Ph. D. Dissertation in Engineering

Analyzing the Patterns of Technology Diffusion, Convergence, and Strategies in the Artificial Intelligence Sector

인공지능 섹터의 기술 확산, 융합, 전략 패턴 분석

February 2023

Graduate School of Seoul National University

Technology Management, Economics, and Policy Program

Soyea Lee

Analyzing the Patterns of Technology Diffusion, Convergence, and Strategies in the Artificial Intelligence Sector

지도교수 황준석

이 논문을 공학박사학위 논문으로 제출함
2023 년 2 월

서울대학교 대학원
협동과정 기술경영경제정책 전공
이소예

이소예의 공학박사학위 논문을 인준함
2023 년 2 월

위 원 장 김 연 배 (인)

부위원장 황 준 석 (인)

위 원 허 은 녕 (인)

위 원 김 원 준 (인)

위 원 윤 현 영 (인)

Abstract

Analyzing the Patterns of Technology Diffusion, Convergence, and Strategies in the Artificial Intelligence Sector

Soyea Lee
Technology Management, Economics, and Policy Program
College of Engineering
Seoul National University

This dissertation empirically explores the potential of AI as a general-purpose technology (GPT) from a multi-dimensional perspective. In particular, considering that AI is in its early stages in the economy as an emerging technology, this dissertation analyzes the impact of this technology based on technological pervasiveness, rather than its impact on the economic factors. This dissertation analyzes technological pervasiveness in terms of three dimensions of analysis level: the knowledge flow, industrial sector, and technology portfolio. Additionally, for each analysis level, this dissertation focuses on three different conceptual backgrounds of technology diffusion, convergence, and strategies in relation to innovation sides of GPTs.

The first study examines how the patterns of GPT-related features of AI appear

depending on AI technology progress and diffusion. The GPT-related features here are focused on the process of the recombination and diffusion of technological knowledge via the concepts of generality, originality, and complementarity. For this analysis, the diffusion process of each technology is constructed as time series data, after which dynamic time warping and time series clustering are used for a pattern analysis of the time series data. Also, the differences are identified in clusters classified according to time and diffusion patterns through an analysis of variance. As a result, it is found that the GPT-related features of AI show a further increase according to the progress of technologies, with the features found to be higher in technologies with high diffusion levels. In particular, among AI technologies, GPT-related features exist at high levels in AI-application types. This study identifies the development patterns of AI as a GPT from the perspectives of the knowledge flows and diffusion.

The second study proposes a new framework to investigate the horizontal and vertical applicability of industrial sectors to AI technology. The framework is a two-way approach which integrates an analytic method on technological classification-based structure data and text-based unstructured data to understand technology convergence into the three aspects: the industrial sector, the technology category, and technology utilization. Network analysis and clustering analysis methodologies are used. As a result, based on the framework, various industrial sectors and common technologies to which AI can be applied are identified, with the results showing that AI technology has horizontal applicability. Also, various technology categories and patterns of utilization in each industry are derived; these results show that AI technology has vertical applicability. Through the proposed framework, this study confirms the applicability of AI as a GPT in the industrial sector.

The third study focuses on the technology portfolio strategies considering the technology supply side of a GPT. In particular, startups can drive the disruptive innovation of a GPT, and technology diversity can contribute to the development of a GPT through innovation on the supply side based on a combination of various types of knowledge. Thus, this study investigates the technology diversity of AI startups in relation to startup investment. Technology diversity is measured by dividing it into two levels, related diversity and unrelated diversity, and firm types are divided into industry-specific and cross-industry AI startups. For the empirical analysis, panel data are constructed based on firm investment and patent information and are analyzed with a fixed effect model. The result shows that technology diversity has a positive relationship with investments in AI startups. Also, the relationship is more positive in the unrelated diversity case and with cross-industry AI firms. This finding suggests that the technology diversity of AI startups can act as a driving force for innovation during the development of AI technology as a GPT with respect to the growth mechanism in the AI sector.

The three studies covered in this dissertation provide useful theoretical and empirical contributions to AI and GPT research given their consideration of different conceptual backgrounds and the different characteristics of GPTs. In addition, the studies make methodological contributions to technological innovation research given its dynamic

pattern analysis method and via the new framework proposed in it. Comprehensively, this dissertation presents policies and strategic implications for the AI technology, AI companies, and AI industry overall.

Keywords: Artificial Intelligence, General-purpose technology, Technology diffusion, Technology convergence, Technology strategy**,** Patent analysis

Student Number: 2016-30267

Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Research backgrounds

A general-purpose technology (GPT) is described as "a single generic technology, recognizable as such over its whole lifetime, that initially has much scope for improvement and eventually comes to be widely used, to have many uses, and to have many spillover effects" (Lipsey et al., 2005). A GPT refers to a new technology which it complementary to previous knowledge, applicable to a wide range of existing industries, and that creates totally new industries (Feldman & Yoon, 2012). Likewise, a GPT has the potential and dynamism to be used in a wide range of industrial sectors, pervading the economy and resulting in increased productivity gains (Bresnahan & Trajtenberg, 1995). Historically, the printing, steam engine, electricity, motor vehicle, computer and internet are considered as GPTs that provided the basis for long-term economic growth (Feldman & Yoon, 2012; Helpman, 1998; Crafts, 2004; Rosenberg & Trajtenberg, 2004; Lipsey et al., 2005).

Recently, scholars have considered that artificial intelligence (AI) technology has the potential to grow as a GPT. AI is considered as a primary disruptive innovation, and is to be expected to complement existing products and processes in broad range of industrial sectors (Girasa, 2020; Omoge et al., 2022; Spanaki et al., 2022; Cockburn et al., 2018; Hötte et al., 2022). AI can drive technological innovations in various industries by improving knowledge creation, spillover, learning and absorption capabilities (Liu et al., 2020). Also, AI has been predicted to increase productivity and complementary innovations, and the results of new technological innovations will create new jobs and encourage more hiring in the long run (Brynjolfsson et al., 2017; Gries & Naudé, 2020; Acemoglu et al., 2022; Hötte et al., 2022).

Several studies have used the patent-based analysis method to measure and evaluate GPTs (Goldfarb et al., 2023; Moser & Nicholas, 2004; Feldman & Yoon, 2012; Petralia, 2020). Patents include information about technological innovations, the properties of the innovations, and the relationships among the innovations (Goldfarb et al., 2023). Thus, patents can be used to measure the quantitative and qualitative aspects regarding the impact of technological developments and the scope of technological fields (Hötte et al., 2022; Daim et al., 2006; Jaffe & de Rassenfosse, 2017).

With regard to AI, prior empirical research has investigated the potential of AI as a GPT using patent data. Hötte et al. (2022) explored the trend of AI technology in terms of GPT characteristics, defined as intrinsic growth, generality, and innovation complementarity. Nardin (2021) examined technological dynamism, pervasiveness, and complementarity to confirm the potential of AI technology as a GPT. Klinger et al. (2018) found that deep learning technology has the features of GPTs, specifically rapid growth, generality, impacts in other fields, and geographical evolution.

From the perspective of the industrial applicability of AI, previous studies have targeted specific industrial areas in relation to AI but have not investigated the impact of AI across various industries. Applied AI research has commonly investigated a specific industrial sector to examine positive effects or applied areas of AI in a target industry, such as healthcare, vehicles, finance, manufacturing, agriculture, and energy, among others (Yu et al., 2018; Houlton, 2018; McKinsey, 2018a; Deloitte, 2016; Liu et al., 2020; Zeba et al., 2021; Issa et al., 2022; Li et al., 2023; Yang et al., 2020). Liu et al. (2020) and Yang et al. (2020) found that AI in the manufacturing sector has had a positive influence on technological innovations and performance. Li et al. (2023) examined how AI effectively utilizes energy resources by forecasting energy usage levels and scheduling resources. Zeba et al. (2021) examined the application of AI in the manufacturing sector, investigating several recent important topics. However, while some reports and studies have run comparisons across industries (Klinger et al., 2018; PWC, 2018; Deloitte, 2018; McKinsey, 2018b), they mainly identified quantitative measures of total amounts. Those studies did not conduct an in-depth analysis of technological applications across various industries.

In terms of firms regarding AI, the literature has focused on how AI can affect and increase firm performance. Wamba-Taguimdje et al. (2020) investigated the effects of AI on the firm performance, focusing on the aspects of organizational performance and process innovations. Chatterjee et al. (2021) proposed a model of how implementing AI can impact firm performance and provide a competitive advantage. Mishra et al. (2022) examined the influence of AI in relation to the profitability and operating efficiency of firms. However, research that seeks to understand GPTs in corporate context is limited to a narrow range of studies (Qiu & Cantwell, 2015). In particular, the empirical literature commonly takes a usage side approach with regard to GPTs, whereas there has been little discussion of firms from the perspective of the supply side, especially inventors and creators of GPTs (Petralia, 2021; Qiu & Cantwell, 2015). However, GPT theory has significantly emphasized the importance of who can create, improve, and complement GPTs (Petralia, 2021; Bresnahan & Trajtenberg, 1995). For the corporate context, combinations and recombinations of variety technologies increase technological opportunities in a fundamental manner and consequently lead to firm growth (Qiu & Cantwell, 2015; Kim & Kogut, 1996; Kogut & Zander, 1992).

Despite the fact that the literature has attempted to investigate the impact of AI on firms or industries while focusing on improved productivity rates, measuring these effects and potential as they relate to GPTs has been difficult given the lack of data related to the results of AI innovations (Alderucci et al., 2019; Webb, 2019). The AI adoption cycle and diffusion stage remain in their early stages, making estimations of the diverse economic impacts of AI challenging (Brynjolfsson et al., 2021; Hötte et al., 2022). Research has suggested that the impact of AI investments is minimal on account of time lags related to implementations and restructuring (Chatterjee et al., 2021; Mikalef & Gupta, 2021).

AI is a promising emerging technology (Goldfarb et al., 2023). The characteristics of emerging technologies are rapid recent growth, a transition process, economic and market potential, and a science-based background (Cozzens et al., 2010). Emerging technologies are described as not only consisting of radical novelty with a prominent impact but also as containing ambiguity and uncertainty (Rotolo et al., 2015). Thus, for emerging technology, it is difficult to assess what the technology brings given its high uncertainty and unpredictability (van Merkerk & van Lente, 2005).

Considering that AI technology is in the early stage in the economy, it is difficult sufficiently to examine the spillover effects on the economy caused by AI (Brynjolfsson et al., 2021; Hötte et al., 2022). Estimating the impact of the productivity of AI takes time Crafts, 2021; Brynjolfsson et al., 2021). In line with this, to investigate the GPT potential of an emerging technology empirically can only be done in a limited manner. Also, while the theoretical literature and the development of models related to GPTs are active areas, relatively few empirical studies and methods have been introduced (Feldman & Yoon, 2012; Thoma, 2009; Goldfarb et al., 2023). Therefore, this dissertation focuses on the pervasiveness of technological aspects in an effort to understand empirically the potential of AI as a GPT, rather than the spillover effects of AI on the economy. Technological pervasiveness is described as the "pervasiveness of innovative activity across technological classes as well as of a variety of knowledge sources" (Cecere et al., 2014). This dissertation investigates technological pervasiveness by analyzing three different dimensions from the perspectives of knowledge flow diffusion, convergence in the industrial sector, and technology portfolio diversity strategies.

The novelty of this dissertation is that it presents empirical evidence of the technological pervasiveness of the AI sector from a multi-dimensional perspective. First, empirical studies related to AI as a GPT have focused on technological features related to GPTs, such as generality and complementarity. There has been scant attention paid to exploring these technological features while expanding into the industrial or firm level regarding AI technology. Thus, this dissertation empirically investigates these technological features from three different angles: the knowledge flow, industrial sectors, and from a technology portfolio perspective. Second, given that empirical studies related to GPTs have been mainly based on a theoretical concept of the GPTs, few empirical studies that take a wider view have been conducted. Thus, for each analysis level, this dissertation considers three innovation sides on which innovations occur in relation to technological pervasiveness. In this regard, each study constructs a conceptual background: first technology diffusion, second convergence, and third strategy. Consequently, this dissertation empirically examines the technological pervasiveness of AI from a multi-dimensional perspective with different analysis levels and innovation aspects to understand GPTs from a new point of view.

1.2 Research purposes and questions

The purpose of this dissertation is to explore the AI sector by analyzing multi-dimensional perspectives while focusing on technological pervasiveness. This dissertation involves three studies, each empirically investigating the technological pervasiveness of AI by analyzing first knowledge flows, second the industrial sector, and finally technology portfolios.

The first study, which is chapter two of this dissertation, aims to explore the patterns of the GPT-related features of AI depending on technological progress and diffusion. The

GPT-related features here are focused on the process of the recombination and diffusion of technological knowledge, which encompass generality, originality, and complementarity, in order to investigate technological pervasiveness in the diffusion trajectory of the knowledge structure. Specifically, to understand technology diffusion from a new perspective, this study explores the diffusion patterns of individual patents depending on technology progress using a novel methodological approach. Previous studies of technology diffusion have mainly focused on comparing technological features according to differences among technology life cycle stages from a macro point of view. Extending earlier works, this study focuses on a comparison of technological features according to differences among the diffusion levels of patents from a micro point of view. This study allows us to understand technological pervasiveness in terms of knowledge flows while analyzing the dynamics of GPT-related features depending on technological progress and diffusion. The research question for the first study is as follows. What differences exists in the GPT-related features according to the technological progress and diffusion of AI?

The second study, which is chapter three of this dissertation, aims to investigate the horizontal and vertical applicability of AI in the industrial sector. This study proposes a new framework by which to understand the patterns of the applicability of AI in the industrial sector. Technological growth of a GPT can be extended beyond technology itself to the scale of all related industries and must be considered in the industrial sector. Previous research suggests a convergence assumption with a sequential and time-series process of convergence events via scientific convergence, technology convergence, market convergence, and industry convergence in that order (Curran et al., 2010; Sick et al., 2019). The framework of this study examines technology convergence with various industrial sectors regarding technology categories and utilization. Particularly, this study allows us to understand technological pervasiveness by providing empirical evidence with various levels of industries and applications. The research question for the second study is as follows. How can we understand the horizontal and vertical applicability of AI in the industrial sector?

The third study, which is chapter four of this dissertation, considers technological pervasiveness from the perspective of the supply side of technology. For an AI technology to become a GPT, a disruptive innovation should appear in various sectors through a radical change of the existing system. Thus, this study focuses on AI startups, as small firms are more adequate when attempting to understand new disruptive innovations (Hacklin et al., 2005; Kassicieh et al., 2002). Additionally, this study focuses on technology diversity strategies. The diversity of invented technologies can lead to innovations on the supply side through various combinations of technological knowledge. Additionally, technology diversity can implement technological pervasiveness through an expansion of AI technology to a diverse range of other technologies. Consequently, the technology diversity strategy of a startup can contribute to the development of a GPT by causing a disruptive innovation on the supply side of the technology. Meanwhile, investment funding for AI startups is essential for their growth. Therefore, this study aims to examine the relationship between technology diversity and AI startup investments. Through this, the finding suggests that technology diversity in relation to the growth of AI startups can act as a driving force, leading to the disruptive innovation of a GPT in various areas. The research question of the third study is as follows: What are the relationships between technology diversity and investment in AI startups?

1.3 Research outline

As shown in Figure 1-1, this dissertation consists of a total of five chapters. Chapter one presents the research background, purpose, and an overall outline of this dissertation. Chapters two, three, and four present the three aforementioned studies of this dissertation. The three studies here investigate empirical evidence of technological pervasiveness in the AI sector considering different analysis levels and conceptual backgrounds. Chapter five presents the overall conclusion and the implications of this dissertation.

Table 1-1 shows an overview of the three studies in chapters two, three, and four. The analysis levels are the knowledge flow, industrial sector, and technology portfolio in each study. For each analysis level, this dissertation focuses on the diffusion, application, and supply side of AI technology based on each different conceptual background, i.e., technology diffusion, convergence, and strategy. Specifically, for chapter two, this study identifies technological pervasiveness through the knowledge diffusion flow with a patent index, specifically focusing on generality, originality, and complementarity. For chapter three, horizontal and vertical applicability in various industrial sectors are investigated to understand technological pervasiveness. For chapter four, the technology diversity strategies of startups are examined considering the technological pervasiveness of the supply side of technology.

This dissertation uses patent data to investigate AI technology. In particular, this dissertation constructs and analyzes various types of datasets of patents, in particular time series data, network data, text data, and panel data, to investigate the purpose of each study while analyzing the information of patent citations, classification codes, abstracts, patent assignees, and other sources. Because the US is one of the leading countries and given that it has the largest number of patent filings in AI techniques, functional applications, and application fields and is also regarded as an essential market in other jurisdictions (WIPO, 2019a), this dissertation targeted US patents of AI technology.

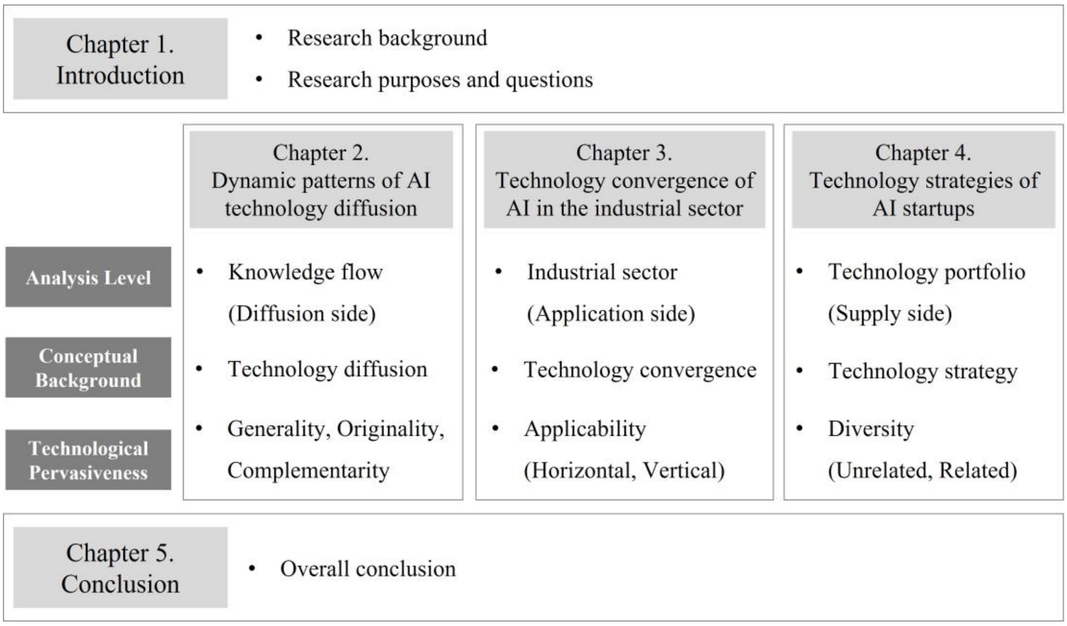

Figure 1-1. Research outline

Table 1-1. Research overview

	Chapter 2	Chapter 3	Chapter 4
Subject	Dynamic patterns of AI technology diffusion: focusing on the patent index	Technology convergence of AI in the industrial sector: insights into a two-way approach	Technology strategies of AI startups: focusing on patent activity and diversity
Research purpose	To explore the patterns of the GPT-related features of AI depending on technological progress and diffusion	To investigate the horizontal and vertical applicability of AI in the industrial sector	To examine the relationship between technology diversity and AI startup investments
Analysis level	Knowledge flow (Diffusion side)	Industrial sector (Application side)	Technology portfolio (Supply side)
Conceptual background	Technology diffusion	Technology convergence	Technology strategy
Technological pervasiveness	Generality, Originality, Complementarity	Applicability (Horizontal, Vertical)	Diversity (Unrelated, Related)
Methodology	- Dynamic time warping - Time series clustering - Analysis of variance	- Network analysis - Clustering analysis	- Multivariate regression analysis - Fixed effect analysis
Data type	- Time series data	- Network data - Text data	- Panel data

Chapter 2. Dynamic patterns of AI technology diffusion: focusing on the patent index

2.1 Introduction

The technology life cycle (TLC) theory broadly explains technological performance according to time or cumulative R&D investment with an S-curve (Gao et al., 2013; Merino, 1990). The two dimensions of the curve are used in relation to various terms and purposes from different perspectives (Taylor & Taylor, 2012). The x-axis of the S-curve, which indicates time or R&D investment, is commonly divided into four stages, as explained below. Ernst (1997) suggested four stages of technological progress, which consist of emergence, growth, maturity, and saturation. Meanwhile, the y-axis of the S-curve, typically indicating performance, was utilized in various approaches in previous studies, which investigated cumulative adoption and technology diffusion, technology improvement performance outcomes, the evolution of patents, and cumulative sales (Taylor & Taylor, 2012; Nieto et al., 1998; Dosi, 1982; Sahal, 1985; Andersen, 1999; Debackere et al., 2002). Generally, performance, which is represented on the y-axis, shows heterogeneous growth depending on time or effort, which appears on the x-axis. Performance shows relatively different outcomes in each stage, slowly increasing in the emerging stage, showing positive marginal progress in the growth stage, negative marginal progress in the maturity stage, and slight progress in the saturation stage (Ernst, 1997). The transition from the emerging to the growth stage requires a large increase in performance, and disruptive innovations occur during this phase, whereas incremental innovation generally occurs from the growth stage to the maturity stage (Hacklin et al., 2005).

To investigate the S-curve, patent data is commonly tracked as an indicator (Gao et al., 2013; Ernst, 1997). Patent data show the evolutionary trends and dynamics of technological growth (Taylor & Taylor, 2012; Haupt et al., 2007). Patent data predict technological success based on the TLC, the diffusion and expansion potential (Altuntas et al., 2015). Extant literature has examined curve-fitting techniques in a relation to patent data to investigate technological progress and performance in terms of the TLC (Lee et al., 2016a).

With regards technology diffusion, patent citations are widely used (Lee et al., 2018; Cheng, 2012). The number of forward citations is used as a measure of technology diffusion, and to trace the spread of technology (Lee et al., 2018; Chang et al., 2007). Numerous studies of patent citations and technology diffusion consider citation networks in their analysis of diffusion trajectories and knowledge transfer process (Chang et al., 2007; You et al., 2017; Huang et al., 2022). The Bass diffusion model has been widely adopted to explain technology diffusion using patent citations (Lee et al., 2018; Cheng, 2012).

In TLC research, patent data are widely used to observe the different characteristics of the TLC stage (Haupt et al., 2007; Gao et al., 2013; Su, 2018). Gao et al. (2013) investigated development trends of thirteen patent indicators to understand different patterns of indicators according to the emerging, growth, maturity, and declining stages. Haupt et al. (2007) divided the life cycle stage into the subcategories of introduction, growth, and maturity to investigate the characteristics of transitions among each stage with patent indicators, discussing backward citations, forward citations, dependent claims, and priorities, among other factors. These findings suggested the significant patent indicators during the transitions among each life cycle stage (Haupt et al., 2007). Su (2018) analyzed the dynamics of patent characteristics to investigate increasing and decreasing trends depending on the TLC.

However, previous studies of the TLC have focused on patent indicators to examine differences among the stages of the TLC while downplaying differences among the patents. In addition, from diffusion studies, patents have been widely used to investigate cumulative diffusion curves or trends in an effort to make technology development forecasts but not to identify differences in the diffusion levels of patents. Thus, there has been scant attention paid to the growth trajectories and trends of different patents and to their diffusion trends according to each patent's characteristics.

To fill this research gap, this study proposes a different angle by which to understand technology growth and the diffusion of recent AI technology. First, this study examines the life cycle and diffusion trajectory of each patent from a micro perspective, not focusing on the TLC stage of the entire technology field. The S-curve commonly targets the accumulated data in a specific technology field, but this curve can be used to consider each single patent trajectory as well. To investigate each single patent trajectory, an exploratory analysis is conducted to find the dynamic diffusion pattern of each patent considering the diffusion level. Second, this study identifies significant patent indices according to both the diffusion level and time. Specifically, this study considers patent indices focusing on the properties of technological pervasiveness to understand the potential of AI to be a general-purpose technology (GPT).

The purpose of this study is to investigate the patterns of the GPT-related features of AI depending on technology progress and diffusion using a patent index of technological pervasiveness. This study proposes following specific research questions. First, what differences exists in the patent indices of technological pervasiveness according to the technological progress phase of AI? Second, what differences exists in the patent indices of technological pervasiveness according to the diffusion level of AI? To this end, this study investigates the diffusion trajectories of each patent while also considering the development trend of technology over time.

Meanwhile, this study utilizes the dynamic time warping (DTW) algorithm and time series clustering to measure pattern similarity levels. DTW, which can clarify similarities among sequential and dynamic time series data, is widely used in the areas of voice recognition (Rabiner & Juang, 1993), motion recognition (Switonski et al., 2019), anomaly detection (Jun, 2011), biological pattern recognition (Cavill et al., 2013), stock market predictions (Zhao et al., 2021; Kim et al., 2018), and computational economics (Franses & Wiemann, 2020). The DTW algorithm has received little attention in the TLC research area and its use to investigate technology growth patterns has rarely been attempted. This approach is a novel way to understand technology diffusion, and the outcomes here present implications from the perspective of an empirical analysis of the research on the TLC and diffusion.

The remainder of this paper is structured as follows. Section 2.2 presents the literature review. Section 2.3 describes the proposed method and research framework. Section 2.4 shows the analysis and results of the study. Section 2.5 presents the discussion and conclusion and suggests future research.

2.2 Literature review

2.2.1 Patent index of technological pervasiveness

The patent index is used to measure technological or economic quality and value (Squicciarini et al., 2013). This study focuses on the properties of technological pervasiveness among various patent indices. To understand the knowledge flow of patents, the information in the patent citation is useful (Lee et al., 2016b). Therefore, the patent index in this study is based on patent citations for consideration of the diffusion process of knowledge in a wide technological scope. Specifically, this study investigates the technology scopes of forward, backward citations, and an originating patent as well.

Generality refers to the extent of how forward citations technology categories spread across various technology areas (Trajtenberg et al., 1997). Generality is what degree forward citations technology categories are spread various technological fields. The generality index has been used to examine GPTs (Hall & Trajtenberg, 2004). GPTs enable the development of subsequent technologies, having a great impact and causing changes in a variety of other technologies and industries (Graf & Menter, 2022).

Originality is defined as the range of the technology area on which a patent depends (Squicciarini et al., 2013). Greater originality means broader technological sources and roots and the integration of diverse knowledge (Trajtenberg et al., 1997). Various studies have used the originality index to measure patent quality and value (Valentini, 2012; Raiteri, 2018; Falk & Train, 2017; Graf & Menter, 2022), and Moser and Nicholas (2004) investigated originality as a characteristic of GPTs. When a patent is built on a diversity of technological bases, the possibility to recombine new knowledge increases (Martinelli et al. 2021). Originality drives generality (Trajtenberg et al., 1992), and originality has a positive correlation with technological dynamism, which refers to the potential to increase efficiency, another GPT characteristic (Martinelli et al., 2021).

Complementarity in this study is measured according to the patent scope. The patent scope as proposed by Lerner (1994) has been developed to measure the technological breadth of a patent based on the patent classification code. A GPT requires a range of different technologies to co-occur, and this co-occurrence is used to measure the innovation complementarity aspects of a technology, as variety of different technologies allows recombinations and improves and helps develop existing technologies and products (Petralia, 2020). When a patent is classified into various categories of technology, the patent requires a variety of inventions from different technological areas in relation to its process or product aspects (Petralia, 2020). This study defines complementarity by assessing the scope of technology of the originating patent.

