AI Journal of Business

Volume 1 Number 2 May 2025

02
9 773058 699007
ISSN 3058-6992

AI Business Academy
www.aiba.or.kr

Content

Papers

Informantions

Development of Machine Learning Models for Predicting Academic Stress in Adolescents

- A Survey on Mental Health Status of Adolescents in 2021 -

Jeong-suk So1 Tae-yeon Oh2

ABSTRACT

In the field of social science research, traditional statistical methods have been predominantly used to analyze the major causes of adolescent stress. Most existing studies focus on this approach. In relation to academic stress, this study aims to develop an analysis model based on artificial intelligence and machine learning-driven statistical methods, utilizing big data analysis. Academic stress has a significant impact on the mental health of adolescents, and predicting it can provide valuable insights into students' stress levels, ultimately suggesting effective ways to alleviate their stress. The collected data was sourced from the 2021 Korea Children and Youth Data Archive survey, where 225 common variables were selected. To identify specific variables influencing adolescent academic stress and to choose the most effective machine learning method, five regression models—Random Forest, Linear Regression, Ridge, Lasso, and Gradient Boosting—were applied to determine significant variables.

The performance of these regression models was evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 Score as assessment metrics. Based on these evaluations, Gradient Boosting was selected as the most effective predictive model. Additionally, an analysis of variable importance was conducted to identify key factors affecting academic stress

☞ keyword : Academic Stress, Machine Learning, Gradient Boosting, Adolescent Mental Health, Big Data Analysis

1. Introduction

According to the Children and Youth Quality of Life 2022 report by Statistics Korea, the largest proportion of leisure time spent by Korean adolescents is on private academies and tutoring, accounting for 47.3%. In contrast, smartphone usage only accounts for 14.1%. This suggests that students do not have sufficient rest and dedicate a significant amount of time to academic activities.

Additionally, the PISA 2015 report by the OECD indicates that the academic-related anxiety index among Korean adolescents is 0.10, which is higher than the OECD average of 0.01. This demonstrates that Korean adolescents experience relatively higher academic stress compared to their peers in other major countries. Notably, 41.9% of students responded affirmatively to the statement, "I feel very tense when studying. Given that the percentage of OECD adolescents who agreed with the same statement is 36.6%, it can be inferred that academic anxiety and stress are particularly severe among Korean students.[1]

According to the 2023 Youth Statistics report published by the Korea Youth Policy Institute under the Ministry of Gender Equality and Family, a survey on depression and stress among adolescents revealed that 41.3% of middle and high school students reported feeling stressed in 2022. This figure represents a 2.5% increase compared to the previous year[2]. Similarly, the 2017 Survey on the Human Rights of Children and Adolescent reported that Korean adolescents suffer from sleep deprivation and academic stress, with their overall happiness levels being the lowest among OECD and major European countries. Among various stress factors, academic stress particularly highlights the significant difficulties adolescents face in their studies[3]. School serves as a crucial environment for adolescents' psychological, social, and cognitive development, helping them form proper values and enhance their social adaptation skills through interpersonal interactions[4]. According to a study by the Korea Institute for Health and Social Affairs, the academic stress level of Korean adolescents is significantly higher than that of their peers in other countries, recorded at 50.5%, compared to the average of 33.3% across 30 surveyed countries[5]. Korean adolescents spend a significant amount of time in school, and much of their academic stress originates from school-related factors. Many reports indicate that schools account for a substantial portion of the stress experienced by students.

This study aims to develop a machine learning model for predicting adolescent academic stress and to analyze specific variables that may influence academic stress. In particular, this research focuses on constructing a predictive model that identifies the key causes of academic stress and assesses their impact, with the goal of accurately forecasting academic stress levels.

The primary objective of this study is to enhance the

predictability of academic stress. While many existing studies have analyzed the causes and effects of academic stress, relatively few have proposed quantitative predictive models. Therefore, this research aims to build an academic stress prediction model using machine learning techniques, with the expectation that this model will enable more accurate predictions of students' academic stress levels.

