Check for updates

Comparing technology convergence of artificial intelligence on the industrial sectors: two-way approaches on network analysis and clustering analysis

Soyea Lee[1] · Junseok Hwang[1] · Eunsang Cho[2]

Abstract

This study investigates technology convergence of AI considering both industrial sectors and technological characteristics with patent data in terms of two-way approaches: IPC-based network analyses and text-based clustering analysis. The IPC-based network analyses, which indicate a top-down approach in this study, focuses on influential technology area with hub nodes and their tie nodes in an IPC-based convergence network. A network centrality analysis is applied to determine the hub nodes which identify notable industrial sectors and influential technology. In addition, an ego-network analysis is conducted to examine the strongly related technology on the hub nodes. Meanwhile, from a bottom-up approach, a text-based clustering analysis is performed and the result shows an applied target of the technology and an integrated form of various technology which are not found from the top-down approach. Consequently, this study suggests new research framework to understand technology convergence based on the industrial sector, influential technology category, and technology application aspects. In line with the findings, this study analyzes technology convergence of AI by the notable industrial sectors: finance/management, medical, transport, semiconductor, game, and biotechnology sector. The results of this study suggest practical implications for AI technology and related industries.

✉ Eunsang Cho
  escho@mmlab.snu.ac.kr

  Soyea Lee
  soyealee@snu.ac.kr

  Junseok Hwang
  junhwang@snu.ac.kr

1 Technology Management, Economics and Policy Program, Seoul National University, Seoul, Republic of Korea

2 Institute of Computer Technology, Seoul National University, Seoul, Republic of Korea

Introduction

The artificial intelligence (AI) software market size in revenue is expected to reach $240.27 billion in 2024 with a compound annual growth rate (CAGR) of 16.7% (IDC, 2020). AI has attracted enormous attention not only in the information and communications technology (ICT) industry but also in a variety of other industries. In the healthcare industry, AI has already begun to transform a variety of aspects, such as offering monitoring, advice to patients and interpretation of scans (Houlton, 2018; Yu et al., 2018). In addition, AI is a key technology for autonomous driving, connectivity, electrification, and the shared mobility trend in the automobile industry (McKinsey, 2018a). Moreover, the expansion of robo-advisors using AI in the finance industry has been utilized (Deloitte, 2016). Likewise, AI is changing the landscape of various industries and applied area has increased gradually.

AI has led to not only technological progress and new innovations but also has the potential to be a general-purpose technology (GPT) (Liu et al., 2021). Research has confirmed how AI can affect technological innovation by improving knowledge creation, knowledge spillover, absorption capabilities, and by increasing investments in R&D, thus explaining the significant relationship between AI and technological innovation and also the positive impact of AI with regard to industry heterogeneity among high- and low-tech sectors (Liu et al., 2020). In addition, relevant study empirically has examined AI, showing that it plays a crucial role in increasing innovation performance at manufacturing enterprises (Yang et al., 2020). Moreover, AI has potential to become a GPT increasing direct productivity and spurring complementary innovations (Brynjolfsson et al., 2017). Because AI can drive innovations and lead to a new paradigm shift by combining various industries, there is a societal need to develop AI while considering its impact on various industries.

Technology convergence has been considered as a tool to drive technological innovation, and interdisciplinary research and the merging of different technologies have therefore increased (Kose & Sakata, 2019). Technology convergence refers to "the process by which two hitherto different industrial sectors come to share a common knowledge and technological base" (Athereye & Keeble, 2000; Rosenberg, 1976). By sharing technological characteristics, the erosion of distinct barriers has been accelerated among various industries (Wang et al., 2019b). Technology convergence leads to industry convergence (Choi et al., 2015; Nystrom, 2008), and industry convergence could only occur with the convergence of technologies (Nystrom, 2008). Therefore, this study attempts to examine the technology convergence of AI considering both technological and industrial perspectives.

Previous studies on the AI have insufficiently investigated the integrated approach considering both the overall industry and individual technology. Applied AI research has generally investigated a specific industrial sector, such as healthcare, vehicle, finance, etc. (Yu et al., 2018; Houlton, 2018; McKinsey, 2018a; Deloitte, 2016). In addition, research on AI patent analysis has analyzed AI technology itself by technological type, firm, and country level (Fujii & Managi, 2018; Tseng & Ting, 2013; WIPO, 2019a). Likewise, those studies generally have not focused on insights across various industries. Meanwhile, some reports on the industrial impact of AI have usually focused on comparing economic impact among industries (PWC, 2018; Deloitte, 2018; McKinsey, 2018b). However, those have not investigated comparing technological aspects according to industries from the perspectives of technology convergence. Therefore, this study attempts to investigate technology convergence of AI considering a set of industries and individual technology.

Research on technology convergence in relation to patent documents has been commonly divided in terms of three perspectives: 1) purpose, 2) methodology, and 3) object of the analysis (Kim & Lee, 2017). In particular, the purposes are two fold, identifying evolutionary trajectory and convergence pattern, and the methodologies are divided into two parts, patent co-citation to examine knowledge flow and patent co-classification to examine convergence phenomenon (Kim & Lee, 2017). Relevant studies of the object of the analysis can be divided according to whether the analysis targets one main technology category or more than two technology categories belonging to heterogeneous industry sectors. In most technology convergence research, these three perspectives have been combined depending on the research questions. The research related to one targeted main technology and the corresponding sub-technologies are as follows. Kim et al. (2014) analyzed the convergence of printed electronics technology based on its element technologies (i.e., device, ink, substrate, circuit, and control) to identify key technologies and their trajectories using co-citation. Han and Shon (2016) analyzed technological convergence in ICT using co-citation to identify crucial roles depending on the period. Wang et al. (2019b) identified emerging topics associated with 3D printing technology depending on time, comparing technology convergence with non-technology convergence environments based on co-classification. Meanwhile, the studies targeted two or more technologies belonging to heterogeneous industry sectors are as follows. Kim and Lee (2017) examine technology convergence in the IT and BT industries to identify key convergence technologies based on co-citation and to forecast future technology convergence. Curran and Leker (2011) analyzed convergence in the areas of NFF and ICT based on co-classification. Kose and Sakata (2019) identified technology convergence in robotics research considering related various sectors extracted from cluster categories such as robot control systems, surgical and medical systems, and automaton in biological and chemistry, among others, based on co-citation.

However, despite their invaluable and meaningful insights, the previous studies have several limitations. First, many of them identify convergence phenomena and trajectories, but there have been insufficient attempts to understand the characteristics of the technology from a multi-dimensional perspective. In other words, few studies have investigated how the technology is actually applied in a relation to the defined convergence phenomenon and/or trajectory. Second, attempts to examine technology convergence from a holistic industrial perspective have been insufficient. That is, many studies have explored convergence while focusing on technology itself and on sub-technologies (e.g., IT and corresponding sub-technologies such as devices and networks) or on combinations of technologies between heterogeneous industries (e.g., IT and BT). However, it is difficult to provide insight from a whole-industry perspective regarding technology.

To overcome these limitations, the paper proposes a two-way approach for technology convergence involving top-down and bottom-up approaches. The top-down approach here attempts to investigate technology convergence from a macro-perspective and to investigate notable industrial sectors and corresponding technology categories. This allows for a comparison of industry-specific technology categories from an all-encompassing perspective of industry. The bottom-up approach here attempts to investigate practical usage instances on a microscope, focusing on technology categories by industry. The integration of these two approaches provides an integrated and multidimensional understanding of technology convergence in terms of industry sectors, technology categories, and technology application levels.

We investigate this two-way approach based on patent documents. In this study, technology convergence is defined as when more than two technologies belonging to different sectors appear in one patent at the same time. If heterogeneous IPCs appear

in one patent, the technology corresponding to each IPC is considered to be converged. In terms of technology convergence, the top-down approach serves to identify the patterns by which two IPCs converge, deriving significant IPCs in technology convergence. On the other hand, the bottom-up approach derives significant keywords regarding the convergence pattern though patent textual data, not covered from an IPC. Specifically, detailed procedures and explanations of the two-way approach are as follows. Using the top-down approach, this study conducts a network analysis in order to identify the central position in the convergence network using IPC codes that describe the technology category as a generally accepted classification scheme. Meanwhile, the bottom-up approach utilizes a clustering analysis, which is commonly used to derive characteristics from numerous of textual data. This approach groups patent documents based on similarities among the patent documents within industrial sectors. Overall, the contributions of the two-way approaches are to identify notable industrial sectors and influential technology categories based on the central position in the AI convergence network from the top-down approach and to identify significant keywords of actual use of the technology within the industrial sectors via the bottom-up approach.

The complementary aspects of the two-way approach are as follows. First, the top-down approach targets structured data, i.e., IPC data, which restricts the discovery of insights other than information in the technology category. In contrast, the bottom-up approach targets unstructured data, i.e., text data, including detailed explanations and information from the patent documents. Second, for unsupervised learning, in which the results of the clustering analysis are not strictly defined, the interpretation of the results is very important. To understand the results of the clustering analysis, the top-down approach, i.e., the network analysis, provides directions pertinent to the technology category.

The novelty of this paper is as follows. First, in order to identify the characteristics of technology convergence, this study attempts to compare the results of the technology categories from the network analyses and the results of the keywords from the cluster analysis. This study presents a new research framework by which to understand the technology convergence by discovering the structure of technology convergence patterns and additionally by investigating practical application aspects in the convergence patterns. Second, the study attempts to analyze AI technology convergence throughout various industries with a holistic and integrated approach that considers significant industries, technology categories, and related applications. In particular, research on general-purpose technology such as AI is crucial from an industry perspective, and the industry-specific AI convergence characteristics identified in this study can have significant implications for all AI-related industries.

The research questions of this study are as follows. (1) What are the notable industrial sectors in technology convergence of AI? (2) To compare the technological characteristics of AI convergence by the industrial sectors, which differential aspects do the two-way approaches have? The remainder of this paper is structured as follows. Section 2 describes the research framework and proposed methodology of this study. Sections 3 briefly shows the dataset in this study. Section 4 shows the results of network centrality analysis, ego-network analysis, clustering analysis, and the two-way approaches. Section 5 presents discussions and conclusions, and also proposes future research directions.

Proposed methodology

Research framework

Figure 1 shows the research framework for this study. A top-down approach focuses on hub nodes and their tie nodes in an IPC-based convergence network. A network centrality analysis is applied to determine the hub nodes (hubs, in short) which identify 'notable industrial sectors' and 'influential technologies'. In addition, an ego-network analysis is done to understand the tie nodes (ties, in short) which are strongly 'related technologies' on hubs. The top-down approach, which means IPC-based network analyses, shows the results on the major technology convergence category with hubs and ties in each industrial sectors. Meanwhile, a bottom-up approach constructs clusters through the process of grouping the whole patents with similar topics. A text-based clustering analysis is performed, and the results show additional information not found in IPC-based network analyses. Generally, a patent is document that explains a new invention which are related to technology category, the problems to solve and how to solve them as well. Thus, text data in a patent document can be understand as information related to the application or method for implementing the technology, in addition to the structured technology categories identified from the IPC codes. Thus, keywords can be extracted from each cluster, and the aspects of technology applications can be examined by linking these aspects with the technology category derived from the network analysis. In this study, a technology application is defined as a practical way in which the technology can be used in the industry and technology categories.

Figure 2 shows the research procedure of this study, and following sections describe the background and provide a detailed explanation of each step. Because the purpose of the top-down and bottom-up approach differ, the range of the analysis dataset in each case is set differently. The top-down approach aims to extract significant industry and industry-specific technology categories using AI patents; thus, this step targets the entire AI patent. On the other hand, the bottom-up approach aims to identify specific and detailed features
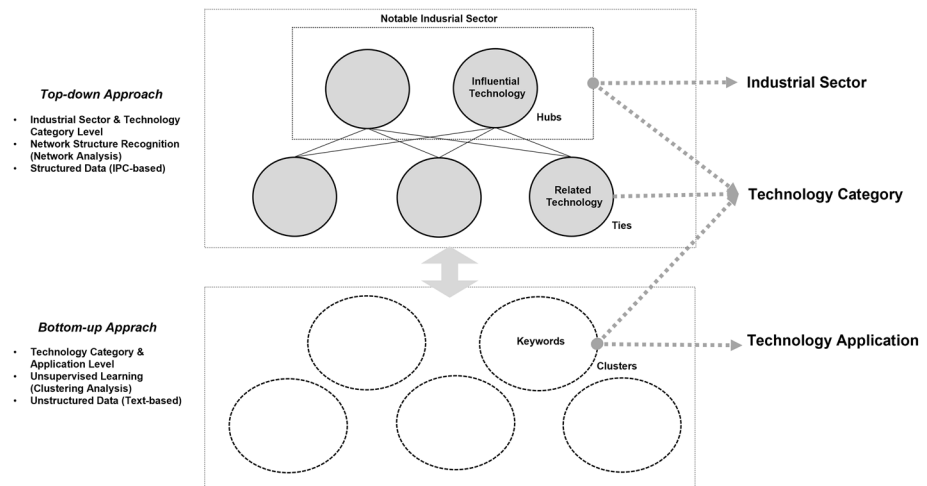

Fig. 1 Research framework
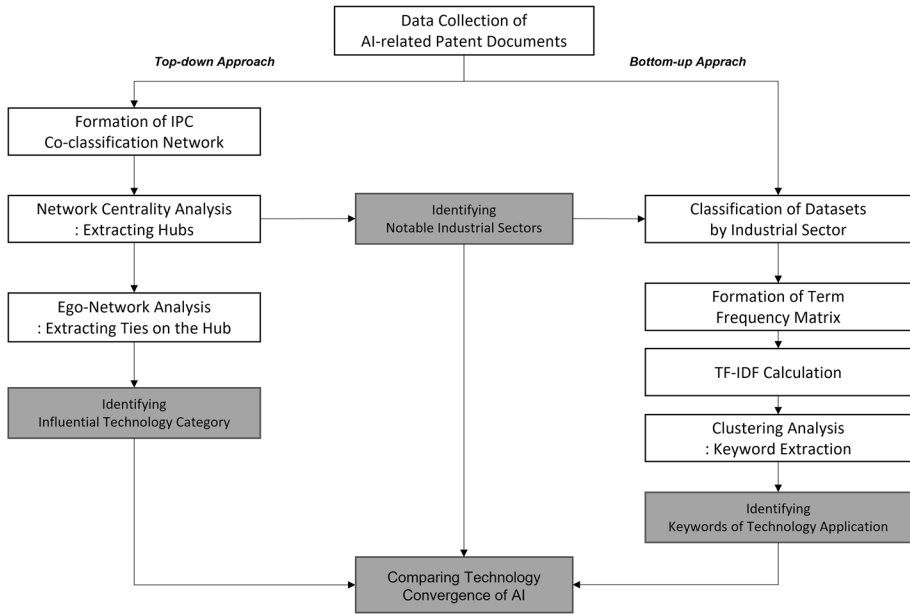
Fig. 2 Research procedure

of technology applications in relation to the technology category regarding notable industrial sectors derived from the top-down approach. Accordingly, it targets AI patents classified by industrial sector. Meanwhile, the study uses abstract data from each patent document for the analysis of the text data.