Meanwhile, to understand the pace of technological progress, this study examines the technology cycle time (TCT). The notion of pace in relation to technological progress was suggested by Ayres (1994), and an indicator of TCT was developed to understand technological progress based on a distance measure. The TCT is defined as the median age of backward citations with respect to the prior of the invention (Kayal, 1999). In line with this, shorter TCTs are considered to indicate fast technological progress, whereas longer TCTs indicate a slow pace of technology progress (Kayal, 1999; Ayres, 1994).

2.2.2 Dynamic time warping

Dynamic time warping (DTW) seeks to find an optimal distance between given sequential time series data (Müller, 2007). The Euclidian distance is simple method to measure the complexity of linear time, whereas it is not adequate when the lengths of the time series are not equal (Li, 2015; Petitjean et al., 2011). Thus, DTW is proposed to measure the similarities between unequal lengths of sequential data.

If there are two sequential time series data regarding the diffusion trajectory of patents A = (a_1, a_2, ..., a_m) and B = (b_1, b_2, ..., b_n), an m by n matrix is devised to compare the time series of lengths m and n. Figure 2-1 shows dynamic time warping in relation to patent data. The distance between two points a_i and b_j is calculated as Equation (1) (Müller, 2007; Li, 2015).

$$d(a_i, b_j) = (a_i - b_j)^2 \quad \text{Eq.(1)}$$

Then, the cumulative distance is calculated while minimizing the distance between the two time series. The equation of DTW can be defined as Equation (2) (Müller, 2007; Li, 2015).

$$D\ (i,j) = d(a_i,\ b_j) + min \begin{Bmatrix} D(i,j-1) \\ D(i-1,j-1) \\ D(i-1,j) \end{Bmatrix} \quad \text{Eq.}(2)$$

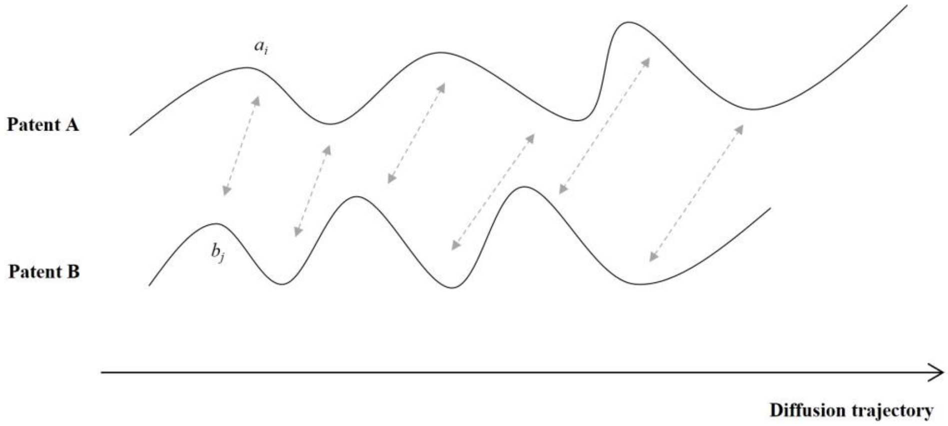

Figure 2-1. Dynamic time warping and patent data

After the similarity is measured based on the DTW, clustering analysis can be conducted. In particular, K-means time series clustering has been widely used to analyze time series data (Liao et al., 2006; Hautamaki et al., 2008; Warren Liao, 2005).

Regarding the patent data in this study, there are different lengths of the diffusion trajectory in each patent. In addition, there are different patterns by year despite the same total diffusion level in each patent. If a patent has innovative value, it would undergo explosive diffusion in a short time. Therefore, to resolve the differences in the lengths and patterns among the diffusion of patents, the DTW algorithm is used.

2.3 Proposed method

2.3.1 Research framework

Figure 2-2 shows the research procedure of this study, and following sub sections provide a detailed explanation. Section 2.3.2 describes data collection and preprocessing. Section 2.3.3 shows the analysis of the time series data including the methods of DTW and time series clustering. Section 2.3.4 describes the calculation of patent indices. Section 2.3.5 briefly describes the method of analysis of variance. Finally, Section 2.4 presents the analysis and results of this study to identify patterns of technology diffusion.

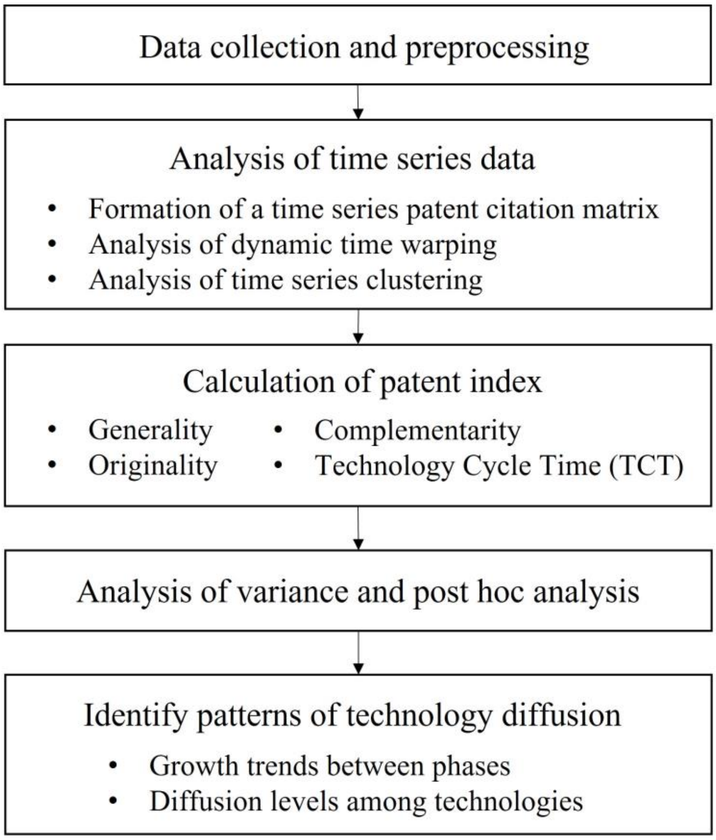

Figure 2-2. Research procedure

2.3.2 Data collection

According to World Intellectual Property Organization (WIPO) (2019a), AI has been categorized into three broad areas: AI techniques, AI functional applications, and AI application fields. Specifically, the AI techniques have been categorized into five sub-areas: fuzzy logic, logic programming, machine learning, ontology engineering, and probabilistic reasoning (WIPO, 2019a). Also, the AI functional applications have been categorized into nine sub-areas: computer vision, control methods, distributed artificial intelligence, knowledge representation/reasoning, natural language processing, planning/scheduling, predictive analytics, robotics, and speech processing (WIPO, 2019a). Furthermore, in this AI categorization, WIPO (2022) presents Cooperative Patent Classification (CPC) and International Patent Classification (IPC) codes which can be matched to the AI category.

The AI patents in this study were based on CPC codes presented by WIPO (2022), as indicated in Appendix A1-1. This study used AI patent as a result of combining the CPC codes and the terms ‘artificial intelligence’. This study does not focus on investigating the entire TLC of the AI technology sector but rather examines the diffusion paths of individual patents from a microscope point of view. Therefore, this study analyzed the diffusion trends of technology of the recent decade from US patents.

The patent data in this study were collected through the United States Patent and Trademark Office (USPTO) Bulk Data and Google Patent Search. As a result of the data collection efforts here, 520,613 patents in total were collected in September of 2022, consisting of 6,852 originating patents, 343,975 backward citation patents, and 169,786 forward citation patents. The originating patents targeted granted patents during the period of 2011 to 2020 for the grant dates. To understand AI technology progress over time, the collected originating patents were divided into phase 1 for the first five years and phase 2 for the five subsequent years. In particular, in 2016, the AlphaGo event occurred, and the division of phases is adequate to understand the changes between the two phases. Finally, a total of 152,603 patents including 1,228 originating patents were the subject of phase 1, and a total of 368,010 patents including 5,624 originating patents were collected in phase 2. Table 2-1 shows a summary of the datasets of each phase in this study.

Table 2-1. Datasets

	All patents	Phase 1	Phase 2
Grant year (originating patents)	From 2011 to 2020	From 2011 to 2015	From 2016 to 2020
Number of originating patents	6,852	1,228	5,624
Number of backward citations	343,975	77,460	266,515
Number of forward citations	169,786	73,915	95,871
Total patents	520,613	152,603	368,010

2.3.3 Analysis of time series data

The first step was the formation of a patent citation matrix. In order to measure the diffusion of individual patents by year as a time series dataset, a matrix was created by aggregating forward citation patents by year for each originating patent. The value of the matrix was determined only by the corresponding year as opposed to a cumulative value, as all of the cumulative values showed an upward pattern and it was difficult to find a trend among the patents. The formation of patent citation matrix was referred to in prior research (Lee et al., 2017), and to extend this line of research, the matrix in this study additionally considered family to family citations of US patents as well, which can reflect the value and influence of a patent better. The matrix generated in this study is shown in Table 2-2, and patents with the same grant year can have different lengths of the diffusion year.

Table 2-2. Patent citation matrix

Patent no	Grant year	Year 1	Year 2	....	…	…	…
P1	GY1	C (P1, Y1)	C (P1, Y2)	…	…	…	C (P2, Yt)
P2	GY2	C (P2, Y1)	C (P2, Y2)	…	…	C (P2, Yt)	
P3	GY3	C (P3, Y1)	C (P3, Y2)	…	C (P3, Yt)		
…	…	…	…	…			

Table 2-3 shows descriptive statistics of the forward citations and the lengths of phases 1 and 2.

Table 2-3. Descriptive statistics of forward citations and lengths

		Phase 1	Phase 2
Forward citations	Mean	56.984528	14.480085
	SD	111.469377	45.528164
	Min	0	0
	Max	1414	1429
lengths	Mean	11.186482	4.104018
	SD	3.791241	2.769019
	Min	7	2
	Max	28	25

The second step was the analysis of DTW. The DTW was applied to measure the similarity between patents with different lengths for each phase, and as a result, a pairwise distance matrix was generated.

The third step was a time series clustering analysis. Based on the pairwise distance matrix, K-means time series clustering was conducted to derive the number of clusters considering the silhouette score. Table 2-4 shows the results of the clustering analysis and the silhouette score. As shown in Table 2-4, this study sets the number of clusters to four because there were outliers in the dataset used. The silhouette score is an index for evaluating the number of clusters comparing the similarity within the cluster and with adjacent clusters for each data. The silhouette score has a value of -1 to 1, and for this score, the higher the value, the better the result of clustering. In this study, the silhouette score of each phase showed a high value between 0.6 and 0.8, indicating that clustering was performed well.

Table 2-4. Results of clustering analysis and silhouette score

	Total Patents	Number of cluster	Silhouette score	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Phase1	1228	4	0.6793	985	44	2	197
				80.21%	3.58%	0.16%	16.04%
Phase2	5624	4	0.7398	4988	88	2	546
				88.69%	1.56%	0.04%	9.71%

2.3.4 Calculation of the patent index

The patent index used here was based on backward and forward citations of an originating patent to consider the diffusion process of knowledge from a wide technological scope.

As shown in Figure 2-3, originality was based on the scope of backward citations compared to the originating patent, and generality was based on the scope of forward citations compared to the originating patent. Complementarity was based on the scope of the originating technology itself. Additionally, to understand the pace of technological progress, TCT was based on the time lags between the backward citations and the originating patent. The technology scope was based on the IPC codes of each patent.

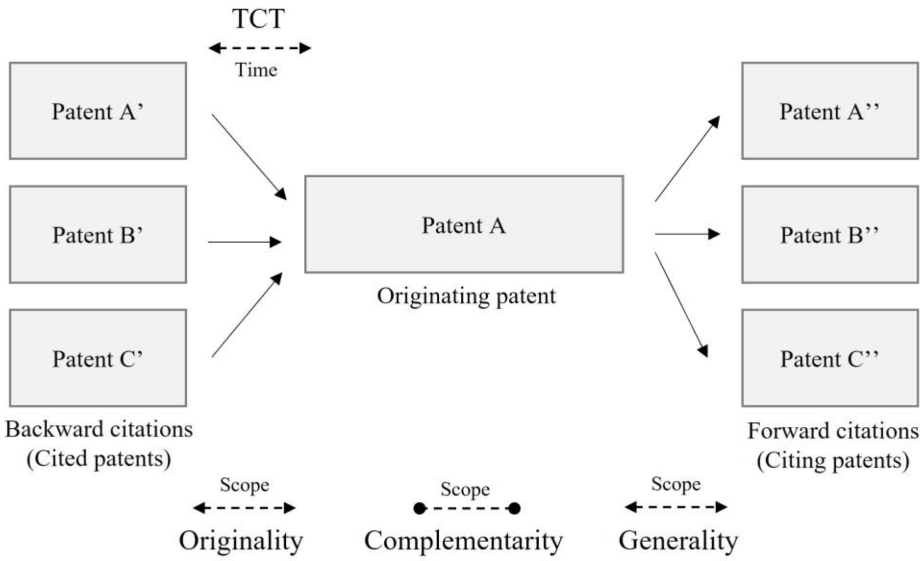

Figure 2-3. Patent index of technological pervasiveness

The generality index in this study referred to earlier work by Trajtenberg (1997). This index is based on the Hirsman-Herfindahl index (HHI) and considers the number and distribution of technology classes of forward citations belongs (Squicciarini et al., 2013). The technology class in this study was based on the 4-digit IPC. The values used in the generality index are between 0 and 1, and a higher value represents more generality. The calculation of the generality index is expressed as follows using Equation (3), referring to the work of Trajtenberg (1997).

$$Generality_i = 1 - \sum_{k=1}^{n} \left(\frac{Citing_{ik}}{Citing_i} \right)^2 \quad \text{Eq.(3)}$$

Where k is the index of the technology class of the citing patents and n is the number of different technology classes of the citing patents.

The originality index in this study was defined similarly to the calculation of the generality index, but the difference was that the technology class of backward citations cited by the originating patent was targeted, as opposed to forward citations citing the originating patent. The calculation of the originality index is determined using Equation (4) (Trajtenberg, 1997).

$$Originality_i = 1 - \sum_{h=1}^{n} \left(\frac{Cited_{ih}}{Cited_i}\right)^2 \quad \text{Eq. (4)}$$

Complementarity in this study followed the measurement of the patent scope proposed by the work of Lerner (1994) and Squicciarini et al. (2013), based on the number of 4-digit IPC classes of the patent. A larger value of the patent scope denotes a broader scope and a higher value of the technological and market potential (Squicciarini et al., 2013). Complementarity is defined as follows using Equation (5), referring to the patent scope proposed by Squicciarini et al. (2013).

$$Complementarity_i = n_i; n \in \{IPC4_1;\ IPC4_2; \dots; IPC4_p;\ IPC4_q; \dots;\ IPC4_n\} \,\&\, IPC4_p \neq IPC4_q \quad \text{Eq. (5)}$$

Where n_i is the number of 4-digit distinct IPC codes of the patent *i*.

The technology cycle time (TCT) here followed Kayal (1999) and was based on the median age of the cited patents. A smaller value of TCT reflects faster progress of the technology. In this study, TCT is determined using Equation (6).

$$TCT_i = Med(\Delta t) \quad \text{Eq. (6)}$$

Where Δt represents the number of years between the grant date of the originating patent *i* and the backward citations, and $Med(\Delta t)$ is the median value of Δt.

2.3.5 Analysis of variance and post hoc analysis

A one-way ANOVA is based on the assumptions of normality, homogeneity of variance, and independence. In this study, the requirement of independence of the clusters was basically satisfied among the clusters, and tests of normality and homogeneity of variance of the patent indices of the clusters were conducted. The Kolmogorov-Smirnov test and the Shapiro-Wilk test were used to test the degree of normality. The homogeneity of variance was tested with the Levene and Bartlett tests. The one-way ANOVA was used when the homogeneity of variance requirement was met, whereas Welch's ANOVA was applied when heterogeneity of the variance was found (Liu, 2015). In a post-hoc analysis, a Bonferroni analysis was conducted in cases showing homogeneity of the variance, whereas the Games-Howell post-hoc test was performed in cases showing heterogeneity of the variance (Shingala & Rajyaguru, 2015).

2.4 Analysis and results

2.4.1 Growth trends between phases

Descriptive statistics pertaining to the patent index for phases 1 and 2 are shown in Table 2-5.

Table 2-5. Descriptive statistics of phases 1 and 2

Phase 1					
	Obs	Mean	SD	Min	Max
Generality	1228	0.6293	0.23066	0	0.948136
Originality	1228	0.590398	0.249249	0	0.941473
Complementarity	1228	1.832248	1.341897	1	16
TCT	1228	7.681189	3.820273	0	46
Phase 2					
	Obs	Mean	SD	Min	Max
Generality	5624	0.453367	0.31194	0	0.949869
Originality	5624	0.607071	0.227077	0	0.949861
Complementarity	5624	2.837304	1.485368	1	17
TCT	5624	6.095484	3.599164	0	24.5

For the normality test of each index for each phase, the Kolmogorov-Smirnov test and the Shapiro-Wilk test were conducted, as noted above, and all four indices of phases 1 and 2 satisfied the normality requirement. In addition, for the homogeneity of variance test, the Levene and Bartlett tests were performed, and it was confirmed that all four indices of phases 1 and 2 showed heteroscedasticity (see Appendix A1-3, #1). Therefore, Welch's T-test was used to investigate the differences in the indices between the two phases. The results of Welch's T-test are shown in Table 2-6.

Table 2-6. Analysis between phases

	t-statistic	*P*-value
Generality	22.59501	2.41E-102
Originality	2.156779	0.03116271
Complementarity	23.31259	3.47E-106
TCT	13.31252	1.43E-38

The result showed that the differences in the generality, originality, complementarity, and TCT were statistically significant between the two phases. Originality and complementarity were higher in phase 2, whereas TCT is shorter in phase 2. Meanwhile, generality was lower in phase 2. Because generality is affected by the number of forward citations, previously developed technologies in phase 1 showed higher generality than those in phase 2.

2.4.2 Diffusion levels among technologies

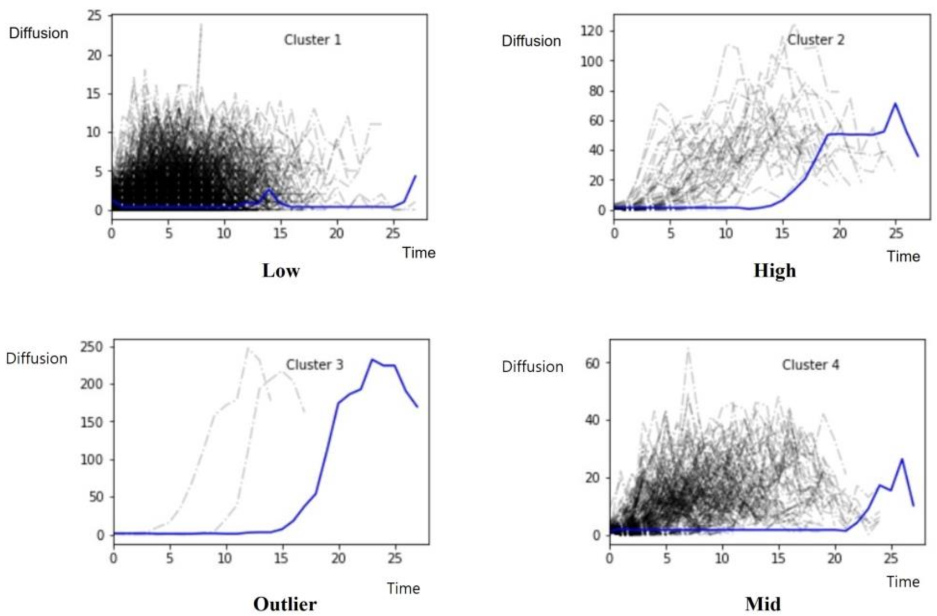

Figure 2-4. Time series clustering in phase 1

The first part of this section described the analysis of the diffusion levels among technologies in phase 1. The results of the time series clustering for patents in phase 1 are divided into four clusters, as shown in Table 2-4. Also, Figure 2-4 shows each cluster, indicating that clusters 1, 2, and 4 are defined as low, high, and mid-level diffusion clusters, respectively. This study analyzed three clusters, except for cluster #3, which was an outlier of the patent dataset.

As a result of the normality test and the homogeneity of variance test of four indices in each cluster, the requirement of normality was met in all cases. Meanwhile, heteroscedasticity was found in the generality, originality, and complementarity assessments (see Appendix A1-3, #2). Thus, Welch's ANOVA analysis and the Games-Howell post-hoc analysis were conducted for these three indices, whereas a one-way ANOVA and a Bonferroni post-hoc analysis were conducted for TCT.

Table 2-7 shows the results of the ANOVA for phase 1, and the Table 2-8 shows the results of the post-hoc analysis for phase 1. Generality was statistically significant, showing a higher value in the high and mid-diffusion-level clusters compared to the low-level cluster. Originality showed statistically significantly higher values in the order of high, mid, and low levels. On the other hand, there were no statistically significant differences among the clusters for the complementarity and TCT. Consequently, technologies with high level of diffusion showed high generality and originality in phase 1.

Table 2-7. Analysis of variance in phase 1

	F-value	*P*-value
Generality	60.17918	1.020421E-18
Originality	35.6587	7.250329E-13
Complementarity	1.087902	0.340838
TCT	2.049213	0.129278

Table 2-8. Post-hoc analysis in phase 1

Games Howell								
	A	B	Mean(A)	Mean(B)	Diff	SE	t-statistic	P-value
Generality	1 (L)	2 (H)	0.6026755	0.759209	-0.15653	0.024159	-6.47938	9.22E-08
	1 (L)	4 (M)	0.6026755	0.731235	-0.12856	0.012928	-9.94437	0
	2 (H)	4 (M)	0.7592087	0.731235	0.027973	0.025184	1.110778	0.511109
Originality	1 (L)	2 (H)	0.5688679	0.739103	-0.17024	0.023708	-7.18038	5.55E-09
	1 (L)	4 (M)	0.5688679	0.662626	-0.09376	0.016624	-5.63996	1.09E-07
	2 (H)	4 (M)	0.7391029	0.662626	0.076477	0.026553	2.880131	0.013837
Complementarity	1 (L)	2 (H)	1.7959391	2.363636	-0.5677	0.393571	-1.44243	0.328419
	1 (L)	4 (M)	1.7959391	1.832487	-0.03655	0.09361	-0.39043	0.91944
	2 (H)	4 (M)	2.3636364	1.832487	0.531149	0.400763	1.325344	0.388415
Bonferroni								
	A	B	Mean(A)	Mean(B)	Diff	SE	t-statistic	P-value
TCT	1 (L)	2 (H)	7.5807107	8.272727	-0.69202	0.573969	-1.1661	0.2438
	1 (L)	4 (M)	7.5807107	8.098985	-0.51827	0.285713	-1.739	0.0823
	2 (H)	4 (M)	8.2727273	8.098985	0.173743	0.617144	0.2864	0.7748

In the second part of this section, this study analyzed the diffusion levels among the technologies in phase 2. The results of the time series clustering of phase 2 are described below. Figure 2-5 presents the diffusion patterns for each cluster, where clusters 1, 2, and 4 are defined as the low, high, and mid-level diffusion clusters, respectively, omitting the outlier cluster #3.

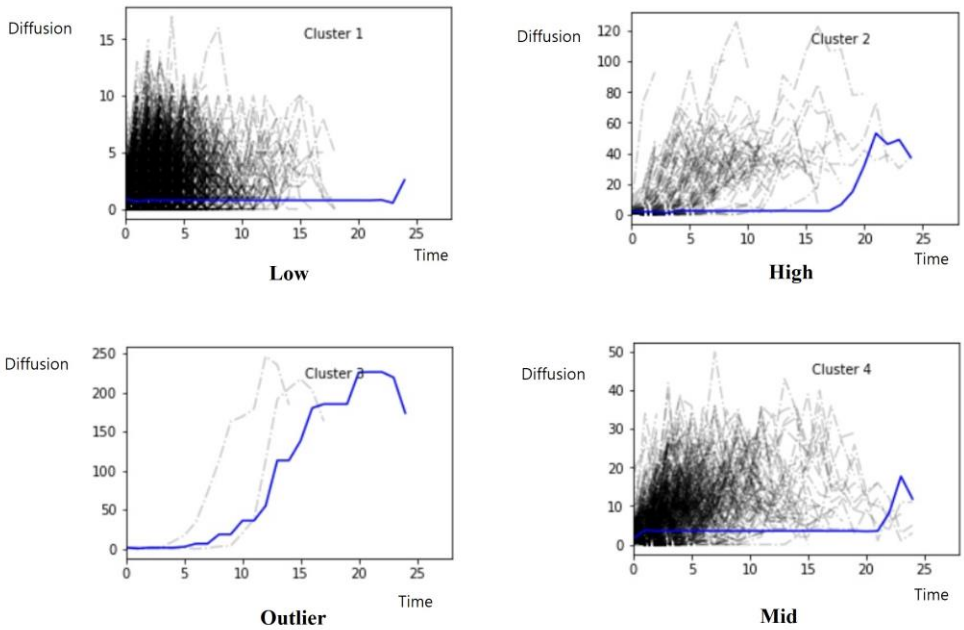

Figure 2-5. Time series clustering in phase 2

Table 2-9. Analysis of variance in phase 2

	F-value	*P*-value
Generality	774.60045	3.181060E-105
Originality	56.550777	1.579680E-20
Complementarity	21.4362	3.546997E-09
TCT	17.43857	1.01E-07

Table 2-10. Post-hoc analysis in phase 2

Games Howell								
	A	B	Mean(A)	Mean(B)	Diff	SE	t-statistic	P-value
Generality	1 (L)	2 (H)	0.419335	0.752299	-0.33296	0.01525	-21.8331	9.99E-15
	1 (L)	4 (M)	0.419335	0.714661	-0.29533	0.008256	-35.7723	1.91E-13
	2 (H)	4 (M)	0.752299	0.714661	0.037637	0.016191	2.32453	0.055937
Originality	1 (L)	2 (H)	0.597968	0.729582	-0.13161	0.016536	-7.95911	1.16E-11
	1 (L)	4 (M)	0.597968	0.669715	-0.07175	0.009446	-7.59577	7.83E-13
	2 (H)	4 (M)	0.729582	0.669715	0.059867	0.018486	3.238549	0.004229
Complementarity	1 (L)	2 (H)	2.773657	3.965909	-1.19225	0.306615	-3.88843	0.000569
	1 (L)	4 (M)	2.773657	3.239927	-0.46627	0.087397	-5.33511	4.05E-07
	2 (H)	4 (M)	3.965909	3.239927	0.725982	0.317652	2.285465	0.062444
Bonferroni								
	A	B	Mean(A)	Mean(B)	Diff	SE	t-statistic	P-value
TCT	1 (L)	2 (H)	5.99579	7.704545	-1.70876	0.401854	-4.4516	0
	1 (L)	4 (M)	5.99579	6.745421	-0.74963	0.168433	-4.6382	0
	2 (H)	4 (M)	7.704545	6.745421	0.959124	0.429833	2.225	0.0264

The normality test and the homogeneity of variance test of four indices in each cluster were conducted (see Appendix A1-3, #3). As a result, Welch's ANOVA analysis and the Games-Howell post-hoc analysis were used to assess generality, originality, and complementarity, while a one-way ANOVA and the Bonferroni post-hoc analysis were used for TCT.