Additionally, this study seeks to provide educational implications by leveraging machine learning techniques. Utilizing machine learning for predicting academic stress facilitates data-driven decision-making and effectively processes a wide range of variables. By applying a machine learning model, this research aims to predict students' academic stress and, based on the findings, derive effective strategies for stress management.

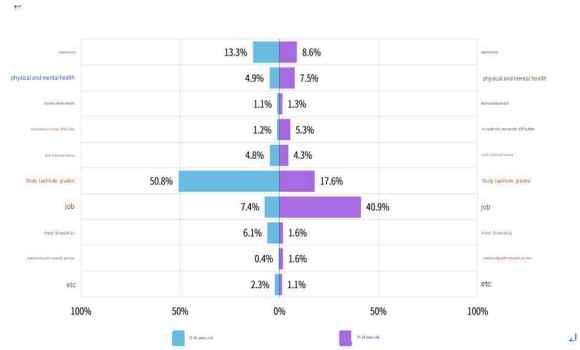
2. Literature Review

2.1 Academic Stress

Stress can be experienced severely due to sudden and unexpected life events. However, it can also have a positive effect by enhancing concentration and preventing tasks from stagnating[3]. Severe and chronic stress depletes mental and physical resources, leading to exhaustion(burnout). This occurs when individuals become excessively fixated on their goals and push themselves too hard, both physically and mentally. It is a condition in which anxiety and tension persist for several weeks or more. In contrast, general stress frequently arises in daily life due to both major and minor tensions or trivial events. Academic stress refers to the negative psychological states such as tension, anxiety, and worry that students experience in relation to grades and studying, among various stress factors encountered in school life[6].

Considering this, academics can be one of the major sources of stress for adolescents in South Korea. Previous studies have also found that Korean adolescents experience academic stress, and that the excessive academic burden is the primary cause of stress for students, aligning with these findings[7].

Especially during adolescence, academic stress has a significant impact on students' personal, social, and academic lives. The potential negative consequences of academic stress include emotional instability, depression, decreased self-esteem, and even physical illnesses, highlighting its seriousness. This type of stress is closely related to the psychological pressure that adolescents experience while engaging in academic activities, representing the mental burden that individual students bear in pursuit of academic achievement. According to a 2022 survey conducted by the Youth Policy Analysis and Evaluation Center targeting 200 male and female adolescents aged 13 to 18, the most concerning issue among teenagers was studies (aptitude and grades), accounting for 50.8%, the highest proportion Figure 1.



(Figure 1. Problems that teenagers are concerned about)
(Data: Youth Policy Analysis and Evaluation Center (by age as of 2022)

2.1.2 Review of Previous Studies

In this study, KISS, RISS, DBpia, and Google Scholar paper search were used as search sources. In previous studies related to academic stress, literature was searched using the keywords "adolescent academic stress," "adolescents," "adolescent stress," "adolescent mental health," and "mental health."

Literature related to previous studies based on the Adolescent Mental Health Survey was searched using the keywords "adolescent mental health," "mental health survey," "adolescent report," and "adolescent stress." For previous studies related to academic stress prediction models, literature was searched using the keywords "machine learning," "prediction model comparison," "machine learning," "random forest," "gradient boosting," and "logistic regression analysis." A search was conducted

on adolescent-based topics, and previous studies on adolescent stress, academic-related research, and adolescent mental health were reviewed, referencing literature that addressed topics and issues related to adolescent stress.

Studies utilizing machine learning techniques remain relatively limited. Therefore, this study focuses on the development of a machine learning model for predicting academic stress, closely examining and referencing the arguments and content of previous studies.

Academic stress among adolescents in South Korea is known to be exceptionally high due to the highly competitive education system focused on college entrance exams. For example, 72.6% of South Korean adolescents identified academic issues as a source of stress, whereas in other countries, only 40–50% of adolescents cited academic problems as a cause of stress[4]. Using structural equation modeling(SEM), a study analyzed data from 2,444 respondents in the 5th-year panel dataset of the Korea Youth Panel Survey(KYPS) for the 4th-grade panel. The results indicated that academic stress and academic performance negatively impact psychological well-being[8]. Concerns and stress caused by declining academic performance can lead to severe frustration and feelings of inferiority, ultimately resulting in a decrease in academic efficiency. This, in turn, causes disappointment in oneself, leading to doubt about studying and a decline in motivation[10]. Additionally, as the intensity of academic stress increases, individuals become more psychologically unstable, leading to depression and even suicide. Given the severity of the problems caused by academic stress, research on this topic has been continuously conducted to date[11].