Data collection

There have been many trials to establish a new patent category in the area of AI. According to previous research, AI has been categorized into three broad areas: big data analytics, vision, and language (Tractica, 2016). In the same vein, to define the scope of our research, we categorized AI software technology into three groups: (1) learning and reasoning, (2) natural language processing, and (3) computer vision. In the previous research, IPC codes were used in selecting and classifying AI technology (Fujii & Managi, 2018; Tseng & Ting, 2013). Comprehensively, we collected the three groups of patents with related IPC codes. The IPC codes for each AI patent group in this study were according to the work of the Korean Intellectual Property Office (KIPO) in 2018 (KIPO, 2018). Table 1 shows the category of the AI technology and IPC codes used in this study. The descriptions of each IPC are referred to the WIPO (WIPO, 2019b).

This study collected AI-related patents which were registered at the United States Patent and Trademark Office (USPTO) from Google Patent Datasets. We constructed a set of standard SQL statements using the Google BigQuery platform for collecting AI-related patents with IPC codes during the period from 2000 to 2019 for the publication dates. In addition, from the Google Patents Search, we collected the forward citations of each patent. The total number of patents collected was 209,212, and one patent was incomplete;

Table 1 Category of AI technology

AI patent group	IPC code	Description
Learning and Reasoning	G06N-003/08	Computer systems based on biological models; Learning methods
	G06N-005/04	Computer systems utilizing knowledge based models; Inference methods or devices
	G06F-019/24	Digital computing or data processing equipment or methods, specially adapted for specific applications; For machine learning, data mining or biostatistics, e.g. pattern finding, knowledge discovery, rule extraction, correlation, clustering or classification
	G06K-009/62	Methods or arrangements for reading or recognising printed or written characters or for recognising patterns; Methods or arrangements for recognition using electronic means
	G06N-003/02	Using neural network models
Natural Language Processing	G06F-017/20	Handling natural language data (speech analysis or synthesis)
	G06F-017/21	Text processing
	G06F-017/27	Automatic analysis, e.g. parsing, orthography correction
	G06F-017/28	Processing or translating of natural language
	G10L-015/	Speech recognition
	G10L-017/	Speaker identification or verification
	G10L-013/	Speech synthesis; Text to speech systems
Computer Vision	G06F-021/32	User authentication; using biometric data, e.g. fingerprints, iris scans or voiceprints
	G06K-009/00	Methods or arrangements for reading or recognising printed or written characters or for recognising patterns, e.g. fingerprints
	G06T-007/50	Image analysis; Depth or shape recovery

Descriptions of each IPC are referred to the WIPO (2019b)

thus, 209,211 patents were selected for this study. Also, we extracted 2517 IPC codes at the group level from the patents.

Formation of IPC co-classification network

To measure technological convergence, previous research commonly used a co-classification network analysis and co-citation network analysis (Curran & Leker, 2011; Kwon et al., 2020). This study constructs a convergence network using co-classification for the following reasons. First, a co-citation is based on the relationships among the patent documents themselves, while co-classification is based on the relationships among the technology classification codes. Although a co-citation analysis is useful to understand knowledge flows (Lee et al., 2016), an IPC co-classification analysis can be a more direct indicator that explains specific technology areas (Choi et al., 2015). Therefore, co-classification is more adequate for the research purpose here, which is to identify technology categories in a convergence network. Second, co-classification is more consistent with the definition of technology convergence in this study. This study defined technology convergence as occurring when more than two technologies belonging to different sectors appear on one patent at the same time, and this definition corresponds to the co-classification concept.

Therefore, in this study, an IPC co-classification network was formed to analyze the convergence using patents. In the patent analysis, each IPC is considered as a node and the relationship between the IPCs as a link, and the weights mean the number of common patents for a pair of IPCs. We formed an undirected and weighted graph to analyze the network. The IPC co-classification network was created at the group-level (e.g. G06K-009) of each patent, a total of 2517 IPCs.

Extracting the hubs through network centrality analysis

Among the various network centrality indicators, degree centrality shows how connected the node is (Jackson, 2008). Degree centrality is an efficient indicator of measuring the power of each node (Borgatti et al., 2013) because a node with many links between other nodes has more advantages and influences on the network. Betweenness centrality shows the mediating role of the network among the nodes. If a node is located on the shortest path between a pair of nodes in the network, the node is considered to be on an advantageous position. Meanwhile, based on those two centralities, the network positions are categorized into four positions: the hub, bridge, core, and periphery (Baek et al., 2014). The hub position means highly connected with others and is important in connecting others, which has both a high degree and betweenness centrality. Therefore, in this study, the degree centrality and betweenness centrality were analyzed to investigate the hub nodes with advantages and influences on the network. From those hub nodes, this study identify the notable industrial sectors and influential technologies.

In terms of technology convergence, degree centrality can measure the direct influence in the technology convergence (Kim et al., 2014), and betweenness centrality is an indicator of the extent of the role of a node as a brokerage, and related to arbitration capabilities in technology convergence (Huang, 2017; Lee et al., 2012). Therefore, from the IPC co-classification network constructed in this study, the degree centrality finds IPCs which play a central role in terms of direct connectivity, whereas the betweenness centrality finds IPCs which play a central role in terms of intermediary connectivity.

In this study, the equation of node degree centrality can be defined as follows (Borgatti et al., 2002; Freeman, 1979).

$$C_D(N_i) = \sum_{j=1}^{g} x_{ij}\, i \neq j,$$

where g is the number of IPCs in the network, x_{ij} is the degree of strength of the relationship between IPC i and IPC j ($0 \leq x_{ij} \leq$ max).

In this study, the equation of node betweenness centrality can be defined as below (Borgatti et al., 2002; Freeman, 1979).

$$C_B(N_i) = \sum_{j<k} \frac{g_{jk}(N_i)}{g_{jk}}, i \neq j \neq k$$

where g_{jk} is the number of shortest paths between IPC j and IPC k, and $g_{jk}(N_i)$ is the number of paths including IPC i in the shortest paths between IPC j and IPC k.

The method for extracting the hubs through network centrality analysis is described as follows. In order to extract influential technology (i.e. hub) in the convergence network, the top 10 percent of the IPCs were considered. Appendix A. shows the distribution of the degree and betweenness centralities, respectively. The distribution shows the form of positive skewness with a long tail on the right. The top 10 percent was within rank 250 and explains 96.3% in the degree centrality (Sum of the top 10 percent centrality measure = 9,528,040 /Sum of the total centrality measure = 9,893,234), and describes 96.7% in the betweenness centrality of the total value (3,413,198 / 3,528,905). Thus, the top 10 percent of technology can represent the influential technology in this study.

Extracting ties on the hubs through ego-network analysis

For the selected hubs in Sect. "Extracting the hubs through network centrality analysis", this section investigated the linked technology. An ego-network consists of a connection between one central node called an ego and other nodes called alters connected to that node. The ego-network was analyzed for each hub, and the strength of the connection with the alter was measured by the tie value which was based on the total number of ties in the ego-network. The nodes with the top 10 tie values were selected to derive the strong-tie in this study. The strong-ties were analyzed to identify the characteristics of the linked technologies among sectors. The linked technologies of the hub in each sector, which are ties, were investigated in terms of common or different technology within the sector compared to other sectors. Meanwhile, technology included in its own sector was not considered. For example, when analyzing the hubs in the medical sector, ties in the medical sector were excluded from the analysis.

Classification of dataset by industrial sector

To classify the patents by industrial sector, we referred to the WIPO IPC and technology concordance table (Schmoch, 2008) which was classified technology into thirty-five fields according to IPC codes. (See Appendix B). In this study, a term 'sector' was commonly used to indicate each 'field' in the IPC-concordance matrix. Among thirty-five sectors, to determine the industrial convergence sector in AI, we excluded sectors that

are directly related with ICT and AI technology. Additionally, we excluded sectors that are specialized for a process or a machine itself and are difficult to identify in a specific industry. Also, in the furniture/game sector, only the game sector was examined in this study because those two are not considered as a same category in common, and a great amount of technology was included in game technology. Thus, the final sixteen sectors to be analyzed in this study were selected and the sectors were as follows: IT methods for management (referred to here as the finance/management sector representing included technology), semiconductors, analysis of biological materials, medical technology, biotechnology, transport, games, environmental technology, organic fine chemistry, pharmaceuticals, civil engineering, food chemistry, nanotechnology, basic materials chemistry, metallurgy, and polymers.

Keywords extraction through clustering analysis

For each industrial sector, there are number of patents and they are expected to have more than a single topic within a sector. To unveil topics consisting of a sector, clustering analysis using a Document-Term Matrix (DTM) could be applicable. From each cluster, keywords representing the core topics could be extracted.

Considering the characteristics of the dataset in this study, each cluster would be very close in an industrial sector. The patents within a sector do share similar topics. It is expected that there can be a degree of overlap among clusters. Also, a DTM of a sector could be a sparse matrix if there are number of clusters which share similar topics. Similar documents (i.e., patents) share a set of terms, and it is differentiated with others.

One of the widely used clustering methods is K-means clustering, which is quite simple and fast, but it has difficulty in handling inherent heterogeneity such that a certain data set is close to more than one cluster (Patel & Kushwaha, 2020). The resultant clusters of K-means clustering are disjoint because a data point is uniquely assigned to the cluster with the closest distance from the centroid which is the cluster center. Due to the disjoint nature, K-means clustering is not fit for clustering patents with similar topics.

Spectral clustering is one of the candidate solutions for patent clustering. It does not rely on distribution of the data. However, the DTM would be sparse and the affinity matrix for the spectral clustering also would be sparse. Thus, the spectral clustering is not suitable with our dataset because the spectral clustering requires a fully connected affinity network.

On the other hand, the Gaussian Mixture Models (GMM) cluster assigns a certain data set to the multivariate normal components maximizing the component posterior probability (Wang et al., 2019a). GMM finds complex patterns and then groups them into cohesive and homogeneous components which are a close representative of real patterns in the data set (Patel & Kushwaha, 2020). GMM is a density-based clustering algorithm which means that a resultant cluster has a high-density region surrounded by low-density regions. Patents with similar topics sometimes cannot be clustered with distinct boundaries, which can be clustered with a density-based model. Also, in GMM a data point can be expressed as a set of probabilities of cluster membership, which means the mixed membership.

GMM is an unsupervised clustering which finds out K Gaussian distributions from the given data, where K is the number of clusters. Thus, a probability density function of GMM $p(\mathbb{X})$ for a D-dimensional vector $\mathbb{X}$ is expressed as a superposition of K Gaussian probability densities (Bishop, 2006.)

$$p(\mathbb{X}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbb{X}|\mu_k, \Sigma_k)$$

where π_k is called the mixing coefficient, which indicates a selection probability of kth Gaussian distribution, $\mathcal{N}(\mathbb{X}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ is kth Gaussian density, $\boldsymbol{\mu}_k$ is a D-dimensional mean vector of kth Gaussian distribution, and $\boldsymbol{\Sigma}_k$ is a $D \times D$ covariance matrix of kth Gaussian distribution. The parameters of the distributions are iteratively updated and converged using the Expectation–Maximization (EM) algorithm. These process means that the iterative estimation of parameters π_k, $\boldsymbol{\mu}_k$, and $\boldsymbol{\Sigma}_k$ for given data $\mathbb{X}$.

The method for the patent clustering is described as follows. First, the abstracts of the patents were tokenized and lemmatized using the Python spaCy library to count terms in different forms in sum. Then a bigram DTM was constructed from the lemmatized text of the patent abstracts after removing stop words including patent-specific common terms such as 'method' and 'apparatus'. To weight the relatively important terms, a bigram Term Frequency-Inverse Document Frequency (TF-IDF) DTM was calculated from the DTM. Based on the TF-IDF DTM, the dimensionality of terms was reduced while preserving the hidden meaning of terms and reducing sparsity of the input DTM by applying Latent Semantic Analysis (LSA) with the target explained variance of 90% (See Appendix C). The patents were then clustered using GMM with the Python scikit-learn library by increasing the number of clusters from 1 to 10. In addition, to prevent EM from converging local maxima, each setting was executed 5 times with different initial random seeds.

The best-fitting models were chosen based on the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values. For a GMM clustering, AIC and BIC were measured to determine the appropriate number of clusters (Burnham & Anderson, 2002). Usually, the lower information criterion indicates the better clustering. Thus, the best-fitting model among different number of clusters could be chosen with the lowest AIC or BIC value. If the lowest AIC model and lowest BIC model were not the same, the resultant clusters of both models needed to be reviewed. In our dataset, however, the lowest AIC model was chosen because the lowest BIC was always the single cluster case.

Then, each resultant cluster had distinct top keywords according to the TF-IDF values. By taking the mean TF-IDF value of each bigram term in a cluster, the top keywords were populated from bigram terms of top mean TF-IDF values.

Dataset

To understand the dataset in this study thoroughly, we analyzed the patent data with the innovation performance indicators, which can compare the level of convergence by industrial sectors classified in the Sect. Classification of Dataset by Industrial Sector. For measuring innovation performance, the number of patents or forward citations were used as indicators (Hagedoorn & Cloodt, 2003; Harhoff et al., 1999; Trajtenberg, 1990; Wartburg et al., 2005). The number of patents means a quantitative aspect in terms of the technological invention of new technology, process, and product. Whereas forward citations analysis implies a qualitative aspect (Hagedoorn & Cloodt, 2003).

The datasets classified into the industrial sectors were analyzed using the innovation performance indicators which were number of patents, CAGR of patent counts, and forward citations per patent (CPP) shown in Table 2. In addition, Fig. 3 shows the trend analysis according to the industrial sector based on the number of patents by the year.