Table 2-9 shows the results of the ANOVA analysis of phase 2, and the Table 2-10 shows the results of the post-hoc analysis of phase 2. As a result of these analyses, differences in the diffusion levels showed statistically significant higher values in the order of the high, mid, and low levels in all four indices, i.e., generality, originality, complementarity, and TCT.

For the third part of this section, an additional analysis was conducted considering the technological properties constituting each cluster. Considering the properties of AI technology, AI patents can be divided into two types: patents of only AI technology and patents that encompass AI application fields. This study divided the patents in each cluster into the AI-only and AI-application types. Among all AI patents, patents containing CPC and IPC codes of AI application fields were classified as AI-application types (see Appendix A1-2), and patents that do not contain these codes were classified as AI-only types.

The results of the analysis are given below. Table 2-11 shows the mean of the patent index according to the AI-only and AI-application types. The normality test and homogeneity of variance test were conducted for each cluster (see Appendix A1-3, from

#4 to #9). Table 2-12 shows the results of ANOVA on all clusters between AI-only and AI-application by phase 1 and 2. Table 2-12 shows the results of the ANOVA between the AI-only and AI-application types for phases 1 and 2 on all clusters. As shown in Table 2-11, the results for each cluster indicated that the patent indices of generality, originality, and complementarity were relatively high in the AI-application types compared to the AI-only types. From the results of the ANOVA shown in Table 2-12, only the high cluster in phase 1 was statistically significant for generality and originality, while for the remaining five clusters, generality, originality, and complementarity are statistically significant. In contrast, for TCT, the ANOVA results showed statistical significance only for the mid-level cluster in phase 1.

Table 2-11. Descriptive statistics of AI-only and AI-application types

Cluster		Type	Obs	Mean(*G*)	Mean(*O*)	Mean(*C*)	Mean(*T*)
Phase 1	H	AI-only	11	0.654	0.644	2.091	8.000
		AI-application	33	0.794	0.771	2.455	8.364
	M	AI-only	80	0.681	0.589	1.525	7.244
		AI-application	117	0.766	0.713	2.043	8.684
	L	AI-only	492	0.568	0.529	1.573	7.392
		AI-application	493	0.638	0.609	2.018	7.769
Phase 2	H	AI-only	23	0.663	0.621	2.522	7.543
		AI-application	65	0.784	0.768	4.477	7.762
	M	AI-only	182	0.631	0.566	2.220	6.863
		AI-application	364	0.757	0.722	3.750	6.687
	L	AI-only	2039	0.371	0.542	2.206	6.049
		AI-application	2949	0.453	0.637	3.166	5.959

Note. *G*: Generality, *O*: Originality, *C*: Complementarity, *T*: TCT

Table 2-12. Analysis of variance between AI-only and AI-application types

			t-statistic	*P*-value
Phase 1	H	Generality	2.857613991	0.006613563
		Originality	2.628947643	0.011914537
		Complementarity	0.398108736	0.692566551
		TCT	0.277816911	0.782515954
	M	Generality	4.173300292	4.52E-05
		Originality	4.130725482	6.30E-05
		Complementarity	3.327114859	0.001050825
		TCT	2.788774071	0.005815156
	L	Generality	4.630651811	4.13411E-06
		Originality	4.956046823	8.48843E-07
		Complementarity	5.795392825	9.86627E-09
		TCT	1.532803026	0.125646151
Phase 2	H	Generality	3.947221236	0.000161122
		Originality	4.374952381	3.39E-05
		Complementarity	2.926897137	0.004379267
		TCT	0.239021108	0.811657902
	M	Generality	8.133910588	1.48696E-14
		Originality	7.845718251	1.01945E-13
		Complementarity	10.7647193	1.36075E-24
		TCT	0.515472518	0.606432279
	L	Generality	9.257043576	3.05034E-20
		Originality	14.41995364	5.23851E-46
		Complementarity	27.89277938	3.2747E-159
		TCT	0.876883915	0.38059194

2.5 Discussion and conclusion

The study investigates how patent indices regarding technological pervasiveness differ depending on technology progress and diffusion in the AI sector. It is found that the GPT-related features show a further increase according to the progress of AI technologies, with GPT-related features find to be higher in AI technologies with high diffusion levels. Consequently, this study presents empirical evidence that the GPT-related features of AI are related to the growth of technology in terms of both the growth trend and diffusion level. Table 2-13 shows a summary of the results of this study.

The results for generality, originality, and complementarity, which are related to a wide range of technological pervasiveness as a GPT, are as follows. First, depending on the progress of AI technologies, the GPT-related features among the AI technologies overall show increased statistical significance. When comparing the technology progress between phase 1 and phase 2, originality and complementarity increase in phase 2, which means that the GPT-related features increase more significantly depending on the technological progress. From these results, this study finds that the direction of AI development is improving the potential to increase knowledge recombination, efficiency, and variety. Second, this study provides evidence that AI technologies with a wide range of technological pervasiveness in the AI technology field show higher diffusion levels. Generality, indicating a wide scope of follow-on inventions, and originality, indicating a diverse technology base for recombinations, are higher for technologies with high diffusion levels in both phase 1 and phase 2. Also, complementarity is higher for technologies with high diffusion levels in phase 2. The diffusion level of patents is related to indirect aspects regarding the scope of forward and backward citations in phase 1; complementarity, which is a direct aspect pertaining to the scope of the originating patents, additionally appears in phase 2.

The results for TCT, related to technological pace, show that it is becoming faster. For technologies overall, TCT is decreasing in phase 2 compared to phase 1. The average value of TCT was 7.68 in phase 1, while it was 6.09 in phase 2. This shows that the technology distance is shortened and that the technology development pace is faster. Additionally, the total amount of technology amounted to 1,228 in phase 1, while this score was 5,624 in phase 2, confirming that the growth of the technology is high. Meanwhile, TCT can be interpreted from the perspective of the longevity of GPTs. In the context of GPTs, high longevity is one of the defining properties. Commonly, the citation lag measured by forward citations of GPT technologies is longer than those of average technologies (Hall & Trajtenberg, 2004; Moser & Nicholas 2004; Feldman & Yoon, 2012). However, this study investigates emerging technologies, for which not enough time has passed to observe forward citations. Thus, this study considers longevity of technology based on backward citations, and the speed of technological development, as defined above. Hence, the results for TCT, related to the time lag of technology as GPTs, are as follows. GPT technologies commonly have a long time lag, and the results of this study show that technologies with a long time lag are advantageous for diffusion. In other words, it is found that technologies citing relatively older previous technology appear to have a high level of diffusion, indicating that high technology longevity is maintained. According to the analysis of phase 2, a longer time lag shows more forward citations and higher diffusion of AI technologies.

In line with this, considering the length of the longevity of GPTs and generality as an index affected by the number of forward citations, relatively older technology corresponding to phase 1 shows higher generality compared to that in phase 2.

Table 2-13. Summary of the results

<table>
<tr><th>Cluster</th><th colspan="2">Comparison results</th></tr>
<tr><td>Phase 2
(Compared to Phase 1)</td><td colspan="2">Generality -
Originality +
Complementarity +
TCT -</td></tr>
<tr><td></td><td>Phase 1</td><td>Phase 2</td></tr>
<tr><td>High
(Compared to Mid)</td><td>Originality +</td><td>Generality +
Originality +
Complementarity +
TCT +</td></tr>
<tr><td>High
(Compared to Low)</td><td>Generality +
Originality +</td><td>Generality +
Originality +
Complementarity +
TCT +</td></tr>
<tr><td>Mid
(Compared to Low)</td><td>Generality +
Originality +</td><td>Generality +
Originality +
Complementarity +
TCT +</td></tr>
</table>

Additionally, the results of analyzing AI patents after dividing them into AI-only and AI-application types are shown in Table 2-14. Table 2-14 shows a summary of the results between the AI-only and AI-application types. First, in each cluster for phases 1 and 2, AI-application types show higher generality, originality, and complementarity than AI-only types. In other words, it is found that AI-application types utilized a wider range of knowledge, included a greater variety of knowledge, and spread to a wider range compared to AI-only types. This result suggests that AI-application types show a higher value of GPT-related features than AI-only types, regardless of the diffusion level. Second, TCT shows an insignificant statistical difference between AI-only and AI-application types. That is, a difference in the speed of technology development between the two categories is not found. Also, it is found that rapid development is progressing for both AI-only and AI-application types in phase 2 compared to that in phase 1.

Table 2-14. Summary of the results between AI-only and AI-application types

Cluster	Comparison results		
		Phase 1	Phase 2
AI-application (Compared to AI-only)	High	Generality + Originality +	Generality + Originality + Complementarity +
	Mid	Generality + Originality + Complementarity + TCT +	Generality + Originality + Complementarity +
	Low	Generality + Originality + Complementarity +	Generality + Originality + Complementarity +

The theoretical implication of this study is that dynamic patterns of the patent index to explain the features of GPT technology are identified in terms of the technology diffusion level and progress. Through these results, empirical evidence of the technological pervasiveness of AI technology is presented. Additionally, the methodological implication of this study is the importance of the adoption of the dynamic time wrapping analysis for technology diffusion and innovation research. Specifically, the analysis of the growth patterns of individual units of technology here provides new insight into understanding patent citation and diffusion trends.

The results of this study show that AI technologies co-evolve with various industries along with technological development and suggest that AI has potential for growth as a GPT. However, legal and institutional policies to prepare for the development of AI as a GPT are insufficient, especially specific policies for coordinating issues among AI and related industries. Oxford Insight (2020; 2021; 2022) has measured AI government readiness indexes around the world, and in the Republic of Korea, the value of the government pillar was found to be lower than those of the United States, Singapore, and Finland. Thus, it is recommended to supplement and reestablish these policies in the Republic of Korea by referring to the law and systems in other relevant countries. Meanwhile, Nieto Council for Economics, Humanities and Social Sciences (2021) has suggested AI and AI legal systems roadmaps for the Republic of Korea. In this regard, this study proposes that issues related to legal liability, profit distribution, negligence, and copyrights among AI and related industries are necessary to consider as more important and urgent short-term tasks in order to promote development in the application of AI in related industries.

The limitations of this study and future research suggestions are as follows. First, AI technology is still progressing and has not completely observed its whole life cycle. In particular, the diffusion trajectory in phase 2 is not sufficiently traced, and the most recent technology is included, meaning that it is necessary to understand the different changes in diffusion and patent indices in phase 2, especially for information related to forward citations. Nevertheless, the results of this study contribute to our understanding of the patterns of diffusion and progress regarding AI technology and recent related trends. In future studies, it will be necessary to trace the next phase, i.e., phase 3, in the future and to investigate the changing patterns of the previous phases as well. Second, according to the clustering results in this study, many patents are concentrated in the low cluster group. This phenomenon stems from the characteristics of the clustering methods, based on the adjacent distance, and also on the characteristics of the patents per se, in which a large number of citations are concentrated in a small group of patents. In addition, in the low-level cluster, there exists not only patents with a low level of diffusion but also very recent patent for which the diffusion pattern is not yet observed. This study does not investigate differences between these two types in low-level clusters. Thus, for low-level clusters, dividing the actual low levels of diffusion of patents and recent patents for which diffusion are not observed is recommended in further research.

Chapter 3. Technology convergence of AI in the industrial sector: Insights into a two-way approach [1]

3.1 Introduction

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 (Yu et al., 2018; Houlton, 2018). In addition, AI is a key technology for autonomous, connectivity, 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., 2021a). 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

[1] The original article of this chapter was published in *Scientometrics.*
Lee, S., Hwang, J., & Cho, E. (2022). Comparing technology convergence of artificial intelligence on the industrial sectors: two-way approaches on network analysis and clustering analysis. *Scientometrics*, 127:407-452.

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 knowledge 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" (Athreye & Keeble, 2000; Rosenberg, 1976). By sharing technological characteristics, the erosion of distinct barriers has been accelerated among various industries (Wang et al., 2019b). Convergence of technologies lead to industrial convergence, and industry convergence could only occur with the convergence of technologies (Nystrom, 2008; Choi et al., 2015). 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 (WIPO, 2019a; Fujii & Managi, 2018; Tseng & Ting, 2013). 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 into three perspectives, which are purpose, methodology, and object of the analysis according to the work of Kim and Lee (2017). In particular, the purposes are two fold, identifying evolutionary trajectory and convergence pattern (Kim & Lee, 2017). 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 belongs 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 Sohn (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 about 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 regarding 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 industrial sector-specific technology categories from an all-encompassing perspective of industry. The bottom-up approach here attempts to investigate practical usage 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 utilization levels.

We suggest 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 approach 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 utilization of the technology within the industrial sectors via the bottom-up approach.

The methodological 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, 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 utilization aspects in the convergence patterns. 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. 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 utilizations. In particular, research on general-purpose technology such as AI is crucial from an industry perspective, and the industrial sector-specific AI convergence characteristics identified in this study can have significant implications for all AI-related industries. Third, the two-way approach proposed in this study is considered as a framework for exploring the potential of AI as a GPT, providing deep insight into horizontal applicability and vertical applicability aspects.

The specific research questions of this study are as follows. First, how does the technology convergence of AI appear regarding the industrial sector, the technology category, and utilization in the two-way approach? Second, how can we understand the horizontal and applicability of AI with the two-way approach? The remainder of this paper is structured as follows. Section 3.2 describes the proposed method and research framework of this study. Sections 3.3 briefly shows the dataset in this study. Section 3.4 shows the results of network centrality analysis, ego-network analysis, clustering analysis. Additionally, Section 3.4 provides insights into the two-way approach in terms of technological applicability. Section 3.5 presents discussions and conclusions, and also proposes future research directions.

3.2 Proposed method

3.2.1 Research framework

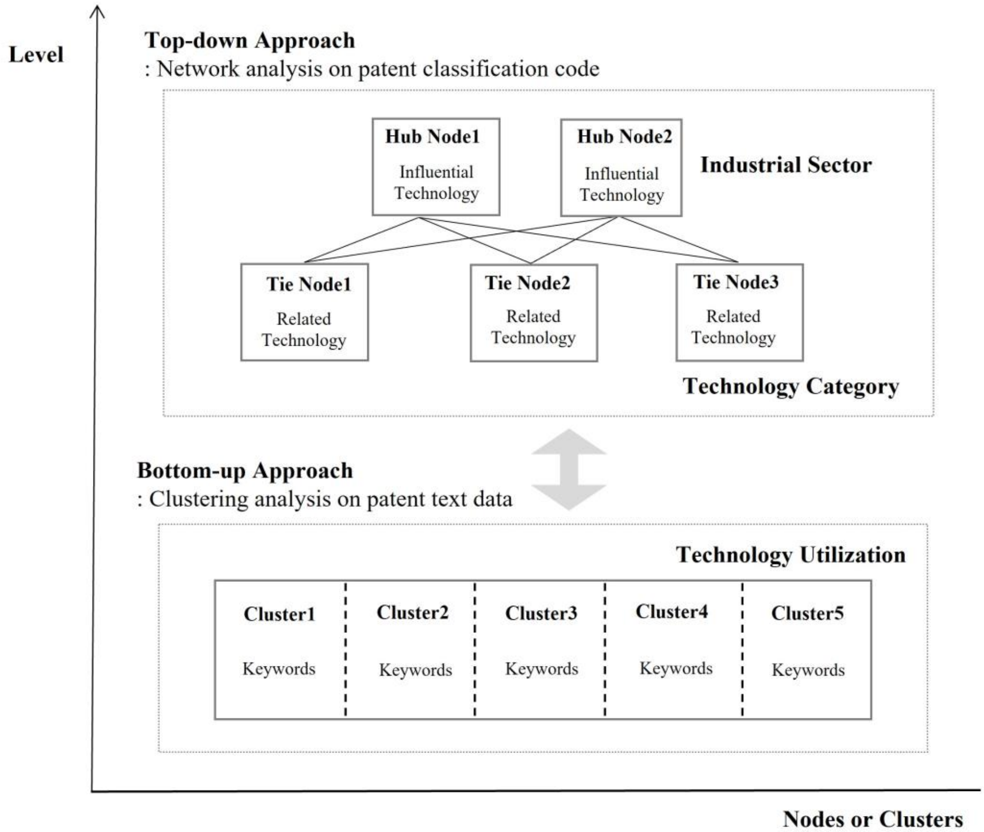

Figure 3-1. Research framework

Figure 3-1 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 utilization can be examined by linking these aspects with the technology category derived from the network analysis. In this study, a technology utilization is defined as an actual way in which the technology used in the industry and technology categories.

Figure 3-2 shows the research procedure of this study, and Sections 3.2.3 describes 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 industrial sector-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 of technology utilizations 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.

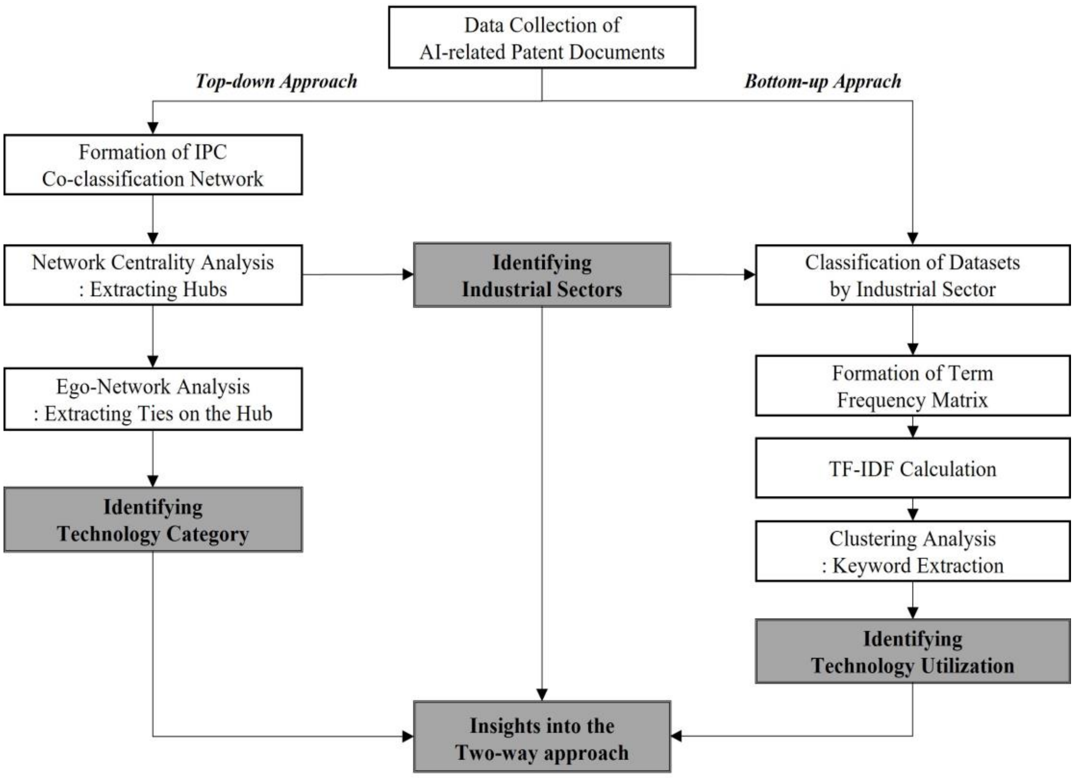

Figure 3-2. Research procedure

3.2.2 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 were according to the work of the Korean Intellectual Property Office (KIPO) in 2018 (KIPO, 2018). Table 3-1 shows the category of the AI technology and IPC codes used in this study. The descriptions of each IPC in this study are referred to the WIPO (WIPO, 2018; 2019b).

Table 3-1. 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
	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

Note. Descriptions of each IPC are referred to the WIPO (2018, 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; thus, 209,211 patents were selected for this study. Also, we extracted 2,517 IPC codes at the group level from the patents.

3.2.3 Methodology

Formation of IPC Co-classification Network

To measure technological convergence, previous research commonly used a co-classification and co-citation 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., 2016b), 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., A01B/22) of each patent, a total of 2,517 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 identifies 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 a node as a brokerage, and related to arbitration capabilities in technology convergence (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 (Freeman, 1979; Borgatti et al., 2002).

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

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 (Freeman, 1979; Borgatti et al., 2002).

$$C_B(N_i) = \sum_{j<k} \frac{g_{jk}(N_i)}{g_{jk}},\ i \neq j \neq k \quad \text{Eq. (2)}$$

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. 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, 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 measured by 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-Technology Concordance Table" (Schmoch, 2008) which was classified technology into thirty-five fields according to IPC codes (see Appendix A2-1). 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, nano-technology, 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 fast and simple, 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 to make a group of cohesive and homogeneous components which are closely representative of the patterns of the dataset (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 $p(\mathbf{x})$ for a D-dimensional vector $\mathbf{x}$ is expressed as a superposition of K Gaussian probability densities (Bishop, 2006.)

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \quad \text{Eq.}(3)$$

where π_k is called the mixing coefficient, which indicates a selection probability of kth Gaussian distribution, $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ 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 $\mathbf{x}$.

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 A2-2). 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.

3.3 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 Section 3.2.3. For measuring innovation performance, the number of patents or citations were widely used as indicators (Hagedoorn & Cloodt, 2003; von Wartburg et al., 2005; Trajtenberg, 1990; Harhoff et al., 1999). The number of patents means a quantitative aspect in terms of the technological invention of new technology, process, and product.

All collected datasets were shown in Figure 3-3. The datasets classified into the industrial sectors were analyzed using the patent count and CAGR of patent count shown in Table 3-2. In addition, Figure 3-4 shows the trend analysis according to the industrial sector based on the new patent count by the year. In terms of patent counts, the finance/management, medical, and transport sectors had a large number of patents, over 6,000, and the semiconductor, games, and biological materials had a patent count of 1,000 to 2,000. The environmental technology, organic chemistry, pharmaceuticals, civil engineering, food chemistry, nano-technology, metallurgy, and polymers had a patent count below 1,000. 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.

Table 3-2. Datasets by the industrial sector in AI patents

Sector	Number of Patents	CAGR (Compound Annual Growth Rate)
Finance/Management	12603	23%
Medical technology	10218	17%
Transport	6426	41%
Semiconductors	1896	9%
Biological materials	1768	13%
Games	1576	27%
Biotechnology	1188	12%
Civil engineering	621	23%
Environmental technology	332	9%
Organic fine chemistry	331	8%
Basic materials chemistry	185	4%
Pharmaceuticals	185	16%
Food chemistry	74	20%
Micro-structure and nano-technology	49	15%
Materials, metallurgy	42	16%
Macromolecular chemistry, polymers	35	16%

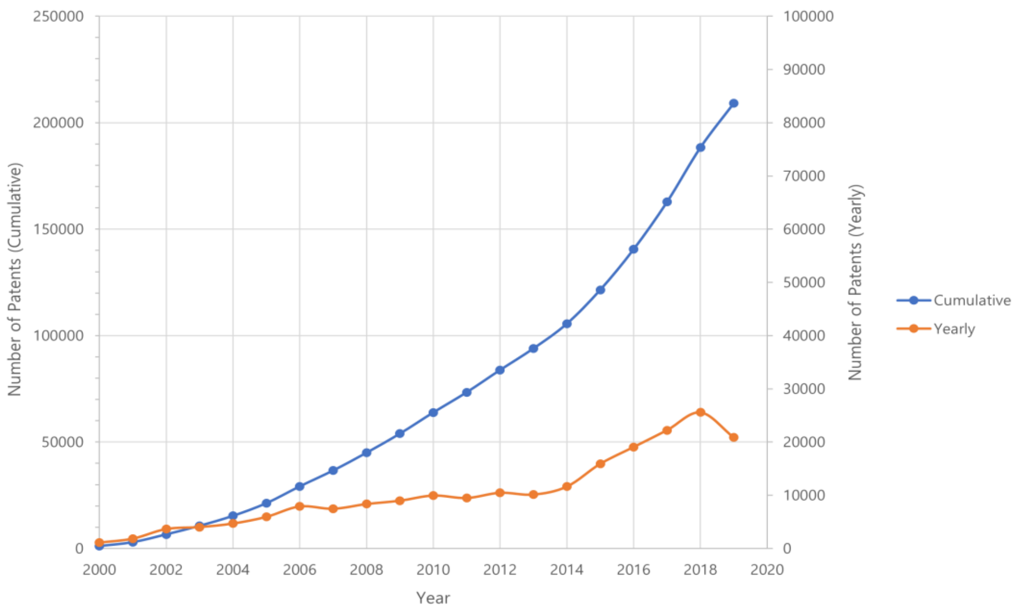

Figure 3-3. Number of AI patents

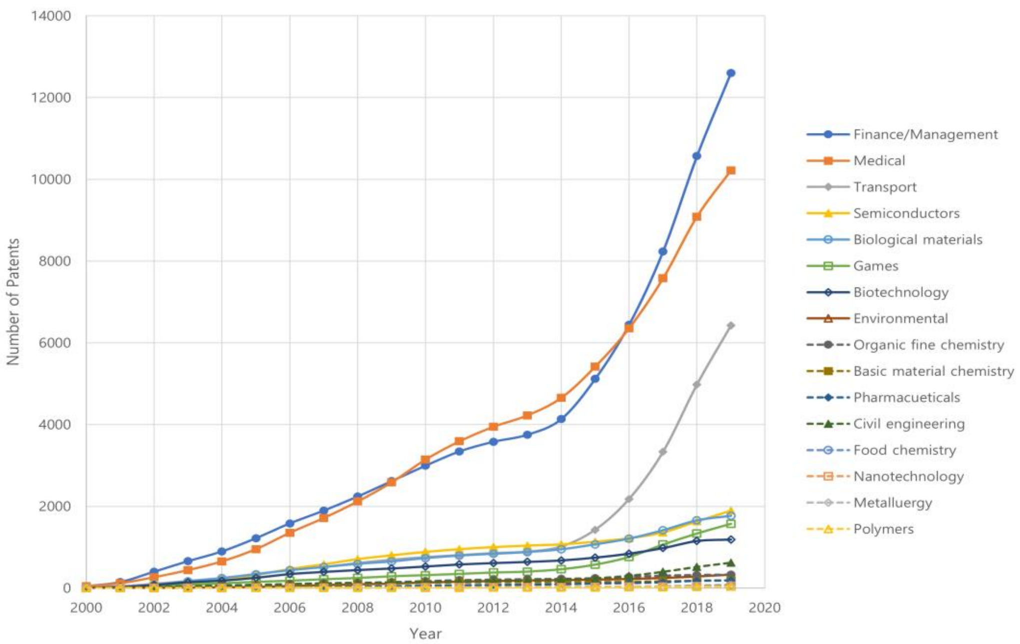

Figure 3-4. Datasets by the industrial sector in AI patents

3.4 Analysis and results

3.4.1 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 3-3) and positioned in the most central in the whole AI convergence network.

