

Identifying Key Elements for Future Education : A Keyword Analysis of Generative Artificial Intelligence in Education Using Text Mining Techniques

Gi-A Kim

ABSTRACT

This study was conducted to identify key elements of future education in relation to generative artificial intelligence (AI), particularly in light of rapid technological advancements and the upcoming implementation of AI digital textbooks in 2025. Using text mining techniques, this research systematically analyzed keywords related to generative AI in education. The study examined 6,170 academic dissertations across various educational research fields from 2022 to 2024.

The research methodology employed key text mining techniques, specifically keyword frequency analysis and keyword network analysis. The results revealed several significant findings. First, the keyword frequency analysis identified high-frequency terms related to generative AI in education, including 'utilization', 'education', 'design', 'prompt', 'class', and 'learning'. Second, the keyword network analysis demonstrated that 'design' exhibited strong correlations with various educational elements, indicating its central role in the educational application of generative AI.

The significance of this study lies in its methodological approach, employing objective text mining techniques to identify key elements of future education. The findings are expected to provide practical implications for establishing educational policies and developing curriculum in the era of generative AI.

☞ keyword : Text Mining, Generative Artificial Intelligence, Education, Keyword Frequency Analysis, Keyword Network Analysis

1. Introduction

1.1 Research Necessity and Objectives

In 2023, the Ministry of Education announced its AI Digital Textbook Roadmap, supporting the transformation of school education through initiatives planned for 2024, including the development of digital pioneer teacher groups, faculty training, and on-site school consulting. As the 2025 implementation approaches, debates surrounding the introduction of AI digital textbooks continue unabated, with heightened interest from educators, students, and parents. This transformation was catalyzed by OpenAI's release of ChatGPT on November 30, 2022, which has generated significant global impact through the emergence of generative artificial intelligence, influencing various industries and sectors. In the education sector particularly, research and support are being provided to utilize this technological innovation as a necessary tool and resource for classroom revolution, preparing for and urgently responding to impending changes.

As part of these response measures, educational programs such as the Digital Sprout Program are being implemented nationwide for elementary, middle, and high school students to experience software and artificial intelligence while developing digital competencies. However, the educational disparity has widened between students at schools that actively understand and apply for these channels and those

who miss out on such educational opportunities. Consequently, there is an emphasized need to reestablish the direction of future education to reduce educational disparities among students and enable personalized learning through verified and standardized curricula based on AI digital textbooks [1][2].

As an alternative solution, generative artificial intelligence is increasingly being utilized as a powerful instructional aid and content resource, with various applications being tested in educational settings. The vision of personalized education for all highlights the importance of 'High-touch High-tech,' describing a future educational landscape where advanced technology assists teachers' tasks, allowing them to focus on their teaching role.

Particularly, in the accelerated digital transformation of education following the COVID-19 pandemic, generative AI has played a crucial role in creating richer and more interactive remote learning environments, strengthening students' motivation as learners, and enabling teachers to function as learning designers supporting active learners.

However, amid these rapid changes, the education sector faces several challenges: First, can generative AI become a sustainable educational technology? Second, are educators equipped with digital and artificial intelligence literacy competencies? Third, ethical concerns regarding the use of AI technology continue to emerge. In this context, establishing the direction of future education and identifying its core elements becomes a crucial task.

Text mining, a technique for extracting useful information and discovering valuable knowledge from unstructured text data, provides a methodology for objectively analyzing large-scale text data. While research utilizing text mining in education is increasing, most studies focus on curriculum analysis or learner feedback analysis.

Against this background, this study aims to systematically analyze education-related keywords associated with generative AI using text mining techniques to identify core elements of future education. This holds significance in several aspects: First, it can suggest future educational directions through objective data analysis. Second, it can provide practical implications for educational policy formation and curriculum development. Third, it can serve as foundational material for effectively utilizing generative AI in educational settings.

Therefore, this study holds significance in presenting future educational directions through objective data analysis based on generative AI-related theses from various academic research fields and deriving concrete implications applicable to actual educational settings. This is expected to serve as a case demonstrating the possibility of data-based approaches in future educational research. Specifically, the purpose of this study is to explore educational discourse on the educational utilization of generative AI and conduct status analysis through text mining.

1.2 Research Questions

This study has established the following research questions to analyze education-related keywords pertaining to generative artificial intelligence and identify core elements of future education through text mining techniques, focusing on academic theses related to “generative artificial intelligence” published from November 30, 2022, to December 31, 2024.

Research Question 1. What are the characteristics of key words appearing in educational discourse related to generative artificial intelligence? 1-1. What are the most frequently occurring keywords throughout the entire analysis period? 1-2. How does the frequency of keyword appearances change across different time periods?

Research Question 2. What characteristics are exhibited in the relationships between education-related keywords associated with generative artificial intelligence? 2-1. Which keywords demonstrate high centrality in the keyword network? 2-2. How are the keywords classified into clusters, and what are the characteristics of each cluster? 2-3. How does the interconnection structure between keywords change across different time periods?

Research Question 3. When synthesizing the results of keyword analysis and network analysis, what are the core elements of future education that emerge?

These research questions will be addressed through sequential and systematic analysis, with the results of each research question interconnecting to ultimately contribute to achieving this study’s objective of identifying core elements of future education.