Table 2 Datasets by the industrial sector in AI patents

Sector	Number of patents	CAGR (Compound annual growth rate (%))	CPP (Citations per patent)
Finance/management	12,603	23	66.87
Medical technology	10,218	17	29.79
Transport	6426	41	29.01
Semiconductors	1896	9	22.20
Biological materials	1768	13	78.76
Games	1576	27	45.47
Biotechnology	1188	12	29.59
Civil engineering	621	23	29.91
Environmental technology	332	9	28.15
Organic fine chemistry	331	8	34.77
Basic materials chemistry	185	4	46.67
Pharmaceuticals	185	16	28.57
Food chemistry	74	20	9.80
Micro-structure and nano-technology	49	15	80.98
Materials, metallurgy	42	16	9.64
Macromolecular chemistry, polymers	35	16	41.14

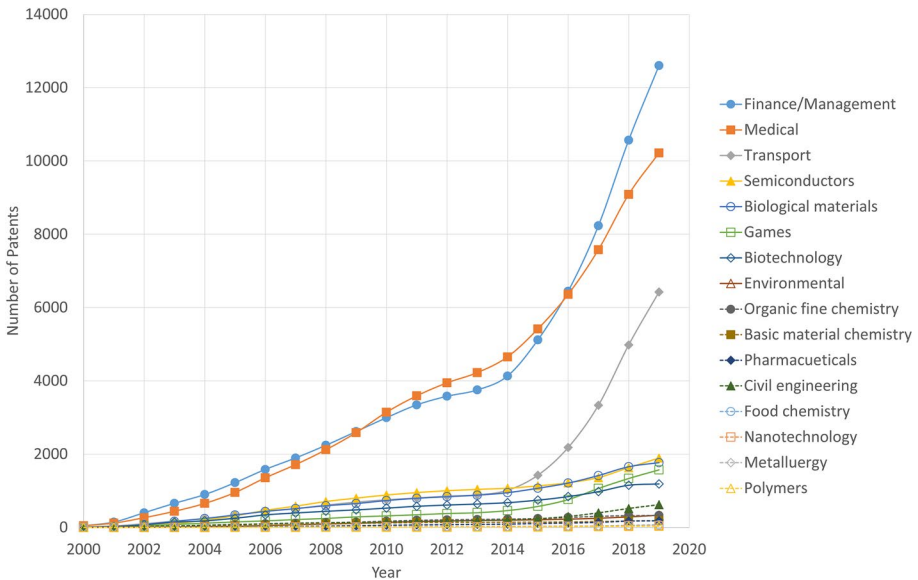**Fig. 3** Trends of datasets by the industrial sector in AI patents

In terms of the total number of patents, the finance/management, medical, and transport sectors had a large patent count, over 6000, and the semiconductor, games, and biological materials had a patent count of 1000 to 2000. The environmental technology, organic chemistry, pharmaceuticals, civil engineering, food chemistry, nanotechnology,

metallurgy, and polymers had a patent count below 1000. From a view of growth rate, the transport sector showed a noticeably high rate as 41% of CAGR. In addition, the game, finance/management, and civil engineering sector showed the high growth rates. In terms of forward CPP by sector, the micro-structure/nanotechnology, biological material, and finance/management sectors have shown significant numbers of citations over other sectors. Meanwhile, the following Sect. Results of network centrality analysis identifies the notable industrial sectors in this study among the sixteen sectors.

Analysis and results

Results of network centrality analysis

This study defined hub nodes as the degree and betweenness centralities within the top 10 percent. Obviously, a majority of hubs corresponded to AI-related or computer technology, such as pattern recognition, image analysis, and data processing. The IPC, G06K-009, which is related to pattern recognition, ranked first both degree and betweenness centralities (see Table 3) and positioned in the most central in the whole AI convergence network.

Meanwhile, among the sixteen sectors in our dataset in the Sect. Dataset, the technologies in finance/management, medical, transport, semiconductor, game, biotechnology, and analysis of biological materials were included in the hub nodes in the AI convergence network. Table 4 shows the results of the hubs in each sector.

In the finance/management sectors, the AI was used for commerce (G06Q-030), payment architectures (G06Q-020), business management systems (G06Q-010, 050), insurance and tax (G06Q-040). In the medical sector, the technologies, involved in diagnosis (A61B-005, 006, 008), examination or testing instruments (A61B-003), and surgical instruments (A61B-017), ranked high. For the transport sector, technologies related to protecting against accidents (B60R-021), optical devices (B60R-001, B60Q-001), vehicle control (B60W-030, 050, 040, 010), monitoring (B60R-011), seat management (B60N-002) and preventing theft (B60R-025) ranked high. In the semiconductor sector, technologies involved in semiconductor devices (H01L-027), manufacturing process (H01L-021), and details of semiconductor device (H01L-023) ranked high. In the game sector, the technologies involved in video games (A63F-013), sports appliances (A63B-071, 069), and controls for exercising (A63B-024) ranked high. For the analysis result of the biological materials sector, investigating and analyzing materials (G01N-033) ranked high, especially in the betweenness centrality, which means this technology has a tendency to mediate other technologies. In addition, measuring or testing processes (C12Q-001) ranked high in biotechnology sector. The biotechnology (C12Q-001) and biological materials (G01N-033), which were divided in the Sect. Dataset, were considered as one sector in this study since biological materials can be included in biotechnology sector.

Consequently, the six sectors were defined as notable industrial sectors in this study. In addition, hub nodes in each sector were determined as influential technology. Figure 4 shows the visualization of the six notable industrial sectors and influential technologies according to Table 3. Especially, for the finance/management sectors, most of the hubs (G06Q-010, 020, 030, 050) were positioned relatively high in both centralities. The medical sector has the highest technology (A61B-005) in both centralities among the hubs.

Table 3 Top 10 rank of the AI technology in AI convergence network

	Rank	IPC Code	Value	Description
Degree Centrality	1	G06K-009	614,610	Methods or arrangements for reading or recognising printed or written characters or for recognising patterns
	2	H04N-005	481,160	Details of television systems
	3	G06F-003	391,088	Input or output arrangements for transferring data
	4	G06F-017	278,642	Digital computing or data processing equipment or methods, specially adapted for specific functions
	5	H04N-007	255,293	Television systems
	6	H04L-012	247,720	Data switching networks
	7	H04N-001	209,014	Scanning, transmission or reproduction of documents
	8	G06T-007	186,469	Image analysis
	9	H04L-009	178,216	Arrangements for secret or secure communication
	10	H04N-021	173,159	Selective content distribution, e.g. interactive television or video on demand
Betweenness centrality	1	G06K-009	1,560,454	Same as above
	2	G06T-007	183,382.2	Same as above
	3	G06F-017	143,697.6	Same as above
	4	G06F-003	103,567.5	Same as above
	5	G10L-015	97,529.1	Speech recognition
	6	G06N-003	84,961.23	Computer systems based on biological models
	7	H04N-005	82,497.74	Same as above
	8	H04N-007	79,561.6	Same as above
	9	G06F-019	62,109.79	Computational chemistry, Chemo informatics, Computational materials science
	10	G06N-005	56,337.88	Computer systems using knowledge-based models

Descriptions of each IPC are referred to the WIPO (2019b)

Table 4 Network centrality analysis on hubs by the industrial sector in AI convergence network

Sector		Hub	Degree Centrality		Betweenness Centrality	
			Rank	Value	Rank	Value
Finance/Mgmt		G06Q-030	40	73,798	22	17,119.15
		G06Q-020	41	73,044	48	6751.979
		G06Q-010	53	49,148	21	18,096.42
		G06Q-050	74	28,710	30	12,771.32
		G06Q-040	163	8673	163	1143.456
Medical		A61B-005	33	81,404	11	39,581.35
		A61B-006	84	23,489	67	4394.374
		A61B-008	142	13,333	202	841.873
		A61B-003	161	9160	153	1251.287
		A61B-017	236	2868	194	886.636
Transport		B60R-021	98	21,874	182	981.005
		B60R-001	131	14,891	61	4837.574
		B60N-002	153	11,183	227	714.849
		B60W-030	157	10,228	123	1637.687
		B60Q-001	159	9605	82	2977.064
		B60W-050	165	8641	115	1878.852
		B60W-040	166	8428	145	1334.781
		B60R-016	168	8126	184	971.496
		B60R-025	172	7670	133	1466.471
		B60R-011	177	6805	76	3538.538
		B60W-010	213	3678	244	654.781
		B64C-039	216	3638	84	2926.836
		B64D-047	227	3032	121	1642.973
Semiconductor		H01L-027	72	30,864	80	3097.656
		H01L-021	178	6770	60	4914.426
		H01L-023	226	3077	141	1361.027
Game		A63F-013	119	18,299	101	2209.875
		A63B-071	215	3646	171	1081.647
		A63B-069	225	3114	186	948.48
		A63B-024	249	2389	116	1858.182
Biotechnology	Biological Materials	G01N-033	95	22,263	16	23,755.15
	Biotechnology	C12Q-001	186	5899	54	5604.225

Results of ego-network analysis

This section analyzed the ego-network formed for the hubs to identify the linked technologies. From the ego-network analysis, the technology classification of strong-ties which were ranked in the top 10 were analyzed in this study. The strong-ties are shown in Table 5 according to each hub. (See Appendix D for each tie values). Obviously, a majority of strong-ties corresponded to common AI technology, such as pattern recognition, image analysis, data processing, and natural language processing. Thus, to determine
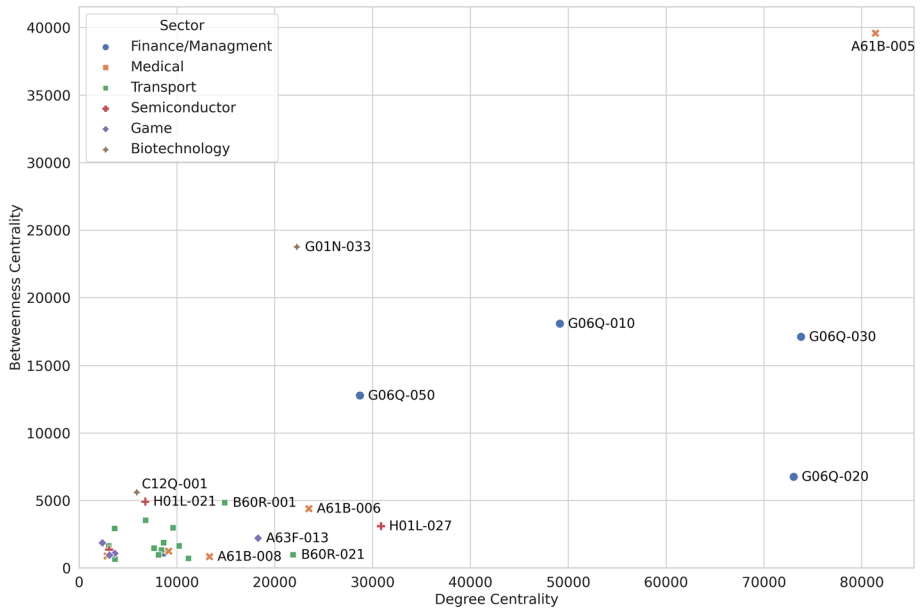
Fig. 4 Hubs according to the industrial sector in the AI convergence network

differential strong-ties within the industrial sectors compared to other industrial sectors, we investigated strong-ties in terms of common or sector-specific technology. The footnotes in Table 5 shows a category of IPCs considering common or sector-specific technology among industrial sectors. Consequently, Table 6 presents a summary of the common ties among the industrial sectors and sector-specific ties by industrial sector.

The result showed that pattern recognition (G06K-009), image analysis (G06T-007), data processing (G06F-017), and speech recognition (G01L-015) were commonly used to almost all industrial sectors, and these technologies correspond to common AI-related technology. Especially, image analysis (G06T-007) was highly ranked in medical, transport, semiconductor, game, and biotechnology sectors. Also, data processing (G06F-017) was highly ranked in finance, medical, game, and biotechnology sectors. The speech recognition (G01L-015) was highly ranked in finance, transport, and game sectors.

On the other hand, in addition to these common AI-related technologies, different ties can be identified for each sector to understand the unique characteristics among the industrial sectors. Table 5 shows the summary ties in each sector. For the finance/management sector, against unauthorized activity (G06F-021) was widely used in payment architecture, commerce, business management systems and insurance (G06Q-020, 030, 040, 050). Moreover, selective content distribution (H04N-021) was highly connected with all. Additionally, speech recognition (G10L-015) technology was appeared in commerce (G06Q-030) and administrative area (G06Q-010). For the medical sector, image analysis (G06T-007) appeared to be commonly used in the sector. Additionally, a variety of image-related technologies such as image generation and enhancement (G06T-005, 001, 017) emerged in the others. In addition, diagnosis-related technology (A61B-005, 006, 007) showed high connectivity with the computational material science (G06F-019), which includes machine learning, data mining, and biostatistics. For the transport sector, image analysis (G06T-007) was shown to be mainly linked to various technologies. Especially, the first representative finding was control

