Knowledge Combination Analysis Reveals That Artificial Intelligence Research Is More Like "Normal Science" Than "Revolutionary Science"

Jieshu Wang
Arizona State University
jwang490@asu.edu

Andrew Maynard
Arizona State University
Andrew.Maynard@asu.edu

José Lobo
Arizona State University
Jose.Lobo@asu.ed

Katina Michael
Arizona State University
Katina.Michael@asu.eduu

Sébastien Motsch
Arizona State University
Sebastien.Motsch@asu.edu

Deborah Strumsky
Jönköping University
D.Strumsky@JU.se

Abstract

Artificial Intelligence (AI) research is intrinsically innovative and serves as a source of innovation for research and development in a variety of domains. There is an assumption that AI can be considered "revolutionary science" rather than "normal science." Using a dataset of nearly 300,000 AI publications, this paper examines the co-citation dynamics of AI research and investigates its trajectory from the perspective of knowledge creation as a combinatorial process. We found that while the number of AI publications grew significantly, they largely follows a normal science trajectory characterized by incremental and cumulative advancements. AI research that combines existing knowledge in highly conventional ways is a substantial driving force in AI and has the highest scientific impact. Radically new ideas are relatively rare. By offering insights into the co-citation dynamics of AI research, this work contributes to understanding its evolution and guiding future research directions.

Keywords: artificial intelligence, scientific research, knowledge combination, bibliographic data, knowledge management, co-citation networks

1. Introduction

Knowledge creation is vital for organizations and society's economy.[1,2] Among the most prominent forms of recorded knowledge are scientific publications. The research and development (R&D) efforts underlying these publications involve information-intensive activities, encompassing tasks like identifying information needs, navigating solution spaces, searching, processing, and validating information to arrive at solutions (Meho & Tibbo, 2003; Wu et al., 2009). Researchers in information systems (IS) have delved into understanding the implications of the knowledge encapsulated in these publications, especially through their citation data. This involves endeavors such as utilizing citation semantics to identify research topics (Brockmann & Roztocki, 2015; Tong et al., 2009), mapping citation networks to predict patent quality (Wang et al., 2009), investigating the implications of citation long tails (Wu et al., 2009), and leveraging co-citation networks to identify central works (Laine, 2009).

Economists and organizational researchers have long emphasized the combination of existing knowledge as a fundamental process of knowledge creation (Nonaka et al., 2000; Schumpeter, 1950). Examining the structures of knowledge combinations can offer insights into the dynamics of invention and innovation, illuminate scientific and technological trends, and provide guidance on predicting and facilitating high-quality knowledge generation (Uzzi et al., 2013). This study initiates by forming co-citation networks from scientific publications to analyze knowledge combinations. It links these combinations to Thomas Kuhn's theory, which differentiates revolutionary from normal science. The hypothesis is that these two types of science may show distinct combinatorial characteristics. Focusing on artificial intelligence (AI) for its transformative potential in science and technology, this research analyzes a dataset of about 300,000 AI publications and their co-citation networks.

[1] This paper is an adaptation of Chapter 2 of Jieshu Wang's doctoral dissertation *Combinatorial Inventions in Artificial Intelligence: Empirical Evidence and Implications for Science, Technology, and Organizations*, submitted to Arizona State University in 2023.

[2] This paper's appendices can be accessed via: https://drive.google.com/drive/folders/1zmJ9bX4DqkscNMjyCXJJyaIYz4e2AjP8?usp=sharing.

URI: https://hdl.handle.net/10125/107058
978-0-9981331-7-1

This research holds strong relevance to the IS domain. It leverages digital data from a web-based knowledge repository, aligning with IS's core focus on information retrieval and processing. Our examination on knowledge search and manipulation through digital data resonates with the main concerns of the IS community — the design and management of information and how it is transformed into knowledge (Borko, 1968). The investigation into combinatorial invention and incremental innovation contributes to the discussion of knowledge management and organizational learning, informing decision-making to facilitate efficient information search, retrieval, and management, fostering competitive advantage (Alavi & Leidner, 2001). Moreover, by tracking the trajectory of AI research, this study sheds light on the evolution of scientific inquiry, aligning with IS's interdisciplinary nature, which evolves by integrating theories and methodologies from diverse domains like computer science and economics (Benbasat & Zmud, 2003).

This paper is structured in the following manner. Section 2 reviews literature and poses research questions. Section 3 describes the methodology and data collection. Section 4 details the dataset and presents the results. Section 5 discusses the implications, while section 6 addresses limitations and future direction.

2. Background

AI research and applications are intrinsically innovative and serve as a tool and source of innovation for R&D in a variety of domains. AI shows tremendous potential to affect the search for information and solutions in many domains ranging from medicine to transportation, from education to manufacturing (Maynard, 2015). In IS research, AI's potential have been extensively explored, encompassing domains such as AI-based group support systems (Siemon et al., 2015), knowledge representation (Fenstermacher, 2005), information retrieval (Pathak et al., 2000), organizational knowledge management (Hine et al., 1994), human-AI cooperation (Schelble et al., 2021), and software engineering (Latinovic & Pammer-Schindler, 2021). AI is widely expected to become a general-purpose technology, serving as a "new method of invention" that will transform the very processes of scientific discovery, invention, and innovation (Hastings, 2023).

The transformative potential of AI for invention can be understood by its capacity to facilitate knowledge creation through a *combinatorial* process. New knowledge can be conceptualized as the recombination of existing knowledge (Arthur, 2009; Schumpeter, 1950; Youn et al., 2015). As societal knowledge accumulates, the scope of possible combinations expands exponentially. Uncovering novel knowledge within such a vast and complex space is akin to a "needle-in-a-haystack" problem, which has become prevalent in science and technology, challenging for individual researchers (Agrawal et al., 2018). AI, however, is poised to address these challenges by excelling in searching complex solution spaces to identify relevant information and viable, thus possibly valuable knowledge combinations (Cockburn et al., 2019).