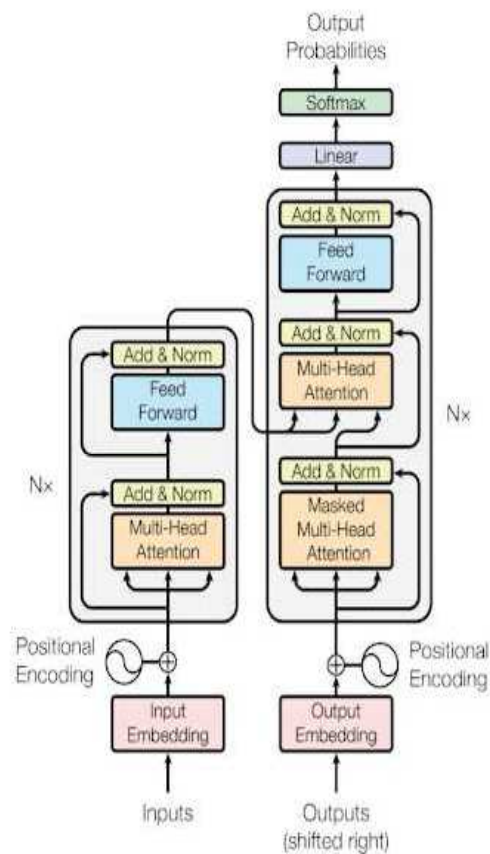
2. Theoretical Background

2.1 Understanding Generative Artificial Intelligence

2.1.1 Concept and Characteristics of Generative Artificial Intelligence

Generative artificial intelligence (generative AI) is a language model that interacts through conversational methods [3]. These systems, based on machine learning technology, are pre-trained large-scale machine learning models capable of autonomously generating new content using provided data. They demonstrate high performance across various applications by comprehensively understanding data and learning complex patterns [4].

Generative AI can generate ideas and content while solving problems by reusing new data with similar characteristics. Beyond text processing, more powerful models have emerged through multimodal model learning, which can learn and process relationships between different types of data [5]. These systems are being utilized across diverse industrial sectors, including natural and meaningful conversations, art, storytelling, software development, image generation, video production, and music creation [6].



(Figure 1) The Transformer - model architecture.
Source : Google Brain(2017),
"Attention Is All You Need"

2.1.2 Current Development Status of Generative Artificial Intelligence

Productive artificial intelligence made a rapid development from November 30, 2022, with the launch of CHATGPT. As of December 31, 2024, the development status of major generated AI models is as follows.

2.1.2.1 Language Models (LM)

(Table 1) Development Status of Generative AI Language Models

Types	Developers	Characteristics
ChatGPT-4o	OpenAI	- Trained using Reinforcement Learning with Human Feedback (RLHF), - Uses training through learning with reinforcement from human input, - Capable of performing complex reasoning and professional tasks[7]
Claude 3.5	Anthropic	- Advanced research and analysis capabilities, - Sophisticated Languages Understand and create[8]
Perplexity	Perplexity AI	AI platform specialized in data management and interpretation[9]

2.1.2.2 Image Generation Model

(Table 2) Development Status of Generative AI Image Generation Models

Types	Developers	Characteristics
Midjourney	Midjourney Inc.	- Text-to-Image Generation Models, - Supports diverse styles ranging from sketches to hyperrealistic images[10]
Leonardo.ai	Leonardo Interactive Pty Ltd.	- AI-Based Design Platform, - AI Canvas, - Intuitive Design Workflow[11]
Imagen 3	Google DeepMind	AI Platform Specialized in Data Management and Interpretation[12]

2.1.2.3 Video Generation Model

(Table 3) Development Status of Generative AI Video Generation Models

Types	Developers	Characteristics
Sora	OpenAI	• Text to Viideo AI Model, • High-Quality V i d e o Generation up to 60 Seconds[7]
runway	Runway AI, Inc.	• Text to Viideo AI Model, • Image to Video AI Model[13]
PixAI	Mewtant Inc.	• AI Art Generator, • Generation o f High-Resolution n Artwork, • Generation of Animation and Pixel Art S t y l e Images[14]

2.1.2.4 Music Generation Model

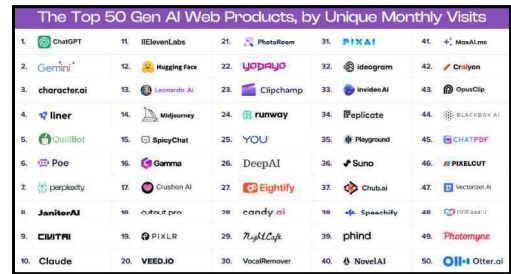
(Table 4) Development Status of Generative AI Music Generation Models

Types	Developers	Characteristics
Suno	Suno, Inc.	• Text-to-Music Generation, • Audio-to-Song Conversion[15]
Lyria	Google DeepMind	• Generation o f High-Quality Vocals, Lyrics, a n d Instrumental Accompanime nt, • Insertion of Imperceptible Watermarks[12]

2.1.2.5 Software Code Generation Models

(Table 5) Development Status of Software Code Generation Models

Types	Developers	Characteristics
GitHub Copilot	GitHub	• Real-Time Code Suggestions and Autocompletion, • Learning and Applying User Coding Styles[16]
Hugging Face	Hugging Face, Inc.	• Model Hub: Providing Over 1 0 0 , 0 0 0 Pre-trained Models, • Datasets: Library for NLP Tasks, • Tokenizers: T e x t Preprocessing Tools, • Spaces: Machine Learning Demo and Application S h a r i n g Platform[17]



(Figure 2) The Top 50 Gen AI Web Products, by Unique Monthly Visits
Source : Andreessen Horowitz, 2024

2.2 Text Analysis

2.2.1 Concept and Procedures of Text Mining

Text mining is a type of data mining that finds useful information from large amounts of data. It is a technique for discovering meaningful patterns and new insights by structuring text, which is unstructured data, into numbers that can be analyzed [18].

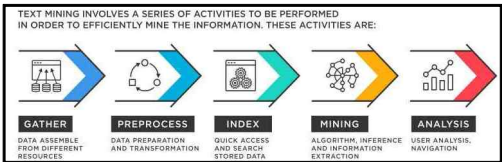


Figure 3) Text mining processing procedure
Source : [REDACTED]

The text mining procedure involves securing source data from various resources, followed by preprocessing steps to improve data quality through the removal of stop words and special characters for analysis. The process generates indices for efficient data retrieval and quick access, then employs mining operations using algorithms and inference to discover key information and patterns, providing valuable insights. In the analysis phase, which supports decision-making, the data is explored and the mining results are utilized to inform conclusions.

2.2.2 Review of Previous Research

A comprehensive review was conducted of various recent reports and published articles to examine the latest trends in generative artificial intelligence. The review of key previous research cases regarding the educational applications of AI and generative AI yielded the following results:

Yu Kyung-sun and Ahn Sung-jin [19] found that Harvard University professors' research demonstrated interactive lessons utilizing generative AI had a positive impact on personalized learning and problem-solving ability improvement.