Table 5 Ties according to each Hub in AI convergence network

Sector	Hubs	Ties
Finance/Mgmt	G06Q-030	G06F-017[c], G06F-021[e], G06K-009[a], G10L-015[d], G06F-003, H04N-001, H04N-021[f]
	G06Q-020	G06F-017[c], G06F-021[e], G06K-009[a], G06F-003, G06K-019, H04N-001, H04N-021[f], B41J-024, B41J-002
	G06Q-010	G06F-017[c], G06K-009[a], G10L-015[d], G06F-003, G06K-007, H04L-029, H04N-021[f], H04W-004
	G06Q-050	G06F-017[c], G06F-019[g], G06F-021[e], G06K-009[a], A61B-005[m], G06F-003, H04L-029, G06F-003, H04L-029, H04N-021[f]
Medical	G06Q-040	G06F-017[c], G06F-021[e], G06K-009[a], G06F-003, H04N-005, H04N-021[f]
	A61B-005	G06F-017[c], G06F-019[g], G06K-009[a], G06F-003, G01J-005, H04N-005, G06T-007[b]
	A61B-006	G06F-019[g], G06K-009[a], G06T-007[b], G01R-033, G06T-001, G06T-005[h], G06T-011[h], G06T-017[h]
	A61B-008	G06F-017[c], G06F-019[g], G06K-009[a], G06T-007[b], G01R-033, G06T-011[h], G06T-017[h]
	A61B-003	G06K-009[a], G06T-007[b], G02B-027, G06F-003, G06T-005[h], H04N-005
	A61B-017	G06K-009[a], G06T-007[b]
Transport	B60R-021	G06K-009[a], G01F-023, G01S-007[j], G01S-013[j], G01S-015[j], G01S-017[j]
	B60R-001	G06K-009[a], G06T-007[b], G01S-015[j], G08G-001[i], H04N-005, H04N-007
	B60N-002	G06K-009[a], G01F-023, G01S-007[j], G01S-013[j], G01S-015[j], G01S-017[j]
	B60W-030	G06K-009[a], G06T-007[b], G01C-021[j], G05D-001[i], G08G-001[i]
	B60Q-001	G06K-009[a], F21S-041, G01C-021[j], G06F-003, G08G-001[i], H04N-007, H04N-021[f], H04W-004
	B60W-050	G06K-009[a], G01C-021[j], G05D-001[i], G08G-001[i], G06F-003, H04N-021[f], H04W-004
	B60W-040	G06F-021[e], G06K-009[a], G01C-021[j], G05D-001[i], G06F-003, G08G-001[i],H04N-021[f], H04W-004
	B60R-016	G06K-009[a], G10L-015[d], G01C-021[j], G01S-015[j], G06F-003, H04N-021[f]
	B60R-025	G06F-021[e], G06K-009[a], G01C-021[j], G06Q-030, G05D-001[i], G06F-003, G07C-005, G08G-001[i], H04N-021[f], H04W-004
	B60R-011	G06K-009[a], G06T-007[b], G01C-021[j], G08G-001[i], H04N-005, H04N-007
	B60W-010	G06K-009[a], G06T-007[b], G05D-001[i], G08G-001[i]
	B64C-039	G06F-017[c], G06K-009[a], G06T-007[b], A01M-001, G05D-001[i], G06Q-010, G08G-005[i], H04N-005, H04N-007
	B64D-047	G06F-017[c], G06K-009[a], G06T-007[b], G05D-001[i], G06Q-010, G06Q-050, G08G-005[i], H04N-005, H04N-007

Table 5 (continued)

Sector	Hubs	Ties
Semiconductor	H01L-027	G06K-009[a], H03M-013, H04L-009[l], H04L-012[l], H04M-001, H04N-005, H04N-007, H04W-008[l], H04W-028[l], H04W-088[l]
	H01L-021	G06K-009[a], G06T-007[b], G01B-011, G01N-021[k], G01R-031, G03F-001, G03F-007, G06T-001
	H01L-023	G06K-009[a], A61B-005[m], G06T-001, G07F-007, G11B-020, H04N-001, H05K-001
Game	A63F-013	G06F-017[c], G06K-009[a], G06T-007[b], G10L-015[d], G06F-003, G06Q-020, G06Q-030, H04N-005, H04N-007, H04N-021
	A63B-071	G06K-009[a], A61B-005 [m], A63B-021, G06F-003, G09B-019[n], H04B-001, G06F-019[g]
	A63B-069	G06K-009[a], A61B-005 [m], B33Y-010, G06F-001, G06F-003, G09B-019[n], H04W-084
	A63B-024	G06K-009[a], G06T-007[b], A61B-005[m], G06F-003, G09B-019[n], H04B-001, H04N-005
Biotechnology	G01N-033	G06F-017[c], G06F-019[g], G06K-009[a], G06T-007[b], G01N-015[k], G01N-021 [k], G06F-007, G06K-007, G06Q-030
	C12Q-001	G06F-017[c], G06F-019[g], G06K-009[a], G06T-007[b], G01J-003, G01N-015[k], G01N-021[k]

[a] Pattern recognition

[b] Image analysis

[c] Data processing methods; handling natural language data

[d] Speech recognition

[e] Against unauthorized activity

[f] Selective content distribution

[g] Computational material science

[h] Image generation; enhancement

[i] Control

[j] Navigation; direction finding

[k] Material analysis

[l] Electric Communication

[m] Diagnostic; identification

[n] Demonstration appliances

Table 6 Hub and ties by the industrial sector in AI convergence network

Sector	Hubs	Strong ties	
		Sector-specific	Common
Finance/Mgmt	Commerce (G06Q-030)	Against unauthorized activity (G06F-021)	Pattern recognition (G06K-009)
	Payment architectures (G06Q-020)	Selective content distribution (H04N-021)	Image analysis (G06T-007)
	Business management systems (G06Q-010, 050)		Data processing methods; handling natural language data (G06F-017)
	Insurance and tax (G06Q-040)		Speech recognition (G01L-015)
Medical	Diagnosis (A61B-005, 006, 008),	Computational material science (G06F-019)	
	Examination or testing instruments (A63B-003),	Image generation; enhancement (G06T-005, 011, 017)	
	Surgical instruments (A06B-017)		
Transport	Protecting against accidents (B60R-021),	Control (G08G-001, G05D-001, G08G-005)	
	Optical devices (B60R-001, B60Q-001)	Navigation; direction finding (G01C-021, G01S)	
	Vehicle control (B60W-030, 050, 040, 010)		
	Monitoring (B60R-011)		
	Seat management (B06N-002) Preventing theft (B60R-025)		
Semiconductor	Semiconductor devices (H01L-027)	Material analysis[a] (G01N-021)	
	Manufacturing process (H01L-021)	Electric communication[b] (H04L-009, 012, H04W-008, 028, 088)	
	Details of semiconductor device (H01L-023)		
Game	Video games (A63F-013)	Measuring diagnostic purpose; identification of person[c] (A61B-005)	
	Sports appliances (A63B-071, 069)	Demonstration appliances[c] (G09B-019)	
	Controls for exercising (A63B-024)		
Biotechnology	Analyzing materials (G01N-033)	Computational material science (G06F-019)	
	Measuring or testing process (C12Q-001)	Material analysis (G01N-021, 015)	

Descriptions of each IPC are referred to the WIPO (2019b)

[a] Especially Highly Ranked in Manufacturing process (H01L-021)

[b] Especially Highly Ranked in Semiconductor devices (H01L-027)

[c] Especially Highly Ranked in Sports/exercises (A63B-071, 069, 024)

technology. The technology seen in transport-related nodes was identified as a traffic control system (G08G-001), which also included general control (G05D-001) and aircraft control (G08G-005). The second finding was directional guidance and detection technology. Navigation (G01C-021) and radio direction-finding (G01S) showed important connections. For the semiconductors sector, manufacturing processes (H01L-021) showed many connections with image analysis (G06T-007). Especially, semiconductor devices (H01L-027) had many connections with the sub-technology of the electric communication technique (H04). In addition, the optical (G01B-011) or photomechanical (G03F-001, 007), and also material analysis (G01N-021) were appeared in manufacturing processes. The game sector was divided into two categories: video games (A63F-013) and sports/exercises (A63B-071, 069, 024). In video games, not only image analysis (G06T-007) and but also speech recognition (G10L-015) showed high connections at the same time. On the other hand, sports/exercise had a high connection with diagnosis or identification (A61B-005). In addition, educational demonstration appliances (G09B-019) were appeared in the sports/exercise area. For the biotechnology sector, the computational material science (G06F-019) was also highly linked, and the analysis of materials (G01N-021, 015) appeared.

Results of clustering analysis

The top 30 keywords were extracted for each cluster, and the results of keywords and TF-IDF values are shown in Appendix E to J. This study selected representative keywords among the top 30, considering the discriminative meaning of each cluster, excluding redundant or common keywords among clusters. Table 7 shows the results of clustering analysis.

Two-way approaches on network analyses and clustering analysis

Based on the research framework in the Sect. ``Research framework", Fig. 5 summarizes the results of this study according to the hubs/ties from network analyses and keywords from clustering analysis according to the Sect. ``Results of network centrality analysis", ``Results of ego network analysis", and ``Results of clustering analysis". As shown Fig. 5, the hubs and ties show the influential technology and strongly related technology in each industrial sectors. In addition, the clusters show keywords regarding the technology application. An analysis of these two aspects can be complementary relative to both and can contribute to a better understanding of the characteristics of the convergence of AI and provide a deeper understanding of the industrial sectors.

The complementary aspects of two-way approach are follows. First, the top-down approach shows only information in the technology category, while the bottom-up approach shows information about practical usage examples via the extracted keywords, which are related to the applied target of the technology. This allows us to understand information about the data utilized in the implementation of the technology. For instance, the financial sector shown in Fig. 5 presents hub technology about the 'payment architecture' (referring to IPC G06Q-020) from the top-down, whereas the keywords of clusters related to 'payment' (cluster #3 in Table 7) provide information about significant applied areas, such as 'user biometric information'. Furthermore, regarding the same technology category among the industrial sectors extracted from the top-down, the results from the bottom-up show different aspects to provide a clearer understanding of the applied areas. For instance, hub technology pertaining to 'image analysis' (G06T-007) is identical among the industrial sectors in the top-down, whereas the bottom-up allows us to identify differences in target data types for each sector used for the

image analysis. The image analysis is mainly utilized for the recognition of electronic documents in the finance sector (cluster #2), projection images (i.e. CT, tomography, X-ray) in the medical sector (cluster #4), patterns of wafers in the semiconductor sector (cluster #1), for object detection in the transport sector (cluster #4), and for biological images (i.e. tissues, cells, specimen) in the biotechnology sector (cluster #1). Second, bottom-up approach gives information about an integrated form of various technology category extracted top-down approach. For instance, the transport sector shown in Fig. 5 shows the each influential technology pertaining to 'control' (G08G-001, G05D-001, G08B-005), 'navigation & direction finding' (G01C-021, G01S), and 'image analysis' (G06T-007) from the top-down. In contrast, the keywords of clusters show an integrated form and application of those technologies in transport, such as 'unmanned aerial vehicle' (cluster #0). For the game sector, the top-down shows individually each influential technology related to 'video games'(A63F-013), 'controls for exercising' (A63B-024), 'image analysis' (G06T-007), and etc., whereas the keywords of clusters show an integrated form, such as 'augment/virtual reality'(cluster #2). Third, the top-down complements the results of the bottom-up when defining the cluster structures and meanings. The bottom-up is a sort of unsupervised learning with unlabeled data, and meanings should be found by means of interpretations of randomly derived results. The results derived from the top-down approach can then serve as common criteria and important indicators for interpreting the outcomes from the bottom-up approach.

Discussion and conclusion

This is an empirical study to understand technology convergence focusing on the case of AI. This study attempts to various analytical methods including network centrality analysis, ego-network analysis, and clustering analysis. As a result, this study identifies the notable six industrial sectors in technology convergence of AI implying the convergence level among industrial sectors in order to answer the first research question of this study. Additionally, to answer the second question in this study, we identify influential technology sectors through a network analysis, whereas keywords related to technology application aspects are found through a cluster analysis according to notable industrial sectors.

The theoretical contributions of this study on technology convergence research are as follows. This study suggests a new two-way approach, consist of the top-down and bottom-up approaches, to understand the characteristics of technology convergence. The framework of this study suggests integrated perspective on technology convergence based on industry, technology category, and related to applications. Consequently, it is possible not only to compare the characteristics of technology convergence by industrial sector, but also to define factors that can be revealed by each of the two-way approaches.

The methodological contributions in this study are as follows. The first is the combination of the network analysis using structured data in patent documents and the cluster analysis using unstructured data in patents. The network analysis provides a direct indicator by which to understand the patterns of technology convergence, while the clustering analysis provides implications related to practical applications considering each defined convergence pattern. Second, this study gives insight to technology convergence research by applying the GMM clustering algorithm. As the technology convergence dataset is a mixture of different categories of technologies, GMM explains well to classify clusters with various means and variances. This is meaningful study since there have been few previous studies related to the GMM clustering regarding technology convergence research.

Table 7 Results of clustering analysis by industrial sectors in AI patents

Sector	Cluster	Number of patents	Representative keywords
Finance/Mgmt	0	767	Content, user, multimedia, advertisement
	1	6583	Use, message, text, language
	2	1053	Electronic document, image
	3	2493	User, biometric, transaction, authentication, payment
	4	1707	Image, capture, processing
Medical	0	564	Medical image processing
	1	2681	User, signal, eye, sensor, control, determine
	2	1418	Image, object, capture, acquire, processing
	3	753	Fingerprint, sensor, identification
	4	414	Projection, ray, image, ct, tomography
	5	4388	Image, tissue, feature
Transport	0	206	Unmanned aerial vehicle
	1	2094	Vehicle, image, sensor, signal, detect, camera, video, control
	2	1331	Vehicle, user, control, parking, voice, autonomous vehicle
	3	476	Drive, assistance, behavior, gaze, eye
	4	840	Object detection
	5	1174	Image capture processing
	6	301	Detect, lane, road, boundary, change
Semiconductor	0	553	Circuit, element, layer, substrate, signal, control, detection
	1	484	Pattern, image, wafer, mask, inspection
	2	270	Light, emit, fingerprint, optical, substrate
	3	184	Defect, inspection, classification
	4	405	Fingerprint sensor
Game	0	301	Voice, audio, speech, control, command, message, communication
	1	587	Motion, exercise, activity, movement, athletic, body, sports
	2	56	Augment/virtual reality
	3	53	User identification access
	4	228	Image, card, face, detect, object, capture
	5	351	Video game, image, gesture, golf, ball, control
Biotechnology	0	854	Determine, measure, detect, feature, genetic
	1	468	Image analysis, tissue, cell, specimen
	2	116	Data identification, real time, pcr
	3	146	Sequence, nucleotide, oligonucleotide, cluster
	4	428	Biological sample, image analyze
	5	657	Gene expression, biomarker, cancer, disease, treatment, diagnosis
	6	287	Cell, blood, image analysis, determine

