Quality & Quantity
https://doi.org/10.1007/s11135-025-02437-8

Exploring semantic relationships and cross-disciplinary influences: case study of information systems and artificial intelligence

Amit K. Shukla[1] · Vagan Terziyan[2] · Timo Tiihonen[2]

Received: 12 March 2025 / Accepted: 3 October 2025

Abstract
Generally, complex problems often require contributions from several research methods or research areas. The simultaneous use of several approaches gives occasion to the cross-fertilization between disciplines, which may even help the individual disciplines to develop. It is thus natural to ask how disciplines influence each other and if this can be studied using the best available methods. We approach this by computing semantic relationships within the publication records and new semantic measures are introduced to study the mutual influence. As an example, we examine the research areas of Artificial Intelligence (AI) and Information Systems (IS). It's often stated that the study of IS is characterized using a wide range of research methodologies to examine and guide areas of interest within the field. This has led the IS community to perceive that it can guide AI by clarifying research questions and helping achieve goals. However, this doesn't establish the core contribution of IS outside its domain. The similarity between the research areas is studied from multiple perspectives, such as conditional dependence, distance, citations/reference analysis, and journal analysis. The analysis and results are also studied with a perspective of bibliometric data from bibliographic indexing platforms such as Web of Science and Scopus. The semantic relationship and bibliometric analyses show that IS and AI concepts co-occur to some degree which reflects the interaction between the research areas. The role of AI has rocketed in recent years with top-cited AI papers having a greater impact on IS than IS papers on AI research.

Keywords Semantic similarity · Conditional probability · Bibliometric · VOSviewer · Information systems · Artificial Intelligence

1 Introduction

Different approaches and even different scientific disciplines are needed to tackle real problems. Thus, various concepts, methods, and disciplines interact, collaborate, and influence each other. Oftentimes, progress in one field triggers interest and development in other fields

Extended author information available on the last page of the article

Published online: 27 October 2025

as well, which results in observable alignment of the fields. There are quantitative measures for alignment between different concepts, distance, or similarity metrics; for example, one relatively easily exploitable method is to use keyword-based index metrics on a database like Google (Cilibrasi and Vitanyi 2007; Cohen and Vitányi 2013). However, such distance metrics alone do not provide inherent information about the type of interaction between the concepts or research approaches, not to mention the primary direction of influence. Moreover, they are subject to limitations of the Google search engine, which does not account for the quality of the web pages and the relevance of the search results. Additionally, these methods are supplemented with a support vector machine classifier which can be computationally expensive (Cilibrasi and Vitanyi 2007; Cohen and Vitányi 2013). Therefore, this work aims to introduce tools or metrics that could reveal more about the role and semantics of interaction between two concepts or research areas.

While cross-disciplinary comparisons could be made between any two relevant fields, the selection of Information Systems (IS) and Artificial Intelligence (AI) is both strategic and topical. Recent studies have identified a significant overlap between these two research fields (Chen et al. 2022; Gupta et al. 2021). AI has become an integral part of intrinsic data analysis, visualization, prediction, and inference, with capabilities of increasingly complex operations. The core AI community is focused on developing novel ideas, algorithms and advancing them towards real-world applications, enhancing innovation and technical progress. Meanwhile, the IS community argues that research in AI is unclear and that IS guides the AI community on a directed path to identify correct research questions and accomplish research (Ågerfalk et al. 2021). This claim, though strong, needs to be validated scientifically. Regardless, it is evident that IS exploits more and more AI capabilities. The field of IS is broadly associated with developing theories where the researchers utilize a multitude of research methodologies to explore/guide a particular area of interest.

If core AI is focused on the "how" aspect of technological invention, IS borrows these inventions and provides a "set of guidelines" on how to use these in a productive manner. The relationship between IS and AI can be metaphorically described as "the minds that plan and the hands that build" (Ågerfalk et al. 2021; Wiederhold 1992). It has long been asserted that IS contributes methodological and contextual grounding to the technical advances of AI, claiming a guiding role in shaping AI's overall research (Ågerfalk et al. 2021; Wiederhold 1992). It's also evident that AI capabilities are usually embedded within IS applications and research, representing a growing and practical convergence. This bidirectional relationship, which is often acknowledged is rarely quantified, and thus necessitates a systematic investigation. So, essentially, these two research areas seem to intersect radically and raise a concern to explore how much research on the core AI domain is influenced or dominated by the research on IS or research methods, and vice-versa? This debate has triggered questions to investigate quantitively and presents the scientific evidence for the impact and semantic relationships between AI and IS research.

The quantitative measures to compute the impact of a research field on another research field may include several publication metrics such as citations, usage, publication trends, and impact factors (Agarwal et al. 2016). Therefore, in this paper, we consider scientific publications as a reliable source for the analysis of the semantic relationship, overcoming the above-stated drawbacks of traditional Google search-based semantic metrics. The publications analysis also represents the semantic relationship between two research areas in how they connect, relate, and refer to each other. For this purpose, we introduce conditional probability-based measures to evaluate semantic relationships between AI and IS. Such a measure has the ability to identify causal relationships between two concepts. We have conducted different ways to extract semantic information about the co-occurrence of different concepts or terms in relationship to various classifications and labelling's of research. Several available classification aspects are considered, such as Web of Science (WOS) categorizations, keywords (from WOS and Scopus), and references/citations to the top cited papers.

The conditional probabilities are used to define similarity as their harmonic mean. Further, we also computed the distance between AI and IS, which is the contra-harmonic mean of the component dissimilarities. In addition, we also studied the relationship between the two fields using Normalized Index Distance (NID), a measure we introduce which is inspired by the well-known Normalized Google Distance (NGD) (Cohen and Vitányi 2013). NGD is used to assess similarity and dissimilarity among keywords based on their co-occurrence in Google search results, while NID adapts this to bibliographic indexing platforms such as WOS and Scopus. NGD is based on raw web search engine result counts, which may include non-academic, ambiguous, or noisy content. Also, web search algorithms are proprietary and may not produce reproducible results. In contrast, NID operates on structured bibliographic indexes such as WOS and Scopus in our case, which allows for more stable, transparent, and domain-specific measurement of co-occurrence between academic fields. Another study we examined, which is based on the journal-based analysis to assess relationship between AI on IS or vice-versa. Lastly, the conclusions from conditional probability-based metrics are then checked against more conventional bibliometric analysis.

The essence of this research work can be compiled as the study to understand the influence of a research field on another with the help of conditional probability-based semantic measures and quantifiable analysis based on scientific publications. Our work draws from theories in scientometrics and interdisciplinary studies. Earlier research work has used co-citation, keyword overlap, and publication trends to understand knowledge transfer and boundary-spanning behavior across different fields (Leydesdorff 2007; Rafols and Meyer 2010). Interdisciplinary research can take several forms including loose borrowing to full integration, and citation-based metrics have been utilized to highlight these relationships. By analyzing the semantic and bibliometric connection between AI and IS, we explore the volume of exchange and also its directionality and intensity, which are central to theories of cross-disciplinary knowledge dynamics. In particular, we consider how disciplines may interact through unbalanced knowledge flows, where one field functions predominantly as a source of technical or epistemic authority. These dynamics have been discussed in terms of epistemic dependence, where fields like IS may adopt methods and insights from AI without reciprocally shaping AI's research agenda. Such patterns are also addressed in boundary-spanning theory, which examines how disciplines import external knowledge while maintaining their own identity and relevance (Carlile 2004; Frickel and Gross 2005). These perspectives motivate our use of semantic asymmetry and bibliometric distance measures to quantify and characterize the directional dynamics between IS and AI. Additionally, our analysis could help academic institutions and policymakers in curriculum design, research evaluation, and strategic planning by highlighting the actual dynamics between foundational technological development and its application or contextualization in ISs.

Our work presents both methodological contributions and empirical findings into the relationship between IS and AI, which are as follows:

1. We introduce a conditional probability-based measure to quantify the semantic asymmetry between scientific fields.
2. We further computed semantic similarity and semantic distance using harmonic and contra-harmonic means of conditional dependencies.
3. A novel NID similarity measure is also introduced, adapted from NGD, to measure semantic proximity using bibliometric platforms such as WOS and Scopus.
4. Our approaches are validated using multiple classification strategies (WOS categories, keyword classification, and reference-based analysis).
5. Research in IS shows increasing semantic dependence on AI, while AI shows minimal referencing of IS work, particularly in core subfields.
6. Highly influential AI papers are significantly cited in IS works, but not vice versa, which indicates asymmetry in intellectual influence.
7. Bibliometric analysis is also conducted for validation that AI-centric concepts dominate influential IS publications, while IS methods play a partial or negligible role in AI progress.

The paper is organized as follows: Sect. 2 investigates the semantic similarity measure between two research areas on the background of standard statistics. A novel conditional dependence-based semantic similarity is introduced, and several other aspects such as distance, citations/reference analysis. Section 3 covers the journal analysis. Section 4 performs the in-depth bibliometric analysis of various performance indicators. Finally, Sect. 5 discusses the outcomes along with limitations and concludes the paper.

2 Analysis of semantic relationships

This section presents manifold dimensions to analyze the semantic relationships between AI and IS. The aim is on identify semantic asymmetry i.e., an imbalance in how the two fields reference or influence each other. We first introduce a conditional probability-based measure, which can formalize the relationship between the dependent entities. This conditional dependence is measured for several types of classifications, such as classified by WOS, classified by keywords (in WOS and Scopus), and classified by references. We then study the semantic distance between the research areas to understand the dissimilarity based on data from WOS and Scopus. The relationship is also explored with respect to the semantic similarity using a NID measure. This section utilizes several related but distinct concepts to quantify the relationship between research domains:

- *Conditional dependence* A directional measure—which forms the basis for identifying semantic asymmetry—that reflects how often one field references the other (e.g., % of IS papers mentioning AI), which is computed using conditional probabilities. It can be interpreted as the frequency with which one research area is mentioned or cited within another.
- *Semantic similarity* A symmetrical measure which represents how frequently terms or concepts co-occur across publications, for instance, by using harmonic means of conditional probabilities.
- *Semantic distance* A derived metric which indicates the separation between two fields,

estimated using a contra-harmonic mean of 1—conditional probabilities.