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

	Rank	IPC Code	Centrality	Description
Degree Centrality	1	G06K-009	614610	Methods or arrangements for reading or recognising printed or written characters or for recognising patterns
	2	H04N-005	481160	Details of television systems
	3	G06F-003	391088	Input or output arrangements for transferring data
	4	G06F-017	278642	Digital computing or data processing equipment or methods
	5	H04N-007	255293	Television systems
	6	H04L-012	247720	Data switching networks
	7	H04N-001	209014	Scanning, transmission or reproduction of documents
	8	G06T-007	186469	Image analysis
	9	H04L-009	178216	Arrangements for secret or secure communication
	10	H04N-021	173159	Selective content distribution, e.g. interactive television or video on demand
Betweenness Centrality	1	G06K-009	1560454	Same as above
	2	G06T-007	183382.2	Same as above
	3	G06F-017	143697.6	Same as above
	4	G06F-003	103567.5	Same as above
	5	G10L-015	97529.1	Speech recognition
	6	G06N-003	84961.23	Computer systems based on biological models
	7	H04N-005	82497.74	Details of television systems
	8	H04N-007	79561.6	Same as above
	9	G06F-019	62109.79	Digital computing or data processing equipment or methods for bioinformatics
	10	G06N-005	56337.88	Computer systems using knowledge-based models

Note. Descriptions of each IPC are referred to the WIPO (2018, 2019b)

Meanwhile, among the sixteen sectors in our dataset in the Section 3.3, 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 3-4 shows the results of the hubs in each sector.

Table 3-4. 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	73798	22	17119.15
	G06Q-020	41	73044	48	6751.979
	G06Q-010	53	49148	21	18096.42
	G06Q-050	74	28710	30	12771.32
	G06Q-040	163	8673	163	1143.456
Medical	A61B-005	33	81404	11	39581.35
	A61B-006	84	23489	67	4394.374
	A61B-008	142	13333	202	841.873
	A61B-003	161	9160	153	1251.287
	A61B-017	236	2868	194	886.636
Transport	B60R-021	98	21874	182	981.005
	B60R-001	131	14891	61	4837.574
	B60N-002	153	11183	227	714.849
	B60W-030	157	10228	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	30864	80	3097.656
	H01L-021	178	6770	60	4914.426
	H01L-023	226	3077	141	1361.027
Game	A63F-013	119	18299	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	G01N-033	95	22263	16	23755.15
	C12Q-001	186	5899	54	5604.225

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 (A63B-003), and surgical instruments (A06B-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 (B06N-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 Section 3.3, 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 3-5 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.

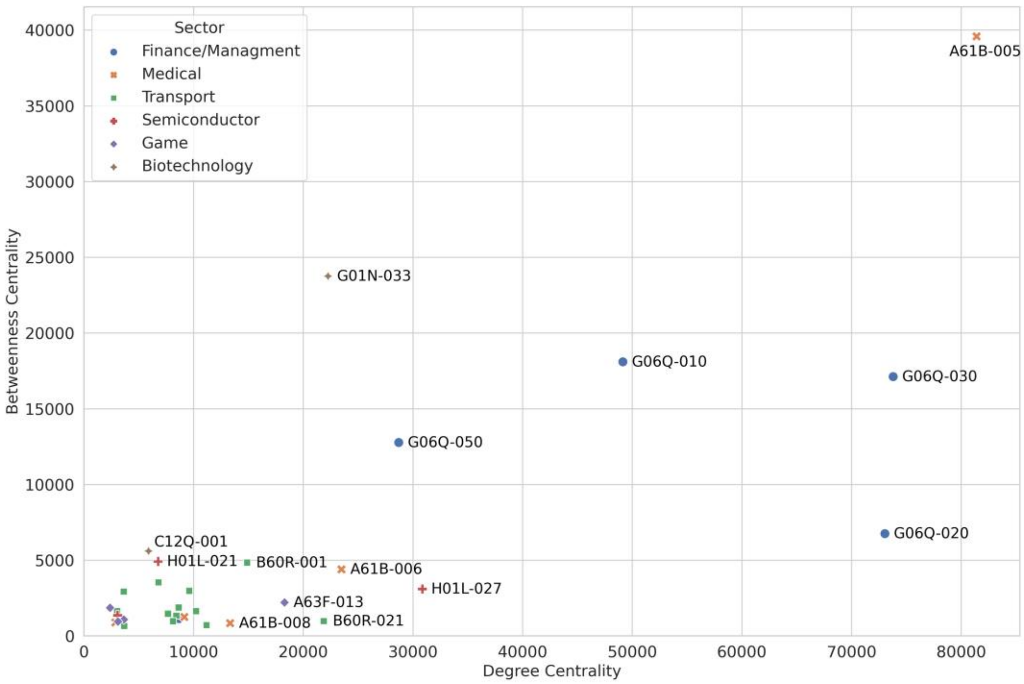

Figure 3-5. Hubs according to the industrial sector in the AI convergence network

3.4.2 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 3-5 according to each hub. (see Appendix A2-3 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 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 sector-specific technology among industrial sectors. Consequently, Table 3-6 presents a summary of the common ties among the industrial sectors and sector-specific ties by industrial sector.

Table 3-5. Ties according to each hub in AI convergence network

Sector	Hubs	Strong 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)]
	G06Q-040	G06F-017 [c)], G06F-021 [e)], G06K-009 [a)], G06F-003, H04N-005, H04N-021 [f)]
Medical	A61B-005	G06F-017 [c)], G06F-019 [g)], G06K-009 [a)], G06F-003, G01J-005, G01R-033, 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-007[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
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[f]

	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 (G06K-009)

b) Image analysis (G06T-007)

c) Data processing methods (G06F-017)

d) Speech recognition (G10L-015)

e) Against unauthorized activity (G06F-021)

f) Selective content distribution (H04N-021)

g) Data processing methods for bioinformatics (G06F-019)

h) Image generation; enhancement (G06T-005, 011, 017)

i) Control (G08G-001, G05D-001, G08G-005)

j) Navigation; direction finding (G01C-021, G01S)

k) Material analysis (G01N-021, 015)

l) Electric communication (H04L-009, H04W-008, 028, 088)

m) Diagnostic; identification (A61B-005)

n) Demonstration appliances (G09B-019)

Table 3-6. Hub and ties by the industrial sector in AI convergence network

Sector	Hubs	Strong Ties	
		Sector-specific	Common
Finance/ Mgmt.	Commerce (G06Q-030) Payment architectures (G06Q-020) Business management systems (G06Q-010, 050) Insurance and tax (G06Q-040)	Against unauthorized activity (G06F-021) Selective content distribution (H04N-021)	Pattern recognition (G06K-009) Image analysis (G06T-007) Data processing methods (G06F-017) Speech recognition (G10L-015)
Medical	Diagnosis (A61B-005, 006, 008), Examination or testing instruments (A63B-003), Surgical instruments (A06B-017)	Data processing methods for bioinformatics (G06F-019) Image generation; enhancement (G06T-005, 011, 017)	
Transport	Protecting against accidents (B60R-021), Optical devices (B60R-001, B60Q-001) Vehicle control (B60W-030, 050, 040, 010) Monitoring (B60R-011) Seat management (B06N-002) Preventing theft (B60R-025)	Control (G08G-001, G05D-001, G08G-005) Navigation; direction finding (G01C-021, G01S)	

Semiconductor	Semiconductor devices (H01L-027) Manufacturing process (H01L-021) Details of semiconductor device (H01L-023)	Material analysis [i.)] (G01N-021) Electric communication [ii.)] (H04L-009, H04W-008, 028, 088)
Game	Video games (A63F-013) Sports appliances (A63B-071, 069) Controls for exercising (A63B-024)	Measuring diagnostic purpose; identification of person (A61B-005) Demonstration appliances (G09B-019)
Biotechnology	Analyzing materials (G01N-033) Measuring or testing process (C12Q-001)	Data processing methods for bioinformatics (G06F-019) Material analysis (G01N-021, 015)

Note. Descriptions of each IPC are referred to the WIPO (2018, 2019b)

i.) Especially Highly Ranked in Manufacturing process (H01L-021)

ii.) Especially Highly Ranked in Semiconductor devices (H01L-027)

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, 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 6 shows the summary of ties in each sector. For the finance/management sector, programs, or data 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 processing (G06T-005, 001, 017) and optical technologies (H04N-005) emerged in the others. In addition, diagnosis-related technology (A61B-005, 006, 007) showed high connectivity with the data processing methods for bioinformatics (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 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 technology related to directional guidance and detection. 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, image-related technology, which are optical (G10B-011) or photomechanical (G03F-001, 007), and also material analysis (G01N-021, G01R-031) 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 registration (G10L-015) showed high connections at the same time. On the other hand, sports/exercise had a high connection with medical technology (A61B-005). In addition, education, or demonstration appliances (G09B-019) were appeared in the sports/exercise area. For the biotechnology sector, data processing methods for bioinformatics (G06F-019) was also highly linked, and technology related to the analysis of materials (G01N-021, 015) appeared.

3.4.3 Results of clustering analysis

The top 30 keyword were extracted for each cluster, and these results of keywords and TF-IDF values are shown in Appendix A2-4 to A2-9. The application of a technology is described by a verb, and a product or process is expressed using a noun. (Taylor & Taylor, 2012). This study selected representative keywords among the top 30, considering the discriminative meaning of each cluster, excluding redundant or common keywords among clusters. Table 3-7 shows the results of clustering analysis.

Table 3-7. 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
	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

3.4.4 Insights into the two-way approach

This study proposes a new framework in the form of a two-way approach to understand technology convergence of AI in the industrial sector overall. As a result, this study investigates technology convergence of AI in different industrial sectors and investigates the technology category and the utilization of the technology. Furthermore, the framework of the two-way approach provides an in-depth understanding of technological applicability from the perspectives of the horizontal and vertical applicability of various industrial sectors.

The two-way approach in this study provides insight by which to understand the technological applicability of AI as a GPT. Figure 3-6 depicts the concept of technological applicability based on the research framework introduced in Section 3.2.1. The main characteristic of a GPT is that it has the potential to expand into a wide range of heterogeneous industrial applications (Feldman & Yoon, 2012). Specifically, Lipsey et al. (2005) suggest two types in relation to this: horizontal applicability and vertical applicability. The horizontal applicability of a GPT refers to whether the technology is applicable and utilized across various industrial sectors, whereas the vertical applicability of a GPT is whether the technology is widely utilized in the subsectors within an industrial area (Feldman & Yoon, 2012; Lipsey et al., 2005). The insight into the two-way approach regarding the horizontal and vertical applicability of AI is explained below.

From the results of this study, AI technology is applied to various industrial sectors and has horizontal applicability. First, according to the hubs and from the top-down approach, the technologies in a range of other industrial sectors were included in the central position in the AI convergence network. In other words, this suggests that applications of AI in various industries occur actively. Finance/management, medical, transport, gaming, semiconductors, and biotechnology were identified as notable industrial sectors of AI in this study. Second, Second, according to the ties from the top-down approach, several common strong ties were identified, in this case pattern recognition (G06K-009), image analysis (G06T-007), data processing methods (G06F-017), and speech recognition (G10L-015). Common strong-tie technologies are variously applied with connections among various industries, as shown in Table 3-6. Third, according to the keywords from the bottom-up approach, the results show horizontal applicability in terms of technology utilization in connection with the results of the top-down approach. The results here show that the same horizontal technology was variously utilized depending on the industrial sectors, as confirmed in Table 3-8. 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 for each sector used for the image analysis. The image analysis is mainly utilized for the identification of electronic documents in the finance sector (cluster #2), analysis of projection images (i.e. CT, tomography, X-ray) in the medical sector (cluster #4), inspection on patterns of wafers in the semiconductor sector (cluster #1), for object detection in the transport sector (cluster #4), and analysis of biological images (i.e. tissues, cells, specimen) in the biotechnology sector (cluster #1).

Also, the results showed that AI technology has vertical applicability as it is applied in a wide range of technology categories with various uses within a specific industrial sector. First, according to the hub and tie nodes in each industrial sector from the top-down approach, it was found that AI technology complements various sector-specific technologies within the industrial sector. As shown in Table 3-6, various sector-specific technologies were identified in each industrial sector in relation to AI. For instance, in the finance/management sector, AI technology is applied to various areas, such as commerce (G06Q-030), payment architectures (G06Q-020), insurance and taxes (G06Q-040). Second, according to keywords from the bottom-up approach, the results show the patterns of technology utilization in connection with the results of the top-down approach within the industrial sector. For instance, the finance/management sector shown in Table 3-6 shows the payment architecture (G06Q-020), and against unauthorized activity (G06F-021) from the results of the top-down approach. In line with this, the keywords of clusters show the various patterns of utilization related to those technologies, such as 'user', 'biometric', and 'authentication' (cluster #3) as shown in Table 3-7.

In conclusion, this study investigated various industrial sectors to which AI technology is applied as well as various technologies to which AI is applied within each industrial sector. Through this, the study presented the results of an empirical analysis showing that AI technology has horizontal and vertical applicability. In particular, the results of the top-down approach based on IPC data and the bottom-up approach based on text data were combined, complementing the interpretation of the results of each analysis in order to understand the applicability of AI in greater detail.

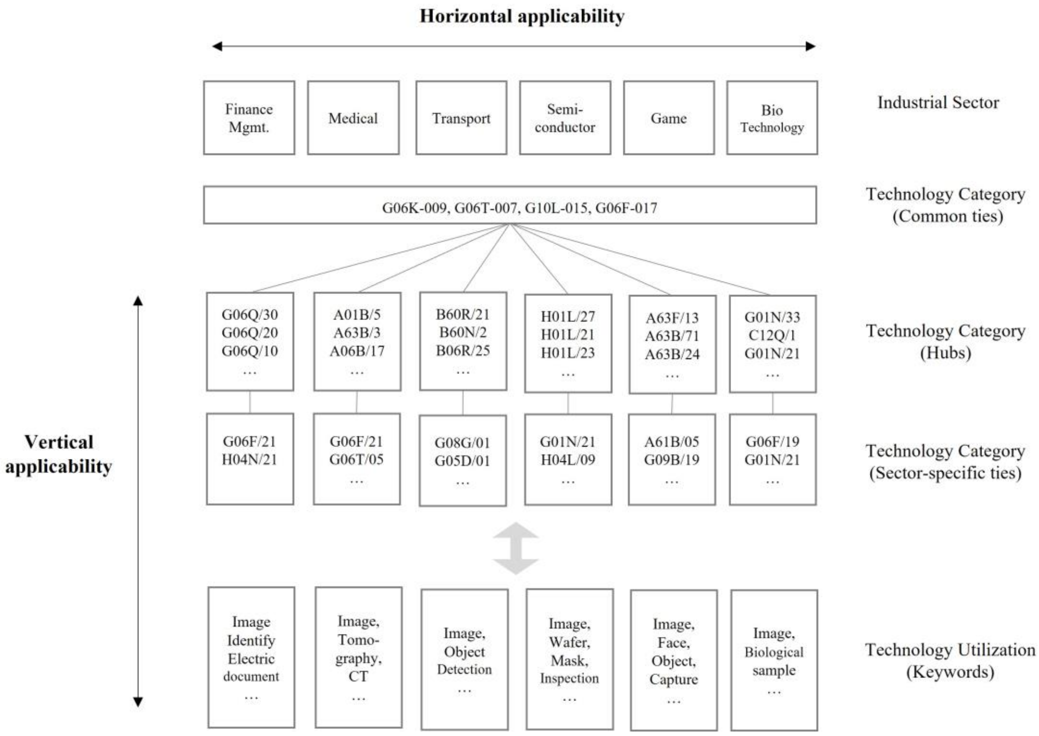

Figure 3-6. Horizontal and vertical applicability of AI

Table 3-8. Summary of horizontal and vertical applicability

		Finance/Mgmt.	Medical	Transport	Semiconductor	Game	Biotechnology
Horizontal applicability	Industrial Sector	Finance/Mgmt.	Medical	Transport	Semiconductor	Game	Biotechnology
	Technology Category (Common Ties)	G06K-009, G06T-007, G06F-017, G10L-015					
Vertical applicability	Technology Category (Hubs)	G06Q-030 G06Q-020 G06Q-010, 050 G06Q-040	A61B-005, 006, 008 A63B-003 A06B-017	B60R-021,001, B60Q-001 B60W-030, 050, 040, 010 B60R-011, B06N-002 B60R-025	H01L-027 H01L-021 H01L-023	A63F-013 A63B-071, 069 A63B-024	G01N-033 C12Q-001
	Technology Category (Sector-specific Ties)	G06F-021 H04N-021	G06F-019 G06T-005, 011, 017 H04N-005	G08G-001 G05D-001 G08G-005 G01C-021 G01S	G01N-021 H04L-009 H04W-008, 028, 088	A61B-005 G09B-019	G06F-019 G01N-021, 015

	Technology Utilization (Keywords)	Image, identify, electric document	Image, tomography, CT	Image, object detection	Image, wafer, mask, inspection	Image, face, object capture	Image, biological sample
		Authentication, payment	Sensor, control determine	Unmanned aerial vehicle	Signal, detection	Augment/Virtual Reality	Detect, feature, generic
		…	…	…	…	…	…

3.5 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. In the results, this study identifies the industrial sector, technology category, and utilization with respect to the technology convergence of AI in order to answer the first research question. Additionally, this study confirms the horizontal and applicability of AI to answer the second question.

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 utilizations. Consequently, it is possible not only to compare the characteristics of technology convergence by industrial sector, but also to define patterns 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 utilization considering each defined convergence pattern. The network analysis complements the results of the clustering analysis when defining the cluster structures and meanings. The clustering analysis is the result of the unsupervised learning with unlabeled data, and meanings should be found by means of interpretations of randomly derived results. The results derived from the network analysis can then serve as common criteria and important indicators for interpreting the outcomes from the clustering analysis. Second, this study gives contribution 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.

From a strategic business perspective, the results of this study can contribute to establishing technological strategic directions for a company. For industrial AI companies, the results of this study identified the influential technologies of each industrial sector. Because the technologies derived from this study exist in an area of high complementarity with AI technology for each industrial sector, it is recommended to review existing business models from various angles within the scope of technology. On the other hand, also from the results of this study, it was found that there was relatively less applied natural language processing in biotechnology and in the medical and semiconductor sectors compared to image analyses. Through the results of this study, it is possible to identify technologies that have been relatively less developed in each industrial sector and to discover new business creation opportunities in new areas in these sectors. For cross-industry AI companies, the results of this study contribute to explorations of scalability opportunities based on the core capabilities of companies. According to the results here, the image analysis based on a material analysis has scalability to wafer pattern analysis during a semiconductor analysis and biological sample pattern analysis in the biotechnology and medical sectors as well. Thus, it is necessary to develop extensible technologies that can be applied across industries.

From a policy perspective, this study suggests balanced AI development for the growth of AI technologies into a GPT. First, 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., 2021a). 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 has found that environmental policies and subsidies by governments to promote inventions of green technologies significantly increased green patent publication counts in China (Fujii & Managi, 2019). 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. Second, it is recommended to establish policies to encourage technology development for privacy and data protection. In the results of this study, technologies related to the protection of privacy or data were rare. Technologies related to authorization were found in the finance sector, whereas this type was not found in the medical and biotechnology sectors, which deal with sensitive personal information. In relation to this, it is necessary to prepare for negative effects that may arise due to the development of AI technology in each industry sector.

The limitations of this study and future research suggestions are as follows. First, this study considered the category of AI as three broad groups of technology based on the IPC codes of patents; learning and reasoning, natural language processing, and computer vision. However, the categories in this study cannot fully explain the overall AI technology scheme. According to previous work, WIPO (2019a) proposes an AI technology category for AI techniques and functional applications, and WIPO (2022) also suggests IPC and CPC codes corresponding to AI techniques and functional applications. However, the present IPC and CPC codes themselves do not completely represent all AI techniques and functional applications, such as machine learning, probabilistic reasoning, predictive analytics, robotics, and others. Thus, future studies should collect and analyze additional data based on a systematic classification of AI technology by utilizing both IPC and CPC codes and key phrases which thus far are not explained by IPC or CPC codes. Second, the results of the text analysis here are simply fragmentary keywords based on TF-IDF. Therefore, there is a limitation in that it is difficult systematically to classify the meanings of corresponding words in the context of patent text, such as words related to technology, processes, products and data. In future studies, understanding text data in patents considering the context and meaning will be needed in research related to natural language processing, such as named entity recognition. Third, with respect to the additional methodological approaches, it is recommended to compare the result of analyzing text data using GMM with other soft clustering algorithm, such as topic modeling and fuzzy modeling. Fourth, 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. Future studies investigate each industrial sector respectively with in-depth understanding, or consider other industries not covered in this study. Also, it is meaningful to examine the background and characteristics on the area where AI convergence occurred low.

Chapter 4. Technology strategies of AI startups: focusing on patent activity and diversity

4.1 Introduction

Previous research suggests that young and small innovative companies are regarded as main drivers of innovative activities and economic growth (Caviggioli et al., 2020; Schneider & Veugelers, 2010). In addition, disruptive innovations are more adequate for small firms considering that seeking market success by bringing new disruptive technologies to the market (Hacklin et al., 2005; Kassicieh et al., 2002). Therefore, investigating the innovative activities and strategies of companies, especially startup firms, is necessary to understand the growth and development of emerging technologies.

The stage of the growth and life cycle of a startup is defined by the J-curve, which refers to the startup financing cycle (Love, 2016). Startups generally confront what can be considered the valley of death in the initial seed stage, go through early- and late-stage funding rounds, and finally enter the public market. The valley of death refers to the unfavorable circumstances any startups encounter early in the life cycle (Gbadegeshin et al., 2022). The survival rate of startups is very low, with only about 20% surviving for five years (Liu et al., 2021b; Song & di Benedetto, 2008). In order to establish the continuous survival and operation of startups, funding by investors is essential for each growth stage and period. Thus, this study considers investment funding as a means of the growth of startups.

Among the many factors influencing startup investments, patents should be considered important to understand a firm's technology strategy. The mainstream of research on startup patents is related to the effects of patent signals on investors and funding. From the perspective of signaling theory (Spence, 1973; Spence, 2002), patents, which are indicative of protected technological and product innovations according to patent rights, act as signal of reducing information asymmetry for investors (Caviggioli et al., 2020; Hsu & Ziedonis, 2013; Mann & Sager, 2007; Nadeau, 2010). Zhang et al. (2019) analyzed three main types of patent signals, technological signals, commercial signals, and legal signals, and examined the effects of these signals on venture capital (VC) amounts. Hoenen et al. (2014) found that the patent signaling value is significant in the early stages of venture capital financing.

Another research stream is related to startup patents along survival or exit routes. The main exit routes are considered to be the initial public offering (IPO) and the merger and acquisition (M&A) route. According to research on the relationship between patents and IPOs, patents positively affect the amount of money gathered during the IPO, the survival rate after the IPO, and the stock returns performance in the long-run (Useche, 2014; Wagner & Cockburn, 2010; Zhang & Zhang, 2020). Exit routes can be divided into bankruptcy, M&A, and voluntary liquidation (Kato et al., 2022). M&A exits are considered to be the typical form with regard to a sale of a firm, representing a desirable outcome for new and young firms (Cotei & Farhat, 2018). From the literature, patents have a positive effect on M&A (Kato et al., 2022; Cotei & Farhat, 2018; Wagner & Cockburn, 2010).

The literature has mainly focused on the number of patents but has neglected various characteristics of patents in a patent portfolio (Caviggioli et al., 2020). Specifically, due to the constraint of a startup itself with a small number of patents, scant attention has been directed toward the issue of patent portfolio diversity. Although research on technology diversity has generally analyzed large firms and has rarely examined small businesses, recent studies suggest diversification strategies for startups in a relation to the RBV (resource-based view) of small firms (Khurana & Farhat, 2021; Coleman et al., 2013; Esteve-Pérez & Mañez-Castillejo, 2008). Although numerous studies of strategic management have examined diversification from a product or market perspective and not from the perspective of technological competence, the diversity of a technology portfolio can account for the capability of a firm to realize technological innovations in terms of the RBV (Lin et al., 2006). Therefore, to fill the research gap, this study considers and examines the diversity of patent portfolios of startups.

In particular, previous research has paid little attention to AI startups despite the fact that investors keenly follow AI startups. According to Tricot (2021), VC investments in AI startups are increasing every year, reaching 21% of all VC investment in 2020, from only 3% in 2012, showing a 28-fold increase in the total amount. Because VC investments provide some understanding of promising industries (Tricot, 2021), this study provides insights into the growth of AI startups and the AI sector as well.

The purpose of this study is to investigate the relationship between patent strategies and investments in AI startups. The first research question is as follows: what are the relationships between patent activity and investments in AI startups? Patent activity in this question considers whether the firms have patents or not and how many they have, if any. Given that a relationship between patent counts and funding amounts was found in other sectors in earlier work, we empirically test this relationship in the AI sector. The second research question asks about the relationships between patent diversity and investments in AI startups. This study attempts to examine the relationship between patent diversity and funding amounts, an area that did not receive attention in previous studies. Patent diversity in this study is divided into two levels, unrelated diversity and related diversity. Additionally, to answer the two research questions, this study examines the moderating effects of the investment funding stage and the type of AI startup firm on the relationships.

The remainder of this paper is structured as follows. Section 4.2 presents the literature review. Section 4.3 shows the methodology. Section 4.4 describes the analysis and results of the study. Section 4.5 presents the discussion and conclusion and suggests future research.

4.2 Literature review

4.2.1 Patent activity and startup investment

Patents of startups are considered to be an intangible asset that can be utilized to measure the valuation of the firm (Lynch, 2021). Through patents, investors can clearly estimate the potential value and growth of the firm. Startups normally depend on external investment capital in order to overcome their financial constraints and to sustain the business and engage in innovative activities (Caviggioli et al., 2020; Atherton, 2012; Macht & Robinson, 2009). Likewise, the funding amounts of investors are recognized as a proxy by which to measure the growth of a startup.

Extant literature shows a positive relationship between patent and VC funding (Caviggioli et al., 2020; Mann & Sager, 2007; Audretsch et al., 2012). VC amounts are positively linked to the number of patents, the patent family, the patentee, and patents cited, whereas no relationship between the scope of the patent, the scope of the firm and the depth of firm can be found (Zhang et al., 2019). According to research on the effects of patent characteristics on VC amounts, public patents are most significant compared to granted, utility, and design patents (Zhang et al., 2019). Meanwhile, from research on patent activity and pre-money valuation, patent applications show a positive relationship with firm value, whereas grant patents show a positive relationship only in the early, pre-revenue stages and the early financing rounds (Greenberg, 2013).

Depending on the financing stage of the life cycle, startups have different requirements in each stage, and the effects of VC investments are different (Flynn &

Forman, 2001). Seed capital is the first financing type used for product research and development, with startup investments then used to produce and sell the product; these two stages are regarded as early stages of investment (Jeng & Wells, 2000). The next stage is considered as the expansion capital stage, which is used to expand manufacturing or the distribution capability of the firm (Jeng & Wells, 2000). Likewise, early stage startups focus on the idea and proof of their concept, while startups in later stages should attempt to expand the market with their already proven concept (Bianco et al., 2022; Davila et al., 2003)

There has been little attention paid to investigating the relationship between patents and financing rounds, and there is no consistent consensus in the research (Caviggioli et al., 2020). Hoenen et al. (2014) found a positive relationship between patents and early stage financing in the high-technology industry, especially comparing the first round of financing and the second round of financing. Hsu and Ziedonis (2013) suggested that patents increase startup valuation more steeply in the early financing rounds among semiconductor startups, whereas Mann and Sager (2007) suggested a greater value of patents in the later stages of financing for software startups. Caviggioli et al. (2020) found positive significant effects between patents and later stages of financing considering low and high patent intensity levels in the industrial sector.

