.

The search period covered the last five years in the Scopus database, specifically from 2020 to 18 February 2025. This timeframe was selected to capture the most recent developments in the field of SEAS, which has experienced accelerated growth and conceptual consolidation in recent years due to advances in explainable AI, autonomous decision-making, and human-centered design. Given the dynamic and emerging nature of this field, a five-year window provides a focused and up-to-date snapshot of the current research landscape while minimizing the inclusion of outdated or less relevant studies. On the other hand, Abramo et al. [36] mention that a bibliometric assessment can be considered relatively stable with a three-year publication period. As Scopus indexes only peer-reviewed literature and automatically removes duplicate entries, no

additional filtering for quality control or redundancy was necessary. To ensure methodological transparency, the following criteria were applied:

Inclusion criteria:

- Articles written in English.
- Published between 2020 and 18 February 2025.
- Indexed in Scopus as research or review articles.
- Explicit focus on SEAS or related domains such as explainable AI, autonomous decision-making, or human-machine interaction, as identified through keyword presence in the title.

Exclusion criteria:

- Non-English publications.
- Documents not classified as articles or reviews (e.g., conference abstracts, editorials, letters).
- Records mentioning relevant terms only superficially without substantive alignment to SEAS.

Thus, the structured application of these criteria leads to the extraction of a set of quality documents and provides a solid basis for the subsequent phases of the bibliometric analysis.

2.5 Evaluation of the Quality of Results

Once the documents had been retrieved, the quality of the data obtained and their relevance to the study were assessed. For this purpose, selection criteria were applied based on previous bibliometric analysis methodologies [37]. After a detailed review of the selected articles, the initial set of publications was reduced to 1380 documents.

2.6 Primary Analysis of Scientific Papers

Relevant data were extracted according to the research question, and the complete search and selection process is depicted in Fig. 1. Publications were analyzed according to Scopus categories, authors, affiliations, publication years, countries/regions, publishers, research areas, and citations per year.

The bibliometric analysis was carried out using a set of specialized data visualization and analysis tools to ensure accuracy, reproducibility, and depth in the interpretation of scientific trends. VOSviewer was selected for its robust capabilities in constructing and visualizing bibliometric networks, such as co-authorship, keyword co-occurrence, and citation analysis. Microsoft Excel facilitated preliminary organization and filtering of the raw data, allowing manual verification and initial descriptive statistics. Python was employed for its flexibility and efficiency in generating customized statistical graphics, which proved especially useful when working with large volumes of data. Along the same lines, MASHA was used, an open-access online platform that facilitates bibliometric analysis based on data extracted from Scopus. This tool makes it possible to explore academic production, the impact of publications and collaborations between authors and institutions, through visual representations such as graphs, co-occurrence networks, and citation analyses.

2.7 Detailed Analysis of Scientific Papers

At this stage, an in-depth analysis of the selected articles was carried out by means of a full-text review. As discussed above, the inclusion criteria required that each article directly address issues related to SEAS that may including explainability features or user interaction models that enhance decision making. The analysis focused on three sections of each article, which are the introduction presenting the motivation and context of the research, the development or methodology outlining the proposed approaches and techniques, and the conclusions describing the main results and implications.

2.8 Writing a Review Report

A review report was prepared based on the results obtained, and a discussion was held.

3 Overview of the Results

Table 2 presents the data we will analyze in this study. These data were extracted from the Scopus platform and processed using MASHA, VOSviewer, and Python for bibliometric analysis.

Table 2: Summary of the main bibliometric data

Description	Results
Main information about data	
Timespan	2020: 18-02-2025
Sources (Journals, Books, etc.)	160
Documents	1380
Annual growth rate (%)	41.38
Document average age	1.90
Average citations per doc	21.97
Citation overview (h-index)	57
Document types	
Article	1298
Review	82
Document contents and affiliation	
Keywords plus (ID)	368
Affiliation contribution rate (%)	76.16
Authors	
Authors	160
Authors with a minimum of 5 works	15
Authors' collaboration	
Single-authored docs	85
Co-authors per doc	4.64
International collaboration (%)	7.39

Analysing the papers from 2020 to February 2025, we see a rapidly expanding field with a high annual growth of 41.38%. With 1380 papers published in 160 sources, the knowledge base on self-explanatory autonomous systems is fully consolidated.

The academic impact is strong, evidenced by an h-index of 57 and an average of 21.97 citations per article, indicating that research in this field generates attention and is widely referenced. However, the average age of the papers (1.90 years) indicates that the field is recent and is still maturing.

Regarding document types, production is concentrated in articles (94%) with a low proportion of reviews (6%), which could indicate a lack of synthesis or meta-analysis studies that structure existing knowledge.

Institutional affiliation shows a contribution rate of 76.16%, which shows established institutions' strong participation. However, international collaboration is low (7.39%), reflecting a tendency towards local or regional research.

In terms of authorship, although the total number of authors is 160, only 15 have produced at least 5 papers in the same discipline, which shows that the field still lacks a consolidated base of prolific researchers. On the other hand, the average number of co-authors per paper (4.64) shows a high degree of collaboration between authors, although with a moderate number of single-authored publications (85 papers).

3.1 Keyword Analysis in the Literature

Keywords are essential to identify trends and relationships within a research field. In this section, we analyse the frequency and connection between key terms used in SEAS studies to identify thematic patterns and the evolution of knowledge in this area.

3.1.1 Main Keywords in the Literature

This section presents the main recurring keywords and their interconnection through co-occurrence analysis, visualizing the thematic structure of the field of study. At the center of the diagram in Fig. 2, five main keywords stand out, which describe SEAS (and their mutual connections with other keywords), namely:

- Machine Learning;
- Artificial Intelligence;
- Explainable AI;
- Deep Learning;
- Human.

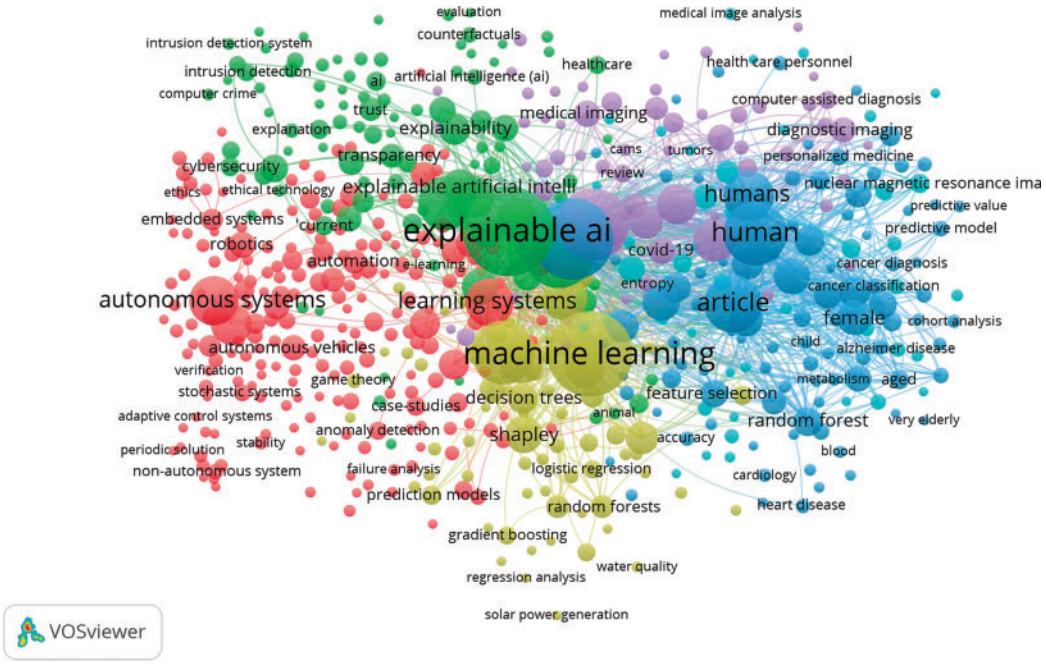

Figure 2: Keyword co-occurrence

The five main keywords in the diagram represent the basis for understanding and applying self-explanatory autonomous systems today. To understand these keywords, we need to describe or define them to continue with this paper.

1. **Machine Learning:** It is a branch of artificial intelligence that develops algorithms and statistical models capable of learning from data and improving their performance without the need for explicit programming [38]. By integrating cognitive science, computer science, and statistics elements, ML enables systems to identify patterns, make predictions, and optimize decisions in different contexts [39]. Its learning process is based on three main approaches: supervised, where the model is trained on labeled data; unsupervised, which discovers structures and relationships without prior information; and reinforcement, where the system learns through interaction with its environment and feedback in the form of rewards or penalties. ML applications span multiple disciplines, including health, where it contributes to the early detection of diseases; business, optimizing operations and improving the user experience through natural language processing; and computer security, robotics, and video game development, among others [40,41]. Its ability to adapt and continuously improve positions it as a relevant technology in the evolution of intelligent systems and the digital transformation of various industries [42,43].
2. **Artificial Intelligence:** It is a multidisciplinary field that develops systems capable of replicating human cognitive functions, such as learning, decision-making, and language processing [44]. Its main approaches include ML, which allows models to improve with data; Deep Learning, based on advanced neural networks; and expert systems, which simulate human reasoning. In healthcare, education, finance, and logistics, AI optimizes diagnostics, personalizes learning, and improves fraud detection [45,46]. However, it faces ethical, transparency, and computational infrastructure challenges. With advances in language processing, computer vision, and robotics, AI continues to evolve, driving innovations and transformations in various industries [47].
3. **Explainable AI:** Beeks to make AI systems more interpretable and understandable to users, enabling greater transparency in decision-making [48]. Its approaches include feature importance analysis, surrogate modeling, LIME, and SHAP, methods that explain how models generate predictions [49]. XAI is fundamental in sectors such as healthcare and manufacturing, where reliability and decision justification are critical [50,51]. However, it faces challenges such as a lack of standardization and difficulty balancing accuracy and interpretability. Despite these barriers, XAI continues to evolve, improving confidence and adoption of AI in high-impact contexts [52,53].
4. **Deep Learning:** It is a branch of machine learning based on multi-layered neural networks, enabling the extraction and hierarchical representation of complex patterns from large volumes of data [54,55]. Its application spans image and speech recognition, natural language processing, medical diagnosis, and autonomous systems. It is notable for its high accuracy and automation capability. However, it faces challenges such as the need for large volumes of labeled data and high computational requirements [56,57]. Tools such as TensorFlow, PyTorch, and Keras have facilitated its implementation. At the same time, current research seeks to improve its scalability and integration with other technologies, expanding its impact in various sectors [58,59].
5. **Human:** The field of study highlights the multifaceted role, from their development to their interaction and use [60]. As creators and developers, they design the algorithms and datasets that shape AI behavior. At the same time, controllers and decision-makers oversee its operation in critical areas such as justice and security [61,62]. In collaborative environments such as Industry 5.0, AI is positioned as an ally that boosts productivity without replacing the human [63]. Also, its integration into everyday life seeks to improve efficiency and reduce risks, always with an ethical and user-centered approach [64].

However, its adoption poses challenges, such as the impact on employment and the need to improve human-IA interaction to foster trust and cooperation [65]. The key to successful implementation lies in transparent, accountable, and people-centered design, ensuring that AI complements and enhances human capabilities rather than replacing them [66].

3.1.2 Frequency of Keywords over Time

In this subsection, a temporal analysis of keywords allows us to evaluate the evolution of interest in different aspects of SEAS. In this case, Fig. 3 shows the frequency with which important terms appear in recent years, identifying terms such as "Explainable AI" and "Machine Learning", which have experienced an exponential increase in 2023 and 2024, consolidating themselves as areas of high interest. The rise of terms such as "Deep Learning" and "Explainable Artificial Intelligence" is evidence of a growing specialization within the field of explainable AI. In contrast, the progressive emergence of methodologies such as "SHAP" and "LIME" indicates a broader adoption of interpretive techniques in AI models. While general concepts such as "Artificial Intelligence" and "XAI" show a more stable trend, the notable rise of specific terms reflects a shift in research direction toward more applied approaches and concrete explanatory tools. This evolution exposes a maturation of the field, where the interpretability of AI is not only consolidating as a central theme but is also driving the integration of methods to make autonomous models more understandable and transparent.

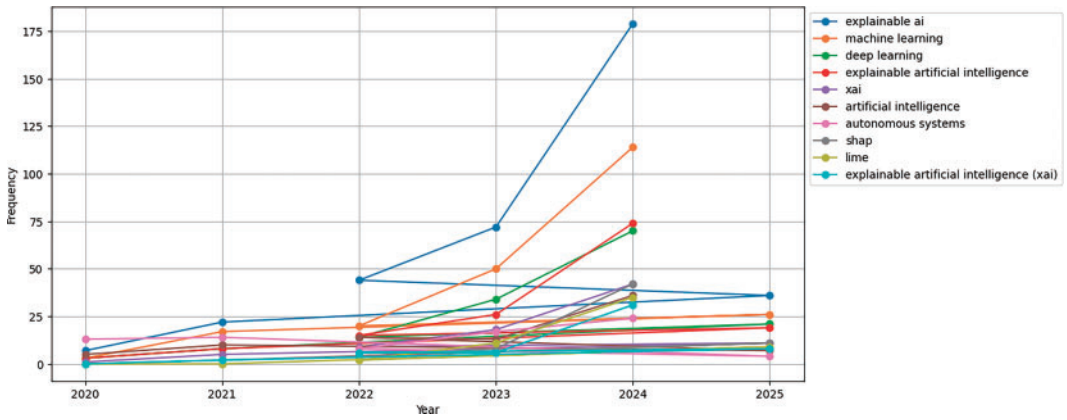

Figure 3: Top 10 most frequent keyword trends

3.1.3 Thematic Evolution and Research Trends

In this way, the above elements are related to what was foreseen in subsection 3.1.1, where, through the Keyword Co-occurrence analysis, a high interrelation between the most used concepts in the recent literature on self-explanatory autonomous systems is evidenced. Likewise, Fig. 4 visualizes the temporal co-occurrence network, shedding light on thematic trends that go beyond citation frequencies and reveal the structural evolution of the field. The most prominent node, "explainable AI", is strongly linked to "machine learning", "learning systems", and "transparency". This reflects an academic shift towards building interpretable and trustworthy models—a response to growing concerns over the "black-box" nature of AI. From a scholarly perspective, this trend drives research on model-agnostic explanations, saliency maps, and concept attribution techniques, establishing XAI as a core area of inquiry rather than a peripheral concern.