2.1 Semantically asymmetric co-occurrence

Assume that we are interested in the relationships between two 'things', A and B, that occur in our global population. Moreover, we assume that these can occur in two semantically different roles, say A1 and A2, and B1 and B2, respectively. To fix the mindset, we also assume that A1 stands for primary classification to A and A2 for secondary classification. We are interested in the co-occurrence of A and B and, in particular, in the possible asymmetry between the co-occurrence between A1 and B2 compared to B1 and A2. The standard conditional probabilities: P(B2|A1) and P(A2|B1) indicate how the secondary classifications are conditional to the primary classification of the other class. Consider now the ratio

$$\frac{P(A2|B1)}{P(B2|A1)} \tag{1}$$

If the classifications A1 and B1 do not have any effect to the occurrence of A2 and B2, the ratio should be equal to P(A2)/P(B2). Consequently, the ratio

$$\frac{P\left(B2\right)P(A2|B1)}{P\left(A2\right)P(B2|A1)} \tag{2}$$

should indicate the base rate corrected asymmetry between the influence of A1 to B2 versus B1 to A2. The semantic meaning of this asymmetry depends, of course, on the semantics of the corresponding primary and secondary classifications.

Equations (1) and (2) provide a conceptual foundation for measuring semantic asymmetry between AI and IS by incorporating base-rate normalization, that is, adjusting for the general prevalence of AI or IS mentions across the entire publication corpus. These formulations are useful in theory to avoid bias caused by uneven keyword frequency distributions. Implementing these base-rate-adjusted formulas would require estimates of P(AI) and P(IS) across the entire corpus, which are not inferable from the current metadata structure due to limitations in keyword consistency and indexing coverage across WOS and Scopus. Therefore, in practice, our investigation is based on the simpler conditional probability-based measure defined in Sect. 2.2 (Eq. (3)), which uses co-occurrence within classified papers without normalizing by global base rates.

2.2 Conditional dependence/semantic similarity

This sub-section introduces a novel metric based on conditional dependence to understand the semantics between AI and IS. We propose to compute the conditional probability among the research fields AI and IS such that we have a measure that quantifies "*How much research on IS dependent on research on AI and vice-versa*." We can simplify this as the probability of AI as a mentioned term in the published articles given IS as a classification of the articles. Mathematically, it can be represented as:

$$\mathrm{P}_{corpus}\left(\mathrm{AI}_{mention}|\mathrm{IS}_{classifaction}\right) = \frac{\mathrm{D}\ (\mathrm{AI}\mid\mathrm{IS})}{\mathrm{D}\,(\mathrm{IS})} \tag{3}$$

Here, P_{corpus} denotes probabilities from *corpus* (WOS or Scopus), D(IS) is total paper classified under IS, and $\mathrm{D}\ (\mathrm{AI}\mid\mathrm{IS})$ is the number of papers that belongs to the publications classified as IS and mention the term AI somewhere in the paper. Thus, $\mathrm{P}_{corpus}\ (\mathrm{AI}_{mention}\ |\mathrm{IS}_{classification})$ signifies a probability value or conditional dependence, which computes how much research on IS depends on the research on AI in that respective corpus. Precisely, based on bibliographic metadata, it measures how often AI terms and concepts are mentioned in IS-related publications and vice versa. This provides a quantifiable view of semantic or citation-based influence between disciplines. A percentage value (%) is also estimated for the same. We have also used the Harmonic Mean among the two asymmetric conditional probabilities to quantify an actual semantic similarity, denoted as SIM(AI, IS). Note that, for paper extraction, the areas of research method, decision science, and information systems are collectively addressed as ISs, and the area of machine learning and artificial intelligence is collectively referred to as AI. Basically, they were chosen as keywords in the indexing platforms.

Several classification aspects are considered to compute conditional probability in Eq. (3), such as: (i) classified by WOS, (ii) classified by keywords (in WOS and Scopus), and (iii) classified by references. The first classification aspect considers the classification label given by the indexing platform itself, which is an ideal case for computing such conditional dependence. WOS assigns categories for AI as "*Computer Science, Artificial Intelligence*", and for IS as "*Computer Science, Information Systems*". Notably, it is extremely difficult to perform the exact classification of papers related to IS and AI. Hence, the number of extracted papers might have some variability, but it's still acceptable in a broader sense. The second classification is used because Scopus does not define separate categories for AI and IS. Hence, we have used this classification approach, where we used the respective keywords (from each research fields) for the classification of the papers in both WOS and Scopus. The last classification aspect we have considered is the citations or the references of the top 10 highly influential papers from AI and IS. Now, we will individually analyze the conditional dependence based on these three types of classification.

(i) Classified by WOS

Figure 1 corresponds to the estimates for the conditional dependence/probability (P_{wos}), its percentage (%) and SIM among AI and IS over the last ten years (2015–2024). For 2024, the percentage of $\mathrm{P}_{wos}(\mathrm{AI}_{mention}|\mathrm{IS}_{classifaction})$ i.e., in %(AI|IS) is 17.21%, while %(IS|AI) is 1.63%. It shows that only 1.63% of papers mention IS among the 18,360 papers classified as AI, however, 17.21% of paper among 28,815 IS papers refer to or mentions AI.

It is observed that the number of IS papers has increased much more than AI papers, however, the role of IS within AI has dropped whereas the occurrence of AI in IS papers has significantly increased. Table A.1 (mentioned in the supplementary material) reflects the above analysis for a combined ten years from WOS. The %(IS|AI) represents a low dependence of AI on IS because from 241,031 papers classified as AI, only 1.63% of the papers mention IS or related terms. On the other hand, around 10.68% of papers in IS referred to AI-centric work. Further, the analysis with conditional probabilities is explored in a wider

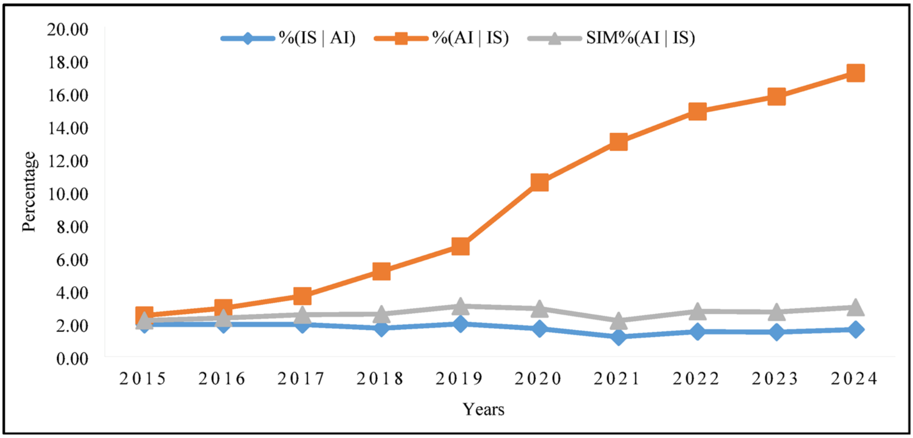

Fig. 1 Line graph of estimates for the conditional dependence (P) and SIM among AI and IS in %, over last 10 years (as per WOS classification)

Table 1 Conditional dependence (in %) of ML for IS

D(ML)	D(IS|ML)	%(IS|ML)
240,884	3931	1.63

Table 2 Conditional dependence (in %) of AI and several AI classes for IS

Class	D(AI, Class)	D (IS|Class)	%(IS|AI, Class)
DL	26,003	238	0.92
SW	859	203	23.63
GAN	3066	4	0.13
AGI	90	0	0.00
CNN	15,506	118	0.76

context where we have computed the vertical and horizontal taxonomies closely related to areas of AI and IS. These computations are performed on WOS as it already has classification labels.

Vertical Taxonomy (class-subclass) driven: We observed the behaviour of dependence of IS on AI (as a whole research domain accumulation ML and AI) and vice-versa. Here, we see the conditional dependence of ML for IS. Table 1 shows that only 1.63% of IS-centric publications are being quoted in papers classified as ML.

Horizontal Taxonomy (same level subclass) driven: Further, we have considered a few widely referenced domains (or classes) within AI, such as Deep Learning (DL), Semantic Web (SW), Generative adversarial networks (GANs), Artificial general intelligence (AGI), Convolutional neural networks (CNNs), and computed their conditional dependence on IS-centric papers. Table 2 shows the conditional dependence (in %) of AI and above AI classes for IS. For D(AI, Class), we have executed the query WOS classification (WC) = (Computer Science, Artificial Intelligence) AND TS = (*Class name as a keyword*)). Among the considered AI classes, DL, CNN, and GAN-related publications referred to or mentioned IS only 0.92%, 0.76%, and 0.13%. For AGI, this dependence is null. The SW-related publications have a considerably higher dependence of 23.63%. After exploring these publications, it is found out that the SW is a concept closely related to the field of ISs because it is concerned with making the information and data on the World Wide Web more accessible and understandable to computers. This can help improve the way ISs work by making it easier to find, share and integrate data between systems.