In line with this, the literature also finds that the effects of patents on firms show heterogeneous results depending on the industrial sectors (Caviggioli et al., 2020; Hall et al., 2014; Nadeau, 2010). One reason is the different level of appropriability in each industry, particularly in relation to whether the patent offers strong protection from imitations (Caviggioli et al., 2020; Hall et al., 2014). In addition, the relationship between patent activity and investments of VC or corporate venture capital (CVC) differs according to the industry. (Dushnitsky & Lenox, 2005; Nadeau, 2010).

Thus, this study investigates the relationship between patent activity levels and funding amounts in AI startups. In particular, the proportion of startups without patents is very large even when the value of such firms is high (Lynch, 2021). Hence, this study examines whether these firms have patents and if so how many patents they have as a measure of patent activity. In addition, this study examines the moderating effects of funding stage and industrial range (referred to here as the firm type) of AI startups on the relationship between patent activity levels and funding amounts.

4.2.2 Patent portfolio diversity

The research on patent portfolios considers perspectives of risk reduction and synergy creation (Appio et al., 2019; Lin et al., 2006). In term of risk reduction, the primary concept is based on the financial portfolio theory, which holds that aggregate portfolio risk can be pooled and total risk reduced compared to the simple sum of the risk of individual investments (Jacobs & Swink, 2011).

The portfolio theory can be conceptualized to the portfolio of product as well, aggregating the uncertainty from the various product pools and reducing the risk (Jacobs & Swink, 2011; Baker et al., 1986). Meanwhile, various strategic management studies suggest the importance of synergy creation aspects, indicating that the integrated patent portfolio overall is more valuable than the sum of each single patent (Appio et al., 2019; Lin et al., 2006).

The research stream on the patent portfolio strategy focuses on the two aspects of patent portfolio specialization and patent portfolio diversity (Appio et al., 2019; Lin et al., 2006). Regarding patent portfolio specialization, this concept is based on the capability-based view and suggests that the core competences of firms can create synergy when the firm focuses on their technological knowledge in relatively few fields (Lin et al., 2006; Barney, 1991). Focusing on a small breadth of technology field has an advantage in that it can offer a dominant position in core technology fields (Lin et al., 2006). In contrast, the patent portfolio diversity strategy seeks to extend the technological scope of the existing knowledge of firm so that it gains various types of heterogeneous knowledge and capabilities (Appio et al., 2019). Studies have shown that a new knowledgebase can emerge from technology diversity, which can help a firm explore new business opportunities, finding that high-tech firms increase their levels of technology diversity over time (Lin et al., 2006; Granstrand et al., 1997). In addition, technology diversity can allow the firm to undergo sustainable business evolution, preventing the negative effects from the lock-in of one specific technology area (Garcia-Vega, 2006).

Extant literature examines various types of relationships between technology portfolio diversity and performance, and the results differ depending on the proxy used to measure diversity and firm performance (Lloyd & Jahera, Jr., 1994). The research broadly uses patent classification codes to measure patent portfolio diversity in an entropy-based manner (Appio et al., 2019; Zabala-Iturriagagoitia et al., 2020; Chen et al., 2010; Chen et al., 2012; Lin et al., 2006). Appio et al. (2019) measured patent portfolio diversity at the three different levels of section, class and the sub-class level of IPCs, showing that only the section level has a significant inverted-U shape for firm profitability. Chen et al. (2012) unrelated diversity and related diversity using UPC, finding that related diversity has a positive effect on innovation performance, while unrelated diversity has an inverse U-shape relationship. Zabala-Iturriagagoitia et al. (2020) measured the concepts of unrelated variety and related variety using IPC codes, suggesting that unrelated variety is a relative effective during the stages of economic expansion, whereas related variety is more effective during stages of economic crisis.

Meanwhile, there has been little attention paid to technology diversity and performance in startups, and no evidence exists, especially for AI startups. Hence, this study investigates the relationship between patent portfolio diversity levels and funding amounts in the AI sector. This study divides patent portfolio diversity into two types: unrelated technology diversity and related technology diversity. Also, this study examines the moderating effects of the funding stage and the firm type of AI startups on the relationship between patent portfolio diversity and the funding amounts.

4.3 Methodology

4.3.1 Data

Crunchbase data is widely used in the research on startups (Żbikowski & Antosiuk, 2021; Bock & Hackober, 2020; Malyy et al., 2021). This study collected a list of AI startups and basic information about the startups from Crunchbase. The list of AI startups in this study consisted of those in the artificial intelligence industry group, particularly in subordinate industries of artificial intelligence, intelligent systems, machine learning, natural language processing, and predictive analytics as existed in Crunchbase. The AI startups here were US-based startups and were founded from 2000 to 2022. Also, in order to ensure representativeness of the sample, this study considered various firm statuses, including unicorn, IPO, M&A, and general startups. A unicorn describes a company on the private market valued at over $1 billion (Brown & Wiles, 2015; Lynch, 2021). While there has been scant research on unicorns in academia, unicorn companies have captured investors' funds, the public's attention, and have brought significant innovations to the public, despite being few in number and despite their uniqueness (Lee & Lin, 2020; Bock & Hackober, 2020; Lynch, 2021). Thus, this study distinguished unicorns from other general startups. All US-based AI unicorn firms and IPO firms were collected given that the total number of AI unicorn and IPO startups was too small. Then, we constructed top listings of AI startups from Crunchbase and collected M&A firms, which have a history of being acquired by other organizations. General firms which were not included as unicorn, IPO, and M&A types were also collected. Next, this study collected basic information about the selected AI startups. Additionally, this study targeted AI startups which had funding from at least once venture capital investment firm. Finally, a total of 505 startups were collected, including 82 unicorns, 39 IPO firms, 178 M&A firms, and 206 other general firms. The classification criteria for the unicorn, M&A, and IPO types depended on the final events of the companies.

In addition, this study collected application patents regarding the 505 firms from a Google Patent Search. The assignee of the patents was matched with legal name from Crunchbase. In addition, in case of the assignee of a patent was re-assigned, the date was reflected as the designated date of the target company's patent. According to the literature, application patents were more positively related to funding amounts and firm value compared to grant patents (Zhang et al., 2019; Greenberg, 2013), and thus this study targeted application patents.

Table 4-1 shows a summary of the datasets used in this study, and Table 4-2 presents the descriptive statistics of the firms. There were 271 firms with application patents and 234 firms without application patents.

Table 4-1. Datasets

		Firm status				Firm type	
	All	General	M&A	Unicorn	IPO	Industry-specific	Cross-industry
Early Stage (Seed, Series A, B)	292	141	132	7	12	192	100
Later Stage (Series C and over)	213	65	46	75	27	133	80
Total	505	206	178	82	39	325	180
Firms with patents							
Early Stage (Seed, Series A, B)	127	52	62	4	9	89	38
Later Stage (Series C and over)	144	37	35	53	19	86	58
Total	271	89	97	57	28	175	96
Firms without patents							
Early Stage (Seed, Series A, B)	165	89	70	3	3	103	62
Later Stage (Series C and over)	69	28	11	22	8	47	22
Total	234	117	81	25	11	150	84

Table 4-2. Descriptive statistics

All firms	Mean	SD	Min	Max
total investment	$149M	$286M	$50K	$3,520M
number of investors	11.76832	8.5226	1	51
number of patents	8.140594	32.0004	0	471
age (year)	8.752475	3.766514	2	23
employees	260.6931	622.9084	5	7500
Firms with patents				
total investment	$200M	$359M	$2M	$3,520M
number of investors	13.01107	9.112256	1	51
number of patents	15.16974	42.47928	1	471
age (year)	9.804428	3.769583	2	23
employees	354.3542	787.0115	5	7500
Firms without patents				
total investment	$88.8M	$141M	$50K	$791M
number of investors	10.32906	7.55116	1	48
number of patents	0	0	0	0
age (year)	7.534188	3.382755	2	22
employees	152.2222	315.5107	5	3000

This study divided AI startups into two broad types of firms using the industry group information from Crunchbase. All of the firms collected for the purpose of this study were assigned to the 'artificial intelligence' industry group, though they could also be assigned to multiple other industry groups. Thus, a total of forty-three industry groups were found from the collected AI firms in this study. This study used the information of assigned industry groups of the AI firms to divide them into the two categories of industry-specific AI firms and cross-industry AI firms. The industry-specific AI firms here are describes as firms which focus on applying AI on specific industries, such as manufacturing, biotechnology, transportation, etc. The cross-industry AI firms in this study refer to firms which focus on developing AI solutions or products across various industries based on ICT technology. In this study, AI firms which were additionally assigned to at least one of the industry groups of the industry-specific list were initially classified as industry-specific AI firms, with the remaining firms classified as cross-industry AI firms (see Appendix A3-1). The industry-specific list was based the case where specific industries other than AI and ICT-related industries were specified as the industry groups. Whereas the ICT-related industries were classified as the cross-industry list considering the characteristics of their development of AI solutions and products, which could not be separated from ICT-related technologies and industries.

4.3.2 Empirical methods

This study constructed two parts of empirical analyses. First, this study explored the relationship between the patent activity levels and funding amounts of AI startups. Second, this study investigated the relationship between patent diversity levels and funding amounts of AI startups. For the first parts, in this study, a multivariate regression analysis was conducted on the variable during the final year of the collected data. The multivariate regression analysis indicated the presence of a relationship between a patent-holding firm (or not) and the funding amount. Also, a fixed-effect analysis was conducted using the variables by year. The fixed-effect analysis illustrated the relationship between patent counts and funding amounts. For the second part, this study conducted a fixed-effect analysis on the variables by year to examine the relationship between funding rounds of startups and patent diversity levels. Additionally, this study tested the moderating effects of the funding stage and the firm type on the relationships considering in both the first and second parts of the model. To consider the minimum number of years for patent publications, a time lag of two years was utilized.

4.3.3 Variables

Table 4-3 shows the definition of each variable. The funding amount (FundAmt), the dependent variable in this study, was calculated as the total amount including, all seed, venture capital, private equity, and other types of funding procured for each funding round. The source of the funds was not classified in this study given that the sources of funds varied depending on the funding round. For research question 1, the independent variables were whether the firm held a patent or not (PatentYN) and the patent count (Patent count), referring to the patent activity in this study. For research question 2, the independent variables were patent diversity, consisting of technology diversity (TD), unrelated technology diversity (UD), and related diversity (RD). The calculation of technology diversity was based on the entropy measure. From previous research, IPC codes were utilized to measure technology diversity, using the widely used 4-digit IPC codes (Appio et al., 2019; Zabala-Iturriagagoitia et al., 2020; Chen et al., 2010). The 3-digit IPC code referred to a wider range of technology diversity (Appio et al., 2019), and this was defined as the unrelated variety of technology (Zabala-Iturriagagoitia et al., 2020). Thus, this study analyzed TD based on 4-digit IPC codes and analyzed UD based on 3-digit IPC codes. Also, the related RD in this study was measured according to the difference between TD and UD (Chen et al., 2012). The equations for TD, UD, and RD are shown in Table 4-3, based on the previous works (Chen et al., 2012; Zabala-Iturriagagoitia et al., 2020, Appio et al., 2019; Chen et al., 2010). P_{it} in the Equation (1) denotes the patent proportion of the 4-digit IPC technology class i within n different 4-

digit IPC technology classes until the given year of t. P_{jt} in the Equation (2) represents the patent proportion of the 3-digit IPC technology class j within m different 4-digit IPC technology classes until the given year of t. The moderating variables were the funding stage and the firm type. A dummy variable for the funding stage (FundStg) was considered in two stages: the early stage and the later stage. In this study, the early stage refers to the seed and series A and B stages, whereas the later stage is the series C stage and over. A dummy variable for firm type (FirmType) was considered in two groups: industry-specific and cross-industry AI firms. Also, the patent count was considered as a moderating variable for research question 2.

Table 4-3. Definition of variables

Type	Variable	Description	Research question
Dependent Variable	$FundAmt_t$	Logarithm of the cumulative funding amounts (USD) in year t	RQ1, RQ2
Independent Variable	*Patent activity*		
	PatentYN	Dummy = 1 if a firm without patents	RQ1
	$PatentCnt_t$	Logarithm of the cumulative number of patent applications in year t	RQ1
	Patent diversity		
	TD_t	Logarithm of the technology diversity of a firm in year t $TD_t = \sum_{i=1}^{n} P_{it} \ln\left(\frac{1}{P_{it}}\right)$ Eq. (1)	RQ2
	UD_t	Logarithm of the unrelated technology diversity of a firm in year t $UTD_t = \sum_{j=1}^{m} P_{jt} \ln\left(\frac{1}{P_{jt}}\right)$ Eq. (2)	RQ2
	RD_t	Logarithm of the related technology diversity of a firm in year t $RD_t = TD_t - UTD_t$ Eq. (3)	RQ2
Moderating Variable	$FundStg_t$	Dummy = 1 if funding rounds are in the early stages (i.e., Seed, Series A, B) in year t	RQ1, RQ2
	FirmType	Dummy = 1 if an industry-specific AI firm	RQ1, RQ2
	$PatentCnt_t$	Logarithm of cumulative number of patent application in year t	RQ2
Control Variable	Age_t	Logarithm of the age of a firm in year t	RQ1, RQ2
	Size	Logarithm of the total number of employees	RQ1
	Investor	Logarithm of the total number of investors	RQ1

4.4 Analysis and results

4.4.1 Patent activity and investment funding

This section presents the results of the relationship between patent activity and funding amounts to answer the first question of this study.

First, this study presented the results of the multivariate regression analysis of the relationship between the total funding amounts of firms with patents compared to those without patents. Firms without patents were the baseline in the variable of PatentYN, and the result showed that firms with patents had relatively high total funding amounts compared to firms without patents, as shown in Table 4-4. A firm in the early stages of funding was the baseline for the variable of FundStg, and this result showed a relatively high funding amount in the later stages of financing compared to the early stages. The variable FirmType indicated that cross-industry AI startups had relatively high funding amounts compared to industry-specific AI startups. The interaction term between PatentYN and FundStg showed a negative sign. This result indicated that patent holding had relatively less of an effect on the funding amount in the later stage compared to the early stage. Meanwhile, the interaction term between PatentYN and FirmType was not statistically significant, whereas the results showed a negative sign. These results indicated that holding a patent had relatively less of an effect on funding amounts in cross-industry firms compared to industry-specific AI firms. Because interaction effects are known to have low statistical power (Aguinis & Gottfredson, 2010), this study added an interpretation of the meaning of the interaction terms, even if not statistically significant.

Table 4-4. Relationship between funding amounts and firms with/without patents

	Dependent Variable: FundAmt		
VARIABLES	Model 1	Model 2	Model 3
PatentYN	0.363***	0.523***	0.438***
	(0.0852)	(0.107)	(0.102)
FundStg	0.828***	1.065***	0.826***
	(0.110)	(0.146)	(0.110)
FirmType	0.0820	0.0980	0.205*
	(0.0833)	(0.0831)	(0.123)
PatentYN × FundStg		-0.407**	
		(0.167)	
PatentYN × FirmType			-0.227
			(0.167)
Age	-0.382***	-0.380***	-0.371***
	(0.103)	(0.103)	(0.103)
Size	0.632***	0.632***	0.635***
	(0.0393)	(0.0391)	(0.0394)
Investors	0.325***	0.323***	0.330***
	(0.0576)	(0.0574)	(0.0577)
Constant	14.44***	14.36***	14.35***
	(0.259)	(0.259)	(0.266)
Observations	505	505	505
R-squared	0.671	0.675	0.672

Note. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

A second analysis was additionally conducted to reveal more specifically the relationship between the number of patents and funding amounts, targeting the firms with patents in the first analysis. The panel regression analysis utilized using variables by year to understand the relationship between the number of patents and the funding amounts. In the result of all model verifications, the VIF scores were all less than 10 (see Appendix A3-2, #2), meaning that multicollinearity was not confirmed. For the panel regression suitability test, the Hausman test was utilized to compare the fixed-effect model and the random-effect model. In the results, the fixed-effect analyses were adequate and thus the fixed-effect analysis was conducted on each of the models.

Table 4-5 shows the results of the second analysis. The relationship between PatentCount and FundAmt was positive. However, from the results of the interaction term between PatentCnt and FundStg showed a negative sign, indicating that PatentCount had relatively less of an effect on FundAmt in the later stages. Consequently, in AI startups with patents, it was found that an increase in the patent count was positively linked to an increase in the funding amount; interestingly, the effect of the patent count decreased in the later stages. The interaction term between PatentCount and FirmType was not significant statistically, but this result showed a negative sign, which indicated that the patent count had relatively less of an effect on the funding amount for cross-industry AI firms.

Table 4-5. Relationship between funding amounts and patent counts

	Dependent Variable: $FundAmt_t$		
VARIABLES	Model 1	Model 2	Model 3
$PatentCount_{t+2}$	0.109***	0.151***	0.126***
	(0.0346)	(0.0402)	(0.0409)
$FundStg_t$	0.625***	0.788***	0.630***
	(0.0702)	(0.105)	(0.0705)
FirmType	-	-	-
$PatentCount_{t+2} \times FundStg_t$		-0.0947**	
		(0.0457)	
$PatentCount_{t+2} \times$ FirmType			-0.0414
			(0.0526)
Age_t	1.483***	1.471***	1.481***
	(0.0608)	(0.0610)	(0.0608)
Constant	14.09***	14.06***	14.09***
	(0.0816)	(0.0832)	(0.0816)
Observations	1,260	1,260	1,260
R-squared	0.646	0.648	0.647
Number of firms	265	265	265

Note. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

4.4.2 Patent diversity and investment funding

This section examines the relationship between patent diversity levels and funding amounts, referring to the second question of this study.

The first analysis examined the relationship between technology diversity and the funding amount with a panel regression analysis. All models were assessed in terms of the VIF score, and all such scores were under 10 without multicollinearity (see Appendix A3-2, #3). A fixed-effect analysis was conducted according to the results of the Hausman test.

Table 4-6 shows the result of the first analysis of this section. The result showed that TD was positively related to FundAmt. All interaction terms for TD were not statistically significant. However, for the interaction term between TD and PatentCount, the result showed a negative direction, indicating that greater technology diversity could have a negative effect on the funding amount for cases with identical patent counts. Additionally, for the interaction term between TD and FirmType, the result showed a positive direction, indicating that the diversity of cross-industry startups had more of an effect on the funding amount compared to this effect for industry-specific AI startups.

Table 4-6. Relationship between funding amounts and technology diversity

	Dependent Variable: $FundAmt_{t}$			
VARIABLES	Model 1	Model 2	Model 3	Model 4
TD_{t+2}	0.133	0.249**	0.138	0.103
	(0.0935)	(0.119)	(0.0978)	(0.112)
$PatentCount_{t+2}$	0.161***	0.212***	0.161***	0.158***
	(0.0608)	(0.0687)	(0.0608)	(0.0610)
$FundStg_{t}$	0.527***	0.526***	0.530***	0.527***
	(0.0840)	(0.0839)	(0.0857)	(0.0840)
FirmType	-	-	-	-
$TD_{t+2} \times PatentCount_{t+2}$		-0.0848		
		(0.0533)		
$TD_{t+2} \times FundStg_{t}$			-0.0164	
			(0.0874)	
$TD_{t+2} \times$ FirmType				0.0641
				(0.134)
Age_{t}	1.430***	1.418***	1.429***	1.432***
	(0.0886)	(0.0888)	(0.0887)	(0.0888)
Constant	14.21***	14.16***	14.21***	14.21***
	(0.135)	(0.138)	(0.136)	(0.136)
Observations	825	825	825	825
R-squared	0.636	0.638	0.636	0.637
Number of firms	196	196	196	196

Note. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

To understand technology diversity specifically, technology diversity was divided into unrelated diversity (UD) and related diversity (RD), as noted above. The second analysis examined the relationship between technology diversity (i.e., UD and RD) and the funding amount via a panel regression analysis.

Table 4-7 shows the result of this analysis. From model 1, which was baseline, UD was positively related to FundAmt at a statistically significant level, whereas RD had a positive relationship, but this was not statistically significant. Models 2 to 4 showed the results of the moderating effect of PatentCount, FundStg, and FirmType, respectively. From model 2, the result showed that RD was positively related to FundAmt. Meanwhile, PatentCount negatively moderated the relationship between RD and FundAmt. However, PatentCount positively moderated this with regard to the relationship between UD and FundAmt. From model 3, though the interaction terms of this model were not statistically significant, FundStg positively moderated the relationship between RD and FundAmt, whereas it negatively moderated this relationship in the case of UD. From model 4, FirmType positively moderated the relationship between RD and FundAmt, showing the same direction for UD and FundAmt.

Table 4-7. Relationship between funding amount and unrelated/related diversity

	Dependent Variable: FundAmt $_t$			
VARIABLES	Model 1	Model 2	Model 3	Model 4
UD $_{t+2}$	0.494***	0.131	0.567***	0.246
	(0.138)	(0.248)	(0.151)	(0.160)
RD $_{t+2}$	0.119	0.583***	0.0280	0.00520
	(0.113)	(0.198)	(0.136)	(0.122)
PatentCount $_{t+2}$	0.0887	0.129*	0.0742	0.0814
	(0.0736)	(0.0754)	(0.0742)	(0.0724)
FundStg $_t$	0.582***	0.575***	0.598***	0.574***
	(0.108)	(0.109)	(0.109)	(0.107)
FirmType	-	-	-	-
UD $_{t+2}$ × PatentCount $_{t+2}$		0.202**		
		(0.0974)		
RD $_{t+2}$ × PatentCount $_{t+2}$		-0.260***		
		(0.0887)		
UD $_{t+2}$ × FundStg $_t$			-0.153	
			(0.131)	
RD $_{t+2}$ × FundStg $_t$			0.185	
			(0.135)	
UD $_{t+2}$ × FirmType				0.651**
				(0.268)
RD $_{t+2}$ × FirmType				0.345*
				(0.196)
Age $_t$	1.279***	1.262***	1.300***	1.288***
	(0.112)	(0.111)	(0.114)	(0.111)
Constant	14.93***	14.88***	14.91***	14.98***
	(0.236)	(0.243)	(0.239)	(0.233)
Observations	516	516	516	516
R-squared	0.623	0.632	0.626	0.638
Number of firms	140	140	140	140

Note. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

4.5 Discussion and conclusion

This study investigates AI firms to understand technological pervasiveness on the supply side. Previous empirical studies of GPT from a corporate aspect mainly presented discussions of productivity and efficiency improvements with regard to the user side for technologies. However, for emerging technologies, it is difficult to measure the results of effect on these technologies, and the results have not been sufficiently observed thus far. Therefore, in order to understand the technological pervasiveness of emerging technologies from a new vantage point, this study explores the technology portfolio strategies of AI firms considering the supply side of the technologies.

Specifically, this study investigates the relationship between the patent strategies and investments of AI startups, as shown in Table 4-8. First, the results show a positive relationship between the patent count and the funding amount. However, in the later stages of AI startups, the results show that the effect of the patent count on the funding amount decreased, with other business performance metric possibly being more important to investors. Extending earlier works about the relationship between patent counts and financing rounds in the literature (Hoenen et al., 2014; Caviggioli et al., 2020; Mann & Sager, 2007), this study finds empirical evidence of less of an effect during later stage in the case of AI startups. Additionally, it is found that cross-industry AI firms with patents negatively moderates the relationship between the patent count and funding amount. In other words, for investors, the visible output of technology capabilities plays a more important role for investments in the case of industry-specific AI firms compared to cross-industry AI startups. Second, technology diversity has a positive relationship with funding amounts. When patent counts are identical, related diversity negatively moderates the effect on the funding amount. From the literature, the number of IPC codes of patents showed a negative effect on VC amounts, and these results proposed a stronger preference by investors for more focused innovations (Caviggioli et al., 2020). However, the results of this study show that unrelated diversity has a stronger positive effect on funding amounts, meaning that the value as a platform role with wide scope of technological capabilities acts as an attractive factor to those investing in AI startups. Furthermore, regarding the firm type, the results show that cross-industry AI firms have a positively effect on the relationship between related diversity and funding amounts compared to industry-specific AI startups, with the same results found for unrelated diversity. A wide range of technology diversity for cross-industry AI firms are found to be more attractive to investors compared to that in industry-specific AI firms.

Table 4-8. Summary of the results

Independent Variable	Dependent Variable	Moderating Variable		
	Funding Amount	Patent Count	Funding Stage	Firm Type
Patent YN	**(+)**		**later stage (-)**	cross-industry AI (-)
Patent Count	**(+)**		**later stage (-)**	cross-industry AI (-)
Technology Diversity	**(+)**			
Unrelated Diversity	**(+)**	**(+)**	later stage (-)	**cross-industry AI (+)**
Related Diversity	**(+)**	**(-)**	later stage (+)	**cross-industry AI (+)**

In particular, the results showing a positive relationship between technology diversity (i.e., unrelated diversity and related diversity) and startup investment as confirmed in this study suggest that the technology diversity of AI startups acts as a driving force for an innovation to become a GPT from the perspective of the overall growth of the mechanisms acting in the AI sector. An increase in technology diversity causes innovations on the technology supply side through the combination of different types of knowledge. This appears because disruptions can occur from the supply side, changing the architecture that determines how technologies are linked rather than the technologies themselves, with innovations occurring as a result of new combinations among existing ideas and technologies (Gans, 2016; Kurz, 2012). Thus, technology diversity in AI startups, which is serves as disruptive innovations, can accelerate disruptive innovations, facilitating to AI to become a GPT. Also, technology diversity provides various opportunities to generate disruptive innovations in a wider range through technological pervasiveness. Meanwhile, previous studies found that venture capital investments have a positive impact on spurring innovations and on macro and micro levels of economic growth (Cheng et al., 2019; Pradhan et al., 2018; Khan et al., 2021). This study confirms that there is a positive relationship between technology diversity and investment funding, which represents the financial resources needed for innovations. Consequently, the results of this study imply that the technology diversity of AI startups can act as the driving force for disruptive innovations in various areas during the growth of AI as a GPT.

This study makes a theoretical contribution to the technology diversity of startups in the field of AI. Considering the properties of GPTs inherent in AI technology, there is a need to realize a greater scope of technology to extend advantageous positions even in the case of small firms. Generally, the importance of technology diversity has been investigated from the perspective of the distributed competences of large firms, and previous research suggests that a firm without a high level of technology concentrates on a small number of innovations and on a specific field of core technology (Granstrand et al., 1997; Lin et al., 2006). However, another study suggests that a broader scope of patents can positively influence the growth, survival, and innovation of a startup (Hegde et al., 2022). Regarding AI, considering the nature of technological pervasiveness, AI should be extended to various areas, and this study provides empirical evidence of the importance of technology diversity at AI startups over both the short and long term. Additionally, from a risk pooling point of view, patent portfolio diversity can be effective for AI startups considering the innovative nature of startups and the disruptive properties of AI technology.

From a business perspective, startups must use different technology strategies depending on their firm type. For cross-industry AI firms, the technology diversity strategy should be based on the competitive core competencies of the company, and they should expand the scope of application of their technologies. The results here show that both related and unrelated technology types of diversity are positively related to investment funding in cross-industry AI firms. Thus, it is necessary to strengthen core technological capabilities based on related diversity while also bolstering the expansion of the firm into various industries based on unrelated diversity. For industry-specific AI firms, priority should be assigned to providing the core AI technology capabilities of the firm rather than focusing on expansion or across-industry diversity. According to the results of the study, the effect of the number of patents on investments is higher than that of cross-industry AI firms, but the total amount of investment is lower than that of cross-industry AI firms. Thus, if industry-specific AI firms sufficiently demonstrate their AI technology capabilities, they can find a comparative advantage with regard to funding investment.