Interest in academic stress has been significantly increasing, establishing itself as a key research area for improving mental health during adolescence. In particular, understanding how academic performance, classroom concentration, and emotional symptoms such as anxiety and depression influence academic stress has emerged as a critical issue[6]. These factors are closely related to the stress adolescents experience at school. By clearly identifying the impact of each factor on academic stress, it is possible to propose better learning environments and mental health improvement strategies.

Therefore, academic performance significantly influences students' stress levels, and dissatisfaction with grades serves as a major cause of academic stress. Additionally, classroom concentration, which refers to students' ability to focus during lessons, directly affects the quality of learning and is also closely related to academic stress. Furthermore, emotional symptoms such as anxiety and depression experienced by adolescents have an even greater impact on academic stress. These factors often interact with one another, further exacerbating academic stress[12]. Therefore, studying how these various factors interact and contribute to academic stress among students is highly important. It has been stated that a significant portion of adolescent issues stem from stress, and a survey on stress experiences among middle and high school students revealed that academic performance issues were the most frequently reported cause of stress for both groups[9]. Academic-related stress has been classified into four categories: decline in academic performance, test anxiety, academic efficiency, and motivation for studying itself. Academic stress arises from the perception that academic activities are too difficult and do not bring happiness. It is a psychological state characterized by feelings of depression, discomfort, anxiety, worry, and mental distress, leading to an overall sense of unhappiness and unease[13].

The stress perception rate among South Korean adolescents is reported to be the highest compared to adolescents in other countries. According to statistics from the 20th(2024) Youth Health Behavior Survey, research findings indicate that stress perception rates increase from 9th grade (middle school) to 12th grade (high school). South Korean adolescents are known to experience the highest levels of academic-related stress, primarily due to a highly competitive, exam-focused education system that emphasizes tests, studying, and grades[14]. This aligns with the notion that stress in adolescent mental health is closely linked to academic-related factors.

3. Research Method

3.1 Study Participants

The Korea Youth Mental Health Survey is a national survey conducted by the Korea Institute for Youth Policy (a government-affiliated research institute under the Prime Minister's Office) to assess the mental health status of South Korean adolescents and utilize the findings for youth mental health policies. This 2021 dataset was collected through a web/mobile survey conducted nationwide with approximately 6,700 adolescents from elementary, middle, and high schools, as well as out-of-school institutions. The survey targeted adolescents who voluntarily agreed to participate, gathering data on their mental health status and policy needs.

Based on this dataset, this study aims to develop a

(Table 1. Question 7. Perceived Stress Scale Questionnaire)

division	There was none at all	There was almost none	There were times	It happened quite often	It happevery often
	0	1	2	3	4
1. How often have you felt upset because something unexpected happened?					
2. How often have you felt like you couldn't do important things well?					
3. How often have you felt anxious or stressed?					
4. How often have you felt confident in your ability to handle personal problems?					
5. How often have you felt that things went your way?					
6. How often have you felt like you couldn't handle everything you had to do?					
7. How often have you felt like you were handling everything well?					
8. How often have you been upset about something you couldn't do anything about?					
9. How often have you thought you were good at managing your time?					
10. How often have you felt like things were too difficult for you to overcome?					

predictive model for adolescent academic stress.

3.2 Measurement Tool

3.2.1 Data Preprocessing and Variable Setting

In this study, data preprocessing was performed on 313 variables collected from the 2021 Korea Youth Mental Health Survey dataset. The missing value analysis revealed that 61 out of the 313 variables contained missing data. Fifty-two variables with a missing rate exceeding 70% were removed, while the missing values in the remaining variables were imputed using the mean value. In machine learning, variables are often referred to as "features." Therefore, in this study, "variables" and "features" will be defined as having the same meaning[15].