Conversely, Zastudil et al. [20] highlighted the duality of generative AI. While it enables higher-order learning activities through coding automation, they raised concerns that excessive dependence on the tool could hinder metacognition necessary for understanding programming concepts.

The University of Leeds [21] case study analyzed the use of generative AI in advertising design production. The results showed that AI-generated images and ideas helped complement the creative process, but noted limitations in that they may lack the depth and expressiveness compared to human-created content.

HAI [22] analyzed the impact of generative AI across various fields including medicine, education, law, work, design, and art from the perspectives of 12 Stanford University professors and global leaders. While the researchers agreed that generative AI would have significant impact across industry and society, they emphasized the need for interdisciplinary cooperation and appropriate regulation to ensure the technology's benefits reach everyone.

Jeon In-sung [23] developed a real-time coding learning support system applying AI and learning analytics based on Entry. This system enables instructors to support individual learning and provide real-time feedback by monitoring learners' progress in real-time. The research showed positive

effects on students' computational thinking, learning motivation, and self-efficacy.

Synthesizing the previous research reveals that while generative AI shows potential as a supplementary learning tool in educational settings, empirical research on its effectiveness is still in early stages. In particular, there is a growing need for systematic research on the impact of generative AI on learners' self-efficacy, metacognition, and problem-solving abilities in computer science education.

This study was conducted to derive core elements of future education through the objective analysis method of text mining, aiming to provide practical implications for future educational policy formation and curriculum development.

3. Research Methodology

3.1 Analysis Subjects

Major academic research information services include RISS by the Korea Education and Research Information Service (KERIS), DBpia by Nurimedia Co., Ltd., KISS by Korean Studies Information Co., Ltd., Korea Citation Index (KCI), and the National Assembly Digital Library. To obtain meaningful data, the collection period was set from November 30, 2022, when ChatGPT was released, through December 31, 2024.

For this study, the Korea Citation Index (KCI) website was selected as the primary source, as it contained the largest number of searchable academic papers.

(Table 6) Status of Thesis Research Related to Generative Artificial Intelligence (Period: 2022-2024)

Source	Site	Number of Papers
Korea Education & Research Information Service	RISS	465
Nuri Media Co., Ltd.	DBpia	1,279
Korean Studies Information Co., Ltd.	KISS	448
Korea Citation Index	KCI	6,170

Source	Site	Number of Papers
National Assembly Library of the Republic of Korea	National Assembly Electronic Library	131
Total		8,493

3.2 Data Collection Method

The research data was collected through keyword searches on the Korea Citation Index (KCI) platform (<https://www.kci.go.kr/>), a system providing access to high-quality academic journal information across various research fields. The search focused on domestic theses and dissertations published between 2022 and 2024, yielding a total of 6,170 academic papers.

For this study, data was collected from the KCI database by identifying papers that included the keyword "generative" in their titles. The initial collection included paper IDs, titles, authors, and publication years for all relevant papers, resulting in data for 6,170 publications. The search results were exported and saved as Excel documents through the platform's bulk download feature.

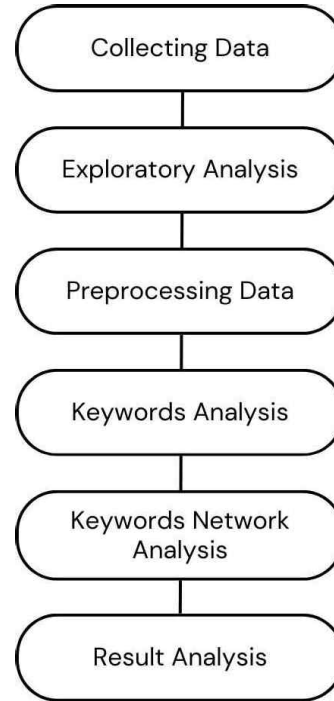
Due to the system's limitation of 5,000 records per download, the data collection process required downloading results by year and then consolidating them into a single Excel document. The consolidated data was then refined by removing the header row and retaining only the paper titles and Korean keywords columns, which were saved as a CSV file. To streamline the analysis, several non-essential columns were removed, including paper ID, foreign language titles, authors, co-authors, journal names, foreign language journal names, publishing institutions, foreign language institution names, registration status, publication year, volume/issue numbers, pages, foreign language keywords, subject areas, citation counts, URLs, and DOIs.

3.3 Analysis Procedure

In this study, text mining techniques were used to analyze the topic keywords that appeared in theses related to generative AI in various academic research fields. 'Text mining' is an analytical methodology that utilizes computer technology to automatically derive meaningful patterns and insights from text data. By systematically analyzing large amounts of text data through automated algorithms, it is possible to uncover potential meanings and relationships that would be difficult to discover by human inspection. In particular, the main advantage of text mining is that the

analysis process is defined by a clear algorithm, which ensures transparency and reproducibility of research and allows other researchers to verify the results by applying the same methodology [24]. The process of preprocessing the texts to analyze the collected unstructured data was performed as shown in Figure 4.

Collecting Data > Exploratory Analysis > Preprocessing Data > Keywords Analysis > Keywords Network Analysis > Result Analysis

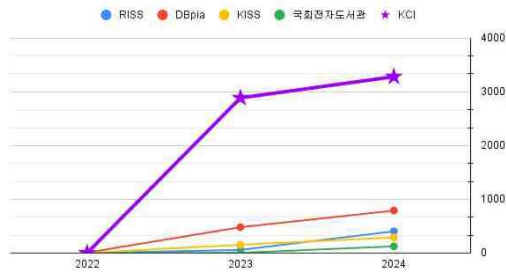


(Figure 4) Data collection and analysis procedures

Exploratory data analysis is the stage of understanding the data collected, and we visualized the number of papers per year to explore the features of the data.