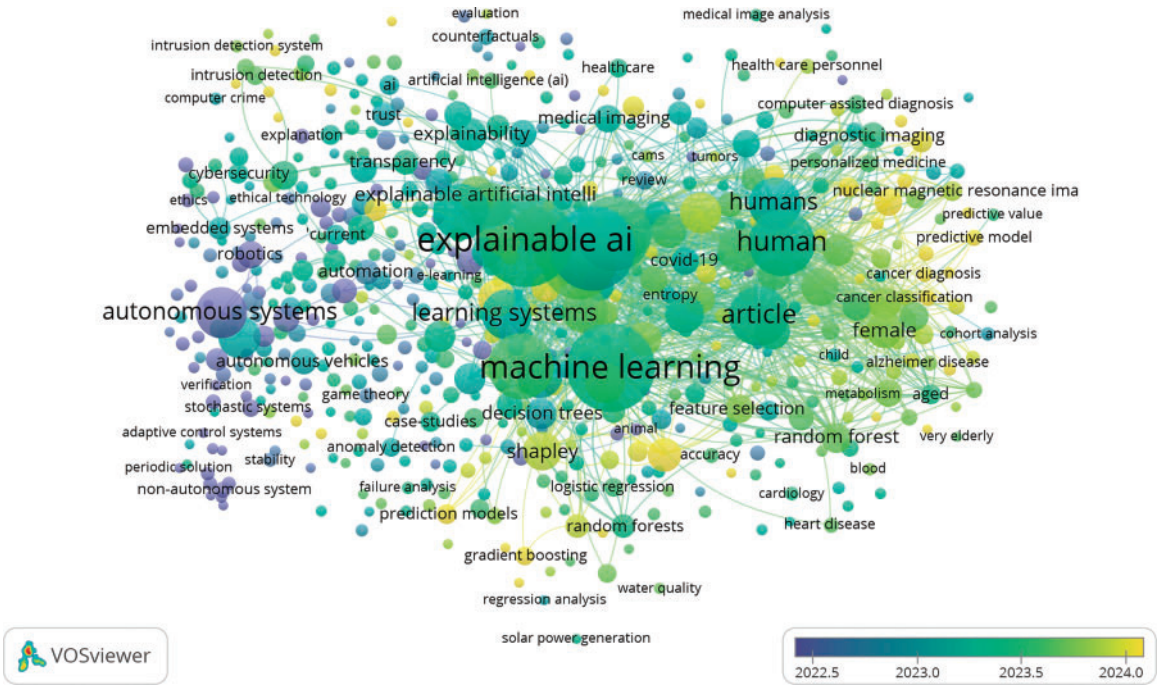

Figure 4: Emerging trends in research

Socially, the growing demand for explainability in artificial intelligence systems is due to an ethical concern focused on guaranteeing impartiality, accountability, and user confidence, especially in sensitive and high-risk contexts. One of the fields where this concern is particularly relevant is healthcare. In the co-occurrence network, the thematic cluster related to medicine links explainable AI with terms such as "medical imaging", "diagnostic imaging", "nuclear magnetic resonance imaging", "cancer diagnosis", and "Alzheimer's disease". These associations evidence a growing expectation, both societal and clinical, that AI models not only support diagnostic decisions but also provide understandable and auditable justifications. For example, approaches such as heat maps applied to medical images, particularly radiological ones, rule-based systems for disease prediction, and MRI analysis in brain tumor detection, are progressively being adopted in real clinical settings [67,68]. Consequently, this trend highlights how XAI research is responding to specific ethical requirements associated with the responsible use of artificial intelligence in healthcare.

In the industrial field, another thematic group in the network includes terms such as "autonomous systems", "robotics", "cybersecurity", and "embedded systems". These results are evidence of an increasingly strategic role for XAI in sectors that require real-time autonomous decisions under conditions of high uncertainty. An illustrative case is that of autonomous vehicles developed by companies such as Waymo or Tesla, where explainability mechanisms allow reconstructing and auditing navigation trajectories after incidents, which is relevant both for continuous improvement and for compliance with road safety regulations [69–71]. Meanwhile, in the field of cybersecurity, platforms such as IBM QRadar incorporate explainability models that allow analysts to understand why an alert was generated, facilitating faster and more reliable decisions in the face of critical threats [72]. On the other hand, explainability in industrial systems plays an important role in building operational trust, especially in environments where humans and machines must interact in a coordinated manner to perform complex tasks. This is necessary in scenarios such as robotic assembly lines or automated logistics processes, where system transparency facilitates monitoring, reduces the margin of error, and improves human-machine collaboration.

These trends show that explainability is not only a technical added value, but an increasingly explicit requirement in emerging regulations, such as the European Union's Artificial Intelligence Act or the ISO/IEC 22989 guidelines, which establish transparency principles for high-risk intelligent systems [73].

Complementing the network-based analysis, Fig. 5 presents a series of word clouds that provide a longitudinal and visually informed perspective on the thematic evolution in the field of explainable artificial intelligence and autonomous systems over the period 2020–2025. Rather than the study only being limited to a descriptive count of keyword frequency, this visualization allows for a deeper reflection on how key concepts consolidate over time, how emerging terms begin to shape new directions, and how the field progressively moves toward semantic refinement and methodological diversification.

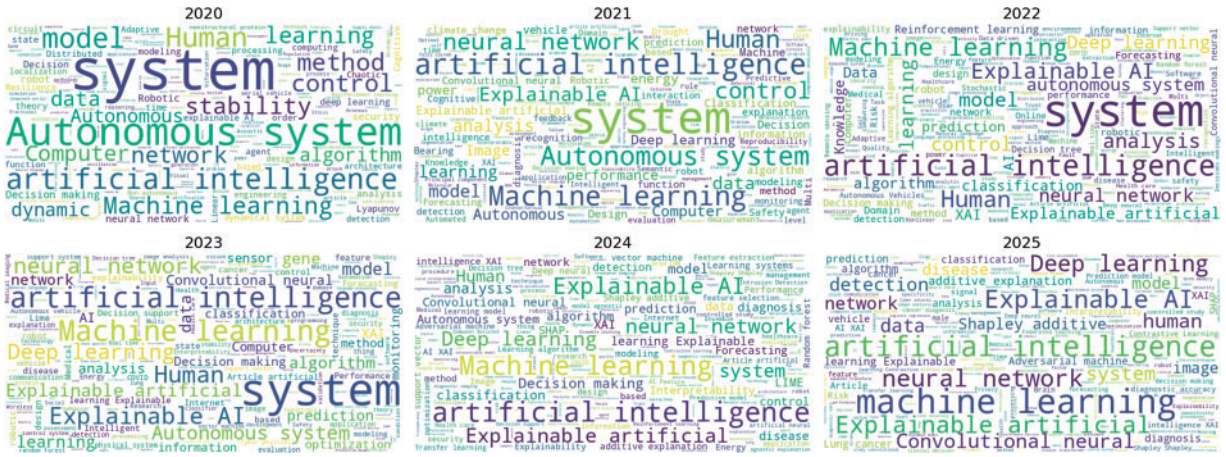

Figure 5: Thematic evolution from 2020 to 2025

In the early years (2020 and 2021), the prominence of foundational terms such as system, autonomous, control, and machine learning evidences an emphasis on infrastructure-level research and algorithmic development. The frequent appearance of terms such as stability, method, and model reflects a research agenda rooted primarily in engineering, systems theory, and optimization concerns, where explainability appears more as a secondary aspect than a central focus.

From 2022 onwards, a thematic shift becomes evident. Keywords such as explainable AI, neural network, decision-making, and human become more relevant, indicating a transition to more human-centered and application-oriented research. This shift points to a process of consolidation of the field, in which explainability is increasingly treated not as an add-on but as an integral component in the architecture and evaluation of autonomous systems.

In 2023, and more markedly in 2024 and 2025, the vocabulary becomes more granular and reveals a conceptual deepening. The emergence of terms such as interpretability, SHAP, LIME, diagnostics, healthcare, and disease denotes the expansion of XAI into specific domains, particularly the biomedical and clinical sectors. This evolution reflects a broader transformation in the field from theoretical models focused on interpretability to practical approaches that respond to ethical, diagnostic, and societal demands.

3.1.4 Keywords and Leading Research Countries

We consider that different regions of the world drive SEAS research. This section analyses how the most frequent keywords are distributed according to the countries that lead the scientific production in several citations. For a better illustration, Fig. 6 shows the close relationship between the keywords and the countries

producing in this field, making it possible to identify the differences in the approaches adopted by each nation in research on SEAS. It is also possible to deduce the thematic priorities that each country assigns within this field, highlighting specific trends and potential variations in the development and application of these systems globally.

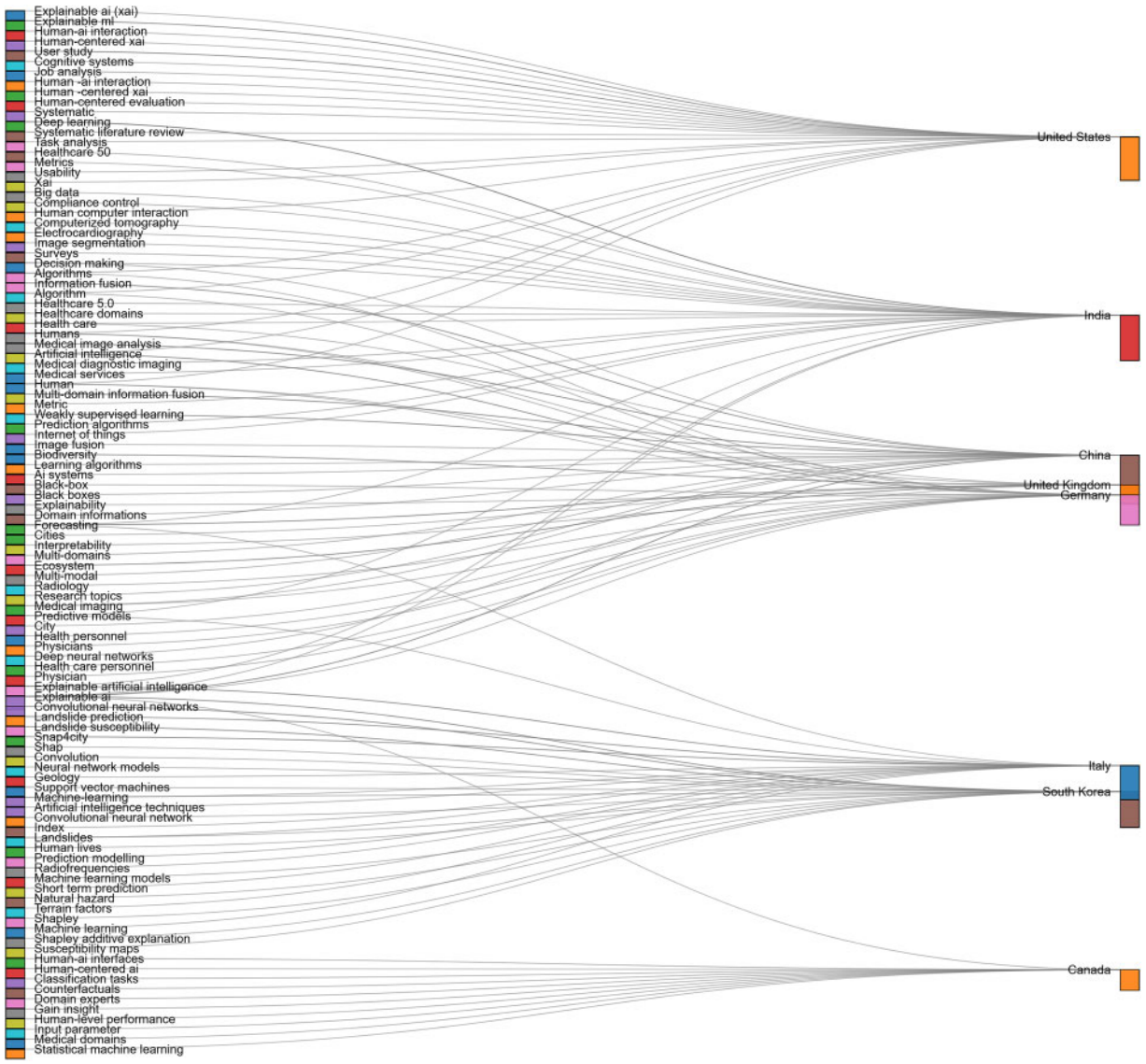

Figure 6: Keyword frequency by country (Top most cited)

The linkage diagram allows for identifying patterns of regional specialization, showing how specific terms are strongly associated with particular countries. For example, the United States stands out as the top contributor, with a broad connection to a diverse spectrum of keywords, indicating a multidisciplinary approach and leadership in generating explainable artificial intelligence knowledge. India and China also link highly to multiple terms, indicating an active role in research, albeit with possible differences in thematic and methodological orientation. Europe, represented by the UK and Germany, exhibits a high degree of

interconnection with specific terms, showing a high rate of participation in the theoretical and applied development of the discipline.