The concept of "revolutionary science" versus "normal science," as elucidated by Thomas Kuhn (1962), distinguishes radical novelty and paradigm shifts (revolutionary science) from incremental knowledge growth built upon well-accepted prior knowledge (normal science). Viewing knowledge creation through the combinatorial perspective, revolutionary ideas would likely exhibit distinct patterns from "normal" knowledge. Revolutionary knowledge may entail relatively novel or infrequent combinations, while normal knowledge would rely heavily on frequently recurring and extensively used combinations. The capabilities of AI in searching combinations have led many people to believe that AI is intrinsically revolutionary (Appenzeller, 2017; Tegmark, 2017; Vatan et al., 2019), and it raises intriguing questions about the extent of its potential for revolution informed by its combinatorial nature. Is AI's combinatorial behavior indicative of its revolutionary character, or does it align with the patterns seen in previous changes in science and technology?

Researchers have been using scientific publications to investigate knowledge evolution in AI research. Some researchers attempt to identify significant trends. Using AI publication datasets, Cockburn et al. (2019) found that the AI community is shifting towards more application-oriented research, while Niu et al. (2016) surveyed where the highest productivity resides in AI. Meanwhile, researchers like Raghupathi and Nerur (1999) examined the author co-citation networks and identified research themes. Others investigated AI within broader knowledge landscapes, exemplified by Frank et al. (2019), who explored the interdisciplinary knowledge flows linked to AI through citations. On the other hand, researchers have explored combinatorial knowledge creation across wider domains. Uzzi et al. (2013) developed methods to identify knowledge combinations. Hofstra et al. (2020) examined keywords combinations, while Strumsky and Lobo (2015), Youn et al. (2015), and Kim et al. (2016) explored how technical components are combined in patents. Nevertheless, no prior research has specifically investigated the field of AI to uncover the dynamics of combinatorial knowledge creation and evolution within it.

This exploratory study aims to investigate *how the characteristics of knowledge combination inform the extent of revolutionary nature in AI research*. The study addresses three sub-questions:

RQ 1 How is existing knowledge combined within the field of AI to generate new knowledge? How novel are those knowledge combinations?

RQ 2 How is knowledge recombination associated with scientific impact?

RQ 3 How do the answers to the preceding RQs illuminate the revolutionary nature of AI research?

3. Method and Data

3.1. Knowledge Recombination Taxonomy

This study employed a variation of the method developed by Uzzi et al. (2013), to identify and measure the novelty of knowledge recombinations in scientific literature.

The *typicality* of a knowledge combination is defined as the frequency of two knowledge components co-appearing together, standardized against all the combinations during a select period. We consider cited journals as representing the previous knowledge combined in a publication. Thus, the pairwise combinations of referenced journals are identified to represent the recombination of existing knowledge. For a co-cited journal pair i and j in year t, its cumulative typicality, or z-score, can be computed using Eq. 1:

$$z_{(i,j),t} = \frac{\sum_{n=1946}^{t} x_{(i,j),n} - \mu_t}{\sigma_t} \quad (1)$$

where $x_{(i,j),n}$ is the observed frequency of journal-pair (i, j) in year n; μ_t is the mean frequency of all pairs up to year t; and σ_t denotes its standard deviation.

Each AI publication is associated with a set of z-scores, describing the standardized cumulative frequency of each journal pair. Two statistical attributes are extracted to characterize its conventional and novel combinations.

The first attribute is ***Tail Conventionality** (TC_p)*, defined as the 80th percentile of the z-scores associated with paper p, featuring the typicality of its right tail, where z-scores are relatively high, and pairs appear more conventional. The TC_p characterizes the paper's tendency to combine conventional pairs. A paper is considered highly conventional if its TC_p is in the upper half of the TC_p of all the papers published previously. The second attribute is ***Tail Novelty** (TN_p)*, defined as the 20th percentile of z-scores of paper p, characterizing its more unusual journal combinations where novelty might dwell. A paper is considered highly novel if its TN_p is less than 0.

A paper falls into one of four categories defined by the Knoweldge Recombination Taxonomy described by Mukherjee et al. (2016): ***Darwin's Tower*** is a category where papers have both high conventionality and novelty (high TC and negative TN). ***Avant Garde*** papers have low conventionality but a high novelty. ***Accepted Wisdom*** papers have high conventionality but a low novelty. ***Platypus*** papers have low conventionality and low novelty. [3]

3.2. Assessing Scientific Impact

Citation counts are considered an indicator for scientific quality and impact, especially at the highly-cited end of the distribution (Phelan, 1999). However, citation counts and citation rates (those received in a given year) do not stay constant over time (Ponomarev et al., 2012). Therefor, it becomes problematic to directly compare the citation counts of two papers published in different years. Instead, researchers often select the mean annual citation rate (or "annual citation") to compensate for this time effect of citation counts (Unger et al., 2018).

This study employs three metrics to gauge papers' scientific impact: total citations, annual citations, and its percentile ranking compared to papers published in the same year. Total citations reflect the paper's overall impact, while annual citations indicate its average impact over time. The percentile ranking compensates for temporal variations in citations. Papers ranking high in citation percentiles are referred to as "hit papers."

3.3. Data

The AI publication dataset for this study was compiled through keyword searching on WOS Core Collection (Fig. 1).[4] Our dataset was collected through a snowball sampling that utilized the "*Topic*" search[5] in WOS in July 2020. Snowball sampling is

[3] A case example is offered in Appendix H to illustrates such categorization process.

[4] According to the institutional subscription available to the authors, the WOS databases searched in this study include Science Citation Index Expanded, Social Science Citation Index, Arts and Humanities Citation Index, and Emerging Sources Citation Index (only 2015-2020).