(Table 7) Theses on generative AI by year of publication (unit: number of sides)

Source	Site	2022	2023	2024
Korea Education & Research Information Service	RISS	2	58	405
Nuri Media Co., Ltd.	DBpia	8	481	790
Korean Studies Information Co., Ltd.	KISS	3	154	291
Korea Citation Index	KCI	0	2,888	3,282
National Assembly Library of the Republic of Korea	National Assembly Electronic Library	1	5	125
Total		14	3,586	4,893



(Figure 5) Line chart of generative AI theses by year of publication

Starting with 14 papers in 2022, we see a 350-fold increase to 4,893 papers in 2024. Of these, KCI shows the steepest increase in data from November 2022 to 2023, while the other platforms have similar slopes, but a steadily growing curve. This suggests that generative AI research will continue to grow in the future.

3.4 Data Preprocessing

Since text data is based on human natural language communication, it requires separate conversion into a structured format for machine analysis by computer [25]. In this conversion process, i.e., the preprocessing stage, a series of systematic procedures are performed to refine and standardize the data into a form that computers can process. This can be considered a core process of converting the complexity and diversity of natural language into structured data that machines can understand and analyze.

This study utilized the Python KoNLPy package in Colab as the development environment. KoNLPy is a Python package for Korean natural language processing. It helps

easily perform sentence separation, morphological analysis, stemming, semantic role extraction, and named entity recognition for Korean natural language processing. 'Keyword frequency analysis' is a method of analyzing the importance of keywords based on the frequency of words appearing in the data, allowing for verification of basic information about the analyzed data.

This research conducted preprocessing as follows: First, as a basic setting for preprocessing, text was separated into individual words from the collected paper titles. Second, to remove stopwords, major Korean particles and endings ('ㅇ', '가', '을', '를'), conjunctions ('및', '또는', '그리고'), and other unnecessary words (English articles, prepositions, etc.) were defined and removed as stopwords. Third, special characters, slashes, brackets, etc. were removed using a refinement function. Fourth, the refined words were recombined into a single string. The analysis was conducted on 4,910 words extracted during the preprocessing process.

3.5 Text Mining

The primary objective of this research is to identify key elements in the educational application of generative AI. Keyword Frequency Analysis enables objective identification of major concepts frequently appearing in academic papers, while Network Analysis allows for systematic examination of organic relationships between these concepts. When compared to other text mining techniques, this methodology most directly aligns with achieving the research objectives.

Considering the characteristics of the data, the 6,170 academic papers under analysis are formal documents written in professional and objective language. Therefore, techniques such as sentiment analysis or topic modeling are not compatible with the nature of this data. In academic papers, explicitly used technical terms and their relationships are more important subjects of analysis than the author's emotions or latent topics.

Keyword Frequency Analysis and Network Analysis are methodologies whose validity and reliability have already been verified in previous educational research. These analytical techniques have been frequently utilized in basic research for curriculum development and educational policy formation, enabling the derivation of empirical results that align with this study's objectives.

In terms of analytical efficiency, while techniques such as text classification or summarization can analyze individual document content more minutely, they may be inefficient in identifying overall trends and patterns in large-scale literature. In contrast, Keyword Frequency

Analysis and Network Analysis can effectively derive core concepts and their relationships from large volumes of documents. Considering these factors collectively, the selection of Keyword Frequency Analysis and Network Analysis as primary methodologies can be considered valid in terms of research objectives, data characteristics, methodological reliability, and analytical efficiency.

3.5.1 Keyword Frequency Analysis

In this research, Python's `re` module was utilized to perform Keyword Frequency Analysis on paper titles. 'Keyword Frequency Analysis' is a fundamental text mining technique that quantitatively measures the frequency of words appearing in text data to understand the main content and characteristics of the text. By enabling objective verification of key word distribution in large volumes of text data, it provides essential baseline data for understanding the overall context and central themes of the text.

The study implemented an automated frequency analysis algorithm based on Python to quantitatively analyze keyword distribution and importance in text data. First, a hash table-based dictionary data structure was used to efficiently calculate word occurrence frequency, and through the `get_word_frequency` function, the frequency of individual words in the preprocessed text corpus was systematically tallied. Specifically, this function performs an iterative algorithm that separates the input text into individual word units (tokens) and updates the cumulative frequency for each word occurrence.

To ensure the reliability of the analysis results, the correspondence between the original text and preprocessed text was verified, and a keyword list sorted in descending order of frequency was generated to identify key terms. Additionally, the lexical diversity of the research text was quantified by calculating the total number of unique words in the corpus. Finally, to ensure reproducibility of the analysis, the preprocessed data was saved in a structured format (CSV), considering potential use in subsequent research.

3.5.2 Keyword Network Analysis

In this research, keyword network analysis was performed using Python through a systematic process consisting of several key stages:

First, in the analysis environment setup phase, essential libraries were installed with particular attention to Korean text readability through the integration of the NanumBarunGothic font. This included automated font

package installation and configuration specifically tailored for the Google Colab environment.

Second, the data preprocessing phase utilized the `pandas` library to load research paper titles from CSV files (including `'preprocessed_research_titles_2023키워드.csv'`) into dataframe format. The analysis employed the `Okt` morphological analyzer from `KoNLPy`, a Korean natural language processing library, with a defined `get_nouns()` function to extract nouns from each research title.

To analyze the frequency of extracted nouns, the `Counter` class from Python's `collections` module was implemented. The process involved consolidating nouns from all research titles into a unified corpus, followed by calculating the occurrence frequency of each word.

For network analysis, the `networkx` library was employed to construct network graphs based on word co-occurrence relationships. To ensure meaningful analysis and reduce noise, only words appearing 50 times or more were selected as analysis subjects. Network edges were added to connect words that co-occurred within the same research title.

Finally, the constructed network was visualized using the `matplotlib` library. The visualization was configured with a 12x8 inch graph size, with each node labeled with its corresponding word to facilitate intuitive understanding of the network structure. This comprehensive analytical approach enabled quantitative understanding of relationships between research keywords and systematic identification of structural characteristics in research trends.

4. Research Findings

4.1 Keyword Analysis Results

A simple keyword frequency analysis was performed to identify the most frequently used keywords in thesis titles, with results shown in Tables 8 and 9. These tables present the top 20 most frequently mentioned words within the paper titles for each period.