Meanwhile, countries such as Italy, South Korea, and Canada show less thematic diversification compared to the primary producers but maintain a significant relationship with specific keywords, which could indicate more specialized approaches or emerging lines of research in these contexts. The distribution of connections in the figure highlights that while research in self-explanatory autonomous systems is global, there are marked differences in the focus of studies by country, which could be influenced by factors such as innovation policies, academic funding, and industrial needs.

3.2 Scientific Output and Sources of Publication

Scientific publications are disseminated through various publishers and journals. This section examines the primary sources that have contributed significantly to developing knowledge in SEAS.

3.2.1 Top Scientific Publishers in the Field

Today, publishers play an essential role in the dissemination of academic articles. This section identifies the ten publishers with the highest volume of publications in the field, highlighting their role in disseminating knowledge. Table 3 provides a detailed analysis of their contribution, showing their influence in the consolidation and development of this emerging discipline. Institute of Electrical and Electronics Engineers Inc. (IEEE) leads the list with 213 publications, representing 23.80% of the total, which reaffirms its position as the reference publisher in explainable artificial intelligence and autonomous systems. Multidisciplinary Digital Publishing Institute (MDPI) shows a significant proportion of publications of 18.66%, which is evidence of its strong presence and contribution to the consolidation of academic literature in this field.

Table 3: Top 10 publishers with the highest scientific output in the field of study

No.	Editorial	Number of works	%
1	Institute of Electrical and Electronics Engineers Inc.	213	23.80
2	Multidisciplinary Digital Publishing Institute	167	18.66
3	Elsevier Ltd.	161	17.99
4	Elsevier B.V.	120	13.41
5	Springer	71	7.93
6	Springer Science and Business Media Deutschland GmbH	44	4.92
7	Nature Research	35	3.91
8	Taylor and Francis Ltd.	31	3.46
9	Association for Computing Machinery	29	3.24
10	John Wiley and Sons Inc.	24	2.68

On the other hand, the publishers Elsevier Ltd. and Elsevier B.V. account for a total of 281 publications (31.04%). Although both are part of the wider Elsevier publishing group, they are registered as separate legal entities and may focus on different publishing or regional operations. Elsevier Ltd is based in the UK, while Elsevier B.V. operates from the Netherlands. In this case, the Scopus database maintains this distinction by attributing publications to the specific entity listed as the publisher. In this case, the Scopus database preserves this distinction by attributing each publication to the specific publishing entity listed as the publisher.

A similar case is observed with Springer, which appears under two different names, Springer (7.93%) and Springer Science and Business Media Deutschland GmbH, the latter ranking second with 44 publications (4.92%), which reflects a possible legal or strategic segmentation in its editorial structure. Meanwhile, publishers such as Nature Research (3.91%), Taylor and Francis Ltd. (3.46%), the Association for Computing Machinery (3.24%), and John Wiley and Sons Inc. (2.68%) show a relevant contribution, though with a lower output compared to the top-ranked publishers, which may indicate a more specialized or narrowly focused approach to particular research areas within the field.

Overall, the distribution of publications shows an intense concentration in IEEE, MDPI, and Elsevier, indicating that these publishers dominate scientific production in self-explanatory stand-alone systems, probably due to their publishing infrastructure, prestige in the academic community, and their ability to attract high-impact research.

3.2.2 Most Influential Scientific Journals and Their Citation Impact

Specialized scientific journals serve as a point of reference for the research community. This subsection identifies the scientific journals with the highest production and relevance in the field of study. Table 4 presents a comparative analysis of the ten most relevant journals in the field of study, considering indicators such as number of publications, total citations received, average impact, and percentage of documents that have been cited at least once. In this case, IEEE Access ranks with 71 publications and 684 citations, consolidating its position as the journal with the highest volume of papers and an average impact of 9.63. The outstanding presence of IEEE Access in the study area positions it as one of the leading platforms for disseminating research. Its leadership in the volume of publications and citations is evidence that it is a reference medium for studies on explainability in autonomous systems, driven by its focus on artificial intelligence, machine learning, and intelligent systems. On the other hand, the Multidisciplinary Digital Publishing Institute has consolidated its position as an important player in the scientific production of SEAS, with a total of 99 publications and 1055 citations in its most representative journals: Applied Sciences, Sensors, Electronics, and Information. This production volume highlights the impact of MDPI in disseminating knowledge in this field with an editorial approach that integrates both theoretical aspects and practical applications. Its contribution has facilitated the publication of studies that address the development and implementation of explanatory methodologies in artificial intelligence.

Table 4: Scientific journals with the highest production and impact

No.	Journal	Publisher	Number of works	Total citations	Average impact	% Cited docs
1	IEEE Access	Institute of Electrical and Electronics Engineers Inc.	71	684	9.63	70.42
2	Applied Sciences (Switzerland)	Multidisciplinary Digital Publishing Institute	27	209	7.74	77.78
3	Scientific Reports	Nature Research	26	195	7.5	76.92
4	Sensors	Multidisciplinary Digital Publishing Institute	18	605	33.61	83.33
5	Electronics (Switzerland)	Multidisciplinary Digital Publishing Institute	14	107	7.64	64.29
6	Computers in Biology and Medicine	Elsevier Ltd.	13	111	8.54	61.54

(Continued)

Table 4 (continued)

No.	Journal	Publisher	Number of works	Total citations	Average impact	% Cited docs
7	Information (Switzerland)	Multidisciplinary Digital Publishing Institute	13	134	10.31	69.23
8	Biomedical Signal Processing and Control	Elsevier Ltd.	12	129	10.75	66.67
9	Expert Systems with Applications	Elsevier Ltd.	12	256	21.33	83.33
10	Multimedia Tools and Applications	Springer	10	33	3.30	70.00

Elsevier Ltd. also maintains a prominent presence in scientific production in the field of study, with 37 publications and 496 citations in three journals: Computers in Biology and Medicine, Biomedical Signal Processing and Control, and Expert Systems with Applications. His contribution focuses on the intersection between explainable artificial intelligence and its applications in the biomedical domain and expert systems, demonstrating the practical impact of explainability in AI beyond the theoretical framework. Finally, Scientific Reports and Multimedia Tools and Applications complete the list, with a broader focus on applied science and technology. In terms of average impact, Sensors (33.61) and Expert Systems with Applications (21.33) stand out, indicating that, although they are not the most widely published journals, the papers in them have a high citation level, which may reflect their relevance in the research community.

To enhance the robustness of the citation analysis and reduce the influence of skewed citation distributions, this study employs the metric *Percentage of Cited Documents* (% Cited Docs). This indicator provides a normalized view of a journal's impact by accounting not only for highly cited works but also for the breadth of citation across its published output. The metric is defined as the ratio, expressed as a percentage, between the number of documents that have received at least one citation and the total number of documents analyzed for each journal, as shown in Eq. (1):

$$\%\text{Cited Docs} = \left(\frac{D_{\geq 1}}{D_{\text{total}}}\right) \times 100 \tag{1}$$

In this formulation, $D_{\geq 1}$ represents the number of documents with one or more citations, and D_{total} denotes the total number of documents considered for the respective journal. The resulting percentage reflects the extent to which a journal's output achieves scholarly visibility.

Consequently, the above metrics offer a complementary perspective on journal performance. While traditional metrics, such as total citations or average impact, may be limited by a few highly cited articles, the percentage of cited articles provides insight into the distribution and consistency of scholarly attention in field-related publications. For example, journals such as *Sensors* and *Expert Systems with Applications*, with a high percentage of cited papers (83.33%), demonstrate a broader citation footprint, evidencing a more balanced and widespread influence. In contrast, journals with similar average impact but lower citation coverage may have their metrics disproportionately driven by a small subset of impactful studies. Therefore, the incorporation of this metric helps to mitigate the distortions inherent to evaluations based exclusively on citation counts, while strengthening the methodological transparency of the analysis.

To enrich the reading of the bibliometric indicators and provide a broader contextual view, a visualization based on the co-occurrence of publication sources is included, which facilitates the identification of patterns of thematic concentration and temporal trends in publications. Fig. 7 represents a co-occurrence network based on publication sources to provide an overview and feed the above information. The size of each node in the visualization indicates the frequency of publications in the respective journal. At the same time, the chromatic scale represents the temporal evolution of the indexed articles covering the period 2022–2024. In this case, it is reaffirmed that high-impact journals, such as IEEE Access, Scientific Reports, and Applied Sciences (Switzerland), account for a significant share of recent publications. On the other hand, publications in multidisciplinary journals, such as Plos One and Heliyon, indicate that interest in these topics is not exclusively limited to computer science but also covers areas such as biomedicine, engineering, and data science. This diversification provides a detailed overview of the distribution of scientific publications in the study area, highlighting the journals with the highest impact and their evolution over time.

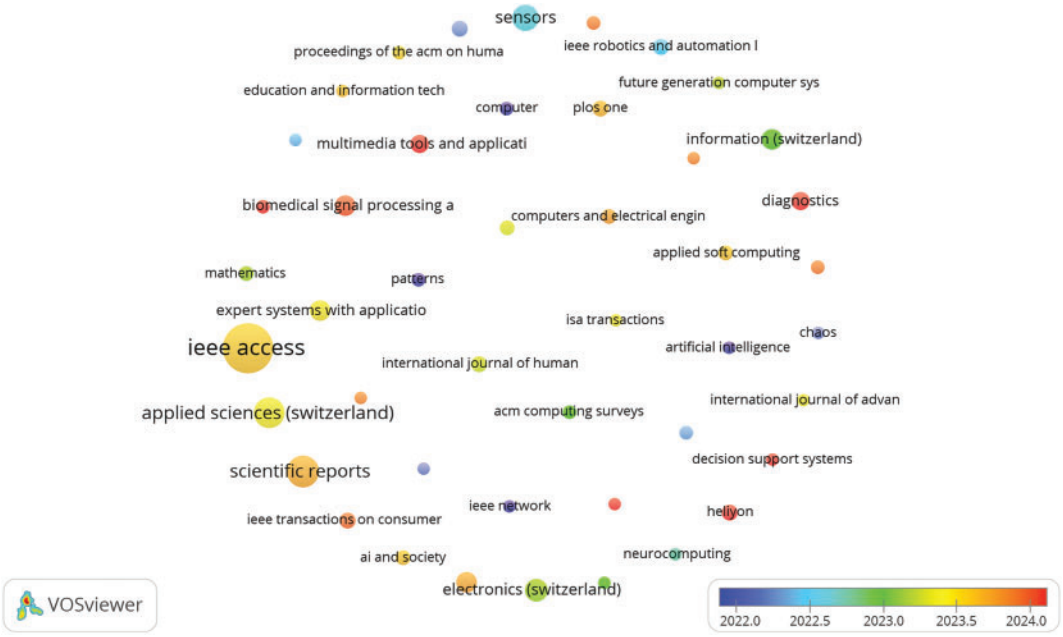

Figure 7: Influential publishers in time

It is important to note that the citation data used in this analysis are derived exclusively from the Scopus database. Citation metrics may vary across databases such as Web of Science or Google Scholar due to differences in indexing criteria, document coverage, update frequency, and inclusion of self-citations. Therefore, while the analysis provides a relevant insight into the influence and impact of scientific journals within the Scopus ecosystem, the results should be interpreted with caution to avoid potential citation bias.

3.3 Impact and Evolution of Research

This section analyses the evolution of citations, academic productivity over time, and the most influential studies in the field.

3.3.1 Evolution of the Citation Rate per Year

The number of citations received per year is a metric for evaluating both the impact and the visibility of a field of research. This section presents a detailed temporal analysis of the evolution of citations, approached from a dual perspective. On the one hand, we examine the absolute values of annual citations, which allow us to identify periods of greater academic intensity. Between 2020 and 2024, 22,258 citations were accumulated, of which 6379 correspond to 2020 (28.65%) and 5337 to 2021 (23.97%), which means that more than half of the total was concentrated in the first two years of the interval. This is followed by 2022 with 4222 appointments (18.96%), 2023 with 4528 (20.33%), and finally 2024 with 1792 (8.05%). This distribution shows an initial stage of strong academic consolidation, followed by a relative stabilization. The figure for 2024 should not be interpreted as a loss of interest, but rather as part of the natural lag in the visibility and citation cycles, widely documented in the scientific literature [74]. The maturation of a scholarly article, from its publication to its full incorporation in new research, may take several years, especially in contexts where scientific production is intense and heterogeneous [75].

Figs 8 and 9 show four complementary metrics that allow a more precise characterization of this temporal evolution. Before analyzing the results, it is important to point out that, although the study horizon formally extends to February 2025, this year has not been considered in the quantitative calculations. This exclusion responds to a methodological decision, since incorporating an incomplete year could introduce distortions in the interpretation. Citation volume is conditioned by exposure time, as has been documented in studies linking thematic maturity with the longer time windows needed to achieve maximum impact [76].