In this study, SPSS (Statistical Package for the Social Sciences) sav format data was used. To enhance analysis efficiency, the data was converted into CSV format using the Pandas library. Additionally, variable names were modified into an interpretable format based on the information provided in the codebook.

3.2.2 Dependent Variable

The Korea Youth Mental Health dataset does not include a variable that directly measures academic stress. However, it was determined that academic stress could be estimated based on items related to "Question 7: Perceived Stress Scale." Based on this, 10 related variables from the scale were analyzed, and academic stress was selected as the dependent variable. The academic stress variable was created by calculating the average values of the related variables, and the detailed information is presented in Table 2.

Some items in the stress scale can have both positive and negative meanings, making it inappropriate to assign the same weight to each item. In the survey questions, some items interpreted "never" as a positive response, while others interpreted "very often" as a positive response.

As shown in Table 1, the items with a positive meaning include. Item 4: "How often have you felt confident in your ability to handle personal problems?" Item 5: "How often have you felt that things were going your way?" Item 7: "How often have you felt that you were coping well with everything?"

Item 9: "How often have you thought you were

managing your time well?"

For these positively interpreted items, reverse coding was applied. Through this process, a high score (4) was converted to a low score (0), and a low score (0) was converted to a high score (4). After applying reverse coding, a new variable called "Academic Stress" was created. The stress score was determined by calculating the average value of the 10 items included in Question 7, and this score was set as the dependent variable.

3.2.3 Independent Variable

When performing machine learning modeling, there are advantages and disadvantages to using all variables versus selecting only certain variables. According to research, when the number of explanatory variables is large, including all variables in the model may lead to unexpected results that differ from the researcher's intentions[16]. When using all variables, the model can learn potential relationships between variables. However, if irrelevant variables or noise are included, the model may become overfitted to the training data, leading to poor performance on test data. On the other hand, using only a subset of variables has the advantage of simplifying the model and increasing training speed by selecting only important features. However, this approach also has a drawback, as it may omit some crucial information.

In this study, machine learning models such as Correlation, Linear Regression, and Random Forest were used to select meaningful variables. Important variables were extracted through each model, and among the duplicated variables, only those deemed significant were selected. As a result, a total of 225 independent variables were finalized, as shown in Table 2.

(Table 2 Independent variable settings)

Variable	value	datail
Question 1: related to ADHD CASS (27 questions)	0	Not at alL (Absolutely not)
	1	A little bit like that
	2	Quite so (It happens often)
	3	Very true (Very common)
Question 2: related to self-harm (5 questions)	1	Yes
	2	No
Question 3: related to child body sculpting (36 questions)	0	No experience
	1	A little difficult
	2	It's very difficult
	3	Very very difficult
Question 4: related to depression (12 questions)	0	Never
	1	A little bit like that
	2	Yes
	3	Quite so
	4	Very much so
Question 6: Suicidal Tendency	0	Never
	1	A little bit like that
	2	Yes
	3	Quite so
	4	Very much so
Question 11: Stress with people around you/ Whether or not you talk openly about your psychological discomfort	1	Not at all
	2	Not really
	3	It's like that
	4	Very much so
Question 12: The extent to which you discuss mental health issues with people around you (8 questions)	1	Not at all
	2	Not really
	3	It's like that
	4	Very much so
	5	No siblings

correlation may lead to the loss of crucial information.