[5] The *Topic* field in WOS searches the title, abstract, author keywords, and *Keywords Plus* of a record. *Keywords Plus* are words or

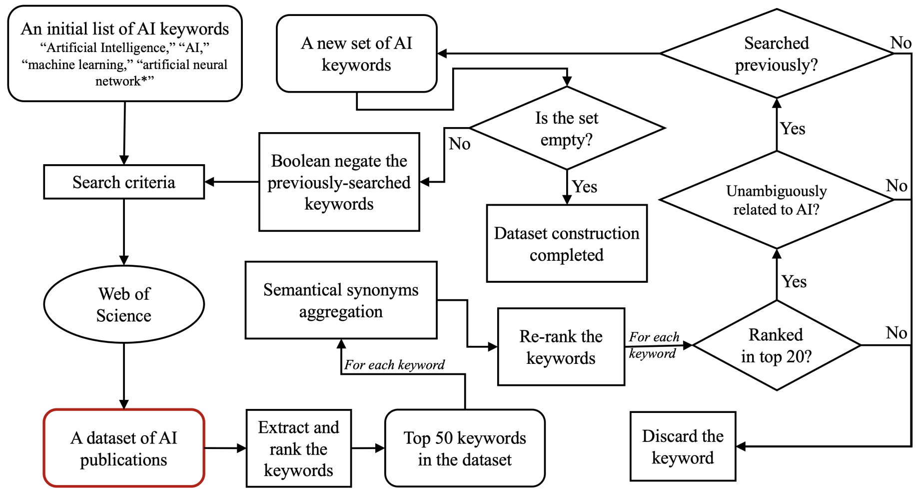


Figure 1: The process of constructing AI research publication dataset.

a sampling technique commonly used in sociology to recruit subjects through existing subjects (Biernacki & Waldorf, 1981). It is also used in bibliometric research to acquire further literature through the references in existing literature (Roetzel, 2018). Here in this study, we uses a snowball sampling process to obtain search terms. Using a "seed set" of four keywords[6], we retrieved a seed dataset, from which more keywords are obtained and fed into subsequent iteration of search. We semantically aggregated synonyms for the top keywords (Appendix B) and heuristically evaluated whether they are unambiguously related to AI (Appendix C). The keywords that may have ambiguous meanings or may direct to non-AI papers are discarded.[7]

4. Results

4.1. Data Description

The compiled dataset contains metadata of 296,378 AI publications spanning from 1946 to 2020. These publications reference 19,474 journals, resulting in nearly 7.8 million distinct journal pairs.

The articles in our dataset are mainly from the fields of engineering and computer science,[8] accounting for nearly 40% each. They are followed by the fields such as chemistry, mathematics, and telecommunications. Among the frequently used keywords, prominent terms include "machine learning," "genetic algorithm," "deep learning," and "artificial neural networks." Noteworthy journals within our dataset include *IEEE Access*, *Expert Systems with Applications*, *Neurocomputing*, and *Sensors*. The paper with the highest citation count is "Random Forests" by Breiman (2001), and most-cited authors include Geofferey Hinton and Yoshua Bengio. [9]

Topic modeling using the Non-negative Matrix Factorization (NMF) technique on the publications' titles revealed ten distinct topics, including data-driven optimization and classification, model design, predictive analytics using neural networks, and AI applications regarding chemical and protein structures. [10]

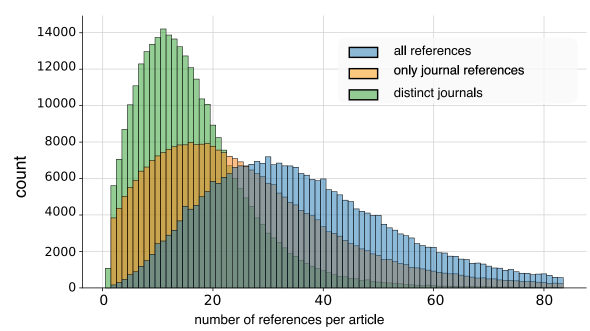


Figure 2: Histogram plot showing the distribution of AI publications' reference count (blue), referenced journal article count (orange), and distinct journal count (green).

The distributions of previous works cited by AI publication are shown in Fig. 2. On average, an AI

phrases generated by an algorithm to identify keywords that frequently appear in the titles of an article's reference but not in the title of the article.

[6]The four seed keywords are "artificial intelligence," "AI," "machine learning," and "artificial neural network*."

[7]Detailed data collection process can be found in Appendix A.

[8]We followed the research areas classified by WOS.

[9]A comprehensive dataset description can be found in Appendix E.

[10]A breakdown of these topics can be found in Appendix F.

publication has 41 references, among which about 28 are journal articles from roughly 16 distinct journals. This implies that AI publications often cite multiple articles from the same journals.

Upon examination of the journal pairs, several observations emerge. First, most pairs were combined only a few times. By the year 2020, approximately 40% of the journal pairs have been cited once, while 73% have no more than 5 co-citations. Secondly, frequently co-cited pairs tend to come from the same or closely related disciplines, for example, *International Journal of Remote Sensing* combined with *Remote Sensing of Environment* (both related to remote sensing), and *Bioinformatics* combined with *Nucleic Acids Research* (both related to biology). [11]

4.2. Exponential Growth of AI Publications

AI papers are growing exponentially (Fig. 3a). Before 1984, annual publications never exceeded 100. It took another ten years for it to grow beyond 1,000. In 2019, over 50,000 publications were recorded.[12]

It is noticeable that there is a broader trend in exponential growth amongst scientific publications, as shown in Fig. 3a. However, the slope of the fitted line for the natural logarithm of the annual count of all scientific publications against time is 0.0484,[13] less than 30% of the magnitude for the coefficient of AI publications. Consequently, the ratio of AI publications compared to scientific publications is increasing, as shown in Fig. 3b. By 2019, AI publications made up 1.73% of all scientific publications.

Fig. 3a shows that there were a few years where the number of AI publications fluctuated and declined slightly, particularly in the early 1970s and late 1980s, corresponding to the historical periods referred to as "AI winters," when funding and interests in AI research were drastically reduced (Howe, 2007). Nevertheless, AI has seen steady growth over the past two decades. Indeed, the world appears to be in the midst of an "AI spring."