Horizontal Taxonomy (super-class) driven: As per WOS, apart from AI and IS, computer science is sub-categorized into five other categories (or classes) such as cybernetics (CYB), hardware & architecture (HA), interdisciplinary applications (IA), software engineering (SE), and theory & methods (TM). The ALL signifies all these categories taken together. Here, we have computed the conditional dependence of these classes on IS and AI, the corresponding values of which are reflected in Table 3.

This analysis indicates that the relationship between IS and AI is a bit stronger in both directions than the corresponding relationship to other fields of AI (although not by order of magnitude). For all these taxonomies, AI is certainly more influential than IS, or more simply, the research on these taxonomies is influenced more by AI (i.e., AI has been referred more). IS, on the other hand, doesn't occur in these taxonomies as such, but it gets inspired by AI. Therefore, this is another motivation for further studying the relationship between AI and IS.

(ii) Classified by keywords

For this keywords-based classification approach, we used the keywords "Artificial Intelligence" and "Machine learning" for AI and "information system" (or "information systems") for IS. Then conditional probabilities are computed from WOS and Scopus databases to understand the dependence. First, we analyze the conditional dependence with respect to the Scopus dataset. Over the ten years, the dependence of IS on AI has been noteworthy, as can be seen from Fig. 2, especially in the last few years, this dependence has increased significantly.

In 2024, around 37% of papers in IS, totaling 10,990, mentions AI or ML, while only ~ 14% of papers in AI referred IS. Since we used keywords for classification in Scopus, the same analysis is studied for WOS. The combined values of the ten years from both the indexes are shown in Table 4. In Scopus, AI dependence is 13.59% on IS, which could be only attributed to the higher number of overall publications, however, around 20.68% of IS publications are dependent on AI. For WOS, the %(IS|AI) is computed as 1.33% while %(AI|IS) is returned as 4.87%. There is a marginal difference here due to the quantity of indexed papers and keywords limitation. Figures 2 and 3 shows the respective line graph for last 10 years and Tables A.2 and A.3 shows the detailed yearly values for Scopus and WOS, respectively.

Table 3 Conditional dependence (in %) of WOS categories for IS and AI

Class	D(Class)	D(IS|Class)	%(IS|Class)	D(AI|Class)	%(AI|Class)
CYB	28,240	354	1.25	1803	6.38
HA	75,607	242	0.32	6170	8.16
IA	214,501	1861	0.87	20,645	9.62
SE	127,574	856	0.67	9240	7.24
TM	151,530	1159	0.76	14,054	9.27
ALL	495,686	4004	0.81	43,028	8.68

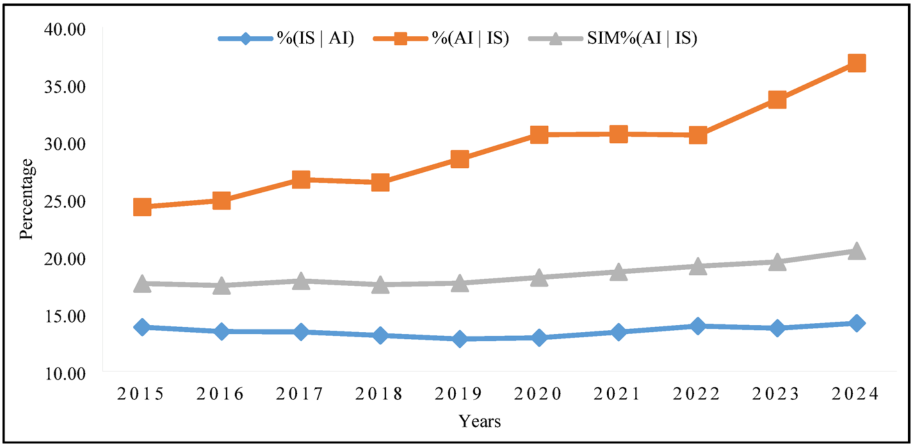

Fig. 2 Line graph of estimates for the conditional dependence (P) and SIM among AI and IS in %, over last 10 years (Scopus-keywords based)

Table 4 Combined estimates for the % conditional dependence (P) and semantic similarity (SIM) among AI and IS

Index	D(AI)	D(IS|AI)	D(IS)	D(AI|IS)	%(IS|AI)	%(AI|IS)	SIM (AI, IS)
Scopus	902,849	122,683	189,461	39,176	13.59	20.68	16.40
WOS	373,929	4979	65,971	3215	1.33	4.87	2.09

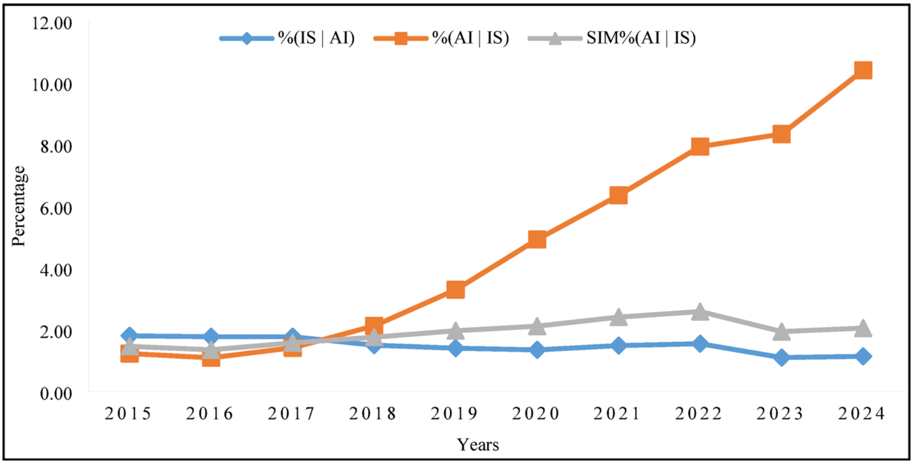

Fig. 3 Line graph of estimates for the conditional dependence (P) and SIM among AI and IS in %, over last 10 years (WOS-keywords based)

From the above two classification aspects, %(AI|IS) signified a relatively high dependence of IS on AI in both WOS and Scopus. Specifically, for Scopus (as per keyword classification), there is significant dependence because, from 189,461 papers classified as IS, as much as 20.68% of the papers mention AI. At the same time, this is 4.87% in the case of

WOS classification, which is again significant as compared to 1.33%. In addition, a more reliable classification approach of WOS also returned significant occurrence of AI in IS-related publications.

(iii) Classified by references

Next, we performed an analysis of the cited references from each of the research areas to assess whether the most influential work in one field (AI or IS) meaningfully impacts the other. This analysis would highlight whether the citation influence observed at the aggregate level is also visible among influential research papers in the respective areas. Essentially, whether high-impact contributions in AI are cited in IS research, and vice-versa. This information could reveal the asymmetry or mutuality of intellectual acknowledgment at the highest levels of each research area. Considering all the cited references would be a complex task, thus, we have considered the top ten papers from each area. Tables A.4 and A.5 shows semantic similarity analysis on the top ten highly cited papers from AI and IS, respectively. For each influential paper, its total citations (TC) is analysed as per the WOS category, and citations referred to by AI, and IS-related papers are also extracted. The TC (or the references) are here considered the classification aspect for that paper. Then, respective conditional probabilities of P(AI|TC) and P(IS|TC) are computed.

For this analysis, its empirically studied how a highly influential paper in one area is influencing (or influenced by) another area. Higher values of P(AI|TC) in Table A.4 make sense as the papers are highly influential AI-related papers. What is interesting is the considerably higher values of P(IS|TC), signifying the higher dependence of IS-related publications on core AI-centric papers. For instance, the most influential paper titled "*Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*" classified under AI, have been referred to 35,269 times, among which ~ 15% of papers are AI-centric while ~ 12% are from IS domain. This dependence of IS on AI-related papers is even higher for the paper "*Federated Machine Learning: Concept and Applications*", which has been referred on only ~ 8% of AI-centric papers while ~ 23% of IS-centric papers.

Inferring the same sentiment, remarkably lower values for P(AI|TC) in Table A.5 indicate the poor dependency of AI publications on IS-centric influential papers. Collectively, only ~ 3% of AI-centric publications referred to the top ten highly influential IS-centric papers.

2.3 Distance measure computation

In addition to similarity, we have also computed the semantic distance between the two research areas. The similarity has been defined as a harmonic mean of the component similarities (i.e., conditional probabilities), hence, distance here is a contra-harmonic mean of the component similarity (i.e., one minus the conditional probabilities). The distance computation may be defined as follows:

$$\begin{aligned}\mathrm{DISTANCE}_{corpus}(X,Y) &= \frac{\left[\overline{\mathrm{P}}_{corpus}\left(\mathrm{X}_{mention}|\mathrm{Y}_{classifaction}\right)\right]^2+\left[\overline{\mathrm{P}}_{corpus}\left(\mathrm{Y}_{mention}|\mathrm{X}_{classifaction}\right)\right]^2}{\overline{\mathrm{P}}_{corpus}\left(\mathrm{X}_{mention}|\mathrm{Y}_{classifaction}\right)+\overline{\mathrm{P}}_{corpus}\left(\mathrm{Y}_{mention}|\mathrm{X}_{classifaction}\right)} \\ &= \frac{\left[1-P_{corpus}\left(\mathrm{X}_{mention}|\mathrm{Y}_{classifaction}\right)\right]^2+\left[1-P_{corpus}\left(\mathrm{Y}_{mention}|\mathrm{X}_{classifaction}\right)\right]^2}{\left[1-P_{corpus}\left(\mathrm{X}_{mention}|\mathrm{Y}_{classifaction}\right)\right]+\left[1-P_{corpus}\left(\mathrm{Y}_{mention}|\mathrm{X}_{classifaction}\right)\right]};\end{aligned} \tag{4}$$