(Table 2 Independent variable settings)

Variable	value	datail
Question 18: related to school life (24 questions)	1	Not at all
	2	Not really
	3	It's like that
	4	Very much so
Question 19: Academic performance	1	Very bad level
	2	Not good enough
	3	middle level
	4	Good level
Question 20: Satisfaction with academic performance	1	Not satisfied at all
	2	I'm not very satisfied
	3	I'm satisfied
	4	Very satisfied
Question 20: Regarding self-awareness (10 questions)	1	Not at all
	2	Not really
	3	It's like that
	4	Very much so

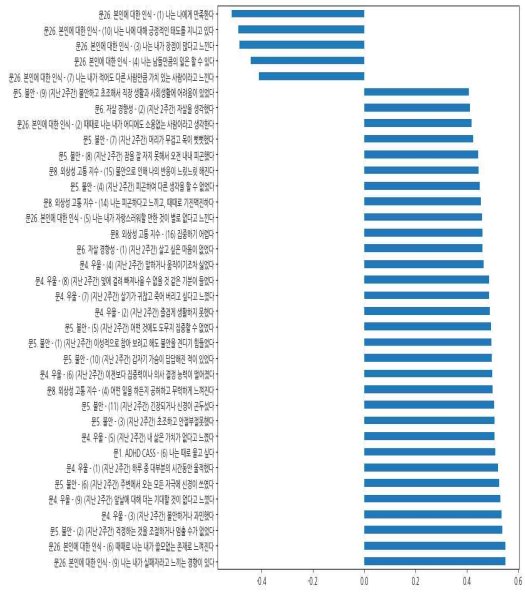
3.2.3.1 Correlation Variable Importance

Correlation is a concept that represents the degree of association between two variables and is commonly used to describe their relationship. In this study, the Pearson Correlation Coefficient, which is the most widely used method for measuring relationships, was utilized to assess the connection between academic stress and other variables.

The Pearson Correlation Coefficient measures the strength of the linear relationship between two numerical attributes. The coefficient ranges from -1 to +1. The Pearson correlation value (or coefficient) between two attributes is denoted by the letter r. If r = 0, it indicates that there is no correlation between the two attributes. If r = +1, it signifies a perfect positive correlation between the two attributes[17].

One advantage of correlation analysis is that it helps identify relationships between variables, allowing for the reduction of redundant information (multicollinearity). However, a drawback is that even if the correlation is low, a variable may still be important due to nonlinear relationships. Thus, excluding variables based solely on

The items related to "Question 7: Perceived Stress Scale," which were used to derive the academic stress variable, were already included in its calculation. Therefore, the correlation analysis was conducted on the remaining variables. The analysis results identified 36 variables with correlation coefficients below -0.4 or above 0.4, as shown in Figure 2.

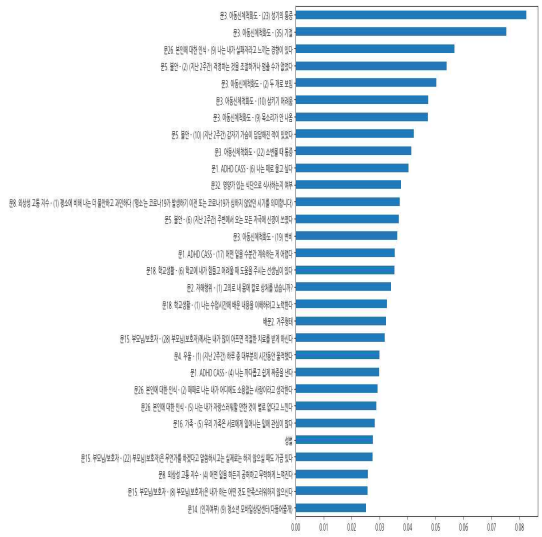


(Figure 2. Importance Criteria of 30 Variables in Correlation Analysis)

3.2.3.2 Linear Regression Variable Importance

The Linear Regression model has the advantage of intuitively identifying the contribution of each variable through regression coefficients. However, since Linear Regression cannot handle nonlinear relationships, it may have limitations when nonlinearity is an important factor.

In this study, to explain the independent variables affecting academic stress, a total of 251 variables (excluding 10 items related to "Question 7: Perceived Stress Scale" from the 261 preprocessed variables) were trained along with the academic stress variable. Subsequently, the importance of independent variables was evaluated based on their regression coefficients, and variables with significant influence were selected. The top 30 variables with the highest regression coefficients are presented in Figure 3.