The contributions and results of this study have practical implications for AI technology and related industries. From a strategic business perspective, AI must be considered as an opportunity to create a new paradigm from a disruptive innovation. For industrial AI companies, it is necessary to consider the current state of technology development in the main stream of industry-specific AI convergence, as posited by the analysis results here, and discover new business creation opportunities in new areas related to AI technology
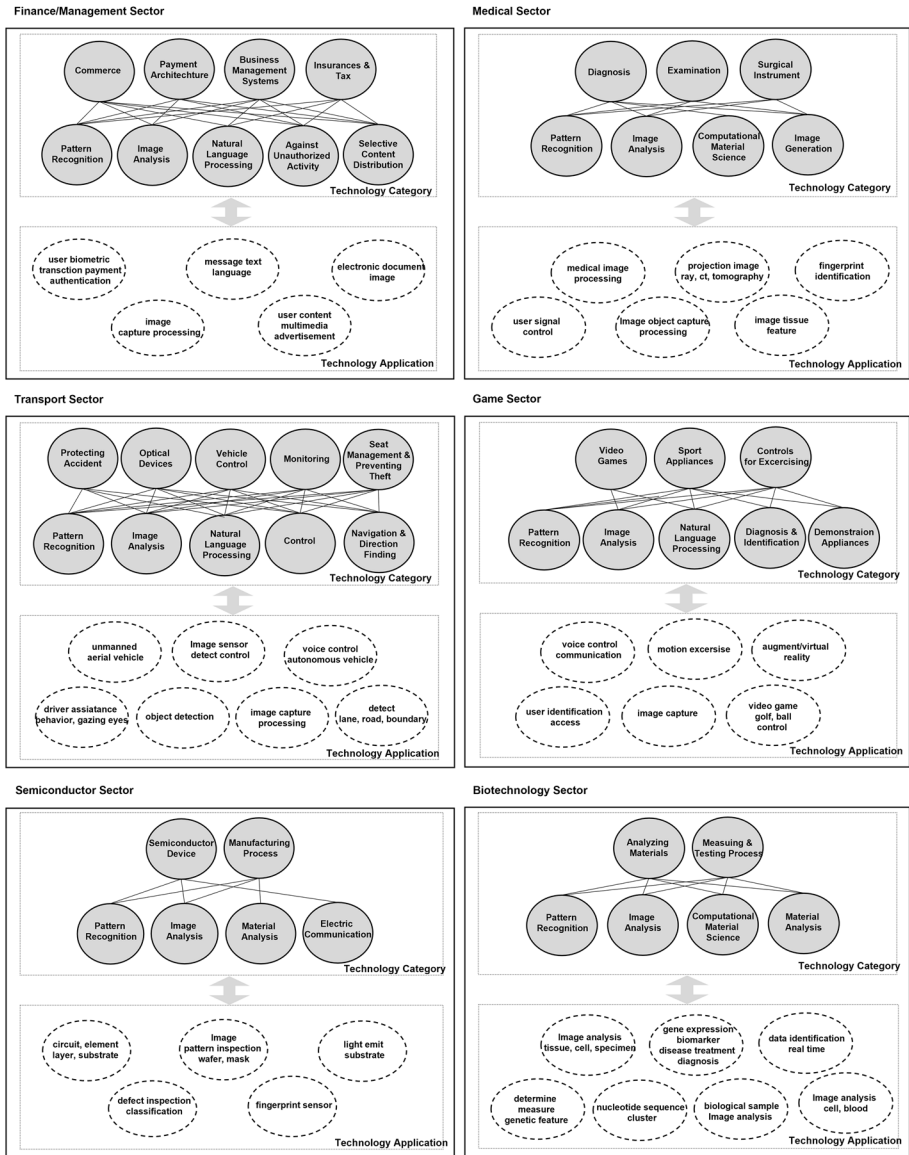
Fig. 5 Summary of hubs, ties and clusters by the industrial sectors in AI patents

that have not been developed thus far. For AI companies, it is necessary to develop extensible general technologies that can be applied to various industries by considering the main stream of AI convergence among industries. Commonly required AI technologies for each industry as found in this study can present insights related to this scalable strategy.

From a policy perspective, it is necessary to establish R&D policies pertaining to AI by industry with respect to sustainable social development beyond pursing short-term technological innovations and growth. The results of this study revealed the status of each sector

regarding AI convergence. While the technological trajectory of the cumulative AI patents has followed the S-curve over the past two decades and is in the growth stage of the technology life cycle (TLC), it appears that there is a difference in the level of growth stage in terms of the industrial sector. In particular, it was confirmed in the results of this study that industries or technologies directly related to the sustainable development do not stand out significantly in AI convergence. For instance, despite the fact that environment-related technology is essential for sustainable growth, the results of this study show that the degree of convergence of AI and environmental technology is still very low. In addition, previous study found that AI has had a significant impact on reducing energy consumption and energy intensity (Liu et al., 2021). However, it appears that the convergence of these sustainable technologies and AI has yet to make notable progress given the current status from the perspective of industry overall. Therefore, government R&D policies should be supported to promote the development of sustainable technologies on AI. Previous research found that enforced environmental policies by governments and subsidies developed to promote the invention of green technologies significantly increased green patent publication counts in China (Fujii and Managi, 2018). Comprehensively, governments in each country should understand the different development statuses with regard to the convergence of AI and various industries and should review R&D strategies considering policies or subsidies to encourage specific sectors in order to realize sustainable growth.

The limitations of this study and future research suggestions are as follows. First, with respect to the additional methodological approaches, it is recommended to compare the result of GMM with other soft clustering algorithm, such as fuzzy modeling and topic modeling. Second, this study considered AI patents based on three categories of technology: learning and reasoning, natural language processing, and computer vision. Thus, future studies should investigate patent data from more various categories and aspects to understand wide range of AI industry and technology. Third, future studies investigate each industrial sector respectively with a deep-depth understanding, or consider other industries not covered in this study. Since this study focuses on the technology and industrial sector with high centrality measure and strong-tie value, the research on the technology and industrial sector which have low centrality and tie values, but are still considered importance is necessary. Also, it is meaningful to examine the background and characteristics on the area where AI convergence occurred low.

Appendices

Appendix A. Distribution of the degree and betweenness centralities
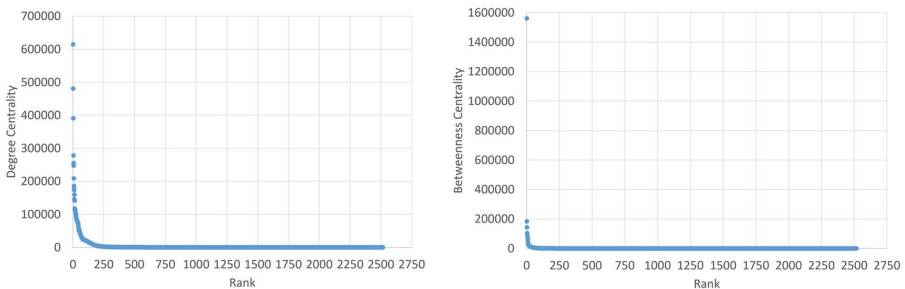
Appendix B. WIPO IPC-technology concordance table (Schmock 2008)

Area	Field		IPC code
Electrical engineering	1	Electrical Machinery, apparatus, energy	F21#, H01B, H01C, H01F, H01G, H01H, H01J, H01K, H01M, H01R, H01T, H02#, H05B, H05C, H05F, H99Z
	2	Audio-visual technology	G09F, G09G, G11B, H04N-003, H04N-005, H04N-009, H04N-013, H04N-015, H04N-017, H04R, H04S, H05K
	3	Telecommunication	G08C, H01P, H01Q, H04B, H04H, H04J, H04K, H04M, H04N-001, H04N-007, H04N-011, H04Q
	4	Digital Communication	H04L
	5	Basic communication processes	H03#
	6	Computer technology	(G06# not G06Q), G11C, G10L
	7	IT methods for management	G06Q
	8	Semiconductors	H01L
Measurement	9	Optics	G02#, G03B, G03C, G03D, G03F, G03G, G03H, H01S
	10	Measurement	G01B, G01C, G01D, G01F, G01G, G01H, G01J, G01K, G01L, G01M, (G01N not G01N-033), G01P, G01R, G01S; G01V, G01W, G04#, G12B, G99Z
	11	Analysis of biological materials	G01N-033
	12	Control	G05B, G05D, G05F, G07#, G08B, G08G, G09B, G09C, G09D
	13	Medical technology	A61B, A61C, A61D, A61F, A61G, A61H, A61J, A61L, A61M, A61N, H05G
Chemistry	14	Organic fine chemistry	(C07B, C07C, C07D, C07F, C07H, C07J, C40B) not A61K, A61K-008, A61Q
	15	Biotechnology	(C07G, C07K, C12M, C12N, C12P, C12Q, C12R, C12S) not A61K
	16	Pharmaceuticals	A61K not A61K-008
	17	Macromolecular chemistry, polymers	C08B, C08C, C08F, C08G, C08H, C08K, C08L
	18	Food chemistry	A01H, A21D, A23B, A23C, A23D, A23F, A23G, A23J, A23K, A23L, C12C, C12F, C12G, C12H, C12J, C13D, C13F, C13J, C13K
	19	Basic materials chemistry	A01N, A01P, C05#, C06#, C09B, C09C, C09F, C09G, C09H, C09K, C09D, C09J, C10B, C10C, C10F, C10G, C10H, C10J, C10K, C10L, C10M, C10N, C11B, C11C, C11D, C99Z
	20	Materials, metallurgy	C01#, C03C, C04#, C21#, C22#, B22#

Area		Field	IPC code
	21	Surface technology, coating	B05C, B05D, B32#, C23#, C25#, C30#
	22	Micro-structure and nano-technolgy	B81#, B82#
	23	Chemical engineering	B01B, B01D-000#, B01D-01##, B01D-02##, B01D-03##, B01D-041, B01D-043, B01D-057, B01D-059, B01D-06##, B01D-07##, B01F, B01J, B01L, B02C, B03#, B04#, B05B, B06B, B07#, B08#, D06B, D06C, D06L, F25J, F26#, C14C, H05H
	24	Environmental technology	A62D, B01D-045, B01D-046, B01D-047, B01D-049, B01D-050, B01D051, B01D-052, B01D-053, B09#, B65F, C02#, F01N, F23G, F23J, G01T, E01F-008, A62C
Mechanical engineering	25	Handling	B25J, B65B, B65C, B65D, B65G, B65H, B66#, B67#
	26	Machine tools	B21#, B23#, B24#, B26D, B26F, B27#, B30#, B25B, B25C, B25D, B25F, B25G, B25H, B26B
	27	Engines, pumps, turbines	F01B, F01C, F01D, F01K, F01L, F01M, F01P, F02#, F03#, F04#, F23R, G21#, F99Z
	28	Textile and paper machines	A41H, A43D, A46D, C14B, D01#, D02#, D03#, D04B, D04C, D04G, D04H, D05#, D06G, D06H, D06J, D06M, D06P, D06Q, D99Z, B31#, D21#, B41#
	29	Other special machines	A01B, A01C, A01D, A01F, A01G, A01J, A01K, A01L, A01M, A21B, A21C, A22#, A23N, A23P, B02B, C12L, C13C, C13G, C13H, B28#, B29#, C03B, C08J, B99Z, F41#, F42#
	30	Thermal processes and apparatus	F22#, F23B, F23C, F23D, F23H, F23K, F23L, F23M, F23N, F23Q, F24#, F25B, F25C, F27#, F28#
	31	Mechanical elements	F15#, F16#, F17#, G05G
	32	Transport	B60#, B61#, B62#, B63B, B63C, B63G, B63H, B63J, B64
Other fields	33	Furniture, games	A47#, A63#
	34	Other consumer goods	A24#, A41B, A41C, A41D, A41F, A41G, A42#, A43B, A43C, A44#, A45#, A46B, A62B, B42#, B43#, D04D, D07#, G10B, G10C, G10D, G10F, G10G, G10H, G10K, B44#, B68#, D06F, D06N, F25D, A99Z
	35	Civil engineering	E02#, E01B, E01C, E01D, E01F-001, E01F-003, E01F-005, E01F-007, E01F-009, E01F-01#, E01H, E03#, E04#, E05#, E06#, E21#, E99Z

Appendix C. Results of DTM (document-term matrix) and LSA (latent semantic analysis)

Sector	Number of patents	Number of bigram terms	Number of LSA-reduced features
Finance/Mgmt	12,603	30,862	3759
Medical	10,218	24,062	3404
Transport	6426	14,524	2177
Semiconductor	1896	5301	647
Game	1576	4495	522
Biotechnology	2956	8640	837

Appendix D. Top 10 Tie Value in Ego-network

Hub	Tie	Value	Hub	Tie	Value	Hub	Tie	Value	Hub	Tie	Value
G06Q-010	G06F-017	5294	G06Q-020	G06K-009	8061	G06Q-030	G06K-009	7737	G06Q-040	G06K-009	1273
	G06K-009	5100		G06F-003	7208		G06F-017	6550		G06Q-020	922
	G06F-003	2404		G06F-017	3223		G06F-003	5832		G06F-017	723
	G06Q-030	2305		G06Q-030	2965		G06Q-020	2965		G06F-003	440
	G06Q-050	1777		G06F-021	2861		H04N-021	2687		G06Q-030	384
	H04W-004	1112		B41J-029	2765		G06Q-010	2305		H04N-021	359
	H04N-021	1074		H04N-001	2761		H04N-001	1805		G06F-021	313
	G10L-015	1041		H04N-021	2284		G06Q-050	1592		G06Q-010	289
	G06K-007	1023		B41J-002	1901		G10L-015	1492		H04N-005	197
	H04L-029	1017		G06K-019	1803		G06F-021	1385		G06Q-050	183
G06Q-050	G06F-017	2759	A61B-005	G06K-009	13,634	A61B-006	G06K-009	5704	A61B-008	A61B-005	2554
	G06K-009	2417		G06T-007	8952		G06T-007	3817		G06K-009	2272
	G06Q-010	1777		A61B-006	3389		A61B-005	3389		G06T-007	2166
	G06Q-030	1592		G06F-019	3092		G06T-011	1138		A61B-006	1099
	G06F-021	1437		A61B-008	2554		A61B-008	1099		G06T-011	449
	G06F-003	1128		G06F-003	2059		G06T-005	889		G06F-019	447
	A61B-005	665		G06F-017	1739		G06F-019	588		G01R-033	305