Analysis of 2023 paper titles revealed that 'generative', the primary research keyword, appeared most frequently with 149 occurrences, followed by 'artificial intelligence, AI' with 140 occurrences, 'utilization' with 65 occurrences, 'research' with 55 occurrences, and 'focus' with 33 occurrences.

In 2024 paper titles, 'generative' again appeared most frequently with 510 occurrences, followed by 'AI, artificial intelligence' with 467 occurrences, 'utilization' with 212 occurrences, 'research' with 176 occurrences, and 'focus' with 142 occurrences.

(Table 8) Keyword Frequency Analysis Results – 2023 Paper Titles

Rank	Keyword	Frequency
1	Generative	149
2	Artificial Intelligence	72
3	AI	68
4	Utilization	65
5	Research	55
6	Focus	33
7	Based	15
8	Methods	12
9	Through	11
10	Image	11
11	Design	9
12	ChatGPT	9
13	Exploration	9
14	Analysis	8
15	Recognition	8
16	Literacy	8
17	Creation	7
18	Education	7
19	Development	7
20	Investigation	7

(Table 9) Keyword Frequency Analysis Results – 2024 Paper Titles

Rank	Keyword	Frequency
1	Generative	510
2	AI	334
3	Utilization	212
4	Research	176
5	Focus	142
6	Artificial Intelligence	133
7	Based	59
8	Analysis	44
9	Methods	44
10	Impact	36
11	Design	33
12	Development	33
13	Education	33
14	Image	28
15	Case Study	27
16	Exploration	27
17	Through	22
18	Recognition	21
19	Investigation	19
20	Production	17

(Table 10) Keyword Frequency Analysis Results – Compare Year-over-Year Rankings

Rank	2023 Paper Titles Keyword	2024 Paper Titles Keyword
1	Generative	Generative
2	Artificial Intelligence	AI
3	AI	Utilization
4	Utilization	Research
5	Research	Focus
6	Focus	Artificial Intelligence
7	Based	Based
8	Methods	Analysis(14→8)
9	Through	Methods
10	Image	Impact
11	Design	Design
12	ChatGPT	Development(19→12)
13	Exploration	Education(18→13)
14	Analysis	Image
15	Recognition	Case Study
16	Literacy	Exploration
17	Creation	Through
18	Education	Recognition
19	Development	Investigation
20	Investigation	Production

This research conducted a comparative analysis of keyword frequency rankings between 2023 and 2024, revealing significant shifts in research priorities and focus areas during this period.

The analysis showed several notable ranking changes. 'Analysis' demonstrated substantial upward mobility, advancing from 14th to 8th position. Similarly, 'Development' improved from 19th to 12th place, while 'Education' moved up from 18th to 13th position. The emergence of new keywords such as 'Impact', 'Case Study', and 'Production' in the 2024 rankings indicates an evolution in research priorities.

The consistent high placement of keywords such as 'Utilization' and 'Research' demonstrates a balanced approach between practical applications and academic inquiry in the field. A particularly significant observation is the disappearance of 'ChatGPT' from the rankings. This change suggests a normalization of interest following the initial surge of attention toward this technology.

The sharp rise in ranking for 'Analysis' indicates a transition toward practical implementation of generative AI technology. This shift, combined with other ranking changes, suggests the field is moving from a focus on the technology itself toward practical applications and real-world implementation.

A keyword frequency analysis was conducted to examine frequently used terms in thesis abstracts' keywords, with

results presented in Tables 11 and 12. These tables display the top 20 most frequently mentioned terms in the paper abstracts' keywords for each year.

In 2023 abstracts, 'generative' appeared most frequently with 127 occurrences, followed by 'artificial intelligence/AI' with 200 occurrences, 'ChatGPT/chatGPT' with 53 occurrences, 'education' with 26 occurrences, and 'utilization' with 13 occurrences.

The 2024 abstracts showed 'generative' as the most frequent keyword with 399 occurrences, followed by 'AI/artificial intelligence' with 608 occurrences, 'education/instruction' with 96 occurrences, and 'Generative/generation' with 108 occurrences.

(Table 11) Keyword Frequency Analysis Results
- Topic words in 2023 Thesis Abstracts

Rank	Keyword	Frequency
1	Generative(Korean-‘생성형’)	127
2	Artificial Intelligence	101
3	AI	99
4	ChatGPT(English)	34
5	Education	26
6	Chat(Korean-‘챗’)GPT	19
7	Image	13
8	Design	13
9	Utilization	13
10	Literacy	12
11	Generative(English)	12
12	Creation	10
13	Writing	10
14	Model	9
15	Prompt	9
16	Large-scale	8
17	Fair Use	7
18	Copyright	7
19	ChatGPT(Korean-‘챗지피티’)	7
20	Language	6

(Table 12) Keyword Frequency Analysis Results
- Topic words in 2024 Thesis Abstracts

Rank	Keyword	Frequency
1	Generative(Korean-‘생성형’)	399
2	AI	370
3	Artificial Intelligence	238
4	Education	75
5	Design	62
6	Generative(English)	47
7	ChatGPT(English)	42
8	Prompt	39
9	Chat(Korean-‘챗’)GPT	32
10	Literacy	29
11	Image	26
12	Digital	25
13	Generative(Korean-‘생성형’) AI	24
14	Model	22
15	Classes	21
16	Writing	21
17	Learning	21
18	Generate	19
19	Research	19
20	GPT(English)	18

(Table 13) Keyword Frequency Analysis Results
- Compare Year-over-Year Rankings

Rank	2023 Paper Topic words	2024 Paper Topic words
1	Generative (Korean-‘생성형’)	Generative (Korean-‘생성형’)
2	Artificial Intelligence	AI
3	AI	Artificial Intelligence
4	ChatGPT(English)	Education(5→4)
5	Education	Design(8→5)
6	Chat(Korean-‘챗’)GPT	Generative (English)(11→6)
7	Image	ChatGPT(English)
8	Design	Prompt(15→8)
9	Utilization	Chat(Korean-‘챗’)GPT
10	Literacy	Literacy
11	Generative(English)	Image
12	Creation	Digital
13	Writing	Generative (Korean-‘생성형’) AI
14	Model	Model
15	Prompt	Classes
16	Large-scale	Writing
17	Fair Use	Learning
18	Copyright	Generate
19	ChatGPT (Korean-‘챗지피티’)	Research
20	Language	GPT(English)

This study conducted a comparative keyword frequency analysis of abstract keywords following the same

year-over-year methodology used for paper titles. The analysis revealed several significant shifts in research focus between 2023 and 2024.