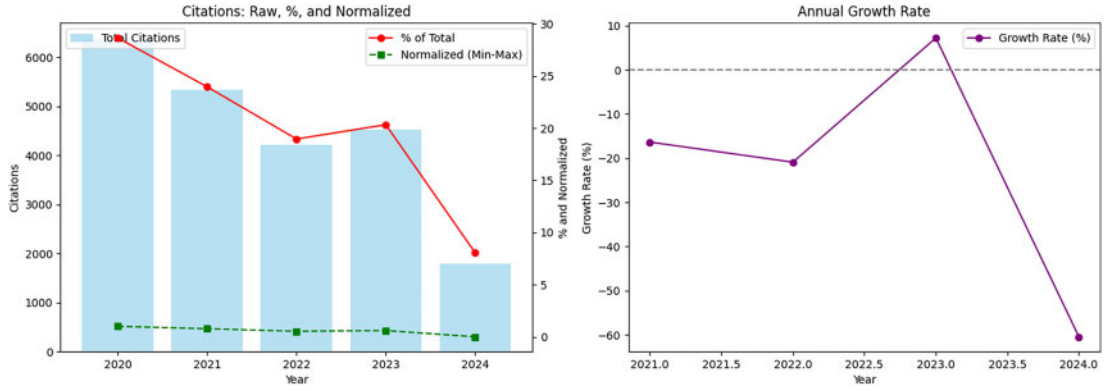

Figure 8: Evolution of scientific production

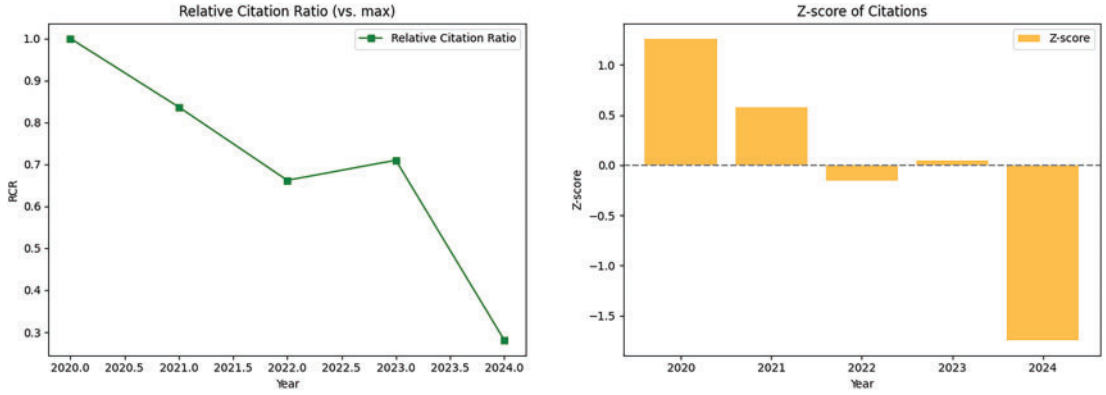

Figure 9: Evolution of scientific production

On the other hand, the Min-Max normalization technique was applied, which made it possible to visualize the relative progression of each year with respect to the best-case scenario observed. Through this metric, a sustained decline is observed from 2020 to 2022, followed by a slight recovery in 2023 and a decline in 2024 that should be interpreted with caution due to its temporal proximity. Rather than representing a deterioration of the field, it could reflect the initial phase of the visibility and citation cycle of the most recent work, in line with findings that highlight how recognition of disruptive or innovative research tends to lag, even when it possesses high potential value [77].

The annual growth rate estimate revealed a continuous decline in 2021 (−16.33%) and 2022 (−20.89%), followed by a modest recovery in 2023 (+7.25%) before a sharper drop in 2024 (−60.42%). This oscillation could be linked to the natural maturation cycle of the field or to the emergence of complementary lines of research that redistribute the focus of attention without necessarily reflecting a loss of relevance. The literature has pointed out that these fluctuations are inherent to citation dynamics in the medium and long term [78].

Metrics such as the relative citation index (RCR) and z-score were also incorporated, which confirm this behavior within the statistical parameters expected in evolving fields. These metrics allow a more nuanced interpretation of the phenomenon, showing that the drop in 2024 does not imply a loss of influence but rather an expected and documented development phase in academic environments where visibility and citation advance at asymmetric rates depending on the time of publication [79].

3.3.2 Scientific Production over Time

The volume of publications in a scientific domain serves as a proxy for its maturation and the degree of scholarly engagement. In this context, we analyzed the temporal distribution of research output between 2020 and February 2025, aiming to identify trends, inflection points, and periods of heightened academic activity. As illustrated in Fig. 10, the field exhibits a marked upward trajectory, with a pronounced peak in 2024, accounting for 42.32% of the total output (584 publications). This surge is indicative of both consolidation and increased academic interest in autonomous systems. The temporal analysis also reveals that 2023 represented a pivotal stage in the consolidation process, with 280 studies published (20.29%). In comparison, 2022 (184 publications; 13.33%) and 2021 (122 publications; 8.84%) showed relatively lower activity, which may reflect the formative phase of the field's conceptual and methodological development.

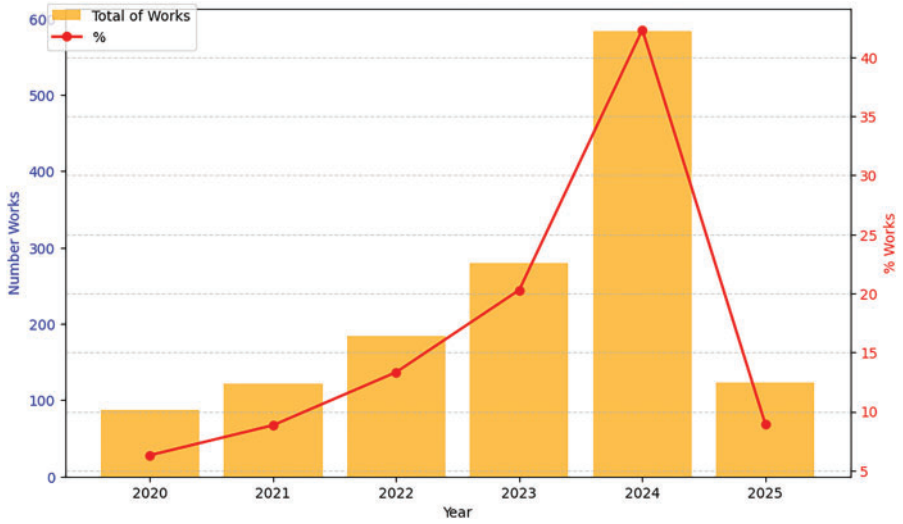

Figure 10: Evolution of scientific production

Although the 2025 data only includes publications through February, the number already totals 123 papers (8.91%), surpassing the 2020 annual total (87 publications; 6.30%). Although preliminary, this early 2025 figure evidences a sustained upward trajectory and continued momentum in this field. To ensure methodological transparency, we emphasize that the 2025 data represent a partial count and should be interpreted as indicative rather than conclusive. Nevertheless, their inclusion highlights the accelerating pace of research and the increasing prominence of the topic within the scientific community.

Beyond the numerical growth, the evolution of scientific production during this period reflects a change in the epistemic structure of the field. There has been a shift from isolated exploratory contributions to a more structured and coherent research agenda in thematic terms. The peak observed in 2024, together with the strong momentum already visible at the beginning of 2025, is evidence that autonomous systems have overcome their initial conceptual phase and have consolidated as a central axis of interdisciplinary research. This trajectory not only affirms the identity of the field but also marks a turning point, where academic interest is increasingly aligned with technological deployment and emerging societal demands.

3.3.3 Areas of Research

The distribution of scientific production in the field of SEAS reflects its varied character and points out the domains where its development has been more outstanding. As shown in Table 5, the predominance of publications in Computer Science and Engineering underlines the technological basis of SEAS research, especially in the design of explainable AI architectures, interpretable models, and real-time reasoning mechanisms embedded in autonomous systems.

Table 5: Distribution of scientific production by research area

No.	Research area	Number of papers	%
1	Computer science	826	35.22
2	Engineering	572	24.39
3	Mathematics	211	9.00
4	Materials science	140	5.97
5	Social sciences	136	5.80
6	Medicine	129	5.50
7	Physics and astronomy	128	5.46
8	Biochemistry, genetics and molecular biology	79	3.37
9	Business, management and accounting	63	2.69
10	Environmental science	61	2.60

This approach is extensively discussed by Trivedi et al. [80] in their vision on Industry 5.0, who highlight the need for XAI approaches tailored to both the domain and the type of user. In the industrial and vehicular domain, Atakishiyev et al. [81] present a compendium on XAI in autonomous driving, where they emphasize the need for explainability to align models with automotive industry trends and requirements. Also, Kuznietsov et al. [70] discuss taxonomies of explanations needed to ensure safety and confidence in these autonomous systems. In parallel, Ahmed et al. [82] perform a systematic review on how XAI is being integrated in Industry 4.0 environments, where the ability to interpret decisions of autonomous systems is key for advanced manufacturing and smart logistics environments. Zablocki et al. [83] review the current challenges in explainability of deep vision-based autonomous driving systems, highlighting the specific needs of the automotive industry in terms of interpretability, user confidence, and regulatory compliance.

Fields such as medicine, physics and astronomy, and environmental sciences contribute a smaller proportion of publications, but they indicate the growing application of SEAS in critical and data-intensive contexts. In Medicine, for example, the integration of explainable AI into clinical decision support systems improves transparency and confidence in AI-assisted workflows, addressing ethical and regulatory considerations. Nasarian et al. [84] propose a responsible collaborative framework between clinicians and AI systems, while Sadeghi et al. [50] provide a detailed review of XAI in healthcare. Singh et al. [85] show how XAI is applied in radiology and diagnostic imaging with an interdisciplinary approach.

On the other hand, areas such as Business, Management, and Accounting remain underrepresented, possibly due to structural challenges in adopting autonomous decision-making frameworks. However, emerging research evidence shows a growing interest in applying SEAS principles to algorithmic governance, explainable financial analysis, and ethical decision making in business operations. Mathew et al. [86] show how XAI can contribute to improved inventory management and resource allocation, anticipating more robust integration in enterprise contexts. In addition, Zhang et al. [87] analyze the role of humans in the explanatory loop in knowledge engineering, which is useful for regulated corporate environments.

3.3.4 Most Cited Articles and Influential Authors

This section presents the most cited articles and the most influential authors in the field of research. Table 6 provides an overview of the papers that have had the most significant impact. In this case, the research by Lundberg et al. [88] (3979 citations) proposes an approach based on game theory to optimize explanations in tree-based models, introducing tools to measure local interactions and analyze the global structure of the model. Its application in the medical field has enabled it to identify risk factors, segment populations, and monitor hospital models, with impacts in multiple domains. Linardatos et al. [89] (1617 citations) emphasize the importance of XAI in the increasing complexity of machine learning models, especially in critical domains such as health, by reviewing and taxonomizing interpretability methods, providing a reference for researchers and practitioners. Complementing this approach, Shin [90] (606 citations) examines the relationship between explainability and user trust, introducing the concept of usability as a relevant factor in the perception of algorithms. Their findings reveal that integrating understandable explanations in AI systems improves trust and transparency. While XAI is applied to deep learning models, Yang et al. [91] (416 citations) review advances in interpretability and propose solutions based on multimodal and multicore data fusion, validated in real clinical scenarios. His contribution extends to designing and evaluating XAI systems, addressing the fragmentation between disciplines such as machine learning, visualization, and human-computer interaction. A systematic review establishes a methodological framework that categorizes design and evaluation objectives in XAI, facilitating its application in various areas.

Table 6: Top 10 most cited articles

Author, Year	Document title	Journal name	Quartile SJR	Citations	%
Lundberg et al. [88], 2020	From local explanations to global understanding with explainable AI for trees	Nature Machine Intelligence	Q1	3979	0.476
Linardatos et al. [89], 2021	Explainable AI: A review of machine learning interpretability methods	Entropy	Q1	1617	0.194

(Continued)

Table 6 (continued)

Author, Year	Document title	Journal name	Quartile SJR	Citations	%
Shin [90], 2021	The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI	International Journal of Human Computer Studies	Q1	606	0.073
Yang et al. [91], 2022	Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond	Information Fusion	Q1	416	0.050
Mohseni et al. [92], 2021	A Multidisciplinary Survey and Framework for Design and Evaluation of Explainable AI Systems	ACM Transactions on Interactive Intelligent Systems	Q1	360	0.043
Dwivedi et al. [93], 2023	Explainable AI (XAI): Core Ideas, Techniques, and Solutions	ACM Computing Surveys	Q1	329	0.039
Saeed et al. [94], 2023	Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities	Knowledge-Based Systems	Q1	307	0.037
Shamim et al. [95], 2020	Explainable AI and mass surveillance system-based healthcare framework to combat COVID-19 like pandemics	IEEE Network	Q1	280	0.034
Holzinger et al. [96], 2021	Towards multi-modal causability with graph neural networks enabling information fusion for explainable AI	Information Fusion	Q1	260	0.031
Chaddad et al. [97], 2023	Survey of explainable AI techniques in healthcare	Sensors	Q1	201	0.024

To complement the above, Mohseni et al. [92] contribution focused on a multidisciplinary framework for designing and evaluating XAI systems, addressing methodological fragmentation. In contrast, Dwivedi et al. [93] provide a comprehensive review of its rationale, techniques, and solutions. Similarly, Saeed and Omlin [94] conducted a meta-analysis on the challenges and future opportunities in the field. In healthcare, Shamim Hossain et al. [95] explored the application of XAI in surveillance systems for pandemic detection, evidencing its relevance in critical decision-making. Holzinger et al. [96] introduced graph neural

networks for multimodal data fusion, improving interpretability in complex scenarios. In contrast, Chaddad et al. [97] analyzed XAI techniques applied to healthcare, reinforcing the importance of transparency in clinical models. Taken together, these studies show the consolidation of XAI as a relevant axis in artificial intelligence, boosting explanatory methodologies and their application in critical contexts such as medicine and epidemiological surveillance.

3.4 Geographic Analysis and Collaboration Networks

Scientific research is a global effort, with multiple institutions and authors collaborating internationally. This section examines the geographic distribution of scientific production and the collaborative networks that have emerged in the field.