Here, the corpus is WOS or Scopus, and X and Y correspond to the research areas. The contra-harmonic mean is used here to compute semantic distance, as it gives more weight to larger dissimilarities. This is appropriate in our context as we aim to emphasize directional imbalance between AI and IS. This mean makes sure that when either direction has a high dissimilarity (i.e., a low cross-reference rate), the resulting distance measure reflects that more strongly. It was empirically found to yield more pronounced distinctions than the arithmetic mean and preserves sensitivity to extreme values. Below, we have computed the % distance between IS and AI from the perspective of WOS (classified by WOS) and Scopus (classified by keywords). From WOS, we compute two distance metric values, one for WOS classification and the other for classification by keywords:

$$\mathrm{DISTANCE}_{WOS}\,(IS, AI) = \frac{[1-0.0163]^2 + [1-0.1068]^2}{(1-0.0163)+(1-0.1068)} = 0.9406 = 94.06\%$$

$$\mathrm{DISTANCE}_{WOS}\,(IS, AI) = \frac{[1-0.0133]^2 + [1-0.0487]^2}{(1-0.0133)+(1-0.0487)} = 0.9693 = 96.93\%$$

However, for Scopus, there is one distance metric value for classification by keywords:

$$\mathrm{DISTANCE}_{Scopus}\,(IS, AI) = \frac{[1-0.1359]^2 + [1-0.2068]^2}{(1-0.1359)+(1-0.2068)} = 0.8300 = 83.00\%$$

The distance values suggest the high degree of separation between IS and AI as per WOS classifications. The dissimilarly as per WOS classification and its keyword classification is 94.06 and 96.93%, which indicates the separability in these two areas and minimum overlap. Interestingly, distance values computed from Scopus on keywords classification, indicates that while these two areas are distant, they still have some overlap. This reflects the wider AI integration within IS-centric research. Figure 4 shows the sematic distance values in percentage between IS and AI.

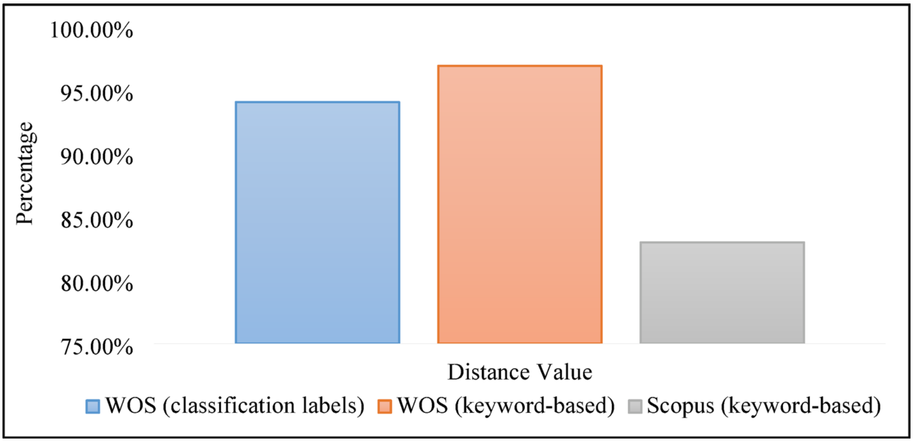

Fig. 4 Semantic distance (%) between IS and AI across WOS and Scopus

2.4 Normalized index (or semantic) distance

This sub-section studies the relationship between the considered research domains (AI and IS) using the already available measure in the literature called NGD (Cohen and Vitányi 2013). NGD is used to assess similarity/dissimilarity among keywords. Since it is exclusively dependent on Google search results, we present NID, where distance is computed with respect to the indexing platforms such as WOS and Scopus. NGD also suffers from several key limitations such as, lack of reproducibility due to proprietary and dynamic search algorithms, and inclusion of non-academic or noisy content. On the other hand, NID adapts the NGD formulation to structured WOS and Scopus databases, ensuring domain-specific relevance, stability, and consistency across queries. This makes NID more suitable for academic and cross-disciplinary research analysis. Mathematically, it is defined as follows:

$$NID\left(x,y\right)=\frac{\max\left\{\log f\left(x\right),\log f\left(y\right)\right\}-\log f\left(x,y\right)}{\log N-\min\left\{\log f\left(x\right),\log f\left(y\right)\right\}} \tag{5}$$

where $f\left(x\right)$, $f\left(y\right)$, and $f\left(x,y\right)$ are the number of hits regarding the search terms x, y, and (x AND y), respectively. N is the total number of papers during the specific period in the indexing platform. We have considered only WOS and Scopus for the analysis of this NID, since there is some vagueness in the resulted outcome of the search queries on Google Scholar. If the valued of NID is close to 0, then the terms occur mainly together, if NID is clearly above 1 this indicates that the terms appear in quite distinct forums and the values around one signal that the terms appear independently of each other. Table 5 shows the NID values between the AI and IS for respective classification criteria. The NID for classification by WOS returns 0.56, while its 0.55 for the classification by keywords in Scopus signifying the tendency AI and IS to occur in shared contexts. With the keyword's classification for WOS, NID is 0.75 which is comparatively high. Note that for the classification by keywords, the "keyword" has been examined anywhere in the test to make is more inclusive.

3 Journal based analysis

Although conditional probability approach set a basis in answering our research question, it might raise concerns on not so concrete classification categories. Therefore, we have performed another set of analysis to reason with the research question. Here, we have taken the ten sought out journals from each of the fields and extracted the publications from another field. For instance, we extracted the number of papers with AI and ML in the journals related to field of IS such as European Journal of Information Systems, Information Systems Journal, etc. The details of the journals from the IS and AI field are mentioned in Tables A.6 & A.7 along with the year wise publication counts from all of the journals. The keywords used for the AI journals were "Artificial Intelligence" and "Machine learning", while the keyword used for IS journals were, "information system", "information systems", "research

Table 5 Normalized index distance

Classification criteria	Indexing platforms	IS and AI
Classification by WOS	WOS	0.56
Classification by keywords	WOS	0.75
	Scopus	0.55

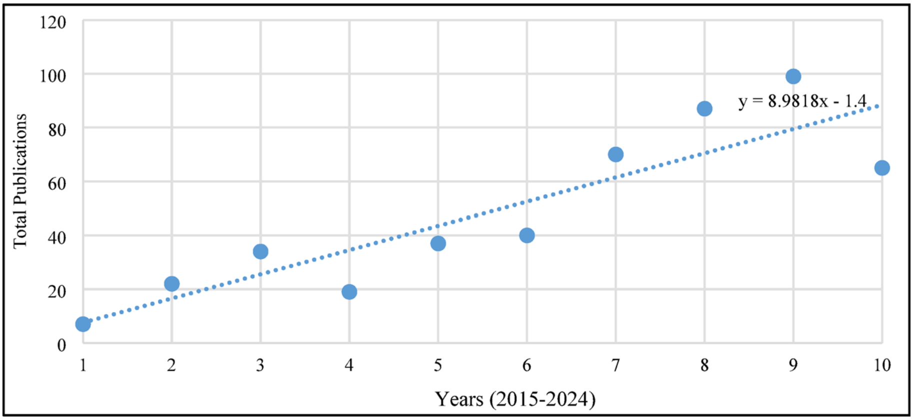

Fig. 5 Linear regression plot of AI related publications in IS-centric journals

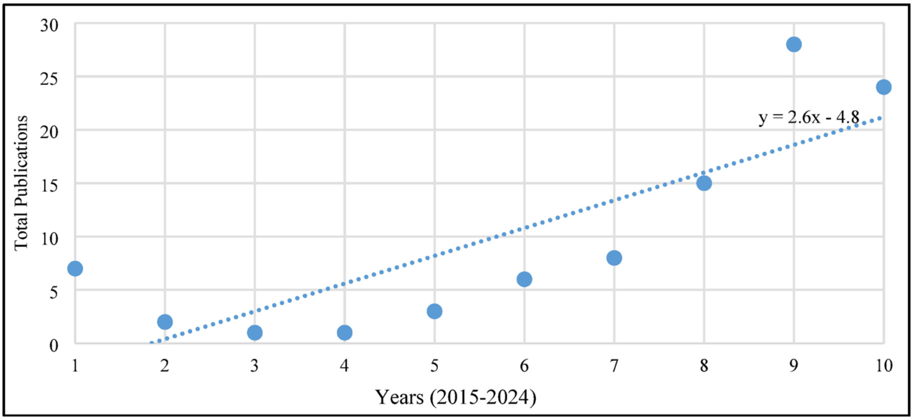

Fig. 6 Linear regression plot of IS related publications in AI-centric journals

method", "research methods", "design science", "design sciences". All these keywords for IS makes the search more comprehensive. These journals are identified with the Field codes of SRCTITLE mentioned in the Scopus. Figures 5 and 6 represents the linear regression graph for the Tables A.6 and A.7, respectively.

An upward trend can be seen in AI-related publications within IS-centric journals (Fig. 4), with the regression equation $y=8.9818x-1.4$ indicating nearly nine additional publications per year. This shows a growing integration of AI and ML related research into IS, especially after 2019, with journals like *Information Systems Research* and *Information Systems,* etc. leading the trend. On the other hand, Fig. 5 depicts a much slower growth of IS-related publications in AI-centric journals, as indicated by the equation $y=2.6x-4.8$. Even though there is a slight increase in recent years, the total publication output remains low, which suggests limited incorporation of IS perspectives in AI research.

4 Bibliometric analysis

To address the argument in the above section, we present a quantitative analysis of bibliometric study where we explore the publications from these two domains and evaluate several publication parameters. The bibliometric analysis section provides a contextual and structural overview (e.g., journal scopes, citation networks, keyword analysis) that supports or also helps to question the interpretations from the above methods. Further, it helps to identify systemic or publication-level indicators of influence or dominance.