4.3. The Skewness of AI Research

AI publications have generated almost 6 million citation counts in total. On average, each publication is cited almost 20 times. However, the distribution of citations received by AI publications is extremely skewed (see Fig 4). This phenomenon has been observed in information science (Wu et al., 2009) and scientific publications in general (Seglen, 1992). We found that 19% of AI publications have not been cited at all. Ten percent have been cited once, and 50% have no more than five citations. Moreover, AI publications can serve as another empirical demonstration of the Pareto principle of the 20/80 rule, which states that roughly 80% of consequences come from 20% of causes (Pareto et al., 1964), as the AI publications ranked among the top 20% in citations make up almost 80% of all citations.

4.4. Conventional Knowledge Driving AI Growth

The category of *Accepted Wisdom*, which combines high conventionality with low novelty, accounts for the largest share (46.4%) in AI research, followed by Darwin's Tower (11.5%). The time series (Fig. 5) reveals a significant trait: *Accepted Wisdom* in AI research has been advancing steadily over the last three decades. By 2020, new publications categorized as *Accepted Wisdom* reached 52.4%. *Avant Garde*, the second largest category, has been stable with a slight decrease in recent years. *Darwin's Tower* that mixes high novelty with high conventionality has been diminishing, from above 20% in the 1990s to only 6% in 2020. This suggests that new knowledge that relies heavily on conventional combinations has become a significant driving force of AI.

4.5. Conventional Knowledge Exerting Greater Impact

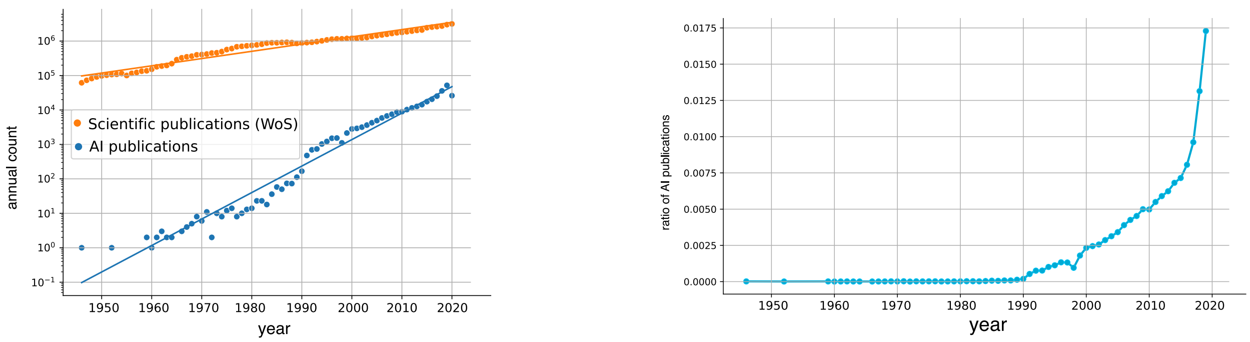
Section 4.3 shows that AI publications' citations exhibits a skewed distribution. The question is whether a paper's category relates to its impact. We found that *Accepted Wisdom* is associated with high impact.

Among the publications highly ranked for citation percentiles of the year, *Accepted Wisdom* occupies an even greater percentage, 67% in the top 1% (Table 1). The higher the group's rank, the larger its share becomes. As addressed in Section 4.4, *Accepted Wisdom* makes up 46% of all AI publications. In the AI publications ranked in the top 10% among the papers in the same year, this category accounts for as high as 60% of publications. In the top 5% and top 1%, this percentage rises to 62% and 67%, respectively. In the 10 most-cited papers, seven belong to *Accepted Wisdom*. In contrast, *Avant Garde* papers decreased relative to other categories, with 35% in all papers yet declining to only 17% in the top 1% group. There is zero *Avant Garde* paper in the top 10 papers. The share of *Darwin's Tower* category remains stable across percentiles, regardless of whether top 1%, 5%, 10%, or all papers are considered.

[11] This is illustrated in Appendix G, where a sample of the most co-cited journal pairs is presented. These pairs often come from the same field, such as bioinformatics, neural science, or chemistry.

[12] OLS regression between years and natural logarithms of annual publication counts resulted in a coefficient of 0.1736 with an R-squared of 0.97 and a p-value of 1.52×10^{-48}.

[13] OLS regression result shows $R^2 = 0.949$, $p = 4.50 \times 10^{-49}$

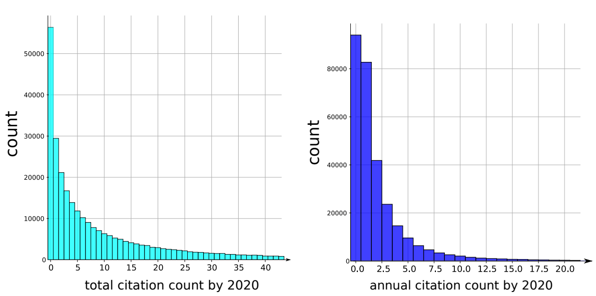


(a) (b)

Figure 3: Time Series of (a) the number of AI publications and the number of scientific publications in general (log scale) and (b) the ratio of AI publications in scientific publications in general.

Table 1: Percentages of Each Category in Top AI publications Regarding Citation Percentile of the Year

Category	**% in all**	**% in top 10%**	**% in top 5%**	**% in top 1%**	**% in top 10papers**	**% in most-cited 10 papers**
Accepted Wisdom	46.39%	60.38%	62.21%	67.34%	100%	70%
Darwin's Tower	11.54%	10.22%	10.23%	8.82%	0%	20%
Platypus	7.40%	6.72%	6.80%	6.77%	0%	10%
Avant Garde	34.68%	22.68%	20.76%	17.07%	0%	0%



(a) (b)

Figure 4: Distribution of AI publications' (a) citation counts and (b) average annual citation counts by 2020.