Hub	Tie	Value	Hub	Tie	Value	Hub	Tie	Value	Hub	Tie	Value
	G06F-019	649		G01J-005	1631		G01R-033	434		G06F-017	294
	H04L-029	645		G01R-033	1547		G06T-001	422		G06T-017	273
	H04N-021	607		H04N-005	1401		G06T-017	342		A61B-034	270
A61B-003	G06K-009	2091	A61B-017	G06T-007	315	B60R-021	B60N-002	3157	B60R-001	G06K-009	2666
	A61B-005	1344		A61B-005	308		G01S-015	2275		B60R-021	1122
	G06T-007	928		G06K-009	306		G01S-007	1705		H04N-005	925
	G06F-003	475		A61B-018	238		B60R-022	1698		G06T-007	764
	G02B-027	386		A61B-006	187		G06K-009	1650		H04N-007	756
	H04N-005	266		A61B-034	153		G01F-023	1368		G08G-001	588
	A61F-009	260		A61B-008	140		G01S-017	1169		B60R-011	492
	A61B-008	167		A61F-002	115		B60R-001	1122		B60N-002	483
	A61M-021	143		A61B-090	88		B60R-016	1055		B60Q-001	420
	G06T-005	140		A61B-010	82		G01S-013	703		G01S-015	346
B60N-002	B60R-021	3157	B60W-030	G06K-009	1723	B60Q-001	G06K-009	1296	B60W-050	G06K-009	818
	G01S-015	1043		G08G-001	967		G08G-001	560		G08G-001	584
	B60R-022	844		G05D-001	770		B60R-001	420		B60W-030	518
	G01S-007	760		B60W-010	731		H04N-021	353		H04N-021	367
	G01F-023	621		G06T-007	618		F21S-041	314		B60W-040	362
	G01S-017	526		B60W-050	518		B60R-021	311		G06F-003	361
	G06K-009	507		B60W-040	440		H04N-007	272		G05D-001	334
	B60R-001	483		G01C-021	394		G01C-021	270		G01C-021	315
	B60R-016	479		B60T-007	237		G06F-003	243		H04W-004	250
	G01S-013	309		B60R-001	234		H04W-004	217		B60W-010	222
B60W-040	G06K-009	832	B60R-016	B60R-021	1055	B60R-025	H04N-021	557	B60R-011	G06K-009	1055
	G08G-001	646		B60N-002	479		G06K-009	490		B60R-001	492
	B60W-030	440		G10L-015	409		G08G-001	381		H04N-005	373

Hub	Tie	Value	Hub	Tie	Value	Hub	Tie	Value	Hub	Tie	Value
	B60W-050	362		G01S-015	379		G06F-003	368		G06T-007	350
	H04N-021	358		G06K-009	302		H04W-004	356		B60R-021	308
	G01C-021	354		G01S-007	299		G01C-021	317		H04N-007	268
	G06F-003	331		G06F-003	297		G06F-021	282		G08G-001	243
	G05D-001	284		B60R-022	272		G06Q-030	238		B60Q-001	170
	H04W-004	253		H04N-021	240		G07C-005	207		B60W-030	146
	G06F-021	212		G01C-021	231		G05D-001	190		G01C-021	119
B60W-010	B60W-030	731	B64C-039	G06K-009	734	B64D-047	G06K-009	544	H01L-027	H04N-005	2883
	G06K-009	510		G05D-001	438		B64C-039	295		H04L-012	1342
	B60W-050	222		B64D-047	295		G05D-001	289		H04N-007	1065
	G05D-001	203		G06T-007	261		G06T-007	214		G06K-009	1045
	G08G-001	200		H04N-005	202		H04N-005	199		H04W-008	854
	G06T-007	178		G08G-005	166		H04N-007	136		H04M-001	738
	B60W-040	143		H04N-007	156		G08G-005	125		H04L-009	734
	B60R-001	116		A01M-001	85		G06Q-010	102		H04W-028	706
	B60T-007	87		G06Q-010	74		G06F-017	64		H04W-088	676
	B62D-015	87		G06F-017	54		G06Q-050	63		H03M-013	628
H01L-021	G06K-009	1387	H01L-023	H01L-021	749	G01N-033	G06K-009	3042	C12Q-001	G06F-019	1679
	H01L-023	749		G06K-009	689		G06F-019	2984		G01N-033	963
	G01N-021	548		H01L-025	176		C12Q-001	963		G06K-009	510
	G03F-001	501		H01L-027	135		G06F-017	934		G01N-021	232
	G06T-007	359		H05K-001	81		G01N-021	869		G06T-007	190
	G01B-011	280		G07F-007	72		G06F-007	824		C12M-001	187
	G06T-001	223		G11B-020	63		G06K-007	761		C12N-015	171
	G03F-007	159		G06T-001	56		G06T-007	715		G06F-017	164
	H01L-027	159		A61B-005	50		G01N-015	557		G01N-015	110

Hub	Tie	Value	Hub	Tie	Value	Hub	Tie	Value	Hub	Tie	Value
	G01R-031	136		H04N-001	50		G06Q-030	510		G01J-003	90
A63F-013	G06K-009	3258	A63B-071	A63B-069	444	A63B-069	A63B-071	444	A63B-024	G06K-009	320
	G06F-003	2155		G06K-009	316		A61B-005	329		A61B-005	254
	H04N-021	1734		A61B-005	302		G06K-009	282		A63B-071	163
	G06T-007	1136		A63B-024	163		G09B-019	144		A63B-069	114
	G06Q-020	865		G06F-003	154		G06F-003	131		G06T-007	94
	H04N-005	861		G09B-019	127		A63B-024	114		G09B-019	71
	G06F-017	739		A63B-021	126		A63B-021	93		G06F-003	66
	G06Q-030	555		A63F-013	87		G06F-001	80		A63F-013	64
	G10L-015	477		G06F-019	85		H04W-084	80		H04B-001	62
	H04N-007	366		H04B-001	83		B33Y-010	79		H04N-005	61

Appendix E. Results of keywords and TF-IDF values (Finance and AI)

Rank	Cluster 0		Cluster 1		Cluster 2		Cluster 3		Cluster 4	
	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF
1	Content	0.198377944	Datum	0.029212789	Document	0.17864845	User	0.087232717	Image	0.177121274
2	User	0.040252441	Base	0.017531206	Code	0.061838046	Information	0.059012376	Object	0.047988414
3	Content Item	0.035167167	Information	0.016409083	Datum	0.059727567	Biometric	0.033614463	Capture	0.041450252
4	Display	0.0325190 78	Use	0.015634365	Sensing	0.05333588	Transaction	0.031552755	Information	0.031399814
5	Item	0.031191526	Customer	0.01473762	Code datum	0.05100205	Authentication	0.028313665	Image datum	0.024372024
6	Page	0.027174535	Provide	0.014297695	Computer	0.035112772	Datum	0.022313967	Product	0.024329546
7	Medium	0.026360775	Product	0.014256802	Indicate datum	0.034483432	Voice	0.020990299	Datum	0.024130782
8	Information	0.025790752	Model	0.014034297	Surface	0.033739238	Provide	0.019496583	Image capture	0.022494829
9	Digital	0.022243898	Message	0.012991128	Print	0.032784042	Service	0.01893352	Capture image	0.021723515
10	Web	0.022192628	Determine	0.012944091	Identity	0.031411633	Base	0.017202477	Display	0.020512878
11	Datum	0.020624464	User	0.012837834	Indicate	0.031348711	Use	0.017170305	Unit	0.020382114
12	Base	0.019255619	Item	0.012461549	Indicative	0.029765584	Communication	0.017162653	Item	0.019177003
13	Provide	0.018451822	Computer	0.012348857	Interface surface	0.028668778	Server	0.016983829	Processing	0.018852817
14	Medium Content	0.017299259	Generate	0.012347353	Electronic document	0.028462684	Card	0.016597788	User	0.017716352
15	Identify	0.016048971	Plurality	0.012329854	Interface	0.027046014	Receive	0.01605443	Base	0.01717543
16	Digital content	0.015065079	Vehicle	0.011934425	Electronic	0.025883016	Electronic	0.015451682	Determine	0.01660745

Rank	Cluster 0		Cluster 1		Cluster 2		Cluster 3		Cluster 4	
	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF
17	Generate	0.01423348	Set	0.011662284	Use	0.023721453	Request	0.015377506	Plurality	0.015681891
18	Request	0.013389292	Identify	0.011510882	Form	0.023544227	Identification	0.015155556	Digital	0.015618133
19	Web page	0.013850067	Receive	0.011483505	Sense	0.023439435	Mobile	0.014984362	Identify	0.015582699
20	Multimedia	0.013385686	Associate	0.011457298	Datum indicative	0.022694584	Second	0.014465021	Use	0.015435008
21	Server	0.012770678	Text	0.011435366	Information	0.022689768	Input	0.013777312	Digital image	0.014805898
22	Receive	0.012665908	Time	0.011424035	Element	0.021656041	Application	0.013754555	Store	0.014418553
23	Audio	0.012664214	Data	0.011018341	Product item	0.021625903	Associate	0.012399184	Recognition	0.013856576
24	Unit	0.012663562	Object	0.01085272	User	0.019911184	Payment	0.012305476	Camera	0.013759393
25	Network	0.012359756	Second	0.010539817	Interactive element	0.019756694	Interface	0.012282748	Medical	0.013214398
26	Determine	0.012244654	Display	0.010466937	Image	0.019695849	Access	0.012200129	Second	0.013186597
27	Advertisement	0.012226459	Service	0.010221166	Transport	0.018801803	Terminal	0.012177946	Process	0.013069365
28	Use	0.012030406	Event	0.010021539	Interactive	0.0185945	Network	0.012159269	Feature	0.012719479
29	Client	0.012022752	Language	0.009718087	Interaction	0.017551788	Computer	0.011668643	Region	0.012668615
30	Signal	0.011840012	Process	0.00952546	Product	0.017465935	Determine	0.011592611	Vehicle	0.012644236

Appendix F. Results of keywords and TF-IDF values (Medical and AI)

Rank	Cluster 0		Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5	
	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF
1	Image	0.12833039	User	0.044566694	Image	0.149681279	Light	0.101583329	Projection	0.119297612	Image	0.052189446
2	Unit	0.106690014	Signal	0.044505499	Image datum	0.027416184	Fingerprint	0.090448725	Ray	0.107269789	Datum	0.027815644
3	Processing	0.08023336	Datum	0.024469763	Datum	0.02627893	Sensor	0.054408846	Image	0.091765201	Region	0.02272248
4	Medical	0.0684575	Eye	0.021456577	Pixel	0.024506636	Finger	0.054060227	Datum	0.049854581	Use	0.02067914
5	Medical image	0.067533979	Sensor	0.021116344	Object	0.023767358	Surface	0.037915458	Projection image	0.049538731	Imaging	0.019572506
6	Image processing	0.06751749	Information	0.020196679	Second	0.023430128	Object	0.035298815	Projection datum	0.049110468	Patient	0.019263426
7	Information	0.051642611	Patient	0.01926218	Capture	0.021532835	Image	0.035106531	Ray Image	0.045568751	Object	0.017973157
8	Display	0.04751378	Base	0.018273905	Display	0.020177258	Light source	0.032372541	Reconstruct	0.04154275	Model	0.017096626
9	Unit configure	0.03899422	Control	0.017792294	Processing	0.019100076	Electrode	0.031413606	Ct	0.040786627	Medical	0.017020779
10	Region	0.033572651	Use	0.017542315	Value	0.018310521	Source	0.030607139	Reconstruction	0.039133382	Tissue	0.016149597
11	Configure	0.032858229	Determine	0.01708166	Information	0.016447049	Layer	0.030330429	Object	0.030652927	Dimensional	0.015811229
12	Datum	0.029559938	Configure	0.015528522	Acquire	0.016266851	Unit	0.029090611	Source	0.024814282	Base	0.015616778

Rank	Cluster 0 Keyword	Mean TF-IDF	Cluster 1 Keyword	Mean TF-IDF	Cluster 2 Keyword	Mean TF-IDF	Cluster 3 Keyword	Mean TF-IDF	Cluster 4 Keyword	Mean TF-IDF	Cluster 5 Keyword	Mean TF-IDF
13	Plurality	0.025654308	Biometric	0.015280724	Set	0.01591795	Identification	0.024699352	Detector	0.024249885	Determine	0.014949172
14	Image datum	0.024733249	Receive	0.015170428	Iris	0.01584503	Capture	0.022607802	Generate	0.022544378	Feature	0.014583906
15	Processing image	0.023606485	Provide	0.014980987	Plurality	0.015392949	Fingerprint sensor	0.020796236	Reconstruct image	0.021832223	Set	0.014091084
16	Acquire	0.021248901	State	0.014854827	Region	0.014996175	Element	0.020529027	Imaging	0.020953452	Point	0.012976287
17	Obtain	0.020051507	Display	0.013994934	Frame	0.014963367	Information	0.020291217	Temperature	0.020715122	Volume	0.012956456
18	Generate	0.019192259	Subject	0.013966003	Base	0.014554158	Contact	0.020132733	Tomography	0.019721685	Provide	0.012928826
19	Second	0.018233156	Detect	0.013610199	Image processing	0.014245878	Pattern	0.019472063	Tomosynthesis	0.019598999	Generate	0.012915165
20	Storage	0.017663728	Individual	0.013408652	Imaging	0.014233494	Array	0.019455031	Use	0.019343465	Second	0.012445498
21	Position	0.016984598	Person	0.013269696	Obtain	0.014162927	Sensing	0.018890653	Compute	0.019306941	Value	0.012395759
22	Circuitry	0.016695628	Speech	0.013112183	Generate	0.013896711	Capacitance	0.018545644	Core temperature	0.018794324	Position	0.011962686
23	Area	0.016131952	Camera	0.013015364	Use	0.013721902	Emit	0.017608535	Body core	0.018794324	Identify	0.011714824
24	Diagnosis	0.016119545	Input	0.012849453	Subject	0.01345284	Eye	0.017537753	Image datum	0.018453915	Structure	0.011460095
25	Extract	0.015930934	Time	0.012812523	Second image	0.013415665	Second	0.017303589	Artifact	0.018279601	Image datum	0.011383725

Rank	Cluster 0		Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5	
	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF
26	Base	0.015386953	Processor	0.012164025	Process	0.013351851	Substrate	0.017104841	Measurement	0.018275404	Subject	0.011382012
27	Section	0.015161255	Voice	0.012074533	Section	0.013123026	Form	0.016517986	Compute tomography	0.018237567	Information	0.011279962
28	Processing circuitry	0.014942322	Physiological	0.011244492	Radiation	0.012503388	Portion	0.016430622	Source point	0.018226013	Display	0.011222254
29	Imaging	0.014822018	Unit	0.011235695	Determine	0.011809622	Circuit	0.015973076	Measurement external	0.018226013	Scan	0.010942954
30	Tomographic image	0.014507676	Response	0.01103956	Image frame	0.011575773	Position	0.015310686	External source	0.017990564	Plurality	0.010866449