Bibliometric study (or scientometric analysis) is the statistical way to analyse the intrinsic information among the academic publications. This study is solely dependent on the metadata of the publications extracted from the indexing platforms such as Scopus or Web of Science (WoS). Although there are many publishers, these two platforms accumulate everyone under same roof. While WOS indexes quality papers published in ranked international journals and conference, Scopus cover a wide set of publishers which makes it more inclusive. All the available indexes were used for the search query, which are: Science Citation Index-Expanded (SCI-E), Social Science Citation Index (SSCI), Arts and Humanities Citation Index (AHCI) and Emerging Source Citation Index (ESCI). The papers are carefully extracted from these indexes using an appropriate set of keywords such that we have a high recall and precision (Shukla et al. 2020). The keywords are explored only in title, abstract, and keywords of the manuscripts. Several performance indicators are used for the bibliometric analysis such as: total papers (TP), TC, total number of citations divided by total papers (CPP), and publication year (PY). The resulting query has all the meta-information about the papers and further analysis is categorised based on citations, authors, research area, journals, countries, etc. For the visual representation, VOSviewer is used as a graphical tool.

In literature, there are basically two types of bibliometric analysis. One is related to the analysis of a publishing source or a journal (Abedin et al. 2021; Janmaijaya et al. 2018; Zurita et al. 2020), which covers all the bibliometric aspects of the journal, such as productive and influential authors, countries, citations, etc. The other type of study is focused on a particular research area or domain (Li et al. 2020; Shukla et al. 2023; Tran et al. 2019), which is very effective in extracting intrinsic information about that area and helps the readers to get an overview (starting point of research). In this study, we have performed both aspects of the bibliometric study to some extent which satisfies the research question.

This section is divided into several sub-sections. First, we individually explored IS and AI related publications along with their combined publications (IS and AI) to investigate the respective influence. Then, few widely studied AI related topics such as CNN, RL, SW, and GAN are studied to understand the impact of IS-centric research. One important aspect to select these AI subfields is the citation volume and research momentum, as they represent some of the most highly cited and rapidly evolving areas within AI over the last decade, as also supported by bibliometric indicators in WOS and Scopus. Also, these are widely acknowledged as foundational or transformative topics within the AI discipline (e.g., CNNs in computer vision, GANs in generative modelling, RL in autonomous systems, etc.). These subfields are increasingly mentioned in IS literature, especially in areas like digital transformation, intelligent systems, and data-driven decision-making. All these strong reasons make them ideal candidates to explore the nature and depth of cross-disciplinary reference.

4.1 Information systems

WOS clearly classifies the paper related to information systems under the category "Computer Science, Information Systems". There are total of 371,885 publications in the ten years from 2015 to 2024, on which we have performed the bibliometric analysis. First, we explore the sources and their scopes where IS papers are published. Table 6 presents the top 10 journals which are publishing IS related papers as per the classification of WOS. Interestingly, around 27% of publications have come from the journals which are majorly focused on AI related publications. Highly productive journal of IEEE Access accounts for 22.41% of TP, which reflects its broader scope that often includes AI-related and multi-disciplinary research. Other influential journals such as the IEEE Internet of Things Journal and Information Sciences also focus heavily on AI and related methodologies. Although, these are multidisciplinary journals and publishes in quantity, we certainly cannot establish this fact that IS publications are influenced and dependent on AI related work. Next, we analysed the highly influenced 2000 papers from this classification. Table A.8 reflects the top 20 most influential papers in IS. Many of the highly cited papers in this list are published in the journals with scope focused on AI and multidisciplinary. The highly cited paper titled "*Federated Machine Learning: Concept and Applications*", receiving 5032 citations is an AI-centric paper. The influential papers are published in IEEE Access, IEEE Internet of Things Journal, and IEEE Communications Surveys and Tutorials, which are known for their focus on emerging technologies, including AI, which indicates the prominence of AI-centric research within the IS domain. In conclusion, the bibliometric analysis of the top 20 highly cited papers reveals a clear AI-centric trend within the IS domain. This reflects the field's adaptation and integration with cutting-edge AI technologies, highlighting the interdisciplinary nature of IS research.

Further, we performed the keyword analysis using the Vosviewer among the 2000 highly influential papers. Notably, we have only considered the keywords classified by the authors themselves to understand their perspective. It will present a picture in term of keywords of how research in IS is referring key terms from AI & ML. Figure 7 shows the keyword analysis among these highly influential 2000 papers. For the ease of representation only top 100 keywords with most frequency is depicted. The link between the keywords is the co-occurrence link, and its strength signifies how often they have occurred together in the paper. The

Table 6 Top 10 productive journals publishing IS-centric papers

S. No.	Journal	TP	%
1	IEEE Access	83,349	22.41
2	Electronics	19,372	5.21
3	Multimedia Tools and Applications	16,656	4.48
4	Information Sciences	9329	2.51
5	IEEE Internet Of Things Journal	9280	2.49
6	Wireless Communications Mobile Computing	5570	1.49
7	CMC Computers Materials Continua	5024	1.35
8	ISPRS International Journal of Geo Information	4819	1.29
9	IEEE Transactions on Information Theory	4520	1.21
10	Journal of Chemical Information and Modelling	4354	1.17

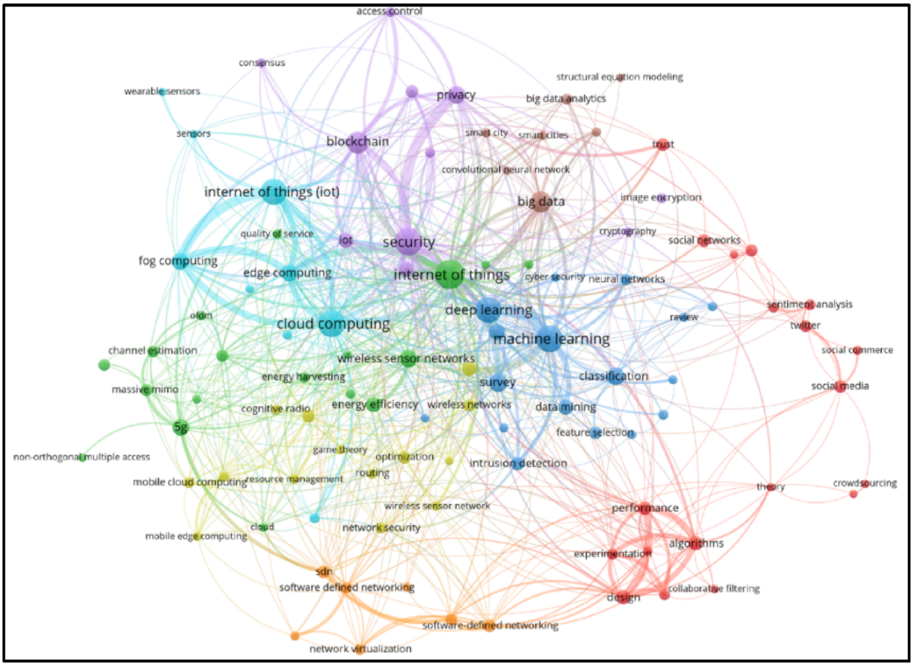

Fig. 7 Author keyword analysis of most influential IS-centric publications

major nodes are machine learning, deep learning, classification, etc., forming one set of clusters which represents more co-occurrence among these keywords. Another prominent cluster is from internet of things, fog computing, edge computing, cloud computing, etc.

4.2 Artificial Intelligence

The classification for AI in WOS is "Computer Science, Artificial Intelligence" which resulted in 241,031 papers in last 10 years. As studied in above section, here we present the perspective of core AI publications. Table 7 reflects the top productive journals publishing AI & ML related papers. Almost all the journals have their scope targeted on AI such as Neurocomputing, Expert Systems with Applications, IEEE Transactions on Neural Networks and Learning Systems, Neural Computing Applications, etc. Only Applied Soft Computing is somewhat related to the multidisciplinary aspect, still the impact of IS is negligible here.

In the top 20 publications (Table A.9), six papers are published in core AI journal of IEEE Transactions on Pattern Analysis and Machine Intelligence, which also accounts for around 48% of the total citations as compared to these 20 papers. It is observed that almost all the papers are breakthrough papers in the AI community which has helped to improve the overall machine intelligence. Not even a single paper is attributed to the research methods or study the idea of ISs. We further investigated the author keywords in the 2000 influential papers, as shown in Fig. 8. Major nodes are deep learning, machine learning, neural networks, optimization, classification etc. Interestingly, these are also the common keywords

Table 7 Top 20 productive journals publishing AI-centric papers

S.No.	Journal	TP	%
1	Neurocomputing	13,180	5.47
2	Expert Systems with Applications	12,890	5.35
3	Journal of Intelligent Fuzzy Systems	9987	4.14
4	Soft Computing	7895	3.28
5	Neural Computing Applications	7829	3.25
6	Applied Soft Computing	7771	3.22
7	Knowledge Based Systems	6489	2.69
8	IEEE Transactions on Neural Networks and Learning Systems	6078	2.52
9	Engineering Applications of Artificial Intelligence	5242	2.18
10	IEEE Transactions on Image Processing	5030	2.09