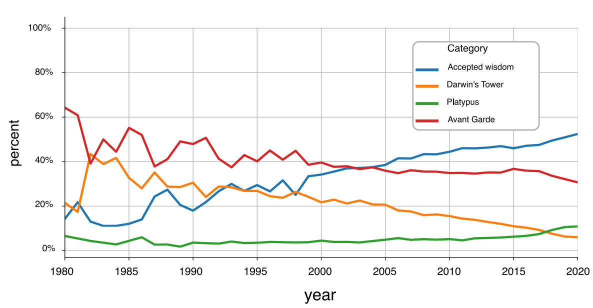


Figure 5: Time Series of AI publications' composition.

Additional evidence for *Accepted Wisdom*'s dominance in the top tiers are presented in figure 6,

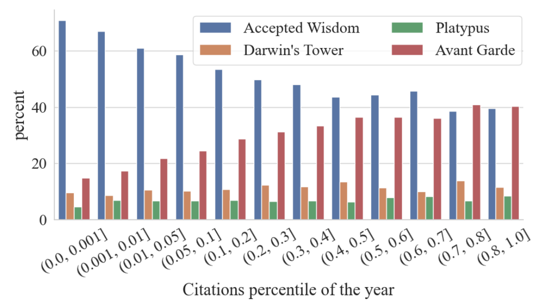


Figure 6: Percentages of each category in AI publications grouped by citation percentile of the year. The leftmost group represents AI publications with the highest citations among its peers in the same year, ranked in the top 0.1%.

where AI publications are arranged into 12 groups based on their citation percentiles. It clearly shows that *Accepted Wisdom* (blue) is increasing towards the left, while the *Avant Garde* category is decreasing.

Moreover, as shown in Table 2, *Accepted Wisdom* accounted for more than half (51%) of all citations, larger than its share in publication count (46%). It is also highly ranked regarding mean and median annual citations. The median citation percentile of *Accepted Wisdom* is 46%. However, *Darwin's Tower* has the highest mean and median citation count. It suggests that *Darwin's Tower* is less skewed in citations than *Accepted Wisdom* and *Darwin's Tower* has fewer papers ranked at the bottom. Fig. J.1a confirms this by showing that the orange curve that represents *Darwin's Tower* has the lowest curvature, indicating a smaller degree of skewness in citations.

To further verify the relationship between the categorization of an AI publication and its scientific impact, a negative binomial regression analysis is conducted. [14] The following regression equation can be assumed:

$$\ln Y = \beta_0 + \beta_1 C + \beta_2 t + \epsilon \tag{2}$$

where Y denotes the number of citation counts an AI publication receives as recorded in the dataset by 2020. C denotes the category of the AI publication (a categorical variable), and t denotes the year of publication. [15] The coefficient β_1 for the four possible values of variable C are computed as 1, -0.29, -0.46, -0.52 for *Accepted Wisdom*, *Darwin's Tower*, *Platypus*, and *Avant Garde* respectively, with Accepted Wisdom as the controlled value and the other three as treatments. All p-values are smaller than 0.001. The coefficients indicate that, among the four categories, an *Accepted Wisdom* tends to have the highest citation counts, followed by *Darwin's Tower* and *Platypus*.

In summary, *Accepted Wisdom* exert a higher impact, and they take the highest share among the articles with the highest citations.

4.6. Accepted Wisdom Has High Hit-rate

We found *Accepted Wisdom* type of AI papers have the highest probability of being ranked in the top tier in terms of citations, annual citations, or citations among its peers of the year. In other words, *Accepted Wisdom* has the highest hit-rate (see Table 3, Fig J.1, and Fig J.2). If hit papers are evenly distributed across the four categories, the hit-rate would be identical across categories (referred to as the background hit-rate). That is to say, the top 10% hit-rate of *Accepted Wisdom* would be the same as that of *Avant Garde*, which would be exactly equal to 10%. However, we found evidence suggesting otherwise. Nearly 13% of papers in *Accepted Wisdom* are ranked in the top 10%. In contrast, only less than 7% of papers in *Avant Garde* are ranked in the top 10% (Table 3). Tests conducted with four different hit rates (10%, 5%, 1%, and 0.1%) confirm the robustness of such an observation — the hit-rate of papers classified under *Accepted Wisdom* consistently surpasses the background hit rate, consistently ranking the highest among the four categories This indicates that *Accepted Wisdom* papers are more likely to receive more citations compared to other papers published in the same year. Nevertheless, it is worth noting that a paper in the *Darwin's Tower* category, which exhibit both high conventionality with high novelty, while having a lower likelihood of ranking in top 10%, 5%, and 1%, have a notably higher chance (0.16%) of achieving a top 0.1% ranking than the background. This finding suggests that *Darwin's Tower* displays a polarized trend. While the average paper in this category performs moderately, there is a heightened probability of these papers excelling and becoming among the best of the best, falling within the top one thousandth.

We further examined the time series of hit rate of each category (Fig. 7). It illustrates that the *Accepted Wisdom* category (blue) has become the category with the highest hit probability since the 2000s. In 2016, almost 15% of *Accepted Wisdom* papers are ranked in the top 10% papers, with a hit rate significantly higher than the background rate (10%).

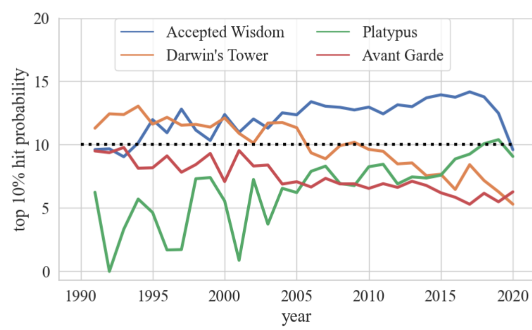


Figure 7: Time series of top 10% hit rate, with the dashed black line representing the background hit rate.