For AI firms, in the short term, cross-industry AI firms need technology diversity across industries, and industry-specific AI firms need technology diversity within an industry. In a long term, AI firms should pursue growth as platform-based companies. The literature shows the technology diversity has a positive relationship with generality, and higher generality enables more diversified technological trajectories of firms to act as a platform (Corradini et al., 2016; Kim & Kogut, 1996). Hence, AI startups should increase the generality of their own technologies by increasing technology diversity and find new business opportunities to establish a platform ecosystem based on their core competences.

From a policy perspective, there is a need the funding for startups that is not for profit-seeking purposes. This study finds with regard to the patterns of startup investment shows that investors prefer to minimize the risk of their investments. It is also found that investments in related diversity are more positive in terms of the funding amount in the later stages compared to the early stages. This indicates that investors prefer the startups that seek stability in the later stages. Also, investment funding in industry-specific AI firms is relatively low compared to that in cross-industry AI firms. This shows that investors are more interested in the development of AI technology rather than the expansion of AI to other industries. However, depending on the growth stage of the startup, the firm should expand their capabilities to various industries and business models in the later stages, and financial support for these adventurous attempts is needed. Also, in order for a certain AI technology to grow as a GPT, startups in various industries related to AI must grow. Governments should not only support the revitalization of investments by private investors of VCs and CVCs but also need to establish policies to ensure a sustainable ecosystem of startups with complementing the tendency toward private benefit-oriented investments by venture capital industries.

The limitations of this study and directions of future research are as follows. First, this study lacks an investigation of differences depending on exit routes and various statuses of startups. Among startups, there are companies with various statuses, such as companies that have achieved successful exit routes (i.e., M&A and IPO) and startups with high corporate value (i.e., unicorns). This study attempted to identify a differentiated pattern according to these statuses, but companies with the IPO or unicorn status were not numerous enough to find statistical meaning in this study, as the total sample size was too small. Thus, for future research targeting AI IPOs or unicorns, it is recommended to discover new insights through simulations or case studies that are not limited by the number of samples. Second, the main limitation of this study is there is a lack of AI startups overall with patents, and many AI startups with few patent counts as well. Among all firms used in this study, there were difficulties in securing a certain level of data, as the total sample size was greatly reduced during the process of verifying the detailed conditions and when refining each research question of the study. At the current stage, the numbers of patents or the growth periods of firms are insufficient for strategic discussions of AI startups. Thus, in future research, it is recommended to expand the research scope by comparing AI startups among different industries or by comparing them among different countries.

Chapter 5. Conclusion

This dissertation investigates the various aspects of the AI sector from the perspective of technological pervasiveness. Empirical evidence of the technological pervasiveness of AI is examined through the theoretical perspectives of technology diffusion, convergence, and strategies considering different levels of analysis of knowledge flows, industrial sectors, and technology portfolios.

Chapter two examines the patterns of AI technology diffusion and development, focusing on a patent index. The patent index as used here is interpreted from the context of the process of knowledge recombinations and the development of GPTs. The theoretical contribution here is that the present study identifies the dynamics of GPT-related features from the perspective of technology diffusion theory. From a methodological point of view, the DTW applied in this study is a new methodological approach to understand the patterns of technology diffusion, making a very useful contribution to technology diffusion research overall. This study suggests implications on policy perspectives regarding AI as a GPT.

Chapter three investigates the technology convergence of AI in terms of industrial sectors, technology categories, and the utilization of the technology. The theoretical contribution of this study is that it presents a new framework by which to understand the evidence pertaining to the industrial applicability of GPTs considering horizontal and vertical applicability perspectives. Additionally, the important methodological contribution is in the combination analysis of supervised learning and unsupervised learning in relation to patents, which is meaningful as it presents methodological results complementary to the empirical analysis. This study proposes implications for businesses that are AI companies and policy perspectives for AI development.

Chapter four analyzes the technology portfolios of AI startups, focusing on the diversity of their technologies. This study considers the technology diversity of startups as the source of disruptive innovations on the supply side and technological pervasiveness in various areas. From the empirical analysis, this study finds a positive relationship between the technology diversity and startup investments, which represent the driving force behind the growth of startups. This makes a theoretical contribution to the research on technology diversity strategies of startups in relation to GPTs. This study provides implications regarding technology portfolio strategies for AI startups and policies pertaining to AI startup investments.

This dissertation contains several limitations and suggests directions for future studies. First, this study considers the impact of AI in terms of technological pervasiveness, though it does not identify the impact of AI adoption or the market and economic sides as well. To understand the potential of a GPT thoroughly, this dissertation suggests future studies related to the spillover effects of user adoption, the market side, and economic value beyond patent data. Also, future studies can analyze business impacts regarding productivity across various sectors. Second, the patent data in this study are only from the US. Despite the fact that US is the representative and leading country of AI technologies, there is a limitation to understanding the worldwide technological progress or development trends of AI here. Therefore, data from other countries should be considered in future studies. Third, despite the great impact of large global IT companies in the AI sector, there is no consideration of such companies in this study. Identifying the strategic directions of AI technologies in large global companies should be considered as an additional future study.

Bibliography

Acemoglu, D., Autor, D. Hazell, J. & Restrepo, P. (2022). AI and jobs: Evidence from online vacancies. *National Bureau of Economic Research,* NBER Working Paper No. 28257.

Alderucci, D., Branstetter, L, Hovy, E., Runge, A., & Zolas, N. (2019). Quantifying the Impact of AI on Productivity and Labor Demand: Evidence from U.S. Census Microdata.

Aguinis, H., & Gottfredson, R. K. (2010). Best‐practice recommendations for estimating interaction effects using moderated multiple regression. *Journal of Organizational Behavior*, 31(6), 776-786.

Altuntas, S., Dereli, T., & Kusiak, A. (2015). Forecasting technology success based on patent data. *Technological Forecasting & Social Change*, 96, 202-214.

Andersen, B. (1999). The hunt for S-shaped growth paths in technological innovation: A patent study. *Journal of Evolutionary Economics*, 9(4), 487-526.

Appio, F. P., de Luca, L. M., Morgan, R., & Martini, A. (2019). Patent portfolio diversity and firm profitability: A question of specialization or diversification? *Journal of Business Research*, 101, 255-267.

Atherton, A. (2012). Cases of start‐up financing: An analysis of new venture capitalisation structures and patterns. *International Journal of Entrepreneurial Behaviour & Research*, 18(1), 28-47.

Athreye, S., & Keeble, D. (2000). Technological convergence, globalisation and ownership in the UK computer industry. *Technovation*, 20(5), 227-245.

Audretsch, D B.., Bönte, W., & Mahagaonkar, P. (2012). Financial signaling by innovative nascent ventures: The relevance of patents and prototypes. *Research Policy*, 41(8), 1407-1421.

Ayres, R. U. (1994). Toward a non-linear dynamics of technological progress. *Journal of Economic Behavior & Organization*, 24(1), 35-69.

Baek, S. D., Kim, K. & Altmann, J. (2014). Role of Platform Providers in Service Networks: The Case of Salesforce.com App Exchange. *2014 IEEE 16th Conference on Business Informatics*, 39-45.

Barney, J. (1991). Firm Resources and Sustained Competitive Advantage. *Journal of Management*, 17(1), 99-120.

Baker, K. R., Magazine, M. J., & Nuttle, H. L. W. (1986). The Effect of Commonality on Safety Stock in a Simple Inventory Model. *Management Science*, 32(8), 982-988.

Bianco, S., Zach, F. J., & Liu, A. (2022). Early and late-stage startup funding in hospitality: Effects on incumbents' market value. *Annals of Tourism Research*, 95, 103436.

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

Bock, C., & Hackober, C. (2020). Unicorns—what drives multibillion-dollar valuations? *Business Research*, 13(3), 949-984.

Borgatti, S., Everett, M., & Freeman, L. (2002). *Ucinet 6 for Windows: Software for Social Network Analysis*. Harvard, MA: Analytic Technologies.

Borgatti, S., Everett, M., & Johnson, J. (2013). *Analyzing social networks*. SAGE Publications: Thousand Oaks. California. USA.

Bresnahan, T. F., & Trajtenberg, M. (1995). General purpose technologies 'Engines of growth'? *Journal of Econometrics*, 65(1), 83-108.

Brown, K. C., & Wiles, K. W. (2015). In Search of Unicorns: Private IPOs and the Changing Markets for Private Equity Investments and Corporate Control. *Journal of Applied Corporate Finance*, 27(3), 34-48.

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.

Brynjolfsson, E, Rock, D. & Syverson, C. (2021). The Productivity J-Curve: How Intangibles Complement General Purpose Technologies. *American Economic Journal: Macroeconomics*, 13 (1): 333-72.

Burnham, K. P. & Anderson, D. R. (2002). *Model Selection and Multi-model Inference: A Practical Information-theoretic Approach*. Springer-Verlag New York.

Caviggioli, F., Colombelli, A., de Marco, A., & Paolucci, E. (2020). How venture capitalists evaluate young innovative company patent portfolios: Empirical evidence from Europe. *International Journal of Entrepreneurial Behaviour & Research*, 26(4), 695-721.

Cavill, R., Kleinjans, J., & Briede, J. (2013). DTW4Omics: Comparing Patterns in Biological Time Series. *PloS One*, 8(8), E71823.

Cecere, G., Corrocher, N., Gossart, C., & Ozman, M. (2014). Technological pervasiveness and variety of innovators in Green ICT: A patent-based analysis. *Research Policy*, 43(10), 1827-1839.

Chang, S., Chang, S., & Guh, W. (2007). Exploring the technology diffusion trajectories and groups of basic patents of business methods: Using the patent citation network. *PICMET '07: Portland International Conference on Management of Engineering & Technology*, 1784-1789.

Chatterjee, S., Rana, N. P., Tamilmani, K., & Sharma, A. (2021). The effect of AI-based CRM on organization performance and competitive advantage: An empirical analysis in the B2B context. *Industrial Marketing Management*, 97, 205-219.

Chen, J. H., Jang, S., & Wen, S. H. (2010). Measuring technological diversification: Identifying the effects of patent scale and patent scope. *Scientometrics*, 84(1), 265-275.

Chen, Y., Shih, C., & Chang, C. (2012). The effects of related and unrelated technological diversification on innovation performance and corporate growth in the Taiwan's semiconductor industry. *Scientometrics*, 92(1), 117-134.

Cheng, A. (2012). Exploring the relationship between technology diffusion and new material diffusion: The example of advanced ceramic powders. *Technovation*, 32(3-4), 163-167.

Cheng, C., Sun, Y., Su, Y., & Yang, S. (2019). Venture capital, innovation, and growth: Evidence from Chinese metropolitan data. *Applied Economics Letters*, 26(7), 549-553.

Choi, J. Y., Jeong, S., & Kim, K. (2015). A study on diffusion pattern of technology convergence: Patent analysis for Korea. *Sustainability*, 7(9), 11546-11569.

Cockburn, I. M., Henderson, R. & Stern, S. (2018). The impact of artificial intelligence on innovation. *National Bureau of Economic Research,* NBER Working Paper No. 24449.

Coleman, S., Cotei, C., & Farhat, J. (2013). A resource-based view of new firm survival: New perspectives on the role of industry and exit route. *Journal of Developmental Entrepreneurship*, 18(1), 1.

Corradini, C., Demirel, P., & Battisti, G. (2016). Technological diversification within UK's small serial innovators. Small Business Economics, 47(1), 163-177.

Cotei, C., & Farhat, J. (2018). The M&A exit outcomes of new, young firms. *Small Business Economics*, 50(3), 545-567.

Cozzens, S., Gatchair, S., Kang, J., Kim, K., Lee, H. J., Ordóñez, G., & Porter, A. (2010). Emerging technologies: Quantitative identification and measurement. *Technology Analysis & Strategic Management*, 22(3), 361-376.

Crafts, N. (2004). Steam as a general purpose technology: A growth accounting perspective. *The Economic Journal*, 114(495), 338-351.

Crafts, N. (2021). Artificial intelligence as a general-purpose technology: an historical perspective. *Oxford Review of Economic Policy*, 37(3): 521-536.

Curran, C., Bröring, S., & Leker, J. (2010). Anticipating converging industries using publicly available data. *Technological Forecasting & Social Change*, 77(3), 385-395.

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

Daim, T. U., Rueda, G., Martin, H., & Gerdsri, P. (2006). Forecasting emerging technologies: Use of bibliometrics and patent analysis. *Technological Forecasting & Social Change*, 73(8), 981-1012.

Davila, A., Foster, G., & Gupta, M. (2003). Venture capital financing and the growth of startup firms. *Journal of Business Venturing*, 18(6), 689-708.

Debackere, K., Verbeek, A., Luwel, M., & Zimmermann, E. (2002). Measuring progress and evolution in science and technology - II: The multiple uses of technometric indicators. *International Journal of Management Reviews : IJMR*, 4(3), 213-231.

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.

Dosi, G. (1982). Technological paradigms and technological trajectories: A suggested interpretation of the determinants and directions of technical change. *Research Policy*, 11(3), 147-162.

Dushnitsky, G., & Lenox, M. J. (2005). When do incumbents learn from entrepreneurial ventures?: Corporate venture capital and investing firm innovation rates. *Research Policy*, 34(5), 615-639.

Ernst, H. (1997). The Use of Patent Data for Technological Forecasting: The Diffusion of CNC-Technology in the Machine Tool Industry. *Small Business Economics*, 9(4), 361-381.

Esteve-Pérez, S. & Mañez-Castillejo, J. (2008). The Resource-Based Theory of the Firm and Firm Survival. *Small Business Economics*, 30(3), 231-249.

Falk, N., & Train, K. (2017). Patent Valuation with Forecasts of Forward Citations. *Journal of Business Valuation and Economic Loss Analysis*, 12(1), 101-121.

Feldman, M. P., & Yoon, J. W. (2012). An empirical test for general purpose technology: An examination of the Cohen-Boyer rDNA technology. *Industrial and Corporate Change*, 21(2), 249-275.

Flynn, D., & Forman, A. M. (2001). Life cycles of new venture organizations: Different factors affecting performance. *Journal of Developmental Entrepreneurship*, 6(1), 41.

Franses, P. H., & Wiemann, T. (2020). Intertemporal Similarity of Economic Time Series: An Application of Dynamic Time Warping. *Computational Economics*, 56(1), 59-75.

Freeman, L. C. (1979). Centrality in social networks conceptual clarification. *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.

Fujii, H., & Managi, S. (2019). Decomposition analysis of sustainable green technology inventions in China. *Technological Forecasting & Social Change*, 139, 10-16.

Gans, J. (2016). The Disruption Dilemma. MIT Press.

Gao, L., Porter, A. L., Wang, J., Fang, S., Zhang, X., Ma, T., Wang, W. & Huang, L. (2013). Technology life cycle analysis method based on patent documents. *Technological Forecasting & Social Change*, 80(3), 398-407.

Garcia-Vega, M. (2006). Does technological diversification promote innovation?: An empirical analysis for European firms. *Research Policy*, 35(2), 230-246.

Gbadegeshin, S. A., Natsheh, A. A., Ghafel, K., Mohammed, O., Koskela, A., Rimpiläinen, A., Tikkanen, J. & Kuoppala, A. (2022). Overcoming the Valley of Death: A New Model for High Technology Startups. *Sustainable Futures*, 4, 100077.

Girasa, R. (2020). *Artificial Intelligence as a Disruptive Technology: Economic Transformation and Government Regulation.* Cham: Springer International Publishing (1st ed.)

Goldfarb, A., Taska, B. & Teodoridis, F. (2023). Could machine learning be a general purpose technology? A comparison of emerging technologies using data from online job postings. *Research Policy*, 52, 104653.

Graf, H., & Menter, M. (2022). Public research and the quality of inventions: The role and impact of entrepreneurial universities and regional network embeddedness. *Small Business Economics*, 58(2), 1187-1204.

Granstrand, O., Patel, P., & Pavitt, K. (1997). Multi-Technology Corporations: Why They Have "Distributed" Rather Than "Distinctive Core" Competencies. *California Management Review*, 39(4), 8-25.

Greenberg, G. (2013). Small Firms, Big Patents? Estimating Patent Value Using Data on Israeli Start-ups' Financing Rounds. *European Management Review*, 10(4), 183-196.

Gries, T. & Naudé, W. (2020). Artificial Intelligence, Income Distribution and Economic Growth. *IZA Discussion Paper* No. 13606.

Hacklin, F., Raurich, V., & Marxt, C. (2005). Implications of technological convergence on innovation trajectories: the case of ICT industry. *International Journal of Innovation and Technology Management*, 2(3), 313-330.

Hagedoorn, J., & Cloodt, M. (2003). Measuring innovative performance: Is there an advantage in using multiple indicators? *Research Policy*, 32(8), 1365-1379.

Hall, B. H. & Trajtenberg, M. (2004). Uncovering GPTS with patent data. *National Bureau of Economic Research*, NBER Working Paper No. 10901.

Hall, B., Helmers, C., Rogers, M., & Sena, V. (2014). The Choice between Formal and Informal Intellectual Property: A Review. *Journal of Economic Literature*, 52(2), 375-423.

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

Harhoff, D., Narin, F., Scherer, F. M., & Vopel, K. (1999). Citation Frequency and the Value of Patented Inventions. *The Review of Economics and Statistics*, 81(3), 511-515.

Haupt, R., Kloyer, M., & Lange, M. (2007). Patent indicators for the technology life cycle development. *Research Policy*, 36(3), 387-398.

Hautamaki, V., Nykanen, P., & Franti, P. (2008). Time-series clustering by approximate prototypes. *2008 19th International Conference on Pattern Recognition*, 1-4.

Hegde, D., Ljungqvist, A., & Raj, M. (2022). Quick or Broad Patents? Evidence from U.S. Startups. *The Review of Financial Studies*, 35(6), 2705-2742.

Helpman, E. (1998), *General Purpose Technologies and Economic Growth.* MIT Press: Cambridge, MA

Hoenen, S., Kolympiris, C., Schoenmakers, W., & Kalaitzandonakes, N. (2014). The diminishing signaling value of patents between early rounds of venture capital financing. *Research Policy*, 43(6), 956-989.

Hötte, K. Tarannum, T., Verendel, V. Bennett, L. (2022). Exploring artificial intelligence as a general purpose technology with patent data: A systematic comparison of four classification approaches. arXiv:2204.10304.

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

Hsu, D. H., & Ziedonis, R. H. (2013). Resources as dual sources of advantage: Implications for valuing entrepreneurial-firm patents. *Strategic Management*

Journal, 34(7), 761-781.

Huang, Y., Li, R., Zou, F., Jiang, L., Porter, A. L., & Zhang, L. (2022). Technology life cycle analysis: From the dynamic perspective of patent citation networks. *Technological Forecasting & Social Change*, 181, 121760.

Issa, H. Jabbouri, R. & Palmer, M. (2022) An artificial intelligence (AI)-readiness and adoption framework for AgriTech firms. *Technological Forecasting & Social Change*. 182. 121874.

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

Jacobs, M. A., & Swink, M. (2011). Product portfolio architectural complexity and operational performance: Incorporating the roles of learning and fixed assets. *Journal of Operations Management*, 29(7), 677-691.

Jaffe, A. B. & de Rassenfosse, G. (2017). Patent citation data in social science research: Overview and best practices. *Journal of the Association for Information Science and Technology*. 68: 1360-1374.

Jeng, L. A., & Wells, P. C. (2000). The determinants of venture capital funding: Evidence across countries. *Journal of Corporate Finance*, 6(3), 241-289.

Jun, B. H. (2011). Fault detection using dynamic time warping (DTW) algorithm and discriminant analysis for swine wastewater treatment. *Journal of Hazardous Materials*, 185(1), 262-268.

Kassicieh, S. K., Kirchhoff, B. A., Walsh, S. T., & McWhorter, P. J. (2002). The role of small firms in the transfer of disruptive technologies. *Technovation*, 22(11), 667-674.

Kato, M., Onishi, K., & Honjo, Y. (2022). Does patenting always help new firm survival? Understanding heterogeneity among exit routes. *Small Business Economics*, 59(2), 449-475.

Kayal, A. (1999). Measuring the Pace of Technological Progress: Implications for Technological Forecasting. *Technological Forecasting & Social Change*, 60(3), 237-245.

Khan, N., Qu, H., Qu, J., Wei, C., & Wang, S. (2021). Does Venture Capital Investment Spur Innovation? A Cross-Countries Analysis. *SAGE Open*, January-March 2021, 1-13.

Khurana, I., & Farhat, J. (2021). The timing of diversification and startup firms' survival: A resource-based perspective. *Industry and Innovation*, 28(10), 1249-1269.

Kim, D., & Kogut, B. (1996). Technological Platforms and Diversification. *Organization Science*, 7(3), 283-301.

Kim, E., Cho, Y., & Kim, W. (2014). Dynamic patterns of technological convergence in printed electronics technologies: Patent citation network. *Scientometrics*, 98(2), 975-998.

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(1), 47-65.

Kim, S. H., Lee, H. S., Ko, H. J., Jeong, S. H., Byun, H. W., & Oh, K. J.(2018). Pattern matching trading system based on the dynamic time warping algorithm. *Sustainability*, 10(12), 4641.

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

Klinger, J., Mateos-Garcia, J. & Stathoulopoulos, K. (2018), Deep learning, deep change? Mapping the development of the Artificial Intelligence General Purpose Technology. arXiv:1808.06355.

Kogut, B., & Zander, U. (1992). Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology. *Organization Science*, 3(3), 383-397.

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

Kurz, H. D. (2012). Schumpeter’s new combinations. *Journal of Evolutionary Economics* 22(5), 871- 899.

Kwon, O. K., An, Y., Kim, M., & Lee, C. (2020). Anticipating technology-driven industry convergence: Evidence from large-scale patent analysis. *Technology Analysis & 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(3), 925-942.

Lee, C., Kim, J., Kwon, O., & Woo, H. (2016a). Stochastic technology life cycle analysis using multiple patent indicators. *Technological Forecasting & Social Change*, 106, 53-64.

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

Lee, C., Kim, J., Noh, M., Woo, H., & Gang, K. (2017). Patterns of technology life cycles: Stochastic analysis based on patent citations. *Technology Analysis & Strategic Management*, 29(1), 53-67.

Lee, W. S., Choi, H. S., & Sohn, S. Y. (2018). Forecasting new product diffusion using both patent citation and web search traffic. *PloS One*, 13(4), E0194723.

Lee, D., & Lin, K. (2020). How to transform sustainable energy technology into a unicorn start-up: Technology review and case study. *Sustainability*, 12(7), 3018.

Lee, S., Hwang, J., & Cho, E. (2022). Comparing technology convergence of artificial intelligence on the industrial sectors: two-way approaches on network analysis and clustering analysis. *Scientometrics*. 127:407-452.

Lerner, J. (1994). The Importance of Patent Scope: An Empirical Analysis. *The Rand Journal of Economics*, 25(2), 319-333.

Li, H. (2015). On-line and dynamic time warping for time series data mining. *International Journal of Machine Learning and Cybernetics*, 6(1), 145-153.

Li, J. Herdem, M. S., Nathwani, J. & Wen, J. Z. (2023) Methods and applications for Artificial Intelligence, Big Data, Internet of Things, and Blockchain in smart energy management, *Energy and AI.* 11. 100208.

Liao, T. W., Ting, C., & Chang, P. (2006). An adaptive genetic clustering method for exploratory mining of feature vector and time series data. *International Journal of Production Research,* 44(14), 2731-2748.

Lin, B., Chen, C., & Wu, H. (2006). Patent portfolio diversity, technology strategy, and firm value. *IEEE Transactions on Engineering Management*, 53(1), 17-26.

Lipsey, R. D., Carlaw, K. I., & Bekar, C. T. (2005). *Economic transformations: general purpose technologies and long-term economic growth*. New York: Oxford University Press.

Liu, H. (2015). *Comparing Welch's ANOVA, a Kruskal-Wallis Test and Traditional ANOVA in Case of Heterogeneity of Variance*. Virginia Commonwealth University. VCU Scholars Compass

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. (2021a). Artificial intelligence and energy intensity in China’s industrial sector: Effect and transmission channel. *Economic Analysis and Policy,* 70, 276-293.

Liu, T., Xing, X., Zhang, Y., Song, Y., Wang, J., & Wang, S. (2021b). Achieving high efficiency and sustainability through new ventures exploration and exploitation strategies: Insight from well-established and emerging technology standards.

Sustainable Cities and Society, 74, 103201.

Lloyd, W. P., & Jahera Jr, J. S. (1994). Firm-diversification effects on performance as measured by tobin's q. *Managerial and Decision Economics*, 15(3), 259-266.

Love, H. (2016) *The Start-Up J Curve: The Six Steps to Entrepreneurial Success*. Greenleaf Book Group Press.

Lynch, K. (2021) Believing in unicorns: how to value unicorn companies and intellectual property while encouraging continued innovation and public disclosure. *Case Western reserve law review*, 72(2), 421.

Malyy, M., Tekic, Z., & Podladchikova, T. (2021). The value of big data for analyzing growth dynamics of technology-based new ventures. *Technological Forecasting & Social Change*, 169, 120794.

Mann, R. J., & Sager, T. W. (2007). Patents, venture capital, and software start-ups. *Research Policy*, 36(2), 193-208.

Martinelli, A. Mina, A. & Moggi, M. (2021). The enabling technologies of industry 4.0: examining the seeds of the fourth industrial revolution, *Industrial and Corporate Change*, 30(1): 161–188.

Macht, S. A., & Robinson, J. (2009). Do business angels benefit their investee companies? *International Journal of Entrepreneurial Behaviour & Research*, 15(2), 187-208.

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.

Merino, D. N. (1990). Development of a technological S-curve for tire cord textiles. *Technological Forecasting & Social Change*, 37(3), 275-291.

Mikalef, P., & Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Information & Management*, 58(3), 103434.

Mishra, S., Ewing, M. T. & Cooper, H. B. (2022). Artificial intelligence focus and firm performance. *J. of the Acad. Mark. Sci.*

Moser, P., & Nicholas, T. (2004). Was Electricity a General Purpose Technology? Evidence from Historical Patent Citations. *The American Economic Review*, 94(2), 388-394.

Müller, M. (2007). *Information Retrieval for Music and Motion* (1st ed.). Berlin, Heidelberg: Springer-VerlagBerlinHeidelberg.

Nadeau, P. (2010). Venture capital investment selection: Do patents attract investors? *Strategic Change*, 19(7-8), 325-342.

Nardin, A. (2021). Artificial Intelligence as a General Purpose Technology: an exploratory analysis of PCT patents (Master' dissertation). Università Ca' Foscari Venezia

National Research Council for Economics, Humanities and Social Sciences (2021). Legal countermeasures needed to promote artificial intelligence convergence by industry.

Nieto, M., Lopéz, F., & Cruz, F. (1998). Performance analysis of technology using the S curve model: The case of digital signal processing (DSP) technologies. *Technovation*, 18(6), 439-457.

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

Omoge, A. P., Gala, P., & Horky, A. (2022). Disruptive technology and AI in the banking industry of an emerging market. *International Journal of Bank Marketing*, 40(6), 1217-1247.

Oxford Insights (2020). Government AI Readiness Index 2020.

Oxford Insights (2021). Government AI Readiness Index 2021.