3.4.1 Leading Research Countries and Their Impact on Citations

Leadership in a research area is usually linked to the quantity and quality of publications produced in each country. Table 7 analyzes the ten most productive countries. Their volume of publications, impact, and number of citations have been evaluated. The United States leads with 194 publications (21.6%) and a total of 6716 citations (43.9%), standing out not only for its high production but also for an average impact of 34.62, which shows that its studies are widely referenced and highly influential in the field. India, with 175 articles (19.5%) and 1194 citations (7.8%), is in second place, although with a significantly lower impact (6.82), indicating a high volume of production, but with a lower average citation per article. China and the United Kingdom occupy the third and fourth positions, with 101 and 100 publications, respectively, showing a balance between production and recognition, with average impacts of 13.72 and 17.2. Germany (83 publications, 1321 citations, impact of 15.92) and Canada (39 publications, 760 citations, impact of 19.49) show a strong presence in the scientific literature, with important impacts on the academic community. Italy (65 publications, impact of 15.03) and South Korea (70 publications, impact of 8.79) maintain a notable production, although with less influence in terms of citations.

Table 7: Top 10 most productive countries

Rank	Country	No. of paper	%	Citations	%	Average impact
1st	United States	194	0.216	6716	0.439	34.62
2nd	India	175	0.195	1194	0.078	6.82
3rd	China	101	0.112	1386	0.090	13.72
4th	United Kingdom	100	0.111	1720	0.112	17.2
5th	Germany	83	0.092	1321	0.086	15.92
6th	South Korea	70	0.078	615	0.040	8.79
7th	Italy	65	0.072	977	0.060	15.03
8th	Canada	39	0.043	760	0.049	19.49
9th	Saudi Arabia	38	0.042	251	0.016	6.61
10th	Australia	34	0.038	331	0.021	9.74

In the lower segment of the ranking, Saudi Arabia (38 publications, impact of 6.61) and Australia (34 publications, impact of 9.74) reflect a relevant contribution, although with a lower volume of publications and lower impact than the leading countries. We can point out that the production is strongly dominated by the United States, China, and the United Kingdom, whose publications are more numerous and have a high

citation and impact. Despite its high productivity, India faces the challenge of improving the influence of its research. The presence of European countries such as Germany, Italy, and the United Kingdom highlights their contribution to high-impact studies. At the same time, Canada stands out for its citation efficiency in its production. These results show how scientific research in this field has become more dynamic, reflecting an increase in academic production and greater specialization and geographical diversification.

To this, it is important to add that Fig.11 complements in a general way the information on the countries that have contributed to the development of scientific production in this field, considering both the number of papers and citations. After Australia (10th), Spain, a country of the European Union, is followed by Bangladesh, with 26 manuscripts. South America shows a low participation in scientific production. Brazil stands out among the region's countries with 10 papers and 70 citations. At the same time, Argentina, Chile, Colombia, Ecuador, and other South American nations show an even more limited representation, with less than 10 papers each. In some cases, this low level of scientific production can be attributed to several factors. First, investment in science and technology in the region is significantly lower compared to other parts of the world, such as North America, Europe, or Asia, which limits the development of large-scale research projects. On the other hand, the lack of integration with international collaboration and funding networks reduces the opportunities for South American researchers to access resources and strategic alliances with prestigious centers.

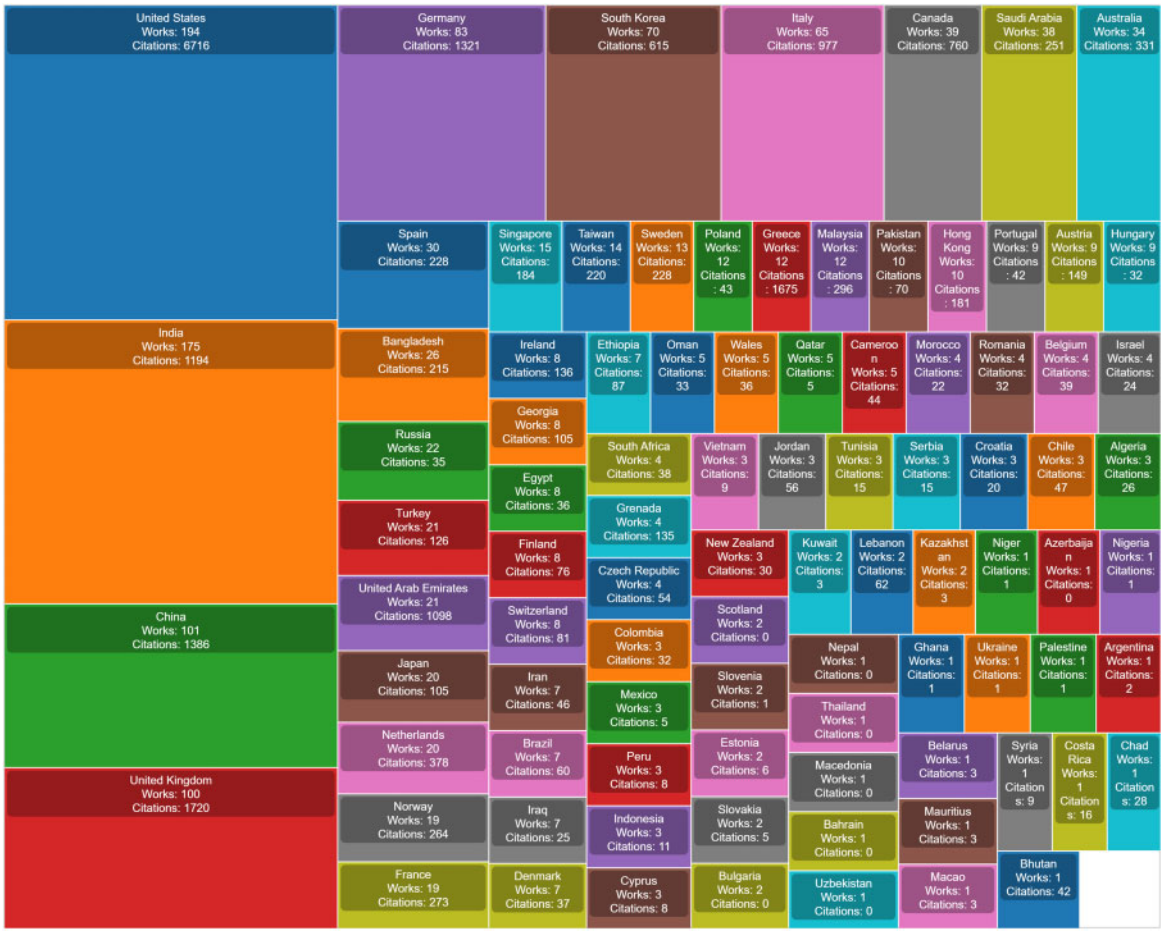

Figure 11: Treemap of scientific production by country

3.4.2 Network for Co-Citation among Authors in the Field

Understanding the intellectual structure of a scientific field is fundamental for tracing the development of knowledge and identifying influential research communities. One of the most robust approaches to achieve this is the analysis of author co-citation. As noted by Nerur et al. [98], this method enables the visualization of conceptual relationships among scholars, helping to identify foundational contributions and dominant schools of thought. In the same vein, González-Valiente et al. [99] emphasize that co-citation analysis offers a powerful means to uncover the thematic organization and socio-intellectual dynamics of a discipline.

Fig.12 illustrates the co-citation network generated from the dataset, showing how frequently authors are cited together in the scientific literature. The resulting map reveals several distinct clusters, each marked by a different color, corresponding to groups of researchers who share similar thematic orientations or methodological frameworks. The red-colored nodes, larger in size, represent the most influential authors in the network, whose work has been extensively referenced and forms the theoretical backbone of the field.

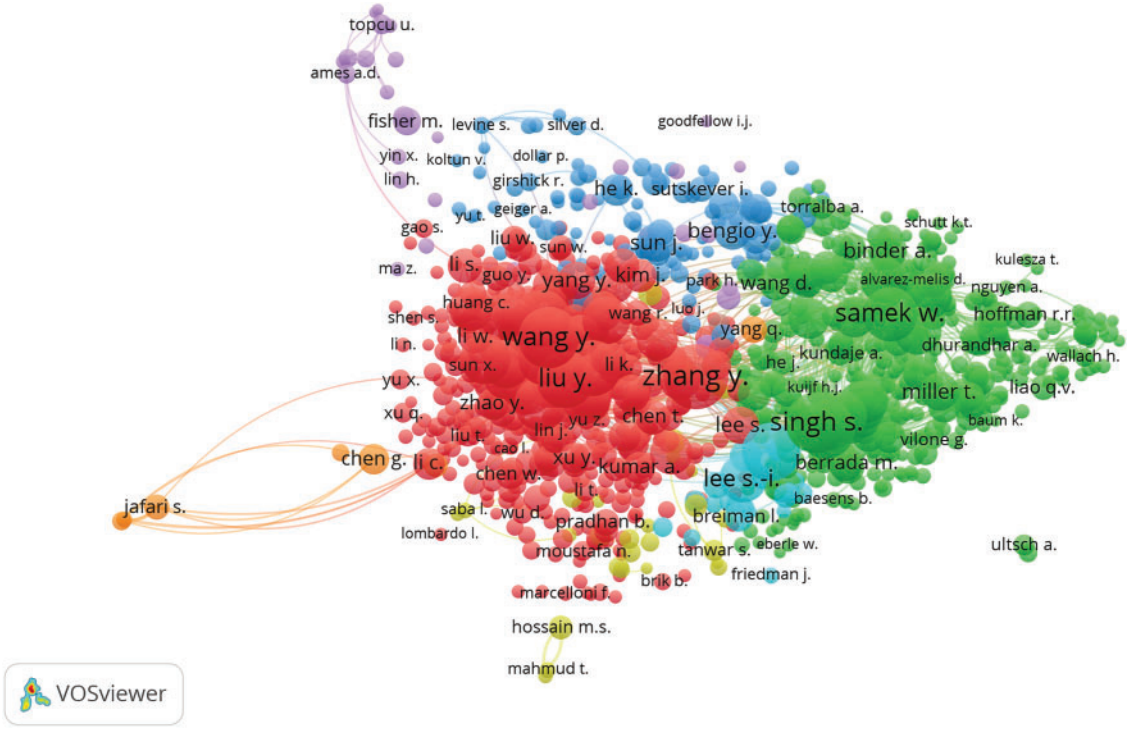

Figure 12: Author co-citation

Four central research communities emerge from the analysis. The red cluster focuses on methodological development and model interpretability, particularly in the context of explainable machine learning. The green cluster addresses the use of deep neural networks and techniques aimed at improving transparency in artificial intelligence. The blue cluster leads research in advanced neural architectures and explainability in generative models. Finally, the purple cluster explores the convergence between robotics and model optimization, reflecting cross-disciplinary engagement between intelligent systems and algorithmic design.

These findings evidence the structure and evolution of the field of study, showing how different research streams converge around methodological approaches, deep neural networks, generative models, and their application in robotics. The density and connection between the clusters reflect a highly interconnected

knowledge dynamic, where specific authors and communities have played a central role in consolidating the discipline. The following section presents a discussion of the results obtained, analyzing their relevance in the current context. The implications of the patterns identified, their impact on the field's evolution, and the emerging trends that could influence future research will be addressed. Also, the remaining challenges and opportunities for developing more integrated and collaborative approaches in the study area will be described.

4 Discussion

After presenting the main results in Section 3, this section engages in a critical discussion of the main findings in light of the research objectives set out in the Introduction. The discussion is structured along six interrelated dimensions. Subsection 4.1 addresses the conceptual consolidation of SEAS and contrasts the results with previous literature. Subsection 4.2 focuses on methodological convergence and current heterogeneity in the use of explanatory techniques. Subsection 4.3 examines geographic asymmetries and structural fragmentation within the research ecosystem. Subsection 4.4 explores the stabilization of citation growth, questioning whether it indicates disciplinary maturity or thematic bifurcation. Subsection 4.5 discusses the absence of integrated theoretical frameworks capable of supporting system-level interpretability. Finally, subsection 4.6 reflects on normative convergence and the institutionalization of explainability in the context of emerging AI governance frameworks.

4.1 Conceptual Consolidation and Divergence with Prior Literature

The bibliometric analysis confirms a rapid and sustained growth in scientific production on SEAS, with an annual increase of 41.38%. This trend reflects an intensifying interest in developing transparent and interpretable AI systems, particularly in domains where reliability, traceability, and ethical oversight are relevant. In alignment with prior studies, this expansion reinforces the central role of Explainable AI in current research. Mahajan et al. [100] highlight the growing concern over model opacity in high-stakes contexts such as autonomous driving, where decision interpretability is essential for regulatory compliance and public trust. In contrast, Sadeghi [50] emphasizes the role of XAI in enhancing user engagement and accountability across decision systems.

However, our study diverges from these perspectives by showing that, while XAI has indeed been consolidated as a central concept, its implementation in SEAS research remains uneven and often domain-specific. The strong recurrence of terms such as "machine learning", "deep learning", and "trust" reflects that the interpretability discourse remains technically centered. Nevertheless, the increasing presence of socially oriented terms indicates a growing but still insufficient integration of ethical and human-centered considerations into the conceptual core of SEAS. This nuanced positioning, not fully addressed in previous literature, points to a fragmented epistemic structure in which SEAS is simultaneously shaped by technical innovation and normative imperatives.

To further situate these findings, other studies also emphasize the shift toward interdisciplinary integration. For example, Confalonieri et al. [101] note that, despite advances in technical methodologies, the lack of a unified theory of explainability limits the field's ability to generalize ideas across domains. Similarly, Mathew et al. [86] review emerging techniques in explainable artificial intelligence aimed at improving the interpretability of AI models, highlighting the ongoing challenges in achieving comprehensive human understanding. Kim et al. [102] explore how explainability can support human-AI interaction, emphasizing the importance of providing users with practical explanations that enhance collaboration with AI systems. In contrast, our analysis provides empirical evidence of this fragmentation, as reflected in thematic clusters and co-occurrence networks, thereby reinforcing the need for a broader, more integrative framework.