Appendix G. Results of keywords and TF-IDF values (Transport and AI)

Rank	Cluster 0 Keyword	Mean TF-IDF	Cluster 1 Keyword	Mean TF-IDF	Cluster 2 Keyword	Mean TF-IDF	Cluster 3 Keyword	Mean TF-IDF	Cluster 4 Keyword	Mean TF-IDF	Cluster 5 Keyword	Mean TF-IDF	Cluster 6 Keyword	Mean TF-IDF
1	Aerial	0.159427233	Vehicle	0.031056294	Vehicle	0.096151063	Driver	0.160409504	Object	0.156138695	Image	0.13412893	Lane	0.279929199
2	Uav	0.145661685	Image	0.030292462	Control	0.047991409	Vehicle	0.058581736	Light	0.064978981	Display	0.077100225	Vehicle	0.072954673
3	Aerial vehicle	0.142521953	Sensor	0.028861171	User	0.036221433	Information	0.03931589	Image	0.058780659	Vehicle	0.069974302	Line	0.053775216
4	Unmanned	0.134573367	Occupant	0.025066266	Parking	0.034826919	Drive	0.037710207	Vehicle	0.04289003	Camera	0.061163007	Travel	0.046231281
5	Unmanned aerial	0.116606092	Datum	0.023229803	Information	0.03387836	State	0.030959532	Detection	0.040101062	Capture	0.04924199	Lane change	0.045317069
6	Flight	0.05441209	Signal	0.019014479	Voice	0.027270317	Assistance	0.028984218	Unit	0.038623791	View	0.042136822	Change	0.041479546
7	Vehicle	0.048996729	Position	0.018430637	Autonomous	0.024645105	Gaze	0.02569218	Detect	0.033569292	Image datum	0.040922548	Image	0.035969317
8	Datum	0.035075087	Determine	0.018028235	Unit	0.024396807	Driver assistance	0.024306248	Object detection	0.027210971	Datum	0.03847697	Road	0.033542202
9	Vehicle uav	0.033635654	Use	0.01795721	Speech	0.022978872	Unit	0.024076081	Region	0.026264719	Unit	0.033593735	Boundary	0.030161741
10	Image	0.033429338	Base	0.01770307	Command	0.022113133	Image	0.023990315	Source	0.023025796	Capture image	0.029347439	Travel lane	0.029410796
11	Landing	0.033115104	Information	0.017147233	Base	0.021048951	FINGERPRINT	0.023723216	Information	0.022428436	Control	0.027182451	Lane Mark	0.029223709
12	Structure	0.031329129	Camera	0.016754367	Recognition	0.019435878	Determine	0.022911094	Light source	0.02222869	Processing	0.024307513	Unit	0.029202966
13	Location	0.027392452	Detect	0.014940307	Input	0.017964097	Vehicle driver	0.022681827	Area	0.021011946	Second	0.023564644	Detect	0.029111632
14	Camera	0.024072838	Video	0.014822091	Display	0.017905456	Driver vehicle	0.021946511	Second	0.020922634	Area	0.019879779	Mark	0.029030678
15	Flight path	0.023507981	Unit	0.014491309	Determine	0.017509464	Driving	0.021296849	Camera	0.020169578	Processor	0.019678465	Control	0.027330377
16	Target	0.023366543	Area	0.014135431	Datum	0.017251797	Base	0.020289442	Determine	0.020018552	Image capture	0.019659964	Lane boundary	0.026474082

Rank	Cluster 0		Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5		Cluster 6	
	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF
17	Use	0.02324508	Second	0.014035532	Drive	0.017200359	Detect	0.02028685	Configure	0.018680393	Configure	0.018959778	Lane line	0.026366581
18	Inspection	0.023187405	Control	0.0139589	Configure	0.016609577	Behavior	0.020003265	Detect object	0.017903007	Image processing	0.018357075	Marking	0.026134196
19	Object	0.022544702	Provide	0.013449327	Vehicle control	0.016406075	Configure	0.019591766	Position	0.017532529	Field	0.017836405	Vehicle lane	0.023587277
20	Unmanned aircraft	0.021658919	Point	0.013389051	Signal	0.016384093	Datum	0.018354341	Base	0.017026624	Field view	0.017425634	Recognize	0.023400175
21	Sensor	0.020468828	Aircraft	0.013294046	Autonomous vehicle	0.016192819	Provide	0.017186052	Target	0.015741914	Video	0.017173459	Information	0.023005102
22	Control	0.019872358	Detection	0.013118654	Controller	0.01603727	Fact check	0.017167639	Capture	0.015699335	Imaging	0.016904874	Determine	0.022772694
23	Information	0.019386384	Configure	0.013108036	Provide	0.015433222	Fact	0.01701888	Dimensional	0.014939229	Region	0.016245116	Lane marking	0.022094068
24	Fly	0.019241628	Seat	0.013065	Sensor	0.015428558	FACE	0.016975401	Object detect	0.014715535	Object	0.016130362	Point	0.021735698
25	Path	0.019021107	Value	0.012675534	Operation	0.015032859	Driver state	0.016829022	Control	0.014583048	Vision	0.015902234	Position	0.020982387
26	Capture	0.018277651	Feature	0.012621097	Receive	0.014855577	Eye	0.016721036	Use	0.01428023	Detect	0.015749778	Detect lane	0.020286683
27	Receive	0.017994064	Receive	0.012565294	Motor	0.014831073	Monitor	0.01668831	Traffic light	0.014278756	Rear	0.015735298	Base	0.019967848
28	Rooftop	0.017598594	Plurality	0.012556094	Detect	0.014382639	Direction	0.016681767	Traffic	0.014135008	Driver	0.015485735	Departure	0.019700223
29	Aircraft	0.017561018	Road	0.012529721	Motor vehicle	0.014134266	Duration	0.016436393	Datum	0.013785172	Process	0.015212184	Vehicle travel	0.019125669
30	Configure	0.017507896	Surface	0.011833398	Communication	0.013981051	Alert	0.016247663	Distance	0.013653862	Road	0.015167621	Datum	0.018656464

Appendix H. Results of keywords and TF-IDF values (Semiconductor and AI)

Rank	Cluster 0		Cluster 1		Cluster 2		Cluster 3		Cluster 4	
	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF
1	Signal	0.044702592	Pattern	0.108125686	Light	0.147184061	defect	0.250480901	sensor	0.096859752
2	Pixel	0.035166852	Image	0.083281964	Display	0.122139957	inspection	0.120098083	layer	0.091713197
3	Circuit	0.0303897	Wafer	0.048132518	Emit	0.069394297	image	0.113537734	fingerprint	0.078581499
4	Image	0.030184402	Datum	0.047816355	Fingerprint	0.068853102	pattern	0.08121038	surface	0.066872685
5	Sensor	0.029509571	Mask	0.041385668	Light emit	0.063671234	inspect	0.052762786	chip	0.065993618
6	Array	0.028699363	Inspection	0.036879544	Layer	0.060014388	detect	0.047682959	fingerprint sensor	0.059097502
7	Element	0.026141355	Position	0.030430035	Panel	0.053804267	candidate	0.040216658	sensing	0.055443751
8	Second	0.024502222	Second	0.026302931	Substrate	0.045932114	pattern inspection	0.038987837	electrode	0.054315777
9	Datum	0.023922522	Reference	0.026121717	Display panel	0.045335622	defect candidate	0.038512054	package	0.052778546
10	Substrate	0.02104948	Edge	0.025437432	Sensor	0.044821112	defect inspection	0.03674768	substrate	0.050240005
11	Light	0.020837783	Semiconductor	0.025205057	Optical	0.042653287	wafer	0.034416902	structure	0.040057451
12	Semiconductor	0.019296188	Object	0.024340352	Pixel	0.03975595	reference	0.033249917	form	0.037756285
13	Line	0.019261878	Use	0.024225027	Electrode	0.036813508	classification	0.032691374	conductive	0.035978669
14	Unit	0.018957725	Inspect	0.024183914	Sensing	0.034345266	value	0.030955755	second	0.035715671
15	Information	0.018206322	Region	0.023990048	Unit	0.033452424	unit	0.030821797	circuit	0.03434731
16	Output	0.018028201	Value	0.023737108	Recognition	0.03314576	reference image	0.029270298	die	0.034211313
17	Layer	0.017821151	Exposure	0.023651016	Dispose	0.031829755	detection	0.028878046	pad	0.02665942

Rank	Cluster 0		Cluster 1		Cluster 2		Cluster 3		Cluster 4	
	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF
18	Object	0.017703665	Measure	0.023420336	Fingerprint recognition	0.030024992	detect defect	0.027800169	connection	0.02640326
19	Use	0.017229621	Obtain	0.02276489	Organic	0.029888434	Sample	0.026861223	Connect	0.024482952
20	Control	0.017111873	Process	0.022333478	Plurality	0.02855115	Defect image	0.026787822	Dielectric	0.024316756
21	Channel	0.016341966	Optical	0.021761987	Identification	0.028415769	Datum	0.026483719	Portion	0.023274139
22	Detection	0.015834037	Unit	0.020266575	Organic light	0.027707207	Inspection condition	0.026323708	Electrically	0.022969956
23	Processing	0.015516258	Design	0.019899876	Array	0.027135693	Image defect	0.024251549	Cover	0.022713028
24	Process	0.014981377	Processing	0.01928304	Element	0.026772149	Condition	0.023292133	Contact	0.02180223
25	Sensing	0.014679848	Correction	0.019276089	Fingerprint identification	0.026619077	Inspection image	0.022985454	Dispose	0.021795028
26	Configure	0.014505143	Plurality	0.019130024	Region	0.026614971	Information	0.022942306	Finger	0.021667264
27	Provide	0.014295471	Measurement	0.018463426	Photosensitive	0.025447119	Obtain	0.022364682	Material	0.021526337
28	Image sensor	0.013897129	Detect	0.018446415	Oled	0.025377127	Compare	0.022319001	Dielectric layer	0.021258494
29	Form	0.013848361	Determine	0.018338073	Source	0.025243756	Use	0.021711746	Plate	0.021160379
30	Cell	0.013842385	Calculate	0.017847281	Light source	0.025209527	Pattern defect	0.02169489	Semiconductor	0.02069056

Appendix I. Results of keywords and TF-IDF values (Game and AI)

Rank	Cluster 0 Keyword	Mean TF-IDF	Cluster 1 Keyword	Mean TF-IDF	Cluster 2 Keyword	Mean TF-IDF	Cluster 3 Keyword	Mean TF-IDF	Cluster 4 Keyword	Mean TF-IDF	Cluster 5 Keyword	Mean TF-IDF
1	Voice	0.080636368	User	0.056927493	Reality	0.232453193	Object	0.43868811	Image	0.171958667	Game	0.067825941
2	Signal	0.050224917	Datum	0.055877063	Augment reality	0.225395519	Capture	0.177600925	Card	0.119606654	Video	0.059295019
3	Audio	0.046373366	Motion	0.04830412	Augment	0.213153391	Information	0.152080695	Face	0.050237501	Object	0.056198027
4	Sound	0.043668412	Sensor	0.045016688	Reality display	0.121915096	Address correspond	0.151615894	Processing	0.040308931	Image	0.0403155
5	Speech	0.042752214	Exercise	0.036717848	World	0.113477717	Information address	0.151615894	Information	0.038907048	User	0.039687433
6	Robot	0.041001165	Event	0.033750717	Passable world	0.091772342	Capture identification	0.150979901	Object	0.035160213	Gesture	0.038438559
7	Control	0.040628489	Activity	0.02634334	Passable	0.091772342	Image	0.150808477	Capture	0.034964776	Player	0.03795768
8	Input	0.03950159	Information	0.025593977	World model	0.091772342	Identification process	0.149741889	Display	0.032678618	Golf	0.034766763
9	Command	0.03571509	Movement	0.023126028	Display	0.091545322	Use access	0.149119244	Section	0.029071137	Camera	0.031573598
10	Character	0.035622645	Use	0.022092086	Individual augment	0.088029668	Process digital	0.148484742	Unit	0.028571637	Depth	0.028391094

Rank	Cluster 0		Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5	
	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF
11	Information	0.03558271	Unit	0.020711602	Model datum	0.085965192	Object database	0.147882605	Capture image	0.028386674	Ball	0.026640296
12	Game	0.033377825	Provide	0.019940811	Waveguide	0.075197316	Communication pertinent	0.147165537	Detect	0.027820283	Virtual	0.026504529
13	Unit	0.029224434	Base	0.019641776	Model	0.06336711	Pertinent object	0.147165537	Information processing	0.026776362	Use	0.026256791
14	Output	0.02911663	Athletic	0.019339101	Individual	0.062621649	Initiate communication	0.147165537	Area	0.025953933	Determine	0.025900724
15	Datum	0.024423045	Toy	0.018716625	Pass	0.061063022	Information initiate	0.147165537	Image processing	0.024842534	Position	0.023614954
16	Message	0.022798916	Analysis	0.018553061	Comprise	0.050209962	Object recognize	0.146653117	User	0.023421889	Base	0.023264422
17	Communication	0.022145106	Tag	0.018017177	Datum	0.044756867	Address	0.146272932	Datum	0.02328261	Information	0.023241233
18	Generate	0.021571009	Body	0.017155502	Planar waveguide	0.044367352	Capture object	0.146000332	Region	0.022670175	Capture	0.022559429
19	Processing	0.020771693	Performance	0.016964766	Planar	0.043032665	Pertinent	0.145906691	Position	0.022422868	Provide	0.021166177