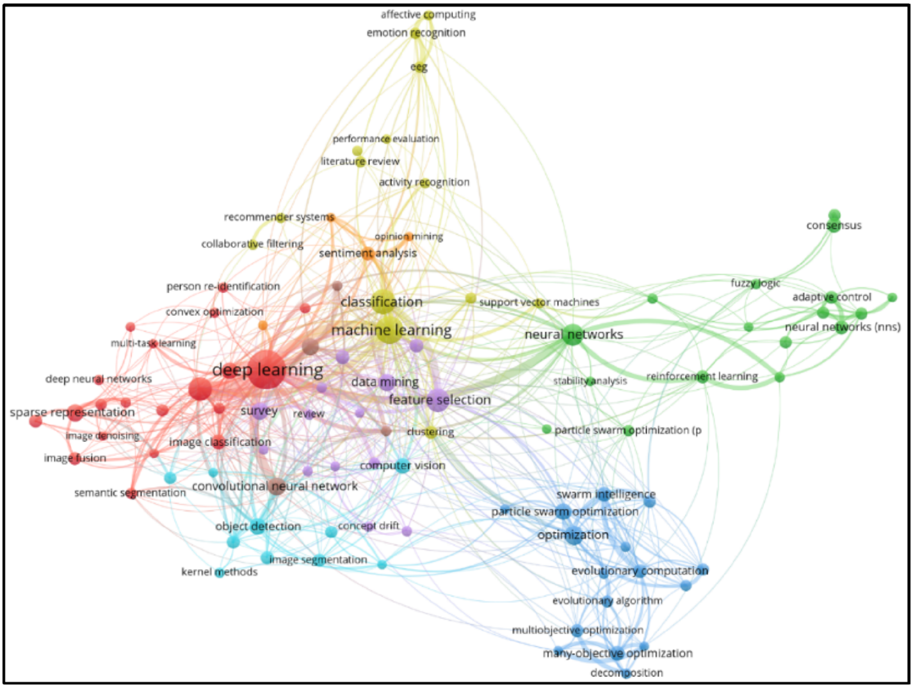

Fig. 8 Author keyword analysis of most influential AI-centric publication

when studied with respect to the IS domain. These common keywords clearly indicate the AI-centric influence on the IS based research.

4.3 Convolutional neural networks

As per WOS, the first paper on CNN was published in 1997, while Scopus published first paper in 1993. The NID value higher than 1 in Table 8 indicates the semantic dissimilarity among CNN and IS domain. NID is computed for the last 10 years and overall years

Table 8 NID values among CNN and IS

Indexes	All Years	Last 10 Years
WOS	1.03	1.07
Scopus	1.04	1.1

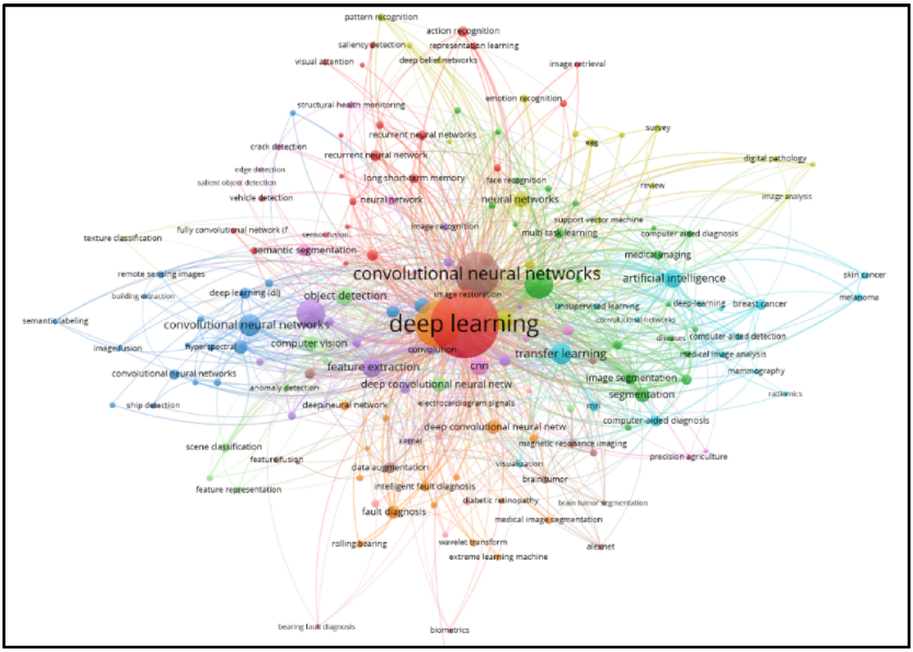

Fig. 9 Keyword analysis in highly cited papers over last 10 years (WOS)

since the inception of CNN. Then, we present bibliometric analysis from both the indexing platforms for last 10 years. Tables A.11 & A.12 shows the top 20 influential papers in CNN (from WOS and Scopus, respectively) to see if there are any influential papers from the IS domain. There is not a single paper, or the publishing source influenced or related to IS. Most influential paper as per both the indexes is "*ImageNet Classification with Deep Convolutional Neural Networks*" and most sought out journal is "*IEEE Transactions on Pattern Analysis and Machine Intelligence*". As per WOS, the first indexed paper was published in 1997. Further, we observe that none of these impactful papers comes from or have inclination from ISs or research methods domains. Moreover, the scope of all the journals targets the mathematical development and applications of AL and ML.

For the keyword analysis, we took the top 2000 highly cited papers from WOS & Scopus in the last 10 years and extracted Figs. 9 & A.11 from the VOSviewer. All the keywords are either CNN centric or applicative. Different colours specify clusters of keywords which signifies that such keywords have been used together most often. For instance, "deep learning" has been used mostly with "long short-term memory", "recurrent neural network", etc. From the analysis, it's evident that any term associated with research methods or design science doesn't belong to any of the cluster, hence, no relation or impact as such in the development of CNN. The same can be observed with keyword analysis from the Scopus database. The relation between the cluster of keywords is also identical as in the case of WOS.

Further, it is important to study the most significant publications which marked the beginning of CNN. It's also crucial to understand in terms of bibliometric study how and where those influential publications have their impact. The original idea of CNN came from the simple and complex sells in human visual cortex. In 1980, a neocognitron model was proposed by Dr. Kunihiko Fukushima in the paper titled "*Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position*", where author proposed mathematical formulations in contrary to the biological cells. Later in 1990's, modern CNNs were proposed in the paper "*Gradient-Based Learning Applied to Document Recognition*" (LeCun et al. 1998). After this, various improvements and additions were performed on modern CNNs. However, until recently in 2012, "*ImageNet Classification with Deep Convolutional Neural Networks*" paper by Krizhevsky et al. (Krizhevsky et al. 2012) performed significantly on labelling pictures in the ImageNet challenge. Since then, CNNs have seen exponential growth in its usage. This paper has received 68,002 citations till June 2024 as per WOS. We have analysed Scopus citations to extract the highly sought out journals and fields/research domain of most cited authors. The WOS citations are explored to extract the WOS categories as its clearly points out the specific subject area where the paper has been published.

The left section of Tables A.13 & A.14 depicts the journals and author information of cited references of the paper "*Gradient-Based Learning Applied to Document Recognition*", while the right section reflects the same information for the paper by Krizhevsky et al. (Krizhevsky et al. 2012). From Table A.13, we observe that almost all of the journals are more catered towards core AI publications, either in terms of mathematical contribution or application specific publications.

In Table A.14, we have also showed the research area of the authors as categorized by the Scopus, available in the bibliometric data. Interestingly, all the authors are categorized broadly under AI or ML. To be more specific, some author classifications were: "Object Detection; Deep Learning; IOU", "Hyperspectral Imagery; Spectroscopy; Image Classification", "Distance Metric; Camera; Reranking", "Collaborative Filtering; Recommender Systems; Factorization", "Manifold Learning; Clustering Algorithm; Dimensionality Reduction", etc.

Figures A.3 and A.4 shows the top 10 categories classified by WOS for the two influential papers in CNN. Interestingly, IS has referred around 16% and 17% in both the paper, respectively. These figures and the highly cited paper in CNN reflect that IS is referring works from AI/ML community, however, the most influential works are produced and referred from AI community only.

4.4 Reinforcement learning

The other core AI area we have covered is RL. Table 9 represents the NID from WOS and Scopus for the last 10 years, which is greater than 1 for all the cases, indicating that both terms are relatively very different.

We then study the most cited 20 papers from WOS and Scopus to find out i) what domain of publications helped in the development of the field of RL, ii) does IS guided the field of RL? From the Tables A.15 & A.16, we find out that paper by Mnih et al. titled "*Human-level*

Table 9 NID values among RL and IS

Indexes	All Years	Last 10 Years
WOS	1.11	1.15
Scopus	1.1	1.19

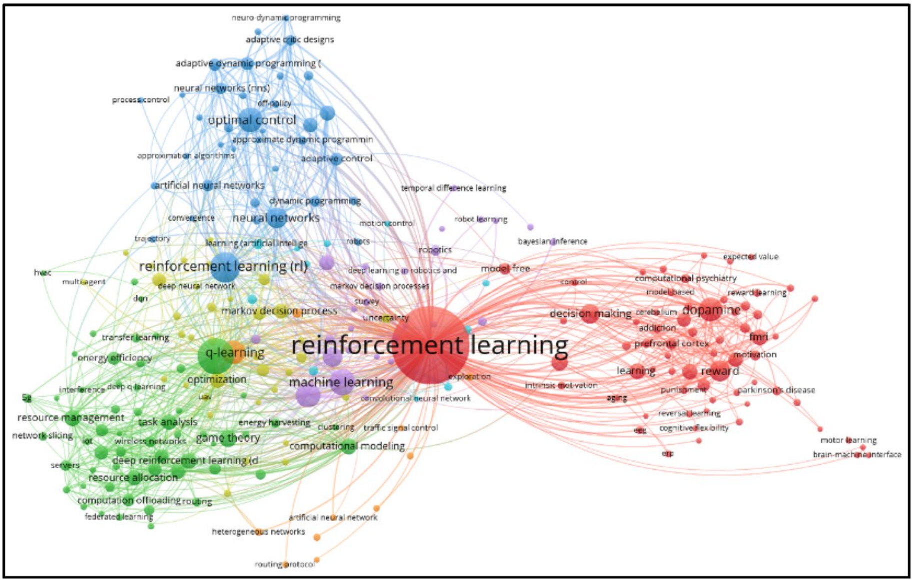

Fig. 10 Keyword analysis of highly cited papers (WOS)

control through deep reinforcement learning” is the highly cited paper in both the indexes. Also, all the papers were from the core AI domain, either proposing novel ML approaches or applied to many application areas. Even the publishing sources are AI-centric.