5. Conclusion and Discussion

This study provides empirical evidence that AI research has largely progressed incrementally and follows conventional scientific inquiry, aligning with the development patterns of normal science. Highly conventional AI publications yield higher scientific impact. Radically new research are rare. Our finding does not contradict AI's revolutionary potential; rather, it portrays a revolutionary technology unfolding through notably "normal" and incremental progress. This observation suggests that revolutionary science and

[14]Because citation is count data, a Poisson regression is generally applicable. However, the variances of citation counts in each category are substantially larger than the mean values (see Appendix K). Due to this over-dispersion, the assumption of Poisson regression is not satisfied. A negative binomial regression is considered a more suitable model to compensate for such an over-dispersion.

[15]The summary of negative binomial regression results of Equation 2 can be found in Appendix L.

Table 2: Selected Citation Features of the Four Categories

Category	Total count	Total citation	Citation percentage	mean citation	median citation	mean annual citation	median annual citation	median citation percentile
Accepted Wisdom	**137,481**	**3,034,185**	**51.31%**	22.07	6	**3.04**	**1.17**	**46%**
Darwin's Tower	34,203	971,780	16.43%	**28.41**	**8**	2.54	1.00	51%
Platypus	21,922	270,211	4.57%	12.33	3	1.97	0.78	52%
Avant Garde	102,772	1,636,783	27.68%	15.93	5	1.86	0.83	52%

Table 3: Hit rates of the Four Categories

Category	Top 10% hit rate	Top 5% hit rate	Top 1% hit rate	Top 0.1% hit rate
Accepted Wisdom	**12.78%**	**6.52%**	**1.47%**	**0.16%**
Darwin's Tower	8.76%	4.36%	0.82%	**0.13%**
Platypus	8.94%	4.49%	0.94%	0.08%
Avant Garde	6.43%	2.93%	0.51%	0.06%

technology can emerge deliberately through routine practices rather than hinging solely on extraordinary breakthrough moments.

The merits of AI's "normal" trajectory should not be underestimated. Throughout history, revolutionary technologies have not always translated into unequivocal societal benefits, and they sometimes rapidly impact the society and lead to unintended consequences before full comprehension. By predominantly relying on incremental and cumulative progress, AI research maintains a pace advantageous for necessary regulation and control, allowing for adjustments that align with human values, emphasizing stability, robustness, and sustainable growth.

Our examination of combinatorial knowledge creation can serve as a model for understanding the dynamics of knowledge integration within IS-related domains. We emphasize incremental knowledge creation, resonating with findings in IS where foundational works often serve as cornerstone references for extended periods (Webster & Watson, 2002). It provides insights into how knowledge can be catalogued, accessed, searched, retrieved, and disseminated in databases and digital libraries. In organizations, combinotorial search can foster R&D and organizational learning. Nevertheless, our findings reveal a potential tension between novel knowledge combinations and academic impact, as the papers featuring conventional pairs tend to receive more citations while highly novel papers are less cited, raising questions about the reward mechanisms in academic publishing, a concern echoed in IS community (Agarwal & Lucas, 2005).

Furthermore, this research offers a valuable literature review tool for researchers, particularly those with interdisciplinary interests. For example, our methodology unveils journal pairs that denote the most prominent interdisciplinary intersection between IS and AI, such as (1) *Information Systems Research (ISR)* + *MIS Quaterly*, (2) *ISR* + *Management Science*, (3) *ACM TOIS* + *Information Processing and Management*, (4) *Expert Systems with Applications* + *Information Systems*, and (5) *ACM TOIS* + *IEEE Transactions on Knowledge and Data Engineering*.

Furthermore, our approach can serve as a tool to identify research gaps and potential opportunities by highlighting journal pairs that are infrequently co-cited, indicating less-explored areas. Here are some examples of underrepresented co-cited journal pairs within the context of AI and IS, pointing to potential research gaps: (1) Information Systems and *Zootaxa*: AI's use in animal species categorization, (2) *The Information Systems Journal (ISJ)* and *The Washington Law Review*: addressing ethical and legal aspects of AI and IS, such as data privacy, (3) *Tourism Economics* and *Information Systems*: IS-related AI applications in optimizing tourism, such as personalized recommendations, (4)*Teaching of Psychology* and *ISR*: AI-driven psychology education, (5) *The ACM Transactions on Asian and Low-Resource Language Information Processing* and *Information Processing and Management*: AI's role in processing underrepresented languages, (6) *Journal of Alzheimer's Disease* and *JMIS*: AI's contribution to medical research, such as early diagnosis and information management.[16]

6. Limitations and Future Work

This study is based only on AI publications that went through standard academic publishing practice particularly the peer-review process of journals. This excludes reports, non-reviewed materials, and informal discussions such as those found on platforms like arXiv, patents, Github

[16] Further discussion and examples can be found in Appendix M.

records, and social media. In particular, limited by our institutional subscription to WOS, conference proceedings were not systematically searched in this study, although our dataset does encompass some conference proceedings if they are recognized by WOS as academic journals.[17] Conferences are increasingly crucial for AI breakthroughs, yet many impactful conference papers eventually find their way into traditional academic journals.[18] In addition, AI researchers' expanding presence on platforms like arXiv and Github, often with code sharing, underscores the need to consider these avenues in future assessments of AI research.

This study is also limited by the search terms used to construct the dataset. The data collection process involves manual and heuristic evaluation regarding whether the terms are unambiguously related to AI, and it might have introduced ambiguities or inaccuracies in synonym aggregation due to the authors' AI knowledge constraints. Moreover, the evolving nature of keywords in AI research were not explicitly captured. Language choice is another limitation; as WOS predominantly comprises English publications, our dataset largely mirrors this bias.[19] Therefore, it will be crucial to include non-English publications in order to achieve a broader, culturally diverse view of AI research.

In addition, our premise that a journal represents a unit of knowledge simplifies the complex nature of academic journals, particularly those spanning diverse themes, such as *Science* and *Nature*. A more nuanced approach could involve keywords for a more granular representations of knowledge (Hofstra et al., 2020). Furthermore, delving into semantic analysis of journal pairs, and systematically understanding their contextual meanings rather than just counting co-citation frequencies, would offer richer insights to researchers interested in AI.