Oxford Insights (2022). Government AI Readiness Index 2022.

Patel, E., & Kushwaha, D. S. (2020). Clustering Cloud Workloads: K-Means vs Gaussian Mixture Model. *Procedia Computer Science,* 171, 158-167.

Petitjean, F., Ketterlin, A., & Gançarski, P. (2011). A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition*, 44(3), 678-693.

Petralia, S. (2020). Mapping general purpose technologies with patent data. *Research Policy*, 49(7), 104013.

Petralia, S. (2021). GPTs and growth: Evidence on the technological adoption of electrical and electronic technologies in the 1920s. *European Review of Economic History*, 25(3), 571-608.

Pradhan, R. P., Arvin, M. B., Nair, M., Bennett, S. E., Bahmani, S., & Hall, J. H. (2018). Endogenous dynamics between innovation, financial markets, venture capital and economic growth: Evidence from Europe. *Journal of Multinational Financial Management*, 45, 15-34.

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

Qiu, R. & Cantwell, J. (2015) Revisit the Classification of General Purpose Technologies (GPTs) in Corporate Innovation Research Using Patent and Patent Citation Data. *Journal of International Technology and Information Management*, 24(2),6.

Rabiner, J. & Juang, B. (1993). *Fundamentals of Speech Recognition,* Englewood Cliffs, N.J. : PTR Prentice Hall

Raiteri, E. (2018). A time to nourish? Evaluating the impact of public procurement on technological generality through patent data. *Research Policy*, 47(5), 936-952.

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

Rosenberg, N., & Trajtenberg, M. (2004). A General-Purpose Technology at Work: The Corliss Steam Engine in the Late-Nineteenth-Century United States. *The Journal of Economic History*, 64(1), 61-99.

Rotolo, D., Hicks, D., & Martin, B. R. (2015). What is an emerging technology? *Research Policy,* 44(10), 1827-1843.

Sahal, D. (1985). Technological guideposts and innovation avenues. *Research Policy*, 14(2), 61-82.

Schmoch, U. (2008). Concept of a Technology Classification for Country Comparison. *WIPO*. June 1-15.

Schneider, C., & Veugelers, R. (2010). On young highly innovative companies: Why they matter and how (not) to policy support them. *Industrial and Corporate Change*, 19(4), 969-1007.

Shingala, M. C., Rajyaguru, A. (2015). Comparison of post hoc tests for unequal variance. International *Journal of New Technologies in Science and Engineering*, 2(5). 22-33

Sick, N., Preschitschek, N., Leker, J., & Bröring, S. (2019). A new framework to assess industry convergence in high technology environments. *Technovation*, 84-85, 48-58.

Song, M., & di Benedetto, C. A. (2008). Supplier's involvement and success of radical new product development in new ventures. *Journal of Operations Management*, 26(1), 1-22.

Spanaki, K., Sivarajah, U., Fakhimi, M., Despoudi, S., & Irani, Z. (2022). Disruptive technologies in agricultural operations: A systematic review of AI-driven AgriTech research. *Annals of Operations Research*, 308(1-2), 491-524.

Spence, M. (1973). Job Market Signaling. *The Quarterly Journal of Economics*, 87(3), 355-374.

Spence, M. (2002). Signaling in Retrospect and the Informational Structure of Markets. *The American Economic Review,* 92(3), 434-459.

Squicciarini, M., Dernis, H., & Criscuolo, C. (2013), Measuring Patent Quality: Indicators of Technological and Economic Value. *OECD Publishing,* Technology and Industry Working Papers. No. 2013/03.

Su, H. (2018). How to analyze technology lifecycle from the perspective of patent characteristics? the cases of DVDs and hard drives. *R & D Management*, 48(3), 308-319.

Switonski, A., Josinski, H., & Wojciechowski, K. (2019). Dynamic time warping in classification and selection of motion capture data. *Multidimensional Systems and Signal Processing*, 30(3), 1437-1468.

Taylor, M., & Taylor, A. (2012). The technology life cycle: Conceptualization and managerial implications. *International Journal of Production Economics*, 140(1), 541-553.

Thoma, G. (2009). Striving for a large market: Evidence from a general purpose technology in action. *Industrial and Corporate Change*, 18(1), 107-138.

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. 1-23.

Trajtenberg, M. (1990). A Penny for Your Quotes: Patent Citations and the Value of Innovations. *The Rand Journal of Economics*, 21(1), 172-187.

Trajtenberg, M. Henderson, R., & Jaffe, A. (1992). Ivory tower versus corporate lab: an empirical study of basic research and appropriability. *National Bureau of Economic Research*. NBER Working Paper No. 4146.

Trajtenberg, M., Henderson, R., & Jaffe, A. (1997). University Versus Corporate Patents: A Window On The Basicness Of Invention. *Economics of Innovation and New Technology,* 5(1), 19-50.

Tricot, R. (2021). Venture capital investments in artificial intelligence: Analysing trends in VC in AI companies from 2012 through 2020. *OECD Publishing*, OECD Digital Economy Papers. No. 319.

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

Useche, D. (2014). Are patents signals for the IPO market? An EU–US comparison for the software industry. *Research Policy,* 43(8), 1299-1311.

Valentini, G. (2012). Measuring the effect of M&A on patenting quantity and quality. *Strategic Management Journal*, 33(3), 336-346.

van Merkerk, R. O., & van Lente, H. (2005). Tracing emerging irreversibilities in emerging technologies: The case of nanotubes. *Technological Forecasting & Social Change*, 72(9), 1094-1111.

von Wartburg, I., Teichert, T., & Rost, K. (2005). Inventive progress measured by multi-stage patent citation analysis. *Research Policy*, 34(10), 1591-1607.

Wagner, S., & Cockburn, I. (2010). Patents and the survival of Internet-related IPOs.

Research Policy, 39(2), 214-228.

Wamba-Taguimdje, S., Fosso Wamba, S., Kala Kamdjoug, J. R., & Tchatchouang Wanko, C. E. (2020). Influence of artificial intelligence (AI) on firm performance: The business value of AI-based transformation projects. *Business Process Management Journal*, 26(7), 1893-1924.

Wang, Z., da Cunha, C., Ritou, M., & Furet, B. (2019a). Comparison 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 & Social Change*, 146, 723-732.

Warren Liao, T. (2005). Clustering of time series data—a survey. *Pattern Recognition*, 38(11), 1857-1874.

Webb, M. (2019). The Impact of Artificial Intelligence on the Labor Market.

WIPO. (2018). Retrieved the version of '2018.01', from https://www.wipo.int/classifications/ipc/ipcpub/?notion=scheme&version=20180101

WIPO. (2019a). WIPO Technology Trends 2019: Artificial Intelligence, 1-154. World Intellectual Property Organization.

WIPO. (2019b). Retrieved the version of the '2019.01', from https://www.wipo.int/classifications/ipc/ipcpub/?notion=scheme&version=2019

0101

WIPO. (2022). Retrieved September 16, 2022, from https://www.wipo.int/tech_trends/en/artificial_intelligence/patentscope.html

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.

You, H., Li, M., Hipel, K., Jiang, J., Ge, B., & Duan, H. (2017). Development trend forecasting for coherent light generator technology based on patent citation network analysis. *Scientometrics*, 111(1), 297-315.

Yu, K., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. *Nature Biomedical Engineering*, 2(10), 719-731.

Zabala-Iturriagagoitia, J. M., Porto Gómez, I., & Aguirre Larracoechea, U. (2020). Technological diversification: A matter of related or unrelated varieties? *Technological Forecasting & Social Change*, 155, 119997.

Żbikowski, K., & Antosiuk, P. (2021). A machine learning, bias-free approach for predicting business success using Crunchbase data. *Information Processing & Management*, 58(4), 102555.

Zeba, G. Dabić, M., Čičak, M., Daim, T. Yalcin, H. (2021). Technology mining: Artificial intelligence in manufacturing. *Technological Forecasting & Social Change*. 171. 120971.

Zhang, L., Guo, Y., & Sun, G. (2019). How patent signals affect venture capital: The evidence of bio-pharmaceutical start-ups in China. *Technological Forecasting & Social Change*, 145, 93-104.

Zhang, Y., & Zhang, X. (2020). Patent growth and the long-run performance of VC-backed IPOs. *International Review of Economics & Finance*, 69, 33-47.

Zhao, F., Gao, Y., Li, X., An, Z., Ge, S., & Zhang, C. (2021). A similarity measurement for time series and its application to the stock market. *Expert Systems with Applications*, 182, 115217.

Appendix 1: Appendix for Chapter 2

Table A1-1. CPC symbols related to AI (Source: WIPO, 2022)

CPC symbols
A61B5/7264, A61B5/7267, A63F13/67, B23K31/006, B25J9/161, B29C2945/76979, B29C66/965, B60G2600/1876, B60G2600/1878, B60G2600/1879, B60W30/06, B60W30/10, B60W30/14, B62D15/0285, B64G2001/247, E21B2041/0028, F02D41/1405, F03D7/046, F05B2270/707, F05B2270/709, F05D2270/709, F16H2061/0081, F16H2061/0084, G01N2201/1296, G01N29/4481, G01N33/0034, G01R31/2846, G01R31/3651, G01S7/417, G05B13/027, G05B13/0275, G05B13/028, G05B13/0285, G05B13/029, G05B13/0295, G05B2219/33002, G05D1/00, G05D1/0088, G06F11/1476, G06F11/2257, G06F11/2263, G06F15/18, G06F17/16, G06F17/2282, G06F17/27, G06F17/28, G06F17/30029, G06F17/30247, G06F17/30401, G06F17/3043, G06F17/30522, G06F17/30654, G06F17/30663, G06F17/30666, G06F17/30669, G06F17/30672, G06F17/30684, G06F17/30687, G06F17/3069, G06F17/30702, G06F17/30705, G06F17/30731, G06F17/30743, G06F17/30784, G06F19/24, G06F19/707, G06F2207/4824, G06K7/1482, G06K9/00, G06N3/00, G06N3/004, G06N5/003, G06N7/005, G06N7/046, G06N99/005, G06T2207/20081, G06T2207/20084, G06T2207/20084, G06T2207/30236, G06T2207/30248, G06T3/4046, G06T9/002, G08B29/186, G10H2250/151, G10H2250/311, G10K2210/3024, G10K2210/3038, G10L15/00, G10L17/00, G10L25/30, G11B20/10518, H01J2237/30427, H01M8/04992, H02H1/0092, H02P21/0014, H02P23/0018, H03H2017/0208, H03H2222/04, H04L2012/5686, H04L2025/03464, H04L2025/03554, H04L25/0254, H04L25/03165, H04L41/16, H04L45/08, H04N21/4662, H04N21/4666, H04Q2213/054, H04Q2213/13343, H04Q2213/343, H04R25/507, Y10S128/924, Y10S128/925, Y10S706/00

Table A1-2. AI application fields (Source: WIPO, 2022)

AI application fields	CPC codes	IPC codes
Agriculture		A01
Arts and humanities		
Banking and finance		G06Q40/00
Business		
Cartography		
Computing in government		G06Q50/26
Document management and text processing		G06F17/21
Education		G09B, G06Q50/20
Energy management	H01J2237/30427, H01M8/04992, H02H1/0092, H02P21/0014, H02P23/0018, H03H2017/0208, H03H2222/04	G21, H02, H01M8/04992, H03H17/02
Entertainment	A63F13/67, A63F2300/00	A63
Industry and manufacturing		G06Q10/06, G06Q10/08, G06Q50/04, G06Q50/28
Law, social and behavioral sciences		
Life and medical sciences	G06F19/24, A61B5/7264, G16H50/20	A61, G06F19/24, G06F19/00, G16H50/20
Military		B63G, G01S19/18, B64D7/00, F41, F42
Networks		
Personal devices, computing and HCI		
Physical sciences and engineering	G06F19/707	C, D, E, F01, F03, F05, F07, F09, F11, F13, F15, F17

Publishing		
Security		G06F21/00, A61B5/117, H04W12/00
Telecommunications	H04L2012/5686, H04L2025/03464, H04L25/0254, H04L25/03165, H04L41/16, H04L45/08, H04N21/4662, H04Q2213/054, H04Q2213/13343, H04Q2213/343, H04R25/507	H04L12/70, H04L25/03, H04L25/02, H04L25/03, H04L12/24, H04L12/751, H04N21/466, H04R25/00
Transportation	B60W30/06, B60W30/10, B60W30/12, B60W30/14, B60G2600/1876, B60G2600/1878, B60G2600/1879, B62D15/0285, B64G2001/247, G06T2207/30248, G06T2207/30236, G06K9/00791, G05D1/00, B64C2201	B60W30/06, B60W30/10, B60W30/12, B60W30/14, B62D15/02, B64G1/24, G06K9/00, G05D1/00

Table A1-3. Results of normality test & homogeneity of variance test

[1] Phase 1 & 2

		Kolmogorov-Smirnov test (P-value)	Shapiro-Wilk test (P-value)
Generality	Phase 1	1.95E-264	8.12E-31
	Phase 2	0.00E+00	0.00E+00
Originality	Phase 1	2.03E-287	8.80E-30
	Phase 2	0.00E+00	0.00E+00
Complementarity	Phase 1	5.21E-135	0.00E+00
	Phase 2	4.61E-314	0.00E+00
TCT	Phase 1	2.86E-42	1.12E-25
	Phase 2	1.26E-227	0.00E+00
		Levene test (P-value)	Bartlett test (P-value)
Generality		9.37E-61	7.79E-37
Originality		2.42E-04	2.01E-05
Complementarity		7.78E-12	8.22E-06
TCT		0.42779	0.00674

[2] Phase 1

		Kolmogorov-Smirnov test (P-value)	Shapiro-Wilk test (P-value)
Generality	H	1.07E-10	6.39E-05
	M	5.98E-50	8.83E-08
	L	7.80E-217	9.38E-28
Originality	H	1.16E-12	0.00136
	M	1.12E-39	1.21E-10
	L	2.40E-218	7.09E-27
Complementarity	H	3.97E-05	4.31E-10
	M	1.76E-31	5.66E-18
	L	5.81E-130	5.32E-42
TCT	H	0.09469	0.30373
	M	3.11E-05	0.16007
	L	1.61E-37	8.66E-26
		Levene test (P-value)	Bartlett test (P-value)
Generality		2.54E-06	1.11E-16
Originality		1.63E-05	9.87E-08
Complementarity		1.75E-02	1.15E-17
TCT		0.60593	0.51537

[3] Phase 2

		Kolmogorov-Smirnov test (P-value)	Shapiro-Wilk test (P-value)
Generality	H	7.18E-20	1.09E-04
	M	8.61E-130	8.74E-18
	L	0.00E+00	0.00E+00
Originality	H	1.64E-22	6.29E-05
	M	1.52E-124	8.47E-19
	L	0.00E+00	0.00E+00
Complementarity	H	8.68E-05	1.34E-09
	M	2.93E-18	5.13E-26
	L	0.00E+00	0.00E+00
TCT	H	1.70E-04	7.71E-04
	M	6.25E-26	1.17E-15
	L	8.99E-197	2.10E-44
		Levene test (P-value)	Bartlett test (P-value)
Generality		1.37E-74	4.62E-78
Originality		4.93E-03	2.90E-07
Complementarity		1.02E-19	1.51E-66
TCT		0.14634	0.22552

[4] Phase 1 (High)

		Kolmogorov-Smirnov test (P-value)	Shapiro-Wilk test (P-value)
Generality	AI-only	0.001217129	0.50448072
	AI-application	2.35E-13	1.94E-06
Originality	AI-only	0.009421852	0.451570749
	AI-application	1.23E-09	0.00846728
Complementarity	AI-only	0.027069435	1.61E-06
	AI-application	5.06E-04	1.90E-08
TCT	AI-only	0.428870319	7.23E-02
	AI-application	1.35E-01	3.33E-01
		Levene test (P-value)	Bartlett test (P-value)
Generality		7.33E-01	0.278928099
Originality		1.91E-01	0.07771763
Complementarity		6.93E-01	0.591397687
TCT		3.52E-01	0.636465226

[5] Phase 1 (Mid)

		Kolmogorov-Smirnov test (P-value)	Shapiro-Wilk test (P-value)
Generality	AI-only	5.56E-19	0.000437377
	AI-application	5.09E-28	3.16E-06
Originality	AI-only	1.19E-20	0.000282798
	AI-application	1.20E-25	1.74E-08
Complementarity	AI-only	1.93E-24	3.37E-12
	AI-application	2.22E-13	2.29E-12
TCT	AI-only	1.60E-02	1.15E-01
	AI-application	5.98E-04	2.10E-01
		Levene test (P-value)	Bartlett test (P-value)
Generality		1.82E-01	1.97E-01
Originality		5.14E-03	2.44E-03
Complementarity		2.48E-04	5.30E-06
TCT		1.95E-01	3.91E-01

[6] Phase 1 (Low)

		Kolmogorov-Smirnov test (P-value)	Shapiro-Wilk test (P-value)
Generality	AI-only	2.77E-106	5.32E-19
	AI-application	5.28E-117	3.23E-21
Originality	AI-only	9.44E-106	1.90E-18
	AI-application	1.08E-122	9.18E-20
Complementarity	AI-only	2.52E-142	0.00000
	AI-application	6.70E-32	1.71E-31
TCT	AI-only	4.93E-16	9.08E-13
	AI-application	2.77E-23	0.00000
		Levene test (P-value)	Bartlett test (P-value)
Generality		0.1659792222387221)	0.382712662
Originality		2.84E-03	8.07E-03
Complementarity		3.53E-07	4.79E-33
TCT		0.462317294	0.395302744

[7] Phase 2 (High)

		Kolmogorov-Smirnov test (P-value)	Shapiro-Wilk test (P-value)
Generality	AI-only	4.34E-06	0.167338282
	AI-application	5.46E-20	4.87E-06
Originality	AI-only	2.33E-05	0.277399093
	AI-application	1.52E-18	7.68E-05
Complementarity	AI-only	4.20E-03	1.04E-06
	AI-application	1.81E-03	2.62E-07
TCT	AI-only	1.96E-01	8.39E-02
	AI-application	1.47E-03	4.07E-03
		Levene test (P-value)	Bartlett test (P-value)
Generality		3.92E-01	0.225138413
Originality		1.13E-01	0.223159457
Complementarity		2.41E-02	0.015039941
TCT		4.49E-01	0.480242127

[8] Phase 2 (Mid)

		Kolmogorov-Smirnov test (P-value)	Shapiro-Wilk test (P-value)
Generality	AI-only	2.66E-32	4.61E-06
	AI-application	2.65E-103	1.31E-14
Originality	AI-only	2.62E-37	1.67E-06
	AI-application	2.13E-86	2.40E-15
Complementarity	AI-only	4.94E-15	1.79E-13
	AI-application	8.11E-18	4.28E-22
TCT	AI-only	1.36E-06	1.08E-05
	AI-application	7.23E-20	4.50E-14
		Levene test (P-value)	Bartlett test (P-value)
Generality		1.12E-06	9.87E-09
Originality		4.04E-09	1.75E-09
Complementarity		2.52E-04	1.08E-15
TCT		6.65E-01	1.67E-01

[9] Phase 2 (Low)

		Kolmogorov-Smirnov test (P-value)	Shapiro-Wilk test (P-value)
Generality	AI-only	0.00E+00	4.25E-40
	AI-application	0.00E+00	0.00E+00
Originality	AI-only	0.00E+00	4.49E-32
	AI-application	0.00E+00	4.76E-44
Complementarity	AI-only	3.31E-224	0.00000
	AI-application	5.04E-137	2.15E-42
TCT	AI-only	6.97E-82	2.65E-31
	AI-application	4.18E-114	0.00000
		Levene test (P-value)	Bartlett test (P-value)
Generality		0.8584886129266383	0.014804795
Originality		3.92E-11	2.37E-08
Complementarity		5.04E-39	6.32E-84
TCT		0.052250236	0.000528827

Appendix 2: Appendix for Chapter 3

Table A2-1. WIPO IPC Technology-Concordance Table (Source: 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#
	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

Table A2-2. Results of DTM (Document-Term Matrix) and LSA (Latent Semantic Analysis)

Sector	# of Patents	# of Bigram Terms	# of LSA-reduced Features
Finance	12,603	30,862	3,759
Medical	10.218	24,062	3,404
Transport	6,426	14,524	2,177
Semiconductor	1,896	5,301	647
Game	1,576	4,495	522
Biotechnology	2,956	8,640	837

Table A2-3. 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	13634	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

	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
	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
	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

Table A2-4. Results of keywords and TF-IDF values (Finance and AI）

	Cluster 0 (767)		Cluster 1 (6583)		Cluster 2 (1053)		Cluster 3 (2493)		Cluster 4 (1707)	
Rank	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF
1	content	0.1984	datum	0.0292	document	0.1786	user	0.0872	image	0.1771
2	user	0.0403	base	0.0175	code	0.0618	information	0.0590	object	0.0480
3	content item	0.0352	information	0.0164	datum	0.0597	biometric	0.0336	capture	0.0415
4	display	0.0325	use	0.0156	sensing	0.0533	transaction	0.0316	information	0.0314
5	item	0.0312	customer	0.0147	code datum	0.0510	authentication	0.0283	image datum	0.0244
6	page	0.0272	provide	0.0143	computer	0.0351	datum	0.0223	product	0.0243
7	medium	0.0264	product	0.0143	indicate datum	0.0345	voice	0.0210	datum	0.0241
8	information	0.0258	model	0.0140	surface	0.0337	provide	0.0195	image capture	0.0225
9	digital	0.0222	message	0.0130	print	0.0328	service	0.0189	capture image	0.0217
10	web	0.0222	determine	0.0129	identity	0.0314	base	0.0172	display	0.0205
11	datum	0.0206	user	0.0128	indicate	0.0313	use	0.0172	unit	0.0204
12	base	0.0193	item	0.0125	indicative	0.0298	communication	0.0172	item	0.0192
13	provide	0.0185	computer	0.0123	interface surface	0.0287	server	0.0170	processing	0.0189
14	medium content	0.0173	generate	0.0123	electronic document	0.0285	card	0.0166	user	0.0177
15	identify	0.0160	plurality	0.0123	interface	0.0270	receive	0.0161	base	0.0172
16	digital content	0.0151	vehicle	0.0119	electronic	0.0259	electronic	0.0155	determine	0.0166

17	generate	0.0142	set	0.0117	use	0.0237	request	0.0154	plurality	0.0157
18	request	0.0139	identify	0.0115	form	0.0235	identification	0.0152	digital	0.0156
19	web page	0.0139	receive	0.0115	sense	0.0234	mobile	0.0150	identify	0.0156
20	multimedia	0.0134	associate	0.0115	datum indicative	0.0227	second	0.0145	use	0.0154
21	server	0.0128	text	0.0114	information	0.0227	input	0.0138	digital image	0.0148
22	receive	0.0127	time	0.0114	element	0.0217	application	0.0138	store	0.0144
23	audio	0.0127	data	0.0110	product item	0.0216	associate	0.0124	recognition	0.0139
24	unit	0.0127	object	0.0109	user	0.0199	payment	0.0123	camera	0.0138
25	network	0.0124	second	0.0105	interactive element	0.0198	interface	0.0123	medical	0.0132
26	determine	0.0122	display	0.0105	image	0.0197	access	0.0122	second	0.0132
27	advertisement	0.0122	service	0.0102	transport	0.0188	terminal	0.0122	process	0.0131
28	use	0.0120	event	0.0100	interactive	0.0186	network	0.0122	feature	0.0127
29	client	0.0120	language	0.0097	interaction	0.0176	computer	0.0117	region	0.0127
30	signal	0.0118	process	0.0095	product	0.0175	determine	0.0116	vehicle	0.0126

Table A2-5. Results of keywords and TF-IDF values (Medical and AI)

	Cluster 0 (564)		Cluster 1 (2681)		Cluster 2 (1418)	
Rank	Keyword	Mean TF-IDF	Keyword (TF-IDF)	Mean TF-IDF	Keyword (TF-IDF)	Mean TF-IDF
1	image	0.1283	user	0.0446	image	0.1497
2	unit	0.1067	signal	0.0445	image datum	0.0274
3	processing	0.0802	datum	0.0245	datum	0.0263
4	medical	0.0685	eye	0.0215	pixel	0.0245
5	medical image	0.0675	sensor	0.0211	object	0.0238
6	image processing	0.0675	information	0.0202	second	0.0234
7	information	0.0516	patient	0.0193	capture	0.0215
8	display	0.0475	base	0.0183	display	0.0202
9	unit configure	0.0390	control	0.0178	processing	0.0191
10	region	0.0336	use	0.0175	value	0.0183
11	configure	0.0329	determine	0.0171	information	0.0164
12	datum	0.0296	configure	0.0155	acquire	0.0163
13	plurality	0.0257	biometric	0.0153	set	0.0159
14	image datum	0.0247	receive	0.0152	iris	0.0158
15	processing image	0.0236	provide	0.0150	plurality	0.0154
16	acquire	0.0212	state	0.0149	region	0.0150
17	obtain	0.0201	display	0.0140	frame	0.0150

18	generate	0.0192	subject	0.0140	base	0.0146
19	second	0.0182	detect	0.0136	image processing	0.0142
20	storage	0.0177	individual	0.0134	imaging	0.0142
21	position	0.0170	person	0.0133	obtain	0.0142
22	circuitry	0.0167	speech	0.0131	generate	0.0139
23	area	0.0161	camera	0.0130	use	0.0137
24	diagnosis	0.0161	input	0.0128	subject	0.0135
25	extract	0.0159	time	0.0128	second image	0.0134
26	base	0.0154	processor	0.0122	process	0.0134
27	section	0.0152	voice	0.0121	section	0.0131
28	processing circuitry	0.0149	physiological	0.0112	radiation	0.0125
29	imaging	0.0148	unit	0.0112	determine	0.0118
30	tomographic image	0.0145	response	0.0110	image frame	0.0116

Table A2-5. (continued)

	Cluster 3 (753)		Cluster 4 (414)		Cluster 5 (4388)	
Rank	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF
1	light	0.1016	projection	0.1193	image	0.0522
2	fingerprint	0.0904	ray	0.1073	datum	0.0278
3	sensor	0.0544	image	0.0918	region	0.0227
4	finger	0.0541	datum	0.0499	use	0.0207
5	surface	0.0379	projection image	0.0495	imaging	0.0196
6	object	0.0353	projection datum	0.0491	patient	0.0193
7	image	0.0351	ray image	0.0456	object	0.0180
8	light source	0.0324	reconstruct	0.0415	model	0.0171
9	electrode	0.0314	ct	0.0408	medical	0.0170
10	source	0.0306	reconstruction	0.0391	tissue	0.0161
11	layer	0.0303	object	0.0307	dimensional	0.0158
12	unit	0.0291	source	0.0248	base	0.0156
13	identification	0.0247	detector	0.0242	determine	0.0149
14	capture	0.0226	generate	0.0225	feature	0.0146
15	fingerprint sensor	0.0208	reconstruct image	0.0218	set	0.0141
16	element	0.0205	imaging	0.0210	point	0.0130
17	information	0.0203	temperature	0.0207	volume	0.0130