Rank	Cluster 0		Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5	
	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF
20	Text	0.019757139	Signal	0.016798959	Virtual	0.042669155	Digital image	0.145766315	Extract	0.022265651	Video game	0.020576439
21	Display	0.019359831	Receive	0.016777097	Doe	0.041106619	Access information	0.145294647	Image datum	0.022040247	Control	0.020411992
22	Data	0.019269211	Processor	0.016497447	Object	0.041042742	Recognize plurality	0.144888864	Camera	0.021327564	Location	0.019835285
23	User	0.019119122	Time	0.016310078	Set	0.040731824	Object capture	0.144751453	Set	0.019801488	Club	0.01952646
24	Receive	0.019008362	Analyze	0.016168934	Augmented	0.0402833	Object use	0.142621069	Base	0.019186749	Second	0.019158768
25	Recognition	0.018232912	Configure	0.016112217	Map point	0.040218541	Plurality object	0.141614644	Second	0.018326374	Display	0.019015473
26	Conversation	0.018006436	Generate	0.015586703	Real world	0.039140088	Database Information	0.14099138	Acquire	0.018265498	Datum	0.018607618
27	Mean	0.017700223	Sport	0.015533802	Set map	0.038909688	Initiate	0.140764922	Generate	0.018245853	Track	0.018386891
28	Operation	0.017157954	Determine	0.015438573	Point	0.037892553	Image object	0.13912926	Hmd	0.018053191	Play	0.017989826
29	Microphone	0.017014063	Example	0.015084341	Reality augment	0.037739146	Correspond object	0.138428774	Point	0.017493771	Computer	0.017731684
30	Provide	0.016361572	Application	0.015024328	Image	0.037487396	Digital	0.134547617	Register	0.016696607	Plurality	0.017222084

Appendix J. Results of keywords and TF-IDF values (Biotechnology and AI)

Rank	Cluster 0 Keyword	Mean TF-IDF	Cluster 1 Keyword	Mean TF-IDF	Cluster 2 Keyword	Mean TF-IDF	Cluster 3 Keyword	Mean TF-IDF	Cluster 4 Keyword	Mean TF-IDF	Cluster 5 Keyword	Mean TF-IDF	Cluster 6 Keyword	Mean TF-IDF
1	Datum	0.036473279	Image	0.156658117	Code	0.173355978	Sequence	0.168139919	Sample	0.103804642	Gene	0.074982537	Cell	0.249888437
2	Use	0.024491862	Object	0.038706815	Product Item	0.165962072	Chip	0.073742299	Biological	0.050898444	Expression	0.054858939	Image	0.057863485
3	Determine	0.018885495	Light	0.034858575	Item	0.154643534	Database	0.069136125	Biological sample	0.032702629	Cancer	0.052521331	Unit	0.034151346
4	Value	0.018420464	Pixel	0.031148781	Product	0.147050332	Oligonucleo-tide	0.063335794	Analysis	0.027891585	Disease	0.046226299	Culture	0.027509779
5	Base	0.017852577	Tissue	0.030449241	Datum	0.123147121	Model	0.044599847	Signal	0.025045956	Invention	0.043425127	Blood cell	0.027130116
6	Set	0.017378239	Cell	0.0268223	Code datum	0.108370305	Information	0.043993096	Nucleic acid	0.024570137	Treatment	0.032086037	Blood	0.026437477
7	Measure	0.016570474	Color	0.02678293	Interface surface	0.105965646	Probe	0.04297725	Nucleic	0.024570137	Patient	0.031713597	Analysis	0.024753323
8	Signal	0.016528994	Capture	0.026582046	Identity	0.100365016	Nucleotide	0.039583624	Acid	0.024219381	Gene Expression	0.03086849	Cell cell	0.023976884
9	Array	0.015775794	Specimen	0.022120576	Interface	0.098022844	Acid	0.03843464	Microbi-ome	0.024005384	Use	0.02939628	Sample	0.023760615
10	Sensor	0.015517946	Processing	0.021232978	Code data	0.090562877	Fiber	0.03829139	Cell	0.023295319	Present invention	0.029353071	Provide	0.020640079
11	Provide	0.014924164	Imaging	0.021168146	Surface	0.087748252	Nucleotide sequence	0.038161032	Provide	0.022441868	Subject	0.028935993	Identify	0.019906672
12	Detect	0.014720327	Sample	0.020879066	Data	0.08171961	Peptide	0.03791799	Composi-tion	0.022221726	Present	0.028095427	Determine	0.019544511
13	Test	0.014624313	Area	0.019834573	Data por-tion	0.080793378	Cluster	0.035695158	Image	0.021939285	Provide	0.028013307	Colony	0.019171383
14	Informa-tion	0.014587642	Analysis	0.01958918	Indicate datum	0.079544553	Array chip	0.035557393	Analyze	0.021192761	Biomarker	0.02696519	Target	0.018813344

Rank	Cluster 0 Keyword	Mean TF-IDF	Cluster 1 Keyword	Mean TF-IDF	Cluster 2 Keyword	Mean TF-IDF	Cluster 3 Keyword	Mean TF-IDF	Cluster 4 Keyword	Mean TF-IDF	Cluster 5 Keyword	Mean TF-IDF	Cluster 6 Keyword	Mean TF-IDF
15	Feature	0.014573884	Obtain	0.01954078	Portion	0.079357652	Nucleic	0.033188994	Use	0.020734013	Relate	0.026797458	Detect	0.018252357
16	Plurality	0.014482166	Value	0.018873391	Sensing	0.079181473	Nucleic acid	0.033188994	Condition	0.019911701	Identify	0.024544193	Feature	0.0180671
17	Analysis	0.013916717	Use	0.018401931	Indicative	0.072968414	Protein	0.031659175	Generate	0.019817458	Drug	0.023791878	Cell image	0.017901893
18	Plant	0.013620567	Process	0.017318653	Sense	0.072515478	Array	0.030876366	Base	0.019252419	Marker	0.023438864	State	0.017561688
19	Time	0.013342118	Plurality	0.017180416	Indicate	0.065400597	Database model	0.029801976	Determine	0.019170932	Invention relate	0.022895685	Cell population	0.017512512
20	Genetic	0.013196075	Biological	0.016232473	Sense code	0.057112256	Probe array	0.028351435	Comprise	0.019022345	Risk	0.021198079	Use	0.017261805
21	Model	0.012928349	Identify	0.016165183	Datum indicative	0.057610649	Sample	0.027982863	Dna	0.018385415	Predict	0.020590024	Target cell	0.016239512
22	Computer	0.011881703	Digital	0.016118842	Indicative identity	0.054971961	Organize information	0.026252613	Blood	0.018227272	Diagnosis	0.019536176	Plurality	0.015968902
23	Product	0.011797208	Image analysis	0.016074127	Identity product	0.0526563	Provide	0.025278891	Subject	0.018011882	Response	0.019291142	Imaging	0.015958532
24	Unit	0.01179494	Colony	0.015875143	Scanning	0.046985715	identify	0.024696997	Detect	0.017252995	Tumor	0.019054098	Information	0.015800876
25	Comprise	0.011793525	Information	0.015852125	Threshold	0.044426225	Use	0.024473405	Particle	0.016378128	Sample	0.018256374	Analysis cell	0.015479632
26	Process	0.011703455	Feature	0.015821996	Time pcr	0.040259911	Information relate	0.023505797	Second	0.016371287	Level	0.017849214	Population	0.015455206
27	Second	0.011584073	Second	0.015565325	Baseline	0.039091157	Organize	0.023057981	Specimen	0.016134434	Determine	0.017335211	Optical	0.015037281
28	Result	0.011465704	Determine	0.015558044	User	0.034381374	Acid sequence	0.022637953	Measurement	0.01609807	Invention provide	0.017306212	Cell culture	0.014921321
29	Protein	0.011311549	Analyze	0.015456845	Adapt	0.033804075	Target	0.022221809	Dataset	0.016004942	Genetic	0.017102623	Step	0.014885546
30	Select	0.01107696	Optical	0.015390779	Real time	0.031954265	Code	0.018777318	Analyte	0.015919984	Base	0.016971216	Tissue	0.013886119

Funding Not applicable.

Data availability USPTO patent data from Google Patent Datasets.

Code availability UCINET, Python custom code.

Declarations

Conflicts of interest The authors declare that they have no conflict of interest.

References

Athereye, S., & Keeble, D. (2000). Technological convergence, globalization and ownership in the UK computer industry. *Technovation, 20*, 227–245.

Baek, S., Kim, K., & Altmann, J. (2014). Role of platform provider in service network evolution: the case of Salesforce.com AppExchange. In *2014 IEEE conference on business informatics*, Geneva, Switzerland, Jul. 39–45.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). *Ucinet 6 for windows: software for social network analysis*. Analytic Technologies.

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2013). *Analyzing social networks*. SAGE Publications.

Brynjolfsson, E. Rock, D., & Syverson, C. (2017). Artificial intelligence and the modern productivity paradox: a clash of expectations and statistics. *National Bureau of Economic Research*. NBER Working Paper No. 24001. http://www.nber.org/papers/w24001

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multi-model inference: a practical information-theoretic approach*. Springer-Verlag.

Choi, J. Y., Jeong, S., & Kim, K. (2015). A Study on diffusion pattern of technology convergence: patent analysis for Korea. *Sustainability, 7*, 11546–11569.

Curran, C. S., & Leker, J. (2011). Patent indicators for monitoring convergence - examples from NFF and ICT. *Technological Forecasting and Social Change, 78*(2), 256–273.

Deloitte. (2016). *The expansion of Robo-advisory in wealth management*. 8/2016, 1–5.

Deloitte. (2018). *State of AI in the Enterprise*. 2nd Edition, 1–25.

Freeman, L. C. (1979). Centrality in social networks conceptual classification. *Social Networks., 1*(3), 215–239.

Fujii, H., & Managi, S. (2018). Trends and priority shifts in artificial intelligence technology invention: a global patent analysis. *Economic Analysis and Policy, 58*, 60–69.

Hagedoorn, J., & Cloodt, M. (2003). Measuring innovative performance: is there an advantage in using multiple indicators? *Research Policy, 32*(8), 1365–1378.

Han, E. J., & Sohn, S. Y. (2016). Technological convergence in standards for information and communication technologies. *Technological Forecasting and Social Change, 106*, 1–10.

Harhoff, D., Narin, F., Scherer, F. M., & Vopel, K. (1999). Citation frequency and the value of patented inventions. *Review of Economics & Statistics, 81*, 511–515.

Houlton, S. (2018). How artificial intelligence is transforming healthcare. *The Prescriber, 29*(10), 13–17.

Huang, J. (2017). An analysis of the intellectual structure of the cloud patents of SaaS. *Technology Analysis and Strategic Management, 29*(8), 917–931.

IDC. (2020). *Worldwide Artificial Intelligence Software Forecast*. 2020–2024, Aug.

Jackson, M. O. (2008). *Social and economic networks*. Princeton University Press.

Kim, J., & Lee, S. (2017). Forecasting and identifying multi-technology convergence based on patent data: the case of IT and BT industries in 2020. *Scientometrics, 111*, 47–65.

Kim, E., Cho, Y., & Kim, W. (2014). Dynamic patterns of technological convergence in printed electronics techniques: patent citation network. *Scientometrics, 98*, 975–998.

KIPO. (2018). https://www.kipo.go.kr/kpo/HtmlApp?c=33001&catmenu=m06_07_06

Kose, T., & Sakata, I. (2019). Identifying technology convergence in the field of robotics research. *Technological Forecasting & Social Change, 146*, 751–766.

Kwon, O., An, Y., Kim, M., & Lee, C. (2020). Anticipating technology-driven industry convergence: evidence from large-scale patent analysis. *Technology Analysis and Strategic Management, 32*(4), 363–378.

Lee, D. H., Seo, I. W., Choe, H. C., & Kim, H. D. (2012). Collaboration network patterns and research performance: the case of Korean public research institutions. *Scientometrics, 91*, 925–942.

Lee, S., Kim, W., Lee, H., & Jeon, J. (2016). Identifying the structure of knowledge networks in the US mobile ecosystem: patent citation analysis. *Technology Analysis and Strategic Management, 28*(4), 411–434.

Liu, J., Chang, H., Forrest, J. Y., & Yang, B. (2020). Influence of artificial intelligence on technological innovation: evidence from the panel data of china's manufacturing sectors. *Technological Forecasting & Social Change., 158*, 120142.

Liu, L., Yang, K., Fujii, H., & Liu, J. (2021). Artificial intelligence and energy intensity in China's industrial sector: effect and transmission channel. *Econometric Analysis and Policy, 70*, 276–293.

McKinsey & Company. (2018a). *Artificial intelligence-automative's new value-creating engine*. January, 1–32.

McKinsey & Company. (2018b). *Notes from the AI Frontier insights from hundreds of use cases*. April, 1–36.

Nystrom, A. (2008). *Understanding change processes in business networks: a study of convergence in Finnish telecommunications 1985–2005*. Ph.D. Dissertation. Åbo Akademi University Press. Finland.

Patel, E., & Kushwaha, D. S. (2020). Clustering cloud workloads: K-means vs Gaussian mixture model. *Procedia Computer Science, 171*, 158–167.

PWC. (2018). *The macroeconomic impact of artificial intelligence*. February, 1–78.

Rosenberg, N. (1976). *Perspectives on Technology*. Cambridge University Press.

Schmoch, U. (2008). Concept of a technology classification for Country comparison. *WIPO*. June 1–15.

Tractica (2016). *Top 15 use cases for artificial intelligence, practical AI use cases for big data, vision, and language applications: strategic analysis and market outlook*. pp.1–23.

Trajtenberg, M. (1990). A penny for your quotes: patent citations and the value of innovations. *Rand Journal of Economics, 21*(1), 172–187.

Tseng, C., & Ting, P. (2013). Patent analysis for technology development of artificial intelligence: a country-level comparative study. *Innovation: Management, Policy and Practice, 15*(4), 463–475.

Wang, Z., Cunha, C. D., Ritou, M., & Furet, B. (2019a). Comparion of K-means and GMM methods for contextual clustering in HSM. *Procedia Manufacturing, 28*, 154–159.

Wang, Z., Porter, A. L., Wang, X., & Carley, S. (2019b). An approach to identify emergent topics of technological convergence: A case study for 3D printing. *Technological Forecasting and Social Change, 146*, 723–732.

Wartburg, I., Teichert, T., & Rost. K. (2005). Inventive progress measured by multi-stage patent citation analysis. *Research Policy, 34*, 1591–1607.

WIPO. (2019a). *WIPO Technology Trends 2019*: Artificial Intelligence, pp. 1–154.

WIPO. (2019b). https://www.wipo.int/classifications/ipc/ipcpub/?notion=scheme&version=20190101

Yang, J., Ying, L., & Gao, M. (2020). The influence of intelligent manufacturing on financial performance and innovation performance: the case of China. *Enterprise Information Systems., 14*(6), 812–832.

Yu, K. H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. *Nature Biomedical Engineering, 2*, 719–731.