A notable change was observed in the positioning of key terms. 'ChatGPT' declined from 4th to 7th position, while 'education' advanced from 5th to 4th position. This shift suggests an expanding application of generative AI technology in educational contexts. Particularly noteworthy was the substantial rise of 'prompt' from 15th to 8th position, reflecting the growing recognition of prompt engineering's importance in AI technology implementation.

The emergence of new keywords such as 'class,' 'learning,' and 'research' in the rankings indicates a concrete movement toward implementing AI technology in educational settings. This trend demonstrates a transition from theoretical frameworks to practical applications in educational environments.

The 2024 keyword analysis results comprehensively indicate that AI technology is progressing beyond theoretical stages toward practical implementation in educational settings. The rise in education-related keywords particularly demonstrates the emergence of educational applications of AI as a primary research focus.

This study applied word cloud visualization techniques to intuitively understand the keyword occurrence patterns derived through frequency analysis. Figures 6 and 7 provide visual representations where the frequency of each keyword is converted into font size, with more frequently occurring words displayed in larger text. This approach effectively illustrates the relative importance of key terms in the research corpus.

The visualization methodology translates raw frequency data into an immediately comprehensible format, where the proportional sizing of words creates a visual hierarchy that reflects their prevalence in the analyzed texts. High-frequency keywords dominate the visual space, while less common terms appear in smaller text, creating an organic representation of the research landscape.



(Figure 6-1) The word cloud visualization of frequently occurring keywords in 2023 thesis titles



(Figure 6-2) The word cloud visualization of frequently occurring keywords in 2024 thesis titles

The word cloud was recreated with meaningful keyword distribution patterns after removing high-frequency terms '생성형' (Generative), 'AI', and '인공지능' (Artificial Intelligence) as stop words. This refined visualization revealed distinct patterns between 2023 and 2024.

The 2023 visualization shows a more dispersed pattern, centered around '교육' (Education) and '활용' (Utilization). Keywords appear in varying sizes and are distributed more sporadically across the visual space, suggesting a broader but less focused research landscape during this period.

In contrast, the 2024 visualization demonstrates a more structured pattern. Key terms such as '활용' (Utilization), '연구' (Research), and '교육' (Education) dominate the central space with larger text sizes. The surrounding keywords show a more balanced distribution, indicating a



(Figure 7-1) The word cloud visualization of frequently occurring keywords in 2023 thesis titles



(Figure 7-2) The word cloud visualization of frequently occurring keywords in 2024 thesis titles

The comparative analysis of word clouds Figures 8 and 9 visualizing thesis abstract keywords reveals a notable

evolution in research focus between 2023 and 2024. This shift demonstrates the field's progression toward greater academic sophistication.

The 2023 visualization prominently features technically-oriented terms in its central region, with keywords such as '데이터' (Data), '학습' (Learning), and '이미지' (Image) dominating the space. This arrangement reflects the period's primary emphasis on technological implementation and technical capabilities.

In contrast, the 2024 visualization shows a significant shift toward more academic terminology. Keywords such as '학습' (Learning), '기반' (Based), and '연구' (Research) occupy central positions, accompanied by an increased presence of scholarly terms. The emergence of research methodology-related terminology suggests a more rigorous academic approach to the field.

A particularly noteworthy development is the increased frequency of academic terms such as '연구' (Research), '평가' (Evaluation), and '분석' (Analysis). This change indicates that research in generative AI is transitioning from an experimental phase to a stage of systematic validation and verification.

This evolution in keyword distribution suggests that thesis research is moving beyond technical implementation toward methodological consideration and validation. The shift reflects growing academic maturity in the field, as evidenced by the increased emphasis on theoretical frameworks and systematic evaluation methods.



(Figure 8) The word cloud visualization for frequently occurring keywords in 2023 thesis abstracts



(Figure 9) The word cloud visualization for frequently occurring keywords in 2024 thesis abstracts

Through keyword network analysis, we visualized the interconnected structure of research topics. The network analysis revealed a complex network where major nodes are interconnected. Nodes positioned in the central area demonstrate particularly high connection density, suggesting that these keywords function as core research themes. These highly central nodes can be interpreted as concepts that form the theoretical foundation of the research field.

The overall network structure exhibits a relatively balanced, distributed form. Most nodes are directly connected to multiple other nodes, indicating active interaction between research topics. Notably, there are clear connections between central and peripheral nodes, demonstrating that new research topics are organically integrating with and evolving from existing research foundations.

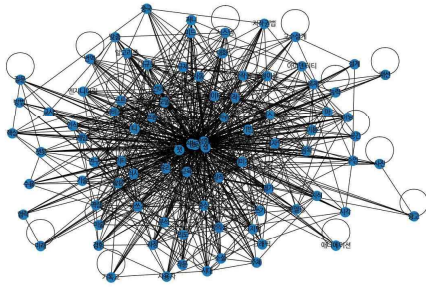
These characteristics of the network structure indicate that the research field is expanding into various subtopics based on a solid theoretical foundation, while maintaining tight interconnections between research topics. This simultaneously demonstrates both the maturity and integration of the research field, providing important implications for predicting future research directions.

In generating co-occurrence matrices for thesis abstract keywords, we conducted sequential analyses with frequency thresholds of 3, 10, and 20 occurrences for 2023 theses, and 3, 10, 20, and 50 occurrences for 2024 theses, as illustrated in Figures 10-16. Examining the network characteristics of high-frequency terms that appeared more than 50 times in the 2024 analysis, we could visually identify relationships between key concepts emerging from research trends.