[17]For instance, papers published on *Proceedings of the Association for Information Science and Technology* are included in our dataset. However, we are uncertain about how WOS classifies certain conference proceedings as journals while excluding others.

[18]For instance, the paper "*ImageNet Classification with Deep Convolutional Neural Networks*" by Krizhevsky et al. (2012) was initially presented at the Conference on Neural Information Processing Systems (NIPS) in 2012, and later submitted and published in the journal of *Communications of ACM*, and this paper is included in our dataset. Similarly, Mnih et al.'s 2015 paper "*Human-level control through deep reinforcement learning*," published on *Nature*, was based on their NIPS paper "*Playing Atari with Deep Reinforcement Learning*" presented in 2013. It is also included in our dataset.

[19]Ninety-eight percent of papers in our dataset are written in English, 0.5% in Chinese, 0.3% in Spanish, and 0.8% in others. Nevertheless, China has become a significant player in AI research. Although many Chinese-speaking authors chose to publish their works in English journals (the dataset used in this research shows that authors based in China have published more AI papers than in other countries, and among the top 10 research institutions that have the most AI publications, five are located in China), publications in the Chinese language are likely to be significant in number.

References

Agarwal, R., & Lucas, H. C. (2005). The Information Systems Identity Crisis: Focusing on High-Visibility and High-Impact Research. *MIS Quarterly*, *29*(3), 381–398. https://doi.org/10.2307/25148689

Agrawal, A., McHale, J., & Oettl, A. (2018, April). *Finding Needles in Haystacks: Artificial Intelligence and Recombinant Growth* (Working paper No. w24541). National Bureau of Economic Research. Cambridge, MA. https://doi.org/10.3386/w24541

Alavi, M., & Leidner, D. E. (2001). Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. *MIS Quarterly*, *25*(1), 107–136. https://doi.org/10.2307/3250961

Appenzeller, T. (2017). The AI Revolution in Science. *Science*, *357*(2017), 16–17. https://doi.org/10.1126/science.aan7064

Arthur, W. B. (2009). *The Nature of Technology: What It Is and How It Evolves*. Free Press.

Benbasat, I., & Zmud, R. W. (2003). The Identity Crisis within the IS Discipline: Defining and Communicating the Discipline's Core Properties. *MIS Quarterly*, *27*(2), 183–194. https://doi.org/10.2307/30036527

Biernacki, P., & Waldorf, D. (1981). Snowball Sampling: Problems and Techniques of Chain Referral Sampling. *Sociological Methods & Research*, *10*(2), 141–163. https://doi.org/10.1177/004912418101000205

Borko, H. (1968). Information science: What is it? *American documentation*, *19*(1), 3–5. https://doi.org/10.1002/asi.5090190103

Brockmann, C., & Roztocki, N. (2015). Topics on Knowledge Management: An Empirical Insight into Articles Published in the International Journal of Knowledge Management. *2015 48th Hawaii International Conference on System Sciences*, 3834–3840. https://doi.org/10.1109/HICSS.2015.460

Cockburn, I. M., Henderson, R., & Stern, S. (2019). The Impact of Artificial Intelligence on Innovation: An Exploratory Analysis. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press. Retrieved November 17, 2020, from http://www.nber.org/books-and-chapters/economics-artificial-intelligence-agenda

Fenstermacher, K. (2005). The Tyranny of Tacit Knowledge: What Artificial Intelligence Tells us About Knowledge Representation. *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, 243a–243a. https://doi.org/10.1109/HICSS.2005.620

Frank, M. R., Wang, D., Cebrian, M., & Rahwan, I. (2019). The Evolution of Citation Graphs in Artificial Intelligence Research. *Nature Machine Intelligence*, *1*(2), 79–85. https://doi.org/10.1038/s42256-019-0024-5

Hastings, J. (2023). *AI for scientific discovery* (First edition.). CRC Press, Taylor & Francis Group. https://doi.org/10.1201/9781003226642

Hine, Goul, & Howorka. (1994). Distributed artificial intelligence: Impacts and implications for future knowledge-based organizations. *1994 Proceedings of the Twenty-Seventh Hawaii International Conference on System Sciences*, *3*, 211–220. https://doi.org/10.1109/HICSS.1994.323350

Hofstra, B., Kulkarni, V. V., Munoz-Najar Galvez, S., He, B., Jurafsky, D., & McFarland, D. A. (2020). The diversity-innovation paradox in science. *Proceedings of the National Academy of Sciences - PNAS*, *117*(17), 9284–9291. https://doi.org/10.1073/pnas.1915378117

Howe, J. (2007, June). History of Artificial Intelligence at Edinburgh. Retrieved August 26, 2020, from http://www.inf.ed.ac.uk/about/AIhistory.html

Kim, D., Cerigo, D. B., Jeong, H., & Youn, H. (2016). Technological novelty profile and invention's future impact. *EPJ Data Science*, *5*(1), 1–15. https://doi.org/10.1140/epjds/s13688-016-0069-1

Kuhn, T. S. (1962). *The structure of scientific revolutions*. University of Chicago Press.

Laine, M. (2009). Virtual Communities: A Bibliometric Analysis. *2009 42nd Hawaii International Conference on System Sciences*, 1–10. https://doi.org/10.1109/HICSS.2009.500

Latinovic, M., & Pammer-Schindler, V. (2021). Automation and Artificial Intelligence in Software Engineering: Experiences, Challenges, and Opportunities: The 54th Hawaii International Conference on System Sciences (T. X. Bui, Ed.). *Proceedings of the 54th Annual Hawaii International Conference on System Sciences, HICSS 2021*, 146–155. https://doi.org/10.24251/HICSS.2021.017

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436. https://doi.org/10.1038/nature14539

Maynard, A. (2015). Navigating the fourth industrial revolution. *Nature Nanotechnology*, *10*(12), 1005–1006. https://doi.org/10.1038/nnano.2015.286

McLevey, J., & McIlroy-Young, R. (2017). Introducing metaknowledge: Software for computational research in information science, network analysis, and science of science. *Journal of Informetrics*, *11*(1), 176–197. https://doi.org/10.1016/j.joi.2016.12.005

Meho, L. I., & Tibbo, H. R. (2003). Modeling the information-seeking behavior of social scientists: Ellis's study revisited. *Journal of the American Society for Information Science and Technology*, *54*(6), 570–587. https://doi.org/10.1002/asi.10244

Minsky, M. L., & Papert, S. (1988). *Perceptrons: An introduction to computational geometry* (Expanded ed). MIT Press.