18	contact	0.0201	tomography	0.0197	provide	0.0129
19	pattern	0.0195	tomosynthesis	0.0196	generate	0.0129
20	array	0.0195	use	0.0193	second	0.0124
21	sensing	0.0189	compute	0.0193	value	0.0124
22	capacitance	0.0185	core temperature	0.0188	position	0.0120
23	emit	0.0176	body core	0.0188	identify	0.0117
24	eye	0.0175	image datum	0.0185	structure	0.0115
25	second	0.0173	artifact	0.0183	image datum	0.0114
26	substrate	0.0171	measurement	0.0183	subject	0.0114
27	form	0.0165	compute tomography	0.0182	information	0.0113
28	portion	0.0164	source point	0.0182	display	0.0112
29	circuit	0.0160	measurement external	0.0182	scan	0.0109
30	position	0.0153	external source	0.0180	plurality	0.0109

Table A2-6. Results of keywords and TF-IDF values (Transport and AI)

	Cluster 0 (206)		Cluster 1 (2094)		Cluster 2 (1335)		Cluster 3 (476)	
Rank	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF
1	aerial	0.1594	vehicle	0.0311	vehicle	0.0962	driver	0.1604
2	uav	0.1457	image	0.0303	control	0.0480	vehicle	0.0586
3	aerial vehicle	0.1425	sensor	0.0289	user	0.0362	information	0.0393
4	unmanned	0.1346	occupant	0.0251	parking	0.0348	drive	0.0377
5	unmanned aerial	0.1166	datum	0.0232	information	0.0339	state	0.0310
6	flight	0.0544	signal	0.0190	voice	0.0273	assistance	0.0290
7	vehicle	0.0490	position	0.0184	autonomous	0.0246	gaze	0.0257
8	datum	0.0351	determine	0.0180	unit	0.0244	driver assistance	0.0243
9	vehicle uav	0.0336	use	0.0180	speech	0.0230	unit	0.0241
10	image	0.0334	base	0.0177	command	0.0221	image	0.0240
11	landing	0.0331	information	0.0171	base	0.0210	fingerprint	0.0237
12	structure	0.0313	camera	0.0168	recognition	0.0194	determine	0.0229
13	location	0.0274	detect	0.0149	input	0.0180	vehicle driver	0.0227
14	camera	0.0241	video	0.0148	display	0.0179	driver vehicle	0.0219
15	flight path	0.0235	unit	0.0145	determine	0.0175	driving	0.0213
16	target	0.0234	area	0.0141	datum	0.0173	base	0.0203
17	use	0.0232	second	0.0140	drive	0.0172	detect	0.0203

18	inspection	0.0232	control	0.0140	configure	0.0166	behavior	0.0200
19	object	0.0225	provide	0.0134	vehicle control	0.0164	configure	0.0196
20	unmanned aircraft	0.0217	point	0.0134	signal	0.0164	datum	0.0184
21	sensor	0.0205	aircraft	0.0133	autonomous vehicle	0.0162	provide	0.0172
22	control	0.0199	detection	0.0131	controller	0.0160	fact check	0.0172
23	information	0.0194	configure	0.0131	provide	0.0154	fact	0.0170
24	fly	0.0192	seat	0.0131	sensor	0.0154	face	0.0170
25	path	0.0190	value	0.0127	operation	0.0150	driver state	0.0168
26	capture	0.0183	feature	0.0126	receive	0.0149	eye	0.0167
27	receive	0.0180	receive	0.0126	motor	0.0148	monitor	0.0167
28	rooftop	0.0176	plurality	0.0126	detect	0.0144	direction	0.0167
29	aircraft	0.0176	road	0.0125	motor vehicle	0.0141	duration	0.0164
30	configure	0.0175	surface	0.0118	communication	0.0140	alert	0.0162

Table A2-6. (continued)

	Cluster 4 (840)		Cluster 5 (1174)		Cluster 6 (301)	
Rank	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF
1	object	0.1561	image	0.1341	lane	0.2799
2	light	0.0650	display	0.0771	vehicle	0.0730
3	image	0.0588	vehicle	0.0700	line	0.0538
4	vehicle	0.0429	camera	0.0612	travel	0.0462
5	detection	0.0401	capture	0.0492	lane change	0.0453
6	unit	0.0386	view	0.0421	change	0.0415
7	detect	0.0336	image datum	0.0409	image	0.0360
8	object detection	0.0272	datum	0.0385	road	0.0335
9	region	0.0263	unit	0.0336	boundary	0.0302
10	source	0.0230	capture image	0.0293	travel lane	0.0294
11	information	0.0224	control	0.0272	lane mark	0.0292
12	light source	0.0222	processing	0.0243	unit	0.0292
13	area	0.0210	second	0.0236	detect	0.0291
14	second	0.0209	area	0.0199	mark	0.0290
15	camera	0.0202	processor	0.0197	control	0.0273
16	determine	0.0200	image capture	0.0197	lane boundary	0.0265
17	configure	0.0187	configure	0.0190	lane line	0.0264

18	detect object	0.0179	image processing	0.0184	marking	0.0261
19	position	0.0175	field	0.0178	vehicle lane	0.0236
20	base	0.0170	field view	0.0174	recognize	0.0234
21	target	0.0157	video	0.0172	information	0.0230
22	capture	0.0157	imaging	0.0169	determine	0.0228
23	dimensional	0.0149	region	0.0162	lane marking	0.0221
24	object detect	0.0147	object	0.0161	point	0.0217
25	control	0.0146	vision	0.0159	position	0.0210
26	use	0.0143	detect	0.0157	detect lane	0.0203
27	traffic light	0.0143	rear	0.0157	base	0.0200
28	traffic	0.0141	driver	0.0155	departure	0.0197
29	datum	0.0138	process	0.0152	vehicle travel	0.0191
30	distance	0.0137	road	0.0152	datum	0.0187

Table A2-7. Results of keywords and TF-IDF values (Semiconductor and AI)

	Cluster 0 (553)		Cluster 1 (484)		Cluster 2 (270)		Cluster 3 (184)		Cluster 4 (405)	
Rank	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF
1	signal	0.0447	pattern	0.1081	light	0.1472	defect	0.2505	sensor	0.0969
2	pixel	0.0352	image	0.0833	display	0.1221	inspection	0.1201	layer	0.0917
3	circuit	0.0304	wafer	0.0481	emit	0.0694	image	0.1135	fingerprint	0.0786
4	image	0.0302	datum	0.0478	fingerprint	0.0689	pattern	0.0812	surface	0.0669
5	sensor	0.0295	mask	0.0414	light emit	0.0637	inspect	0.0528	chip	0.0660
6	array	0.0287	inspection	0.0369	layer	0.0600	detect	0.0477	fingerprint sensor	0.0591
7	element	0.0261	position	0.0304	panel	0.0538	candidate	0.0402	sensing	0.0554
8	second	0.0245	second	0.0263	substrate	0.0459	pattern inspection	0.0390	electrode	0.0543
9	datum	0.0239	reference	0.0261	display panel	0.0453	defect candidate	0.0385	package	0.0528
10	substrate	0.0210	edge	0.0254	sensor	0.0448	defect inspection	0.0367	substrate	0.0502
11	light	0.0208	semiconductor	0.0252	optical	0.0427	wafer	0.0344	structure	0.0401
12	semiconductor	0.0193	object	0.0243	pixel	0.0398	reference	0.0332	form	0.0378
13	line	0.0193	use	0.0242	electrode	0.0368	classification	0.0327	conductive	0.0360
14	unit	0.0190	inspect	0.0242	sensing	0.0343	value	0.0310	second	0.0357
15	information	0.0182	region	0.0240	unit	0.0335	unit	0.0308	circuit	0.0343
16	output	0.0180	value	0.0237	recognition	0.0331	reference image	0.0293	die	0.0342

17	layer	0.0178	exposure	0.0237	dispose	0.0318	detection	0.0289	pad	0.0267
18	object	0.0177	measure	0.0234	fingerprint recognition	0.0300	detect defect	0.0278	connection	0.0264
19	use	0.0172	obtain	0.0228	organic	0.0299	sample	0.0269	connect	0.0245
20	control	0.0171	process	0.0223	plurality	0.0286	defect image	0.0268	dielectric	0.0243
21	channel	0.0163	optical	0.0218	identification	0.0284	datum	0.0265	portion	0.0233
22	detection	0.0158	unit	0.0203	organic light	0.0277	inspection condition	0.0263	electrically	0.0230
23	processing	0.0155	design	0.0199	array	0.0271	image defect	0.0243	cover	0.0227
24	process	0.0150	processing	0.0193	element	0.0268	condition	0.0233	contact	0.0218
25	sensing	0.0147	correction	0.0193	Fingerprint identification	0.0266	inspection image	0.0230	dispose	0.0218
26	configure	0.0145	plurality	0.0191	region	0.0266	information	0.0229	finger	0.0217
27	provide	0.0143	measurement	0.0185	photosensitive	0.0254	obtain	0.0224	material	0.0215
28	Image sensor	0.0139	detect	0.0184	oled	0.0254	compare	0.0223	Dielectric layer	0.0213
29	form	0.0138	determine	0.0183	source	0.0252	use	0.0217	plate	0.0212
30	cell	0.0138	calculate	0.0178	light source	0.0252	pattern defect	0.0217	semiconductor	0.0207

Table A2-8. Results of keywords and TF-IDF values (Game and AI)

	Cluster 0 (301)		Cluster 1 (587)		Cluster 2 (56)		Cluster 3 (53)	
Rank	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF
1	voice	0.0806	user	0.0569	reality	0.2325	object	0.4387
2	signal	0.0502	datum	0.0559	augment reality	0.2254	capture	0.1776
3	audio	0.0464	motion	0.0483	augment	0.2132	information	0.1521
4	sound	0.0437	sensor	0.0450	reality display	0.1219	address correspond	0.1516
5	speech	0.0428	exercise	0.0367	world	0.1135	information address	0.1516
6	robot	0.0410	event	0.0338	passable world	0.0918	capture identification	0.1510
7	control	0.0406	activity	0.0263	passable	0.0918	image	0.1508
8	input	0.0395	information	0.0256	world model	0.0918	identification process	0.1497
9	command	0.0357	movement	0.0231	display	0.0915	use access	0.1491
10	character	0.0356	use	0.0221	individual augment	0.0880	process digital	0.1485
11	information	0.0356	unit	0.0207	model datum	0.0860	object database	0.1479
12	game	0.0334	provide	0.0199	waveguide	0.0752	communication pertinent	0.1472
13	unit	0.0292	base	0.0196	model	0.0634	pertinent object	0.1472
14	output	0.0291	athletic	0.0193	individual	0.0626	initiate communication	0.1472
15	datum	0.0244	toy	0.0187	pass	0.0611	information initiate	0.1472
16	message	0.0228	analysis	0.0186	comprise	0.0502	object recognize	0.1467
17	communication	0.0221	tag	0.0180	datum	0.0448	address	0.1463

18	generate	0.0216	body	0.0172	planar waveguide	0.0444	capture object	0.1460
19	processing	0.0208	performance	0.0170	planar	0.0430	pertinent	0.1459
20	text	0.0198	signal	0.0168	virtual	0.0427	digital image	0.1458
21	display	0.0194	receive	0.0168	doe	0.0411	access information	0.1453
22	data	0.0193	processor	0.0165	object	0.0410	recognize plurality	0.1449
23	user	0.0191	time	0.0163	set	0.0407	object capture	0.1448
24	receive	0.0190	analyze	0.0162	augmented	0.0403	object use	0.1426
25	recognition	0.0182	configure	0.0161	map point	0.0402	plurality object	0.1416
26	conversation	0.0180	generate	0.0156	real world	0.0391	database information	0.1410
27	mean	0.0177	sport	0.0155	set map	0.0389	initiate	0.1408
28	operation	0.0172	determine	0.0154	point	0.0379	image object	0.1391
29	microphone	0.0170	example	0.0151	reality augment	0.0377	correspond object	0.1384
30	provide	0.0164	application	0.0150	image	0.0375	digital	0.1345

Table A2-8. (continued)

	Cluster 4 (228)		Cluster 5 (351)	
Rank	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF
1	image	0.1720	game	0.0678
2	card	0.1196	video	0.0593
3	face	0.0502	object	0.0562
4	processing	0.0403	image	0.0403
5	information	0.0389	user	0.0397
6	object	0.0352	gesture	0.0384
7	capture	0.0350	player	0.0380
8	display	0.0327	golf	0.0348
9	section	0.0291	camera	0.0316
10	unit	0.0286	depth	0.0284
11	capture image	0.0284	ball	0.0266
12	detect	0.0278	virtual	0.0265
13	information processing	0.0268	use	0.0263
14	area	0.0260	determine	0.0259
15	image processing	0.0248	position	0.0236
16	user	0.0234	base	0.0233
17	datum	0.0233	information	0.0232

18	region	0.0227	capture	0.0226
19	position	0.0224	provide	0.0212
20	extract	0.0223	video game	0.0206
21	image datum	0.0220	control	0.0204
22	camera	0.0213	location	0.0198
23	set	0.0198	club	0.0195
24	base	0.0192	second	0.0192
25	second	0.0183	display	0.0190
26	acquire	0.0183	datum	0.0186
27	generate	0.0182	track	0.0184
28	hmd	0.0181	play	0.0180
29	point	0.0175	computer	0.0177
30	register	0.0167	plurality	0.0172

Table A2-9. Results of keywords and TF-IDF values (Biotechnology and AI)

	Cluster 0 (854)		Cluster 1 (468)		Cluster 2 (116)		Cluster 3 (146)	
Rank	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF
1	datum	0.0365	image	0.1567	code	0.1734	sequence	0.1681
2	use	0.0245	object	0.0387	product item	0.1660	chip	0.0737
3	determine	0.0189	light	0.0349	item	0.1546	database	0.0691
4	value	0.0184	pixel	0.0311	product	0.1471	oligonucleotide	0.0633
5	base	0.0179	tissue	0.0304	datum	0.1231	model	0.0446
6	set	0.0174	cell	0.0268	code datum	0.1084	information	0.0440
7	measure	0.0166	color	0.0268	interface surface	0.1060	probe	0.0430
8	signal	0.0165	capture	0.0266	identity	0.1004	nucleotide	0.0396
9	array	0.0158	specimen	0.0221	interface	0.0980	acid	0.0384
10	sensor	0.0155	processing	0.0212	code data	0.0906	fiber	0.0383
11	provide	0.0149	imaging	0.0212	surface	0.0877	nucleotide sequence	0.0382
12	detect	0.0147	sample	0.0209	data	0.0817	peptide	0.0379
13	test	0.0146	area	0.0198	data portion	0.0808	cluster	0.0357
14	information	0.0146	analysis	0.0196	indicate datum	0.0795	array chip	0.0356
15	feature	0.0146	obtain	0.0195	portion	0.0794	nucleic	0.0332
16	plurality	0.0145	value	0.0189	sensing	0.0792	nucleic acid	0.0332
17	analysis	0.0139	use	0.0184	indicative	0.0730	protein	0.0317

18	plant	0.0136	process	0.0173	sense	0.0725	array	0.0309
19	time	0.0133	plurality	0.0172	indicate	0.0654	database model	0.0298
20	genetic	0.0132	biological	0.0162	sense code	0.0577	probe array	0.0284
21	model	0.0129	identify	0.0162	datum indicative	0.0576	sample	0.0280
22	computer	0.0119	digital	0.0161	indicative identity	0.0550	organize information	0.0263
23	product	0.0118	image analysis	0.0161	identity product	0.0527	provide	0.0253
24	unit	0.0118	colony	0.0159	scanning	0.0470	identify	0.0247
25	comprise	0.0118	information	0.0159	threshold	0.0444	use	0.0245
26	process	0.0117	feature	0.0158	time pcr	0.0403	information relate	0.0235
27	second	0.0116	second	0.0156	baseline	0.0391	organize	0.0231
28	result	0.0115	determine	0.0156	user	0.0344	acid sequence	0.0226
29	protein	0.0113	analyze	0.0155	adapt	0.0338	target	0.0222
30	select	0.0111	optical	0.0154	real time	0.0320	code	0.0188

Table A2-9. (continued)

	Cluster 4 (428)		Cluster 5 (657)		Cluster 6 (287)	
Rank	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF	Keyword	Mean TF-IDF
1	sample	0.1038	gene	0.0750	cell	0.2499
2	biological	0.0509	expression	0.0549	image	0.0579
3	biological sample	0.0327	cancer	0.0525	unit	0.0342
4	analysis	0.0279	disease	0.0462	culture	0.0275
5	signal	0.0250	invention	0.0434	blood cell	0.0271
6	nucleic acid	0.0246	treatment	0.0321	blood	0.0264
7	nucleic	0.0246	patient	0.0317	analysis	0.0248
8	acid	0.0242	gene expression	0.0309	cell cell	0.0240
9	microbiome	0.0240	use	0.0294	sample	0.0238
10	cell	0.0233	present invention	0.0294	provide	0.0206
11	provide	0.0224	subject	0.0289	identify	0.0199
12	composition	0.0222	present	0.0281	determine	0.0195
13	image	0.0219	provide	0.0280	colony	0.0192
14	analyze	0.0212	biomarker	0.0270	target	0.0188
15	use	0.0207	relate	0.0268	detect	0.0183
16	condition	0.0199	identify	0.0245	feature	0.0181

17	generate	0.0198	drug	0.0238	cell image	0.0179
18	base	0.0193	marker	0.0234	state	0.0176
19	determine	0.0192	invention relate	0.0229	cell population	0.0175
20	comprise	0.0190	risk	0.0212	use	0.0173
21	dna	0.0184	predict	0.0206	target cell	0.0162
22	blood	0.0182	diagnosis	0.0195	plurality	0.0160
23	subject	0.0180	response	0.0193	imaging	0.0160
24	detect	0.0173	tumor	0.0191	information	0.0158
25	particle	0.0164	sample	0.0183	analysis cell	0.0155
26	second	0.0164	level	0.0178	population	0.0155
27	specimen	0.0161	determine	0.0173	optical	0.0150
28	measurement	0.0161	invention provide	0.0173	cell culture	0.0149
29	dataset	0.0160	genetic	0.0171	step	0.0149
30	analyte	0.0159	base	0.0170	tissue	0.0139

Appendix 3: Appendix for Chapter 4

Table A3-1. Category of the firm type

Firm type	Industry group from Crunchbase
Industry-specific	Administrative, Advertising, Agriculture and Farming, Biotechnology, Commerce and Shopping, Community and Lifestyle, Consumer Goods, Design, Education, Energy, Financial Services, Food and Beverage, Gaming, Government and Military, Health Care, Lending and Investments, Manufacturing, Media and Entertainment, Music and Audio, Natural Resources, Payments, Professional Services, Real Estate, Sales and Marketing, Sports, Sustainability, Transportation, Travel and Tourism, Video, Science and Engineering (Biotechnology, Biometric, Semiconductor, Neuroscience)
Cross-industry	Apps, Artificial Intelligence, Consumer Electronics, Data and Analytics, Hardware, Information Technology, Internet Services, Messaging and Telecommunication, Mobile, Navigation and Mapping, Platforms, Privacy and Security, Science and Engineering, Software

Table A3-2. Results of VIF (Variance Inflation Factor) score

[1] Patent and funding amount

	Model 1		Model 2		Model 3	
Variable	VIF	1/VIF	VIF	1/VIF	VIF	1/VIF
PatentYN	1.15	0.867279	1.85	0.541222	1.64	0.608911
Age	1.34	0.747547	1.34	0.747507	1.35	0.743188
Size	1.73	0.57881	1.73	0.578807	1.73	0.576959
Investors	1.26	0.792083	1.26	0.791912	1.27	0.789084
FundStg	1.88	0.532747	3.36	0.297252	1.88	0.532662
FirmType	1.02	0.982925	1.02	0.976776	2.22	0.451241
PatentYN × FundStg			3.67	0.272826		
PatentYN × FirmType					2.76	0.362833
Mean VIF	1.4		2.03		1.83	

[2] Patent counts and funding amount

	Model 1		Model 2		Model 3	
Variable	VIF	1/VIF	VIF	1/VIF	VIF	1/VIF
PatentCount	1.09	0.913988	2.11	0.474642	1.62	0.616916
FundStg	1.36	0.733453	3.09	0.323543	1.37	0.730843
FirmType	1.03	0.971835	1.04	0.964473	2.46	0.406266
Age	1.36	0.736847	1.38	0.726047	1.37	0.727751
PatentCount × FundStg			4.23	0.236591		
PatentCount × FirmType					3.1	0.322485
Mean VIF	1.21		2.37		1.99	

[3] Technology diversity and funding amount

	Model 1		Model 2		Model 3		Model 4	
Variable	VIF	1/VIF	VIF	1/VIF	VIF	1/VIF	VIF	1/VIF
TD	1.3	0.769237	4.84	0.206517	1.87	0.535093	1.79	0.558835
PatentCount	1.41	0.708197	1.79	0.557735	1.41	0.70721	1.42	0.706104
FundStg	1.31	0.762297	1.31	0.761506	1.32	0.758146	1.34	0.748846
FirmType	1.05	0.949509	1.07	0.933279	1.07	0.938844	1.05	0.949
Age	1.26	0.792385	1.28	0.780496	1.26	0.79063	1.29	0.77269
TD × PatentCount			6	0.166576				
TD × FundStg					1.65	0.604774		
TD × FirmType							1.47	0.681805
Mean VIF	1.27		2.72		1.43		1.39	

[4] Unrelated/related diversity and funding amount

	Model 1		Model 2		Model 3		Model 4	
Variable	VIF	1/VIF	VIF	1/VIF	VIF	1/VIF	VIF	1/VIF
UD	1.37	0.729988	8.45	0.118411	2.39	0.418419	1.84	0.542917
RD	1.26	0.792471	7.68	0.130241	2.19	0.455881	1.85	0.539274
PatentCount	1.45	0.687925	1.57	0.636868	1.47	0.680149	1.47	0.679733
FundStg	1.45	0.689339	1.45	0.687931	1.82	0.54966	1.47	0.679365
FirmType	1.08	0.92241	1.11	0.900765	1.1	0.911836	1.77	0.564821
Age	1.41	0.710459	1.43	0.698396	1.42	0.705677	1.41	0.709331
UD × PatentCount			9	0.111081				
RD × PatentCount			7.83	0.127786				
UD × FundStg					2.85	0.350601		
RD × FundStg					2.68	0.373715		
UD × FirmType							2.56	0.390598
RD × FirmType							2.35	0.42636
Mean VIF	1.34		4.81		1.99		1.84	

Abstract (Korean)

본 학위 논문은 인공지능(Artificial Intelligence, AI) 기술의 범용기술로서 발전 가능성에 대하여 다각도적 관점에서 실증 분석한 연구이다. 특히, 기술 발전 초기 단계에 있는 신흥기술의 특성을 고려하여 해당 기술로 인한 경제적 파급 효과보다는 기술적침투성(Technological Pervasiveness)을 기반으로 이해하고자 한다. 본 학위 논문에서는 기술적침투성을 지식 흐름, 산업 섹터, 기술 포트폴리오의 세 가지 차원의 분석 레벨로 확장하여 다각도로 분석한다. 또한, 각 분석 레벨별로는 범용기술로 인한 혁신 발생 측면에 주목하여, 기술 확산, 융합, 전략의 개념적 배경을 중심으로 새로운 관점에서 범용기술을 해석한다. 본 학위 논문은 다음의 세 가지 연구로 구성된다.

첫번째 연구에서는 인공지능 기술 발전 및 확산에 있어, 범용기술적 특성의 변화와 차이에 대한 패턴을 살펴본다. 여기서 범용기술적 특성으로는 기술 지식의 조합과 확산 과정에 주목하여, 일반성, 독창성, 상보성을 고려한다. 분석 방법으로는 개별기술의 확산 과정을 시계열데이터로 구축하고, 시계열데이터의 패턴 분석을 위해 동적시간왜곡 및 시계열 클러스터링 방법론을 사용한다. 또한, 시간 및 확산 패턴에 따라 구분된 클러스터는 분산 분석을 통해 유의미한 차이를 검증한다. 연구 결과, 기술 발전에 따라 인공지능 기술의 범용기술적 특성이 증가하는 것으로 나타났고, 확산 수준이 높은 인공지능 기술에서 해당 특성이 더욱 높은 것으로 나타났다. 특히, 인공지능 기술 중에서는 적용기술에서 범용기술적 특성이 높게 나타났다. 본 연구를 통하여, 인공지능 기

술의 범용기술로서 발전 패턴을 기술의 지식 흐름 및 확산 차원에서 이해할 수 있다.

두번째 연구에서는 인공지능 기술의 다양한 산업으로의 적용가능성을 구체적으로 분석할 수 있는 새로운 프레임워크를 제안한다. 본 연구에서 제안하는 프레임워크는 기술 융합을 개념적 배경으로 하여, 산업 섹터, 기술 범주, 기술 활용의 세가지 측면에 주목한다. 해당 프레임워크는 특허 데이터의 기술 분류 기반 정형데이터 및 텍스트 기반 비정형데이터 분석을 통합한 양방향적 접근 방법론으로, 네트워크 분석 및 클러스터링 분석 방법론을 활용한다. 연구 결과, 인공지능 기술이 적용된 다양한 산업 섹터 및 공통 기술 범주를 확인하였고, 수평적 적용가능성이 있음을 확인하였다. 또한, 각 산업 내에서 도출된 산업별 인공지능 적용 기술 범주와 기술 활용에 대한 다양한 적용 패턴을 도출하고, 수직적 적용가능성이 있음을 확인하였다. 본 연구에서 제시하는 프레임워크를 통하여, 인공지능 기술의 범용기술로서 적용가능성을 산업 섹터 차원에서 이해할 수 있다.

세번째 연구에서는 기술 공급자 측면에서의 기술 포트폴리오에 주목한다. 특히, 스타트업은 파괴적 혁신을 이끄는 주체가 될 수 있으며, 기술 다각화는 다양한 지식의 조합을 기반으로 한 공급 측면의 혁신을 통해 범용기술로의 발전에 기여할 수 있다. 본 연구에서는 인공지능 스타트업의 기술다각화 전략을 탐구하고, 스타트업 성장을 가능하게 하는 스타트업 투자와의 관계를 분석한다. 본 연구에서 기술다각화는 관련기술다각화와 비관련기술다각화로 구분하여 측정하고, 기업유형은 산업특정기업 및 산업공통기업으로 구분한다. 분석

방법으로는 스타트업의 특허 및 투자 정보 기반의 패널 데이터를 고정 효과 모형을 중심으로 분석한다. 연구 결과, 인공지능 스타트업의 기술다각화가 스타트업 투자와 긍정적인 관계가 있음을 확인하였다. 또한, 해당 관계는 비관련 기술다각화일수록, 산업공통기업일수록 더욱 크게 영향을 미치는 것으로 나타났다. 본 연구를 통하여, 인공지능 섹터의 거시적인 성장 매커니즘 관점에서 인공지능 스타트업 기술다각화가 범용기술로서 발전에 혁신 동력으로 작용할 수 있음을 시사한다.

본 학위 논문에서 다루는 세 가지의 연구는 서로 다른 개념적 배경과 범용기술적 특성과의 연계를 통하여 범용기술 및 인공지능 섹터 연구에 이론적, 실증적 기여를 갖는다. 또한, 연구에서 제안하는 신흥기술의 성장을 이해하는 동적 패턴 분석 방법 및 프레임워크는 기술 혁신 연구에 방법론적으로 기여한다. 본 연구 결과는 인공지능 섹터의 기술, 산업 및 기업에 대한 이해와 발전에 기여할 수 있는 정책적, 전략적 시사점을 제시한다.

주요어 : 인공지능, 범용기술, 기술확산, 기술융합, 기술전략, 특허분석
학 번 : 2016-30267