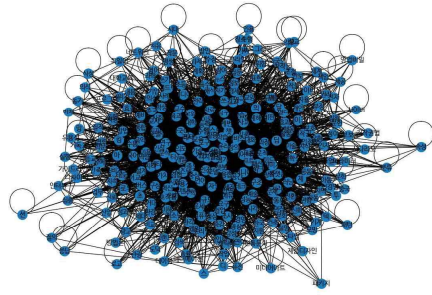
The network exhibits a fully connected structure where all nodes are interconnected, indicating that the analyzed research studies address closely related topics [26]. The structure displays seven nodes intricately connected by edges, forming a complex network. Each node represents a key term frequently appearing in research titles, while the edges between nodes indicate the co-occurrence of these keywords within the same research titles.

Notably, nodes positioned at the network's center demonstrate high connectivity with other keywords, suggesting their role as fundamental concepts within the research field. This central positioning and high degree of interconnectedness imply that these terms serve as pivotal elements linking various research themes.

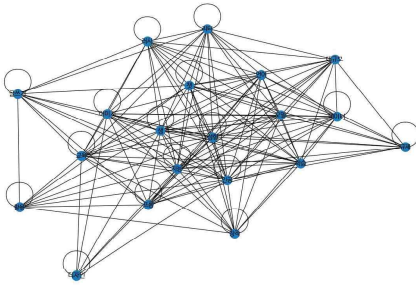
4.2 Keyword Network Analysis Results



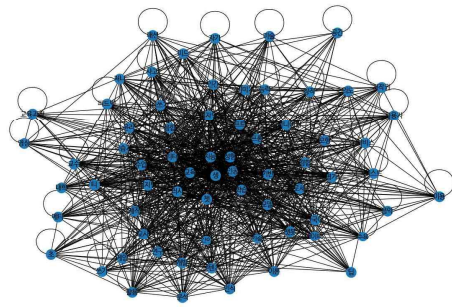
(Figure 10) Topic word frequency keyword network analysis of 2023 dissertation abstracts
- 3 or more Co-occurrences



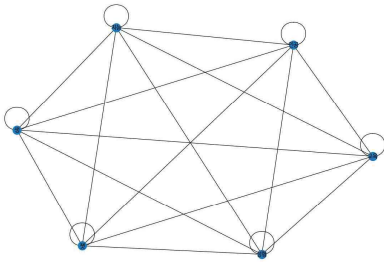
(Figure 13) Topic word frequency keyword network analysis of 2024 dissertation abstracts
- 3 or more Co-occurrences



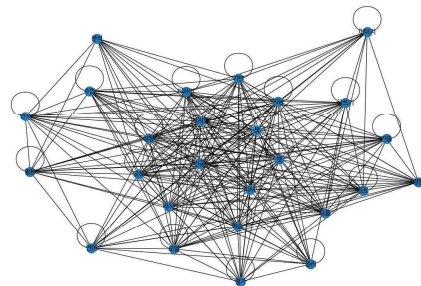
(Figure 11) Topic word frequency keyword network analysis of 2023 dissertation abstracts
- 10 or more Co-occurrences



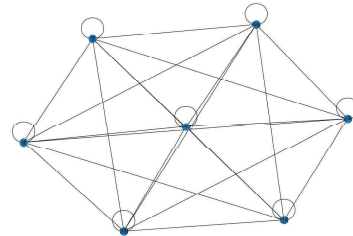
(Figure 14) Topic word frequency keyword network analysis of 2024 dissertation abstracts
- 10 or more Co-occurrences



(Figure 12) Topic word frequency keyword network analysis of 2023 dissertation abstracts
- 20 or more Co-occurrences



(Figure 15) Topic word frequency keyword network analysis of 2024 dissertation abstracts
- 20 or more Co-occurrences



(Figure 16) Topic word frequency keyword network analysis of 2024 dissertation abstracts
- 50 or more Co-occurrences

5. Conclusion and Recommendations

5.1 Summary

This research was conducted to identify key elements related to the educational application of generative artificial intelligence, preceding the introduction of AI digital textbooks in 2025. We performed keyword analysis using text mining techniques on 6,170 academic papers published between 2022 and 2024.

The keyword frequency analysis revealed 'generative', 'AI', and 'utilization' as primary keywords. Through network analysis, we were able to identify the relationships between these keywords and their cluster structures. Notably, the keyword 'design' emerged as a central element, demonstrating strong correlations with various educational components.

5.2 Educational Implications

Based on our analysis, the following educational implications emerge:

First, regarding the sustainability of generative AI, our research indicates that this technology will serve as a fundamental driving force for transforming educational paradigms rather than a temporary trend. The implementation of generative AI in educational settings requires establishing systematic utilization strategies beforehand. This is supported by the prominence of keywords such as 'utilization' and 'design' in our findings.

Second, teacher education and professional development are essential components. Considering the rapid advancement of generative AI technology, the high co-occurrence frequency of the 'literacy' node among keywords suggests a need for continuous teacher training and support. This finding emphasizes the importance of ongoing professional development programs to ensure effective implementation.

Third, in terms of ethical utilization, our analysis revealed distinct node clusters forming around 'ethics' and 'copyright.' The high betweenness centrality of these ethics-related keywords empirically demonstrates that ethical considerations play a central role in the educational application of generative AI.

5.3 Research Limitations and Recommendations

The limitations of this study and recommendations for future research can be categorized into four key areas.

The first limitation concerns the temporal scope of our

research, which covers only the period from 2022 to 2024. This relatively short timeframe restricts our ability to forecast the long-term impact of generative AI technology in education. Future research would benefit from longitudinal studies that can better capture the evolving patterns and sustained effects of this technology in educational contexts.

The second limitation stems from the technical constraints of text mining analysis. While this methodology provided valuable insights, it may not have fully captured contextual meanings and latent semantic structures within the analyzed texts. Future studies should incorporate qualitative research methods alongside quantitative analysis to achieve a more nuanced and comprehensive understanding of AI integration in education.

The third limitation relates to the lack of empirical validation in actual educational settings. Our analysis, while thorough in terms of literature review, does not include substantial evidence from practical implementations. Future research should prioritize empirical case studies that can systematically evaluate both the benefits and limitations of generative AI in real educational environments.