4.2 Methodological Convergence and Heterogeneity in Explanatory Approaches

The frequent appearance of methods such as SHAP and LIME signals methodological convergence around post hoc explainability techniques. These findings echo recent literature [103,104], confirming their relevance in high-risk contexts such as healthcare and autonomous vehicles. However, this study identifies a critical contrast: while the academic community widely references these techniques, their actual deployment remains inconsistent. Adoption varies not only across sectors but also geographically.

Furthermore, the absence of standardized metrics for comparing the quality of explanations limits the ability to assess their effectiveness objectively. This concern is echoed by Donoso-Guzmán et al. [105], who advocate for a comprehensive, human-centered evaluation framework that integrates explanation properties and user experience metrics.

Bridging this gap requires the development of shared evaluation frameworks that go beyond algorithmic performance and account for human interpretability, regulatory compliance, and operational constraints. Our bibliometric analysis, through the mapping of thematic clusters and co-occurrence networks, provides empirical evidence that more integrative and standardized approaches are needed in XAI.

4.3 Geographic Asymmetries and Structural Fragmentation

The geographic distribution of research is notably concentrated in the United States, China, and the United Kingdom. These countries dominate both publication output and citation impact, consolidating their leadership in SEAS research. However, this centralization contrasts sharply with the underrepresentation of Latin America, Africa, and parts of Southeast Asia. The international collaboration rate (7.39%) remains low, suggesting that, despite global interest in SEAS, scientific cooperation continues to be structurally limited. This imbalance reflects broader disparities in research infrastructure and access to funding.

As evidenced by previous interdisciplinary studies [106,107], these asymmetries can hinder the creation of inclusive and globally applicable standards for AI explainability. Selenica [108] further argues that the scientific system in the Global South is constrained by limited access to international funding and collaborations, perpetuating a cycle of exclusion. Similarly, Leslie and Perini [109] highlight persistent data and governance asymmetries in AI policy frameworks that disadvantage regions such as Latin America and Africa.

Moreover, this analysis demonstrates that dominant global AI assessment frameworks and indices often overlook the realities of local technology systems and infrastructures. In regions marked by disparity, accountable and explainable AI must be tailored to specific contexts, as universal standards can mask or misrepresent regional priorities. Our study contributes to this debate by offering a data-driven explanation of collaboration density and geographic concentration in the SEAS literature, highlighting the importance of policy interventions aimed at democratizing participation and increasing visibility in explainability research.

4.4 Stabilization in Citation Growth: Maturity or Thematic Bifurcation?

One of the most notable patterns is the apparent stabilization in citation growth over recent years. This trend may reflect a maturity phase in the field's theoretical foundations, with less emphasis on exploratory or conceptual articles and more on domain-specific implementation. Alternatively, it could signal a thematic bifurcation, where research diverges into specialized subfields with lower citation interconnectivity. Similar phenomena have been documented in adjacent AI domains, such as ethical reasoning and autonomous decision-making [110,111].

Recent bibliometric analyses support both interpretations. Costa and Frigori [112] identify shifts in citation networks in AI that resemble phase transitions, where periods of expansion give way to stabilization,

often coinciding with structural specialization. In this sense, our findings contribute to this debate by quantifying citation plateaus in SEAS and contextualizing them within analogous transitions in adjacent domains. This invites further meta-analytic scrutiny to determine whether SEAS is consolidating as a coherent discipline or dispersing into domain-driven application clusters.

4.5 Theoretical Integration and System-Level Interpretability

The findings of our bibliometric analysis reveal a critical gap in the integration of theoretical frameworks that unify interpretability, auditability, and adaptability within SEAS. Although substantial attention has been given to algorithmic transparency at the model level, there is a notable neglect of explanation design at the system level, where SEAS must function as cohesive, context-aware agents. This oversight becomes particularly problematic in real-world applications, where autonomous systems must continuously adapt to dynamic environments and interface meaningfully with human stakeholders.

Recent works emphasize that interpretability should not be confined to isolated model outputs but must extend to end-to-end system behaviors, enabling humans to trace decisions across multiple interacting subsystems. For instance, Li et al. [113] argue that trustworthy AI requires an ecosystem-level approach that aligns with legal, cognitive, and social dimensions of interpretability. Their framework proposes a layered architecture that integrates interpretability and auditability not only into the algorithms but also into the entire operational process that could be used for future SEAS design.

On the other hand, we believe that advancing the interpretability of SEAS requires interdisciplinary collaboration that articulates technical, cognitive, and normative knowledge. Veitch and Alsos [114] point out that the integration of human factors engineering enables the tailoring of explanations to different cognitive profiles, increasing user confidence and operational safety in complex environments, as is the case with maritime AI. At the normative level, Ziethmann et al. [115] indicate that legal informatics is crucial to ensure that explanatory mechanisms meet standards of accountability and compliance, especially in high-risk applications. In this context, our study contributes to the emerging debate by identifying research groups that promote integrative approaches while exposing the persistent theoretical fragmentation across disciplines, revealing a strategic opportunity to develop more cohesive, technical, cognitively refined, and normatively aligned SEAS.

4.6 Institutionalization and Regulatory Convergence

The progressive institutionalization of explainability, as reflected in instruments such as the EU AI Act and ISO/IEC 22989, underscores a growing regulatory commitment to embedding principles of transparency, accountability, and fairness within autonomous intelligent systems. However, our bibliometric analysis reveals that despite this normative momentum, research on SEAS remains unevenly aligned with regulatory trajectories. For instance, while recent contributions such as Schneeberger et al. [116] emphasize the need for conceptual clarity and harmonized legal frameworks, our co-citation and keyword analyses suggest that the SEAS literature rarely engages directly with these institutional developments. Similarly, although Dey [117] highlights the fragmented operationalization of XAI in sectors such as energy and infrastructure, our findings indicate that these domains are also underrepresented in SEAS research clusters.

This regulatory-scientific disconnect is further substantiated by our bibliometric evidence, which reveals that technological developments in SEAS often outpace regulatory codification and implementation. The intellectual structure uncovered in our analysis shows strong research concentration in domains such as healthcare and autonomous driving, sectors where legal frameworks and public scrutiny are more mature. In contrast, fields like smart homes, agriculture, and education remain peripheral within the SEAS research network despite their growing reliance on autonomous decision-making systems. This thematic asymmetry

likely reflects sector-specific disparities in regulatory pressure and deployment readiness. For instance, Lakshmi et al. [118] advocate for expanding the sustainability impact of explainable AI into domains such as waste management and precision agriculture, areas that are conspicuously underrepresented in SEAS discourse, as confirmed by our cluster mapping.

In parallel, the recent surge in LLMs introduces a technological inflection point and a new vector for research integration. These models offer promising avenues for enhancing the interpretability and accessibility of SEAS through natural language explanations, especially in environments where human-AI interaction is persistent and multimodal. Yet, they also complicate normative compliance due to their high computational demands, opaque internal representations, and limited auditability. Gadekallu et al. [119] frame this tension as central to the evolution of Industry 5.0 systems. Our contribution extends this view by emphasizing that LLMs not only serve as technical augmentations to SEAS but also as catalysts for interdisciplinary convergence, linking advances in natural language processing with longstanding concerns in explainability, user-centered design, and regulatory alignment.

Taken together, these findings position our bibliometric study as a strategic lens through which to identify both synergies and discontinuities in the SEAS knowledge ecosystem. By surfacing underexplored domains, fragmented conceptual linkages, and thematic blind spots, we highlight the need for a more integrative research agenda that embeds interpretability as a foundational design principle underpinned by shared ontologies, auditable protocols, and cross-sectoral alignment with regulatory imperatives.

5 Conclusions

This study offers a systematic cartography of the evolving research landscape in SEAS, revealing not only topical concentrations but also the structural dynamics shaping the field. Beyond tracing growth patterns in scientific production, the bibliometric approach employed here exposes the epistemic contours that govern how explainability is conceptualized, operationalized, and institutionally framed across domains. Rather than simply highlighting thematic gaps, the analysis underscores the uneven integration of interpretability as a foundational principle in autonomous systems research. By revealing the fragmentation of research communities and the underrepresentation of certain application areas and theoretical linkages, the study contributes to a more reflexive understanding of where SEAS research stands and where it may be strategically expanded. Future efforts should not only aim to bridge technical and regulatory developments but also promote a more cohesive knowledge architecture that supports cumulative, interdisciplinary, and policy-relevant advances in explainable autonomy.

5.1 Theoretical Implications

The results show that SEAS research is based on the convergence between machine learning, knowledge representation, and causal reasoning. The field's evolution has demonstrated that the interpretability of autonomous models is not only a technical problem but also a conceptual one, driving the development of hybrid approaches that combine data-driven models with symbolic explanatory structures.

Co-citation analysis reveals the existence of multiple methodological approaches in SEAS, suggesting a fragmentation of the field. This theoretical diversity underscores the need to develop a standard taxonomy to assess explainability, facilitating the comparison of models and their implementation in industrial contexts.

5.2 Practical Implications

The impact of SEAS on the industry is reflected in its growing application in sectors such as healthcare, security, and automation. However, the implementation of these systems faces barriers related to the

accessibility and understanding of explanations by non-expert users. Many explainable models require a high level of technical knowledge, which limits their adoption in operational environments.

The development of explainability tools should focus on generating intuitive interfaces adapted to different levels of expertise, ensuring that the explanations provided are interpretable and actionable. The low rate of international collaboration in SEAS research also suggests that, without global standards, the interoperability of these systems across different sectors will be limited. On the other hand, the bibliometric analysis shows that emerging regulation in artificial intelligence is beginning to influence SEAS development, indicating that the field's evolution will depend on its alignment with transparency, ethics, and security requirements.

These findings can serve as a basis for developing international collaborative policies in SEAS, fostering transnational research networks that reduce the field's fragmentation. Furthermore, the results can guide the formulation of standards to ensure the interoperability and applicability of SEAS in critical industries.

5.3 Limitations and Future Research

This study has some limitations that should be considered in future research. First, the exclusive reliance on the Scopus database may have excluded relevant publications indexed in other repositories such as Web of Science, IEEE Xplore, or ACM Digital Library, potentially affecting the representativeness and completeness of the bibliographic corpus. Second, the temporal scope of the study spans from January 2020 to February 2025, thus covering only a partial segment of the current year. While this may limit the capture of full-year publication trends for 2025, including early 2025 data was deliberate. It allows the study to reflect the most up-to-date scientific developments and research inflections, particularly relevant in a rapidly evolving and technologically sensitive field such as SEAS. As such, this choice strengthens the study's currency, even if it slightly constrains longitudinal consistency for the final year.

Although the temporal analysis shows a possible stabilization of the field, this phenomenon could be explained by a shift in the research direction towards more specialized applications rather than generalist studies on AI explainability. Future studies could further analyze whether this trend represents a consolidation of knowledge in SEAS or a shift towards new emerging areas within explainable AI.

Future research should focus on:

1. Evaluate how emerging AI regulation will impact SEAS design in different regions.
2. Develop standardized metrics and validate their applicability in industrial settings through case studies.
3. Explore hybrid approaches that combine machine learning with symbolic models to improve explainability.

The standardization of metrics for explainability and validation of SEAS in real environments will be a relevant factor in the future to ensure their effective integration in critical sectors. Without a unified framework to assess the transparency and reliability of these systems, their large-scale adoption could be limited. In this sense, progress in the field will depend on technological evolution and cooperation between researchers, regulators, and industry. Greater synergy among these actors will facilitate the creation of global standards that ensure the technical feasibility of SEAS while promoting their reliability, accessibility, and alignment with ethical and regulatory principles, thus enabling their responsible implementation in society.

Acknowledgement: Special acknowledge to my thesis supervisors (M. Gil & M. Albert) and to Computers, Materials & Continua for their support in improving the paper.

Funding Statement: This work has been partially funded by the Programa Nacional de Becas y Crédito Educativo of Peru and the Universitat de València, Spain.