Similar to the CNN analysis, we also studied the most productive and influential authors from WOS and Scopus who have contributed to the field of RL. And once again, it is found out that almost all the authors have research interests in the field of AI or ML. For instance, productive authors, as per WOS, such as Zhang Y, Wang Y, etc. have research expertise in the field of AI and RL (as per Google Scholar). Both the productive and influential authors have contribution in core AI only. Similar is the case with Scopus index as well.

Figures 10 and A.5 shows an interesting pattern as the clustering and symmetry among the keywords is almost similar in both the indexes, WOS and Scopus. “*Reinforcement learning*” being the centre node is connected to three or four cluster with sets of keywords used most frequently by the AI researchers. The other prominent nodes in the clusters are, optimal control, deep reinforcement learning, dopamine, q-learning, machine learning, etc. Again, there is no keyword from the domain of IS reflecting that in the influential publications, IS has no role to play in the development.

Further, we also performed keyword analysis of the references of the highly cited paper in RL, since its quite impossible to point out at a single publication to perform analysis due to its rich history. Thus, we have considered the highly cited paper, which is “*Human-level control through deep reinforcement learning*” as per Scopus and WOS. We will present the results of Scopus only to avoid redundancy. With respect to the keyword analysis, the

Table 10 NID values among GAN and IS

Indexes	Last 10 years
WOS	1.17
Scopus	1.03

top 200 keywords (with constraint that minimum number of occurrences of a keyword is 17) is depicted using the VOSviewer. Figure A.6 represents the significant contribution of keywords used within core AI domain. The prominent nodes are DL, RL, learning systems, stochastic systems.

A similar study on SW is performed and described in Section SM-II of supplementary material.

4.5 Generative adversarial nets

We have considered one more, quite new, and significant core AI research area of GANs (Goodfellow et al. 2014). For this sub-topic, we have considered a different approach to reduce the repetition of the results. Here, we first analysed the original publication of GAN and present the bibliometric analysis of that publication to understand how and which area of research has mot influence on its growth. Before that, NID from WOS and Scopus for the keywords "generative adversarial networks" & "information systems" is shown in Table 10.

Note that the keywords for the bibliometric analysis are considered as "generative adversarial nets" or "generative adversarial networks", as first paper used the terms "nets", not "networks". First and the most influential paper on GANs was first published at "arxiv.org", not at some highly impacted journal or some international conference. Eventually, it was published at Advances in Neural Information Processing Systems (mitpress). As per Scopus, it has gathered 43,241 citations till the mid of 2024. We have included the results of Scopus only for the GANs as it is all inclusive of sources

We have further analysed the cited articles with respect to the citing journals, authors, and subject areas. For journals and authors, we have broadly classified them to AI or IS. Table A.20 represents the journals/international conferences, which have cited the original publication of GAN in which *Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics* tops the list with 1496 publications, followed by *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (TP=810), IEEE Access (TP=803), and Proceedings of the IEEE International Conference on Computer Vision (TP=439).* All the sources which have cited the original paper of GAN's are from the core AI background. Table A.21 reflects the top 20 authors with the number of papers they have used the original GAN paper as a reference and their specialization as per Scopus categorization. This helps to identify the authors broadly into AI or IS. Clearly, its visible that all the authors belong to the core AI community with their specialization such as computer vision, object detection, DL, medical domain analysis, hyperspectral imagery, etc.

5 Discussion, limitation and conclusion

This study examines the significance of IS outside its domain and compares it with the field of AI due to its widespread presence. To compliment and validate the question, we introduce a conditional dependence-based metric and provided quantitative results from the publication analysis of last ten years. The widely accepted bibliometric indexing platform, i.e., WOS and Scopus were considered for data retrieval. The proposed metric is majorly dependent on the precise classification of the papers. These are the following observed outcomes:

1. When classified with labels assigned by WOS, there is significant dependence from IS-centric papers (10.68%) to the AI-centric papers. It is observed that the number of IS papers has increased much more than AI papers, however, the role of IS within AI has dropped whereas the role of AI in IS papers has significantly increased.
2. Another classification by keyword analysis is also performed on WOS and Scopus. More noticeable than the above WOS labels classification, we observed that from 189,461 papers classified as IS, as much as 20.68% of the papers mention AI while only ~ 13% of AI-centric papers mentioned IS.
3. Further, a reference or citation-based analysis (classification by reference) is performed on the top ten highly cited papers from AI and IS. A considerably higher conditional probability in % signified that the paper which are referring AI related publications are from IS background. While significantly lower %(AI|TC) values showed that there are very few papers from AI which are citing the highly influential IS papers. Comparing the accumulative citations from Tables 6 and 7, we can see that both in IS and AI about 14–25% of the citations carry the same label, whereas there are ~ 11% of IS citations to top AI papers but about only ~ 4% AI citations to top IS papers (which is quite significant fivefold difference in citation rates).
4. A distance computation resembling contra-harmonic mean is computed for possible classification aspects. The more reliable classification by WOS labels showed 94.06% and 83.00% distance between IS and AI-centric research as per WOS and Scopus, respectively.
5. A different perspective where a linear regression analysis is performed on highly influenced journals from AL and IS. It revealed the publication rate of 4.45 of AI related papers in IS-centric journals, however, the reverse is only 0.19.

The above semantic similarity measures are followed by the detailed bibliometric analysis of publications among both the domain. First, the domain of IS and AI are studied with this perspective, and several bibliometric entities are observed to validate the hypothesis. Most productive publication sources (journals) are extracted to understand the scope of the publications from each domain publications. AI centric papers are naturally published in journals such as Neurocomputing, Expert Systems with Applications, etc. which are focused on novel AI algorithms and its applications. On the other hand, most productive journals in IS-centric research are IEEE Access, Multimedia Tools and Applications, and Information Sciences, which are multidisciplinary, if not IS or AI-centric completely. In the most influential IS-centric publication list, we find few AI centric papers such as *Federated Machine Learning: Concept and Applications,* and *A Review on Multi-Label Learning Algorithms.* There is not a single publication focused or influenced by IS domain in the list of most influential AI-centric papers. Interestingly, deep learning, machine learning and classification, etc., are the common keywords used by the authors from both the domains.

It is important to recognize that neither AI nor IS are monolithic fields. Each consists of diverse sub-disciplines that may interact differently across domains. For example, areas such as SW or Explainable AI (XAI)(Gunning 2017) may show relatively more overlap with IS concepts like ontology design or decision support systems. However, subfields such as deep reinforcement learning (RL) may have limited overlaps. Although our primary analysis considers AI and IS at the field level, we included subfield-level analyses (e.g., CNNs, GANs, RL, SW) to begin addressing this heterogeneity. We analysed these core AI research areas to explore if research methods or IS-centric research has any influence. Here, for GAN analysis, we also assessed the bibliometric analysis of citations of the inaugurating paper. For instance, the first publication on GAN's was *Generative adversarial nets* which has received more than 40,000 citations. We analysed these citations and found out the category of most referred journals is AI-centric and all the authors citing this paper belonged AI domain only. Despite this, a more systematic mapping of sub-disciplinary interactions remains an important direction to explore. For a balanced view, it is worth exploring how core IS subfields engage with AI. For instance, design science often adopts AI approaches to develop innovative frameworks, while decision support systems use ML for predictive modeling and optimization. Information retrieval also extensively incorporates AI techniques such as NLP and vector embedding models. Although a detailed analysis of these IS subfields was beyond the scope of this study, we shall investigate these interactions more explicitly to better understand the diversity of AI integration across the IS landscape.

In conclusion, we have established the fact that the significant contributions in AI have been obtained without giving an interest to the research methods and related research policies and procedures invented by IS researchers. The dominance of AI-related topics and papers among the most cited papers indicates the that the IS research is gradually integrating AI approaches to address key research questions. This suggests the intersection of IS with AI, where AI methodologies are being used to enhance ISs, which has essentially led to the modern innovations in areas like digital transformation, 6G wireless systems, and edge computing, etc. This also implies that critical role played by AI-centric research leading to the evolution of IS as a discipline. In recent years, the role of AI has grown significantly, and the co-occurrence of highly cited AI papers in IS research is far greater than the co-occurrence of highly cited IS papers in AI research. Our findings from the asymmetric referencing between AI and IS, align with broader observations in interdisciplinary research, where "methodological importation" often flows from technical to applied domains (Klein (2010)). While IS borrows from AI to operationalize complex systems, the reverse does not occur, reflecting a unidirectional mode of interdisciplinarity. This dynamic supports theoretical claims that not all cross-field interactions are reciprocal and that power asymmetries often shape knowledge transfer in science and technology domains (Leydesdorff 2007). Our findings also align with theoretical perspectives on epistemic dependence and unidirectional interdisciplinarity. Asymmetric referencing patterns, with IS drawing heavily from AI but not vice versa, suggest that IS may function as a knowledge-consuming field, relying on the methodological authority of AI without shaping its intellectual trajectory. This reflects broader patterns observed in interdisciplinary research, where cognitive hierarchies and institutional dynamics often lead to one-sided knowledge transfer (Wagner et al. 2011).