Fourth, there is a need for practical research to establish ethical guidelines. For example, in its 2021 "Recommendations for the Ethical Use of AI", UNESCO proposed a global framework to address ethical issues in the use of AI in education [27]. Responsibility for ethical issues that may arise in the educational use of generative AI should be clarified and appropriate oversight and evaluation systems should be established. This is essential to strengthen the ethical accountability of educational institutions and build a sustainable AI education ecosystem.

This is expected to make generative AI a trusted tool in education, ensuring both learners' rights and the quality of education.

References

- [1] Ministry of Education, "2023 Digital Education White Paper", 2023.
- [2] Ministry of Education, "Major Policy Action Plan 2024 - Addressing Social Challenges through Education Reform", 2024
- [3] Roberto Gozalo-Brizuela, Eduardo C. Garrido-Merchan, "ChatGPT is not all you need. A State of Art Review of large Generative AI models", <https://arxiv.org/abs/2301.04655>, 2023.

- [4] Amazon. (2024, December 30). "What Is Generative AI?". AWS. https://aws.amazon.com/ko/what-is/generative-ai/?nc1=f_cc
- [5] Ray, S. "Advanced Machine Learning Techniques in AI Development", *AI and Machine Learning Journal*, Vol. 29, No. 4, pp. 321-337, 2023.
- [6] Kim, Jung-Ah, Kang, Doo-Sik, and Lee, Yong-Cheol, "A Study on Educational Utilization of Generative AI - Focusing on the Use of ChatGPT", *Journal of Information Education*, Vol. 112, pp. 691-704, 2023.
<http://dx.doi.org/10.14352/jkaie.2023.27.6.691>
- [7] OpenAI. (2024, December 30). "Introducing ChatGPT". OpenAI. <https://openai.com/index/chatgpt/>
- [8] ANTHROPIC. (2024, December 30). "Meet Claude". ANTHROPIC. <https://www.anthropic.com/claude>
- [9] Perplexity AI, Inc. (2024, December 30). "Perplexity AI Description". Perplexity AI. <https://www.perplexity.ai/pro>
- [10] Midjourney. (2024, December 30). "Midjourney About". Midjourney. <https://www.midjourney.com/home>
- [11] Leonardo Interactive Pty Ltd® . (2024, December 30). "Enhance your AI Art with Leonardo.Ai". Leonardo.Ai. <https://leonardo.ai/>
- [12] Google DeepMind. (2024, December 30). "Imagen_3_Tech_Report". Google DeepMind. https://storage.googleapis.com/deepmind-media/imagen/imagen_3_tech_report_update_dec2024_v3.pdf#page=26
- [13] Runway AI, Inc. (2024, December 30). "Runway Research". Runway. <https://runwayml.com/>
- [14] Metanomaly Pte. Ltd. (2024, December 30). "About the XL model at PixAI". Pixai.Art. <https://support.pixai.art/en/articles/9291432-pixai%E3%81%A7%E3%81%AE%E3%83%A2%E3%83%87%E3%83%AB%E3%81%AB%E3%81%A4%E3%81%84%E3%81%A6>
- [15] Suno, Inc. (2024, December 30). "Suno About". SUNO. <https://suno.com/about>
- [16] GitHub, Inc. (2024, December 30). "Preview upcoming features for GitHub Copilot". GitHub. <https://github.com/features/copilot>
- [17] Hugging Face. (2024, December 30). "Welcome to Inference Providers on the Hub". Hugging Face. <https://huggingface.co/blog/inference-providers>
- [18] Ho-Yeon Gil, "A Study on Korean Unused Words List for Text Mining", Seowon University, 2018.
- [19] Kyung-Sun Yoo and Sung-Jin Ahn, "Effects of Training on College Students' Academic Self-efficacy, Metacognition, and Problem-solving Skills Using Generative Artificial Intelligence", *Journal of Computer Education*, Vol. 27, No. 4, pp. 13-20, 2024. <https://doi.org/10.32431/kace.2024.27.4.002>
- [20] Zastudil, C., Rogalska, M., Kapp, C., Vaughn, J., & MacNeil, S., "Generative AI in Computing Education: Perspectives of Students and Instructors", In *2023 IEEE Frontiers in Education Conference (FIE)*, 2023.
- [21] University of Leeds, "Case studies: How is Generative AI being used at Leeds? In Generative AI Leeds", Retrieved from <https://generative-ai.leeds.ac.uk/case-studies>, 2023.
- [22] HAI, "Generative AI: Perspectives from Stanford HAI", *Human-Centered Artificial Intelligence*, 2023
- [23] Inseong Jeon, "Development of Block based SW-AI Education Teaching and Learning Support System Using Artificial Intelligence and Learning Analytics", *Korea National University of Education*, 2023.
- [24] Jong-Hee Park, Eun-Jung Park and Dong-Jun Cho, "An Automated Textual Analysis of North Korean New Year's Speeches(1946-2015)", *Journal of the Korean Political Science Association*, Vol. 49, No. 2, pp. 27-61, 2015.
<http://dx.doi.org/10.18854/kpsr.2015.49.2.002>
- [25] Eunhye Ham, Yeerim Yoo, "Characterization of College Students' World Understanding Essay-type Assessment Answers by Performance Level Using Text Mining Techniques", *Educational Evaluation Research*, Vol. 35, No. 4, pp. 687-717, 2022.
- [26] NetworkX developers. (2024, December 30). "NetworkX documentation". NetworkX. <https://networkx.org/>
- [27] UNESCO Korean Commission, "Commentary on the UNESCO Ethical Recommendations on Artificial Intelligence (AI)", 2021.

● Author Biography ●



Gi-A Kim

Adjunct Professor at Shinhan University

Graduated from Seoul School of Integrated Sciences & Technologies (aSSIST), Graduate School of AI, Department of AI Advanced Studies, Major in AI Big Data (Master of Engineering)

Research Interests : AI, Big Data Analytics, Natural Language Processing, Generative AI, Educational Data Mining etc.

E-mail : sdream.pm@gmail.com