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Oscar Peña-Cáceres and Henry Silva-Marchan; methodology, Oscar Peña-Cáceres; software, Elvis Garay-Silupú; validation, Darwin Aguilar-Chuquizuta, Henry Silva-Marchan and Oscar Peña-Cáceres; formal analysis, Oscar Peña-Cáceres; investigation, Oscar Peña-Cáceres and Henry Silva-Marchan; resources, Darwin Aguilar-Chuquizuta; data curation, Elvis Garay-Silupú and Darwin Aguilar-Chuquizuta; writing—original draft preparation, Oscar Peña-Cáceres; writing—review and editing, Oscar Peña-Cáceres and Henry Silva-Marchan; visualization, Darwin Aguilar-Chuquizuta and Elvis Garay-Silupú; supervision, Darwin Aguilar-Chuquizuta; project administration, Oscar Peña-Cáceres; funding acquisition, Darwin Aguilar-Chuquizuta and Henry Silva-Marchan. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: This manuscript is a bibliometric study, so the data are temporary and cannot be reproduced in other studies.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Islam MR, Ahmed MU, Barua S, Begum S. A systematic review of explainable artificial intelligence in terms of different application domains and tasks. Appl Sci. 2022;12(3):1353. doi:10.3390/app12031353.
2. Ziesche F, Klos V, Glesner S. Anomaly detection and classification to enable self-explainability of autonomous systems. In: Proceeding of 2021 Design, Automation and Test in Europe (DATE); 2021 Feb 1–5; Grenoble, France. p. 1304–9. doi:10.23919/DATE51398.2021.9474232.
3. Xie Y, Pongsakornsathien N, Gardi A, Sabatini R. Explanation of machine-learning solutions in air-traffic management. Aerospace. 2021;8(8):224. doi:10.3390/aerospace8080224.
4. Parra-Ullauri JM, García-Domínguez A, García-Paucar LH, Bencomo N. Temporal models for history-aware explainability. In: Proceedings of the 12th System Analysis and Modelling Conference, SAM 2020; 2020 Oct 19–20. Online. p. 155–64. doi:10.1145/3419804.3420276.
5. Kumar Dutta A, Albagory Y, Rahaman Wahab Sait A, Mohamed Keshta I. Autonomous unmanned aerial vehicles based decision support system for weed management. Comput Mater Contin. 2022;73(1):899–915. doi:10.32604/cmc.2022.026783.
6. Stanly A, Aruna K. Autonomous systems revolutionizing health insurance industry: achieving operational excellence in services. In: Modeling, simulation, and control of AI robotics and autonomous systems. Hershey, PA, USA: IGI Scientific Publishing; 2024. p. 131–51. doi:10.4018/979-8-3693-1962-8.ch008.
7. Whig P, Velu A, Nadikattu RR, Alkali YJ. Role of AI and IoT in intelligent transportation. Artificial intelligence for future intelligent transportation: smarter and greener infrastructure design. Palm Bay, FL, USA: Apple Academic Press; 2024. p. 199–220. doi:10.1201/9781003408468-8.
8. Agrawal TK, Hanson R, Sultan FA, Johansson MI, Andersson D, Stefansson G, et al. Automating loading and unloading for autonomous transport: identifying challenges and requirements with a systems approach. IFIP Adv Inf Commun Technol. 2023;691:332–45. doi:10.1007/978-3-031-43670-3_23.
9. Fey G, Fränzle M, Drechsler R. Self-explanation in systems of systems. In: Proceedings of the IEEE International Conference on Requirements Engineering; 2022 Aug 15–19; Melbourne, VIC, Australia. p. 85–91. doi:10.1109/REW56159.2022.00023.
10. Aurangzeb S, Aleem M, Khan MT, Anwar H, Siddique MS. Cybersecurity for autonomous vehicles against malware attacks in smart-cities. Cluster Comput. 2024;27(3):3363–78. doi:10.1007/s10586-023-04114-7.
11. Kiraz M, Sivrikaya F, Albayrak S. A survey on sensor selection and placement for connected and automated mobility. IEEE Open J Intell Transp Syst. 2024;5(7):692–710. doi:10.1109/OJITS.2024.3481328.
12. Chavan A, Ambilpure S, Chhapra U, Gawde V. Autonomous home-security system using internet of things and machine learning. In: Lecture notes on data engineering and communications technologies. Cham, Switzerland: Springer International Publishing; 2020. Vol. 46, p. 498–504. doi:10.1007/978-3-030-38040-3_56.

13. Li F, Lu Y. Human-AI interaction and ethics of AI: how well are we following the guidelines. In: Chinese CHI'22: Proceedings of the Tenth International Symposium of Chinese CHI; 2022 Oct 22–23; Guangzhou, China, p. 96–104. doi:10.1145/3565698.3565773.
14. Zhou J, Chen F, Berry A, Reed M, Zhang S, Savage S. A survey on ethical principles of AI and implementations. In: 2020 IEEE Symposium Series on Computational Intelligence, SSCI 2020; 2020 Dec 1–4; Canberra, ACT, Australia, p. 3010–7. doi:10.1109/SSCI47803.2020.9308437.
15. Kozhevnikova M, Karpova SV. Artificial intelligence: subject and object; [Iskusstvennyi intellekt: sub'ekt i ob'ekt]. Etnograficeskoe Obozrenie. 2020;2020(1):80–79. doi:10.31857/S086954150008759-4.
16. Noorwali A, Awais Javed M, Zubair Khan M. Efficient UAV communications: recent trends and challenges. Comput Mater Contin. 2021;67(1):463–76. doi:10.32604/cmc.2021.014668.
17. Tekkesinoglu S, Habibovic A, Kunze L. Advancing explainable autonomous vehicle systems: a comprehensive review and research roadmap. ACM Trans Hum-Robot Interact. 2025;14(3):1–46. doi:10.1145/3714478.
18. Chaal M, Ren X, BahooToroody A, Basnet S, Bolbot V, Banda OAV, et al. Research on risk, safety, and reliability of autonomous ships: a bibliometric review. Saf Sci. 2023;167(2):106256. doi:10.1016/j.ssci.2023.106256.
19. Ho JS, Tan BC, Lau TC, Khan N. Public acceptance towards emerging autonomous vehicle technology: a bibliometric research. Sustain. 2023;15(2):1566. doi:10.3390/su15021566.
20. Azam M, Hassan SA, Che Puan O. Autonomous vehicles in mixed traffic conditions—a bibliometric analysis. Sustain. 2022;14(17):10743. doi:10.3390/su141710743.
21. Gandia RM, Antonialli F, Cavazza BH, Neto AM, Lima DAD, Sugano JY, et al. Autonomous vehicles: scientometric and bibliometric review. Transp Rev. 2019;39(1):9–28. doi:10.1080/01441647.2018.1518937.
22. Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. How to conduct a bibliometric analysis: an overview and guidelines. J Bus Res. 2021;133(5):285– 96. doi:10.1016/j.jbusres.2021.04.070.
23. Marvi R, Foroudi MM. Bibliometric analysis: main procedure and guidelines. Researching and analysing business: research methods in practice. 1st ed. London, UK: Routledge; 2023. p. 43–54. doi:10.4324/9781003107774-4.
24. Haddow G. Bibliometric research. In: Research methods: information, systems, and contexts. 2nd. Hull, UK: Chandos Publishing; 2018. p. 241–66. doi:10.1016/B978-0-08-102220-7.00010-8.
25. Reshi AA, Shah A, Shafi S, Qadri MH. Big data in healthcare—a comprehensive bibliometric analysis of current research trends. Scalable Comput. 2023;24(3):531–49. doi:10.12694/scpe.v24i3.2155.
26. Baas J, Schotten M, Plume A, Côté G, Karimi R. Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. Quant Sci Stud. 2020;1(1):377–86. doi:10.1162/qss_a_00019.
27. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews; [A declaração PRISMA 2020: diretriz atualizada para relatar revisões sistemáticas]; [Declaración PRISMA 2020: una guía actualizada para la publicación de revisiones sistemáticas]. Rev Panam Salud Publica/Pan Am J Public Health. 2022;46:e112. doi:10.26633/RPSP.2022.112.
28. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. Updating guidance for reporting systematic reviews: development of the PRISMA 2020 statement. J Clin Epidemiol. 2021;134(1):103–12. doi:10.1016/j.jclinepi.2021.02.003.
29. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol. 2005;8(1):19–32. doi:10.1080/1364557032000119616.
30. Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. FASEB J. 2008;22(2):338–42. doi:10.1096/fj.07-9492LSF.
31. Mongeon P, Paul-Hus A. The journal coverage of Web of Science and Scopus: a comparative analysis. Scientometrics. 2016;106(1):213–28. doi:10.1007/s11192-015-1765-5.
32. Peña-Cáceres O. MASHA: an online platform for metrics, analysis, science, hub and analytics. CISAI; 2025. doi:10.5281/zenodo.14933500.
33. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion. 2020;58(3):82–115. doi:10.1016/j.inffus.2019.12.012.

34. Alonso JM, Castiello C, Mencar C. A bibliometric analysis of the explainable artificial intelligence research field. In: Information processing and management of uncertainty in knowledge-based systems. Theory and foundations. Cham, Switzerland: Springer International Publishing; 2018. Vol. 853, p. 3–15. doi:10.1007/978-3-319-91473-2_1.
35. Sharma C, Sharma S, Sharma K, Sethi GK, Chen HY. Exploring explainable AI: a bibliometric analysis. Discov Appl Sci. 2024;6(11):615. doi:10.1007/s42452-024-06324-z.
36. Abramo G, D'Angelo CA, Cicero T. What is the appropriate length of the publication period over which to assess research performance? Scientometrics. 2012;93(3):1005–17. doi:10.1007/s11192-012-0714-9.
37. Kitchenham BA, Mendes E, Travassos GH. Cross versus within-company cost estimation studies: a systematic review. IEEE Trans Softw Eng. 2007;33(5):316–29. doi:10.1109/TSE.2007.1001.
38. Shaw J. How machine learning aids material selection. Manuf Eng. 2024;172(3):26.
39. Kramer O. Machine learning. Stud Comput Intell. 2017;679:65–72. doi:10.1007/978-3-319-52156-5_8.
40. Dayal M, Gupta M, Gupta M, Bara AR, Chaubey C. Introduction to machine learning methods with application in agriculture. In: Applying drone technologies and robotics for agricultural sustainability. Hershey, PA, USA: IGI Global; 2023. p. 184–203. doi:10.4018/978-1-6684-6413-7.ch012.
41. Sathya D, Sudha V, Jagadeesan D. Application of machine learning techniques in healthcare. In: Research anthology on machine learning techniques, methods, and applications. Hershey, PA, USA: IGI Global; 2022. p. 1294–310. doi:10.4018/978-1-6684-6291-1.ch067.
42. Kumar H, Hasija Y. Machine learning in medical image processing. In: Information and Communication Technology for Intelligent Systems (ICTIS 2020). Singapore: Springer; 2021. Vol. 195, p. 377–83. doi:10.1007/978-981-15-7078-0_35.
43. Thakur UK. The role of machine learning in customer experience. In: Handbook of research on ai and machine learning applications in customer support and analytics. Hershey, PA, USA: IGI Global; 2023. p. 80–9. doi:10.4018/978-1-6684-7105-0.ch005.
44. Bastawrous A, Cleland C. Artificial intelligence in eye care: a cautious welcome. Community Eye Health J. 2022;35(114):13.
45. Darda P, Pendse MK. The impact of artificial intelligence (AI) transformation on the financial sector from the trading to security operations. Shaping cutting-edge technologies and applications for digital banking and financial services. 1st. New York, NY, USA: Productivity Press; 2025. p. 322–39. doi:10.4324/9781003501947-20.
46. Cappello G, Defeudis A, Giannini V, Mazzetti S, Regge D. Artificial intelligence in oncologic imaging. In: Multimodality imaging and intervention in oncology. Cham, Switzerland: Springer; 2023. p. 585–97. doi:10.1007/978-3-031-28524-0_24.
47. Sharma N, Jindal N. Emerging artificial intelligence applications: metaverse, IoT, cybersecurity, healthcare—an overview. Multimed Tools Appl. 2024;83(19):57317–45. doi:10.1007/s11042-023-17890-6.
48. Sardar TH, Das S, Pandey BK. Explainable AI (XAI): concepts and theory. In: Medical data analysis and processing using explainable artificial intelligence. 1st ed. Boca Raton, FL, USA: CRC Press; 2023. p. 1–18. doi:10.1201/9781003257721-1.
49. Apicella A, Giugliano S, Isgrò F, Prevete R. SHAP-based explanations to improve classification systems. CEUR Workshop Proc. 2023;3518:76–86.
50. Sadeghi Z, Alizadehsani R, CIFCI MA, Kausar S, Rehman R, Mahanta P, et al. A review of explainable artificial intelligence in healthcare. Comput Elect Eng. 2024;118(5):109370. doi:10.1016/j.compeleceng.2024.109370.
51. Nyrup R, Robinson D. Explanatory pragmatism: a context-sensitive framework for explainable medical AI. Ethics Inf Technol. 2022;24(1):13. doi:10.1007/s10676-022-09632-3.
52. Khakurel U, Rawat DB. Evaluating explainable artificial intelligence (XAI): algorithmic explanations for transparency and trustworthiness of ML algorithms and AI systems. In: Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications IV; 2022 Apr 3–7; Orlando, FL, USA. Vol. 12113. doi:10.1117/12.2620598.
53. Rizzo M, Veneri A, Albarelli A, Lucchese C, Nobile M, Conati C. A theoretical framework for AI models explainability with application in biomedicine. In: CIBCB 2023—20th IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology; 2023 Aug 29–31; Eindhoven, Netherlands. p. 1–9. doi:10.1109/CIBCB56990.2023.10264877.