In this paper, we presented a novel semantic similarity metric and empirically prove the evidence that the core AI community has been contributing independently of the IS community. Different bibliometric results have claimed that IS community has shown interest in the AI research areas and methods (Abdel-Karim et al. 2021). In this work, we strongly assert that the best results in AI could be obtained without giving an interest to the research methods and related research policies and procedures invented by IS researchers. This fact is strongly backed by the bibliometric analysis of both the domains and presented in the form of various bibliometric parameters. Our conditional probability-based approach can be applied to any two research areas to find the similarity between them.

However, our analysis is subject to several limitations, which are associated with the nature of bibliographic metadata and classification schemes. Access to full-text corpora remains a critical bottleneck for precise semantic and citation analyses. Further iteration of this work could benefit from collaborations with open-access platforms such as CORE, arXiv, etc. or publishers supporting large-scale text mining initiatives. Such access would enable more accurate base-rate normalization, sentiment-aware citation analysis, and context-aware semantic mapping, ultimately leading to deeper insights into cross-disciplinary influence. Another limitation of our work is that the keywords used in our study for IS and AI may be used inconsistently across papers or disciplines, which leads to possible over or underestimation of domain association. The WOS and Scopus differ in scope, coverage, and classification granularity, which may affect how papers are categorized or cited, and introduces variability in comparative measures. It may also help to understand the higher IS-AI overlap observed in Scopus compared to WOS, and the low reverse dependence as seen in Table 2, where only specific AI subfields such as SW showed clear links to IS. Also, it is noteworthy that, our approach of citation-based metrics does not account for the context or sentiment of references. Papers may cite other papers from another discipline for a variety of reasons such as to support a claim, to critique it, or simply as background. Our method does not distinguish between these purposes, since such information is not typically available in bibliometric datasets. Having said that, the metrices used in our work provides an insight into the *volume* of cross-disciplinary referencing. While our metrics reflect patterns of cross-referencing, they do not reveal the nature or depth of influence, such as whether references indicate adoption, critique, or brief mention. As such, terms like "influence" or "dependence" should be interpreted as referential frequency rather than epistemic alignment. More specifically, a mention may reflect adoption, critique, historical framing, or even tangential referencing. Thus, our findings are best interpreted as indicative of semantic and bibliographic dependence rather than direct intellectual lineage. We see this as complementary to deeper qualitative studies of knowledge transfer.

Findings from our work confirm a rising frequency of AI references in IS publications. Our approach shows practical adoption, such as applying machine learning or other AI approaches for decision support or data mining. Moving ahead, it is important to distinguish between instrumental use of AI tools and epistemological integration of AI methods into IS theories and methodology, however, it would imply a shift in how IS researchers conceptualize problems, formulate models, or define validity. Further research works could explore this distinction through full-text analysis or structured expert evaluation of IS research outputs that cite AI literature. In addition, future work shall explore more normalized comparisons of semantic dependence by incorporating base-rate adjustments using full-text corpora or term-frequency indexes. Furthermore, future research shall extend this work using citation context analysis or NLP techniques on full texts including sentiment or citation purpose analysis, to provide a more nuanced understanding of these interdisciplinary dynamics. Also, we shall evaluate semantic similarity using ontologies to define the distance between terms/concepts and text analysis to estimate the semantic relatedness between units of language.

Supplementary Information The online version contains supplementary material available at https://doi.org

/10.1007/s11135-025-02437-8.

Author contributions All authors equally contributed to the conceptualization, formulation, and writing. All authors read and approved the final manuscript.

Funding Open Access funding provided by University of Vaasa. Not applicable.

Declarations

Conflict of interest The authors have no conflicts of interest to declare.

References

Abdel-Karim, B.M., Pfeuffer, N., Hinz, O.: Machine learning in information systems-a bibliographic review and open research issues. Electron. Mark. **31**(3), 643–670 (2021)

Abedin, B., Jafarzadeh, H., Olszak, C.M.: Thirty six years of information systems management: a bibliometric and thematic analysis. Inf. Syst. Manag. **38**(2), 151–164 (2021)

Agarwal, A., Durairajanayagam, D., Tatagari, S., Esteves, S.C., Harlev, A., Henkel, R., Roychoudhury, S., Homa, S., Puchalt, N.G., Ramasamy, R., Majzoub, A., Ly, K.D., Tvrda, E., Assidi, M., Kesari, K., Sharma, R., Banihani, S., Ko, E., Abu-Elmagd, M., Gosalvez, J., Bashiri, A.: Bibliometrics: tracking research impact by selecting the appropriate metrics. Asian J. Androl. **18**(2), 296 (2016)

Ågerfalk, P., Conboy, K., Crowston, K., Eriksson Lundström, J., Jarvenpaa, S., Mikalef, P., Ram, S. Artificial intelligence in information systems: State of the art and research roadmap. Commun. Assoc Inf. Syst. (2021).

Carlile, P.R.: Transferring, translating, and transforming: an integrative framework for managing knowledge across boundaries. Organ. Sci. **15**(5), 555–568 (2004)

Chen, Y., Jensen, S., Albert, L.J., Gupta, S., Lee, T.: Artificial intelligence (AI) student assistants in the classroom: designing chatbots to support student success. Inf. Syst. Front. (2022). https://doi.org/10.1007/s10796-022-10291-4

Cilibrasi, R.L., Vitanyi, P.M.: The google similarity distance. IEEE Trans. Knowl. Data Eng. **19**(3), 370–383 (2007)

Cohen, A.R., Vitányi, P.M.: Normalized google distance of multisets with applications. Preprint at http://arxiv.org/abs/1308.3177 (2013)

Frickel, S., Gross, N.: A general theory of scientific/intellectual movements. Am. Sociol. Rev. **70**(2), 204–232 (2005)

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Bengio, Y. Generative adversarial nets. Adv. Neural Inf. Process. Syst. **27** (2014)

Grapov, D., Fahrmann, J., Wanichthanarak, K., Khoomrung, S.: Rise of deep learning for genomic, proteomic, and metabolomic data integration in precision medicine. OMICS **22**(10), 630–636 (2018)

Gunning, D. Explainable artificial intelligence (xai). Defense advanced research projects agency (DARPA), nd Web, **2**(2):1 (2017).

Gupta, S., Modgil, S., Meissonier, R., Dwivedi, Y.K.: Artificial Intelligence and information system resilience to cope with supply chain disruption. IEEE Trans. Eng. Manag. (2021).

Janmaijaya, M., Shukla, A.K., Abraham, A., Muhuri, P.K.: A scientometric study of neurocomputing publications (1992–2018): an aerial overview of intrinsic structure. Publications **6**(3), 32 (2018)

Klein, J.T.: A taxonomy of interdisciplinarity. Oxf. Handb. Interdiscip. **15**(6), 15 ((2010))

Krizhevsky, A., Sutskever, I., Hinton, G.E. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. **25** (2012).
LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
Leydesdorff, L.: Betweenness centrality as an indicator of the interdisciplinarity of scientific journals. J. Am. Soc. Inf. Sci. Technol. **58**(9), 1303–1319 (2007)
Li, Y., Xu, Z., Wang, X., Wang, X.: A bibliometric analysis on deep learning during 2007–2019. Int. J. Mach. Learn. Cybern. **11**(12), 2807–2826 (2020)
Rafols, I., Meyer, M.: Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience. Scientometrics **82**(2), 263–287 (2010)
Shukla, A.K., Banshal, S.K., Seth, T., Basu, A., John, R., Muhuri, P.K.: A bibliometric overview of the field of type-2 fuzzy sets and systems [discussion forum]. IEEE Comput. Intell. Mag. **15**(1), 89–98 (2020)
Shukla, A.K., Seth, T., Muhuri, P.K.: Artificial intelligence centric scientific research on COVID-19: an analysis based on scientometrics data. Multimed. Tools Appl. **82**(21), 32755–32787 (2023)
Tran, B.X., Vu, G.T., Ha, G.H., Vuong, Q.H., Ho, M.T., Vuong, T.T., La, V.P., Nghiem, K.P., Nguyen, H.L.T., Latkin, C.A., Tam, W.W.S., Cheung, N.M., Nguyen, H.T., Ho, C.S.H., Ho, R.C.M., Ho, R.C.: Global evolution of research in artificial intelligence in health and medicine: a bibliometric study. J. Clin. Med. **8**(3), 360 (2019)
Wagner, C.S., Roessner, J.D., Bobb, K., Klein, J.T., Boyack, K.W., Keyton, J., Rafols, I., Börner, K.: Approaches to understanding and measuring interdisciplinary scientific research (IDR): a review of the literature. J. Informetr. **5**(1), 14–26 (2011)
Wiederhold, G.: The roles of artificial intelligence in information systems. J. Intell. Inf. Syst. **1**(1), 35–55 (1992)
Zurita, G., Shukla, A.K., Pino, J.A., Merigó, J.M., Lobos-Ossandón, V., Muhuri, P.K.: A bibliometric overview of the Journal of Network and Computer Applications between 1997 and 2019. J. Netw. Comput. Appl. **165**, 102695 (2020)

Authors and Affiliations

Amit K. Shukla[1] · **Vagan Terziyan**[2] · **Timo Tiihonen**[2]

✉ Amit K. Shukla
amit.shukla@uwasa.fi

Vagan Terziyan
vagan.terziyan@jyu.fi

Timo Tiihonen
timo.tiihonen@jyu.fi

[1] School of Technology and Innovations, University of Vaasa, Wolffintie 34, 65200 Vaasa, Finland

[2] Faculty of Information Technology, University of Jyvaskyla (Agora), Box 35, 40014 Jyvaskyla, Finland