Mukherjee, S., Uzzi, B., Jones, B., & Stringer, M. (2016). A New Method for Identifying Recombinations of Existing Knowledge Associated with High-Impact Innovation. *Journal of Product Innovation Management*, *33*(2), 224–236. https://doi.org/10.1111/jpim.12294

Niu, J., Tang, W., Xu, F., Zhou, X., & Song, Y. (2016). Global Research on Artificial Intelligence from 1990–2014: Spatially-Explicit Bibliometric Analysis. *ISPRS international journal of geo-information*, *5*(5), 66. https://doi.org/10.3390/ijgi5050066

Nonaka, I., Toyama, R., & Konno, N. (2000). SECI, Ba and Leadership: A Unified Model of Dynamic Knowledge Creation. *Long Range Planning*, *33*(1), 5–34. https://doi.org/10.1016/S0024-6301(99)00115-6

Pareto, V., Bousquet, G.-H., & Busino, G. (1964). *Cours d'économie politique*. Droz.

Pathak, P., Gordon, M., & Fan, W. (2000). Effective information retrieval using genetic algorithms based matching functions adaptation. *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, 8 pp. vol.1–-. https://doi.org/10.1109/HICSS.2000.926653

Phelan, T. J. (1999). A compendium of issues for citation analysis. *Scientometrics*, *45*(1), 117–136. https://doi.org/10.1007/bf02458472

Ponomarev, I. V., Williams, D. E., Lawton, B. K., Cross, D. H., Seger, Y., Schnell, J., & Haak, L. (2012). Breakthrough Paper Indicator: Early detection and measurement of ground-breaking research. *Proceedings of the 11th international conference on current research information systems*, 295–304.

Raghupathi, W., & Nerur, S. (1999). Research themes and trends in artificial intelligence: An author co-citation analysis. *Intelligence (New York, N.Y. : 1999)*, *10*(2), 18–23. https://doi.org/10.1145/309697.309703

Roetzel, P. G. (2018). Information overload in the information age: A review of the literature from business administration, business psychology, and related disciplines with a bibliometric approach and framework development. *Business research (Göttingen)*, *12*(2), 479–522. https://doi.org/10.1007/s40685-018-0069-z

Schelble, B., Flathmann, C., Canonico, L.-B., & Mcneese, N. (2021). Understanding Human-AI Cooperation Through Game-Theory and Reinforcement Learning Models. https://doi.org/10.24251/HICSS.2021.041

Schumpeter, J. A. (1950). *Capitalism, Socialism, and Democracy* (3d ed.). Harper.

Seglen, P. O. (1992). The Skewness of Science. *Journal of the American Society for Information Science*, *43*(9).

Siemon, D., Eckardt, L., & Robra-Bissantz, S. (2015). Tracking Down the Negative Group Creativity Effects with the Help of an Artificial Intelligence-Like Support System. *2015 48th Hawaii International Conference on System Sciences*, 236–243. https://doi.org/10.1109/HICSS.2015.37

Strumsky, D., & Lobo, J. (2015). Identifying the sources of technological novelty in the process of invention. *Research policy*, *44*(8), 1445–1461. https://doi.org/10.1016/j.respol.2015.05.008

Tegmark, M. (2017). *Life 3.0: Being human in the age of artificial intelligence*. Alfred AKnopf.

Tong, T., Dinakarpandian, D., & Lee, Y. (2009). Literature Clustering using Citation Semantics. *2009 42nd Hawaii International Conference on System Sciences*, 1–10. https://doi.org/10.1109/HICSS.2009.294

Unger, J. M., Barlow, W. E., Tangen, C. M., Ramsey, S. D., Thompson, I. M., Klein, E. A., LeBlanc, M., Blanke, C. D., Goodman, P. J., Minasian, L. M., Nghiem, V. T., & Hershman, D. L. (2018). The scientific impact and value of large, NCI-sponsored randomized phase III cancer chemoprevention trials. *Cancer epidemiology*, *55*, 117–122. https://doi.org/10.1016/j.canep.2018.06.005

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. *Science*, *342*(6157), 468–472. https://doi.org/10.1126/science.1240474

Vatan, Sharma, A., & Goyal, S. (2019). Artificial Intelligence on the Move: A Revolutionary Technology. *International journal of recent technology and engineering*, *8*(4), 12112–12120. https://doi.org/10.35940/ijrte.D7293.118419

Wang, J.-C., Chiang, C.-H., & Lin, S.-W. (2009). Network Structure of Innovation: Can Brokerage or Closure Predict Patent Quality? *2009 42nd Hawaii International Conference on System Sciences*, 1–10. https://doi.org/10.1109/HICSS.2009.322

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, *26*(2), xiii–xxiii. Retrieved August 29, 2023, from https://www.jstor.org/stable/4132319

Wu, L.-L., Luesukprasert, L., & Lee, L. (2009). Research and the Long Tail: A Large-Scale Citation Analysis. *2009 42nd Hawaii International Conference on System Sciences*, 1–10. https://doi.org/10.1109/HICSS.2009.368

Youn, H., Strumsky, D., Bettencourt, L. M. A., & Lobo, J. (2015). Invention as a combinatorial process: Evidence from US patents. *Journal of the Royal Society, Interface*, *12*(106), 20150272. https://doi.org/10.1098/rsif.2015.0272