54. Sharma D. Deep learning without tears: a simple introduction. Resonance. 2020;25(1):15–32. doi:10.1007/s12045-019-0919-9.
55. Narayanan N, Arjun KP. Introduction to deep learning. In: Artificial intelligence for precision agriculture. Boca Raton, FL, USA: Auerbach Publications; 2024. doi:10.1201/9781003504900-2.
56. Zhou X. Application of deep learning in ocean big data mining. J Coast Res. 2020;106(sp1):614–7. doi:10.2112/SI106-139.1.
57. Bhargavi K. Deep learning architectures and tools: a comprehensive survey. Deep learning applications and intelligent decision making in engineering. Hershey, PA, USA: IGI Global; 2020. p. 55–75. doi:10.4018/978-1-7998-2108-3.ch002.
58. Badiger M, Mathew JA. Retrospective review of activation functions in artificial neural networks. In: Proceedings of Third International Conference on Communication, Computing and Electronics Systems. Singapore: Springer; 2022. Vol. 844, p. 905–19. doi:10.1007/978-981-16-8862-1_59.
59. Dargan S, Kumar M, Ayyagari MR, Kumar G. A survey of deep learning and its applications: a new paradigm to machine learning. Arch Comput Methods Eng. 2020;27(4):1071–92. doi:10.1007/s11831-019-09344-w.
60. Scotti V. Artificial intelligence. IEEE Instru Meas Mag. 2020;23(3):27–31. doi:10.1109/MIM.2020.9082795.
61. Zibner J. Subjects' relevance within an AI-included creative process. Jusletter IT. Editions Weblaw; 2019. [cited 2025 May 10]. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85073062598&partnerID=4&md=ebbf9dc27f3ec1b511695a566175d001.
62. Moses LB. Artificial intelligence: affordances and limits in the context of judging. J Proc R Soc New South Wales. 2024;157:123–9.
63. Pathak S, Solanki VK. Impact of internet of things and artificial intelligence on human resource development. In: Intelligent systems reference library. Cham, Switzerland: Springer; 2021. Vol. 193. p. 239–67. doi:10.1007/978-3-030-57835-0_19.
64. Mamad M, Chichi O. Towards a human-centred artificial intelligence in the age of industry 5.0: a cross-country analysis. In: 2024 11th International Conference on Wireless Networks and Mobile Communications (WINCOM); 2024 Jul 23–25; Leeds, UK. p. 1–6. doi:10.1109/WINCOM62286.2024.10655038.
65. Berberian B, Le Guillou M, Pagliari M. Communicating AI intentions to boost human AI cooperation. CEUR Workshop Proc. 2023;3456:145–9.
66. Airaj M. Ethical artificial intelligence for teaching-learning in higher education. Educ Inform Technol. 2024;29(13):17145–67. doi:10.1007/s10639-024-12545-x.
67. Champendal M, Müller H, Prior JO, Dos Reis CS. A scoping review of interpretability and explainability concerning artificial intelligence methods in medical imaging. Eur J Radiol. 2023;169:111159. doi:10.1016/j.ejrad.2023.111159.
68. Zeineldin RA, Karar ME, Elshaer Z, Coburger J, Wirtz CR, Burgert O, et al. Explainability of deep neural networks for MRI analysis of brain tumors. Int J Comput Assist Radiol Surgl. 2022;17(9):1673–83. doi:10.1007/s11548-022-02619-x.
69. Atakishiyev S, Salameh M, Yao H, Goebel R. Towards safe, explainable, and regulated autonomous driving. In: Explainable artificial intelligence for intelligent transportation systems. 1st ed. Boca Raton, FL, USA: CRC Press; 2023. p. 32–52. doi:10.1201/9781003324140-2.
70. Kuznietsov A, Gyevnar B, Wang C, Peters S, Albrecht SV. Explainable AI for safe and trustworthy autonomous driving: a systematic review. IEEE Trans Intell Transp Syst. 2024;25(12):19342–64. doi:10.1109/TITS.2024.3474469.
71. Dong J, Chen S, Miralinaghi M, Chen T, Li P, Labi S. Why did the AI make that decision? Towards an explainable artificial intelligence (XAI) for autonomous driving systems. Transp Res Part C Emerg Technol. 2023;156(4):104358. doi:10.1016/j.trc.2023.104358.
72. Tilbury J, Flowerday S. Humans and automation: augmenting security operation centers. J Cybersecur Priv. 2024;4(3):388–409. doi:10.3390/jcp4030020.
73. Oviedo J, Rodriguez M, Trenta A, Cannas D, Natale D, Piattini M. ISO/IEC quality standards for AI engineering. Comput Sci Rev. 2024;54(1):100681. doi:10.1016/j.cosrev.2024.100681.

74. Luwel M, Van Eck NJ, Van Leeuwen T. Characteristics of publication delays over the period 2000–2016. In: Evaluative informetrics: the art of metrics-based research assessment. Cham, Switzerland: Springer International Publishing; 2020. p. 89–114. doi:10.1007/978-3-030-47665-6_4.
75. Stremersch S, Verniers I, Verhoef PC. The quest for citations: drivers of article impact. J Mark. 2007;71(3):171–93. doi:10.1509/jmkg.71.3.171.
76. Dorta-González P, Dorta-González MI. Impact maturity times and citation time windows: the 2-year maximum journal impact factor. J Inform. 2013;7(3):593–602. doi:10.1016/j.joi.2013.03.005.
77. Jiang H, Zhou J, Ding Y, Zeng A. Overcoming recognition delays in disruptive research: the impact of team size, familiarity, and reputation. J Inform. 2024;18(4):101549. doi:10.1016/j.joi.2024.101549.
78. Schvirck E, Lievore C, Rubbo P, Herrera Cantorani JR, Pilatti LA. Invisible publications: a study of academic productivity in the Web of Science database. Rev Esp Doc Cient. 2024;47(1):e375. doi:10.3989/redc.2024.1.1454.
79. Majhi S, Sahu L, Behera K. Practices for enhancing research visibility, citations and impact: review of literature. Aslib J Inf Manag. 2023;75(6):1280–305. doi:10.1108/AJIM-11-2023-532.
80. Trivedi C, Bhattacharya P, Prasad VK, Patel V, Singh A, Tanwar S, et al. Explainable AI for industry 5.0: vision, architecture, and potential directions. IEEE Open J Ind Appl. 2024;5(3):177–208. doi:10.1109/OJIA.2024.3399057.
81. Atakishiyev S, Salameh M, Yao H, Goebel R. Explainable artificial intelligence for autonomous driving: a comprehensive overview and field guide for future research directions. IEEE Access. 2024;12(3):101603–25. doi:10.1109/ACCESS.2024.3431437.
82. Ahmed I, Jeon G, Piccialli F. From artificial intelligence to explainable artificial intelligence in industry 4.0: a survey on what, how, and where. IEEE Trans Industr Inform. 2022;18(8):5031–42. doi:10.1109/TII.2022.3146552.
83. Zablocki E, Ben-Younes H, Pérez P, Cord M. Explainability of deep vision-based autonomous driving systems: review and challenges. Int J Comput Vis. 2022;130(10):2425–52. doi:10.1007/s11263-022-01657-x.
84. Nasarian E, Alizadehsani R, Acharya UR, Tsui KL. Designing interpretable ML system to enhance trust in healthcare: a systematic review to proposed responsible clinician-AI-collaboration framework. Inf Fusion. 2024;108(1):102412. doi:10.1016/j.inffus.2024.102412.
85. Singh SK, Virdee B, Aggarwal S, Maroju A. Incorporation of XAI and deep learning in biomedical imaging: a review. Polytech J. 2025;15(1):1. doi:10.59341/2707-7799.1845.
86. Mathew DE, Ebem DU, Ikegwu AC, Ukeoma PE, Dibiaezue NF. Recent emerging techniques in explainable artificial intelligence to enhance the interpretable and understanding of AI models for human. Neural Process Lett. 2025;57(1):16. doi:10.1007/s11063-025-11732-2.
87. Zhang B, Meroño Peñuela A, Simperl E. Towards explainable automatic knowledge graph construction with human-in-the-loop. In: Frontiers in artificial intelligence and applications. Amsterdam, Netherlands: IOS Press; 2023. p. 274–89. doi:10.3233/FAIA230091.
88. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell. 2020;2(1):56– 67. doi:10.1038/s42256-019-0138-9.
89. Linardatos P, Papastefanopoulos V, Kotsiantis S. Explainable ai: a review of machine learning interpretability methods. Entropy. 2021;23(1):1–45. doi:10.3390/e23010018.
90. Shin D. The effects of explainability and causability on perception, trust, and acceptance: implications for explainable AI. Int J Hum Comput St. 2021;146(83):102551. doi:10.1016/j.ijhcs.2020.102551.
91. Yang G, Ye Q, Xia J. Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: a mini-review, two showcases and beyond. Inf Fusion. 2022;77(1):29– 52. doi:10.1016/j.inffus.2021.07.016.
92. Mohseni S, Zarei N, Ragan ED. A multidisciplinary survey and framework for design and evaluation of explainable AI systems. ACM Trans Interact Intell Syst. 2021;11(3–4):24. doi:10.1145/3387166.
93. Dwivedi R, Dave D, Naik H, Singhal S, Omer R, Patel P, et al. Explainable AI (XAI): core ideas, techniques, and solutions. ACM Comput Surv. 2023;55(9):194. doi:10.1145/3561048.
94. Saeed W, Omlin C. Explainable AI (XAI): a systematic meta-survey of current challenges and future opportunities. Knowl Based Syst. 2023;263(3):110273. doi:10.1016/j.knosys.2023.110273.
95. Shamim Hossain M, Muhammad G, Guizani N. Explainable AI and mass surveillance system-based healthcare framework to combat COVID-I9 like pandemics. IEEE Netw. 2020;34(4):126– 132. doi:10.1109/MNET.011.2000458.

96. Holzinger A, Malle B, Saranti A, Pfeifer B. Towards multi-modal causability with graph neural networks enabling information fusion for explainable AI. Inf Fusion. 2021;71(7639):28– 37. doi:10.1016/j.inffus.2021.01.008.
97. Chaddad A, Peng J, Xu J, Bouridane A. Survey of explainable AI techniques in healthcare [Review]. Sensors. 2023;23(2):634. doi:10.3390/s23020634.
98. Nerur SP, Rasheed AA, Natarajan V. The intellectual structure of the strategic management field: an author co-citation analysis. Strateg Manag J. 2008;29(3):319–36. doi:10.1002/smj.659.
99. González-Valiente CL, León Santos M, Arencibia-Jorge R, Noyons E, Costas R. Mapping the evolution of intellectual structure in information management using author co-citation analysis. Mob Netw Appl. 2021;26(6):2374–88. doi:10.1007/s11036-019-01231-9.
100. Mahajan Y, Sharma M, Singh I. Real-world applications of explainable AI in healthcare. In: Federated learning and privacy-preserving in healthcare AI. Hershey, PA, USA: IGI Global; 2024. p. 158–78. doi:10.4018/979-8-3693-1874-4.ch011.
101. Confalonieri R, Coba L, Wagner B, Besold TR. A historical perspective of explainable artificial intelligence. WIREs Data Min Knowl Discov. 2021;11(1):e1391. doi:10.1002/widm.1391.
102. Kim SSY, Watkins EA, Russakovsky O, Fong R, Monroy-Hernández A. "Help Me Help the AI": understanding how explainability can support human-AI interaction. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. CHI '23; 2023 Apr 23–28; Hamburg, Germany. 250 p. doi:10.1145/3544548.3581001.
103. Salih AM, Raisi-Estabragh Z, Galazzo IB, Radeva P, Petersen SE, Lekadir K, et al. A perspective on explainable artificial intelligence methods: SHAP and LIME. Adv Intell Syst. 2025;7(1):2400304. doi:10.1002/aisy.202400304.
104. Bhatnagar S, Agrawal R. Understanding explainable artificial intelligence techniques: a comparative analysis for practical application. Bull Electr Eng Inform. 2024;13(6):4451–5. doi:10.11591/eei.v13i6.8378.
105. Donoso-Guzmán I, Ooge J, Parra D, Verbert K. Towards a comprehensive human-centred evaluation framework for explainable AI. In: Explainable artificial intelligence. Cham, Switzerland: Springer Nature; 2023. Vol. 1903, p. 183–204. doi:10.1007/978-3-031-44070-0_10.
106. Hutson J, Plate D. Disrupting algorithmic culture: redefining the human(ities). In: Generative AI in teaching and learning. Hershey, PA, USA: IGI Global; 2024. p. 1–30. doi:10.4018/979-8-3693-0074-9.ch001.
107. Abbonato D, Bianchini S, Gargiulo F, Venturini T. Interdisciplinary research in artificial intelligence: lessons from COVID-19. Quant Sci Stud. 2024;5(4):922–35. doi:10.1162/qss_a_00329.
108. Selenica E. The scientific system in the Global South in an emerging multipolar world. Glob Soc Educ. 2025;23(2):393–409. doi:10.1080/14767724.2023.2209513.
109. Leslie D, Perini AM. Future shock: generative AI and the international AI policy and governance crisis. Harvard Data Science Review. 2024. doi:10.1162/99608f92.88b4cc98.
110. Jedlickova A. Ensuring ethical standards in the development of autonomous and intelligent systems. IEEE Trans Artif Intell. 2024;5(12):5863–72. doi:10.1109/TAI.2024.3387403.
111. Machado J, Sousa R, Peixoto H, Abelha A. Ethical decision-making in artificial intelligence: a logic programming approach. AI. 2024;5(4):2707–24. doi:10.3390/ai5040130.
112. Costa AA, Frigori RB. Complexity and phase transitions in citation networks: insights from artificial intelligence research. Front Res Metr Anal. 2024;9:1456978. doi:10.3389/frma.2024.1456978.
113. Li B, Qi P, Liu B, Di S, Liu J, Pei J, et al. Trustworthy AI: from principles to practices. ACM Comput Surv. 2023;55(9):177–46. doi:10.1145/3555803.
114. Veitch E, Alsos OA. Human-centered explainable artificial intelligence for marine autonomous surface vehicles. J Mar Sci Eng. 2021;9(11):1227. doi:10.3390/jmse9111227.
115. Ziethmann P, Stieler F, Pfrommer R, Schlögl-Flierl K, Bauer B. Towards a framework for interdisciplinary studies in explainable artificial intelligence. In: Artificial intelligence in HCI. Cham, Switzerland: Springer Nature; 2024. Vol. 14734, p. 316–33. doi:10.1007/978-3-031-60606-9_18.
116. Schneeberger D, Röttger R, Cabitza F, Campagner A, Plass M, Müller H, et al. The tower of babel in explainable artificial intelligence (XAI). In: Machine learning and knowledge extraction. Cham, Switzerland: Springer Nature; 2023. Vol. 14065, p. 65–81. doi:10.1007/978-3-031-40837-3_5.

117. Dey MT. Explainable artificial intelligence (XAI): integration, policy frameworks, and applications in critical domains and renewable energy. In: Advances in environmental engineering and green technologies. Hershey, PA, USA: IGI Global; 2024. p. 333–62. doi:10.4018/979-8-3693-7822-9.ch012.
118. Lakshmi D, Tiwari RS, Dhanaraj RK, Kadry S. Explainable AI (XAI) for sustainable development: trends and applications. Boca Raton, FL, USA: CRC Press; 2024. [cited 2025 May 10]. Available from: https://books.google.es/books?id=p2tg0AEACAAJ.
119. Gadekallu TR, Maddikunta PKR, Boopathy P, Deepa N, Chengoden R, Victor N, et al. XAI for industry 5.0-concepts, opportunities, challenges and future directions. IEEE Open J Commun Soc. 2025;6(11):2706–29. doi:10.1109/OJCOMS.2024.3473891.