(Figure 3. Importance Criteria of 30 Variables in Linear Regression)

(Figure 4. Importance Criteria of 30 Variables in Random Forest)

3.2.3.3 Random Forest Variable Importance

Random Forest was utilized as a machine learning technique capable of detecting nonlinear relationships and assessing interaction effects between variables. This model has the advantage of effectively calculating variable importance and was deemed useful for capturing relationships not explained by Correlation and Linear Regression models.

Similar to the Linear Regression model, 251 variables (excluding 10 items related to "Question 7: Perceived Stress Scale" from the 261 preprocessed variables) were trained along with the academic stress variable to explain the independent variables affecting academic stress.

Subsequently, based on the variable importance scores generated by the Random Forest model, the top 30 variables with significant influence were selected, and the results are presented in Figure 4.

3.3 Analysis Methods and Procedures

All analyses in this study were conducted using Python in the Google Colab environment. The Pandas library was used for data processing and analysis, while the scikit-learn module was utilized for machine learning modeling to develop a predictive model for adolescent academic stress.

Machine learning, also known as automated learning, refers to the process of extracting knowledge from data [18]. It is particularly useful for problems that humans can solve but are too large in scale, as well as for problems that are mathematically definable but extremely complex, making them difficult to address[19].

Machine learning can be categorized based on the dependent variable of the data into supervised learning, unsupervised learning, and reinforcement learning. Supervised learning is a learning method that utilizes given data and corresponding labels to make predictions and explore important variables in the process. In other words, the algorithm is provided with both inputs and expected outputs, and it learns to find a way to generate the desired output from the given input. Unsupervised learning, on the other hand, is a method of discovering useful patterns from data, where the algorithm is provided with input data but no corresponding output labels [18].

Reinforcement learning is a learning method in which a defined agent interacts with a specific environment, takes actions based on its current state, and analyzes the results to learn the optimal decision-making strategy that maximizes

future rewards. Reinforcement learning involves an iterative process where a machine repeatedly makes choices and receives feedback, aiming to maximize the long-term rewards [20].

Predictions using supervised learning can be categorized into classification and regression, depending on the use of input and output. In this study, regression was chosen as the predictive approach to estimate the continuous values that influence academic stress. The classification of machine learning methods based on learning approaches is presented in Table 3[19].

(Table. 3 Classification of Machine Learning Based on Learning Methods)
Source: Lee Jung-won, 2022, Machine Learning-Based Big Data Analysis for Frost Occurrence Prediction Modeling

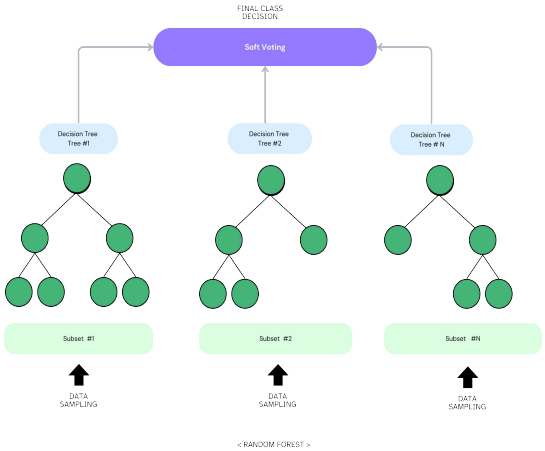
Category	Method	Prediction Approach	
Supervised Learning	Using Input and Output	Classification	Output is Yes/No
		Regression	Predicts continuous data such as weight, height, or numerical values (Value)
Unsupervised Learning	Using Input Only	Clustering	Uses clustering models such as k-means to predict output values based on input

In this study, the supervised learning method of machine learning was applied, and a regression algorithm was chosen to identify the key variables influencing stress. To train the model, machine learning analysis methods were selected based on previous studies and regression methodologies that demonstrated high performance. Compared to traditional statistical analysis methods, machine learning modeling exhibits higher predictive accuracy and enables the simultaneous analysis of hundreds of variables to explore factors influencing the dependent variable. Research in this area has been actively conducted[21]. This study aimed to apply various machine learning techniques to explore the factors influencing academic stress in adolescents. Among the selected five machine learning models, Linear Regression was used as the baseline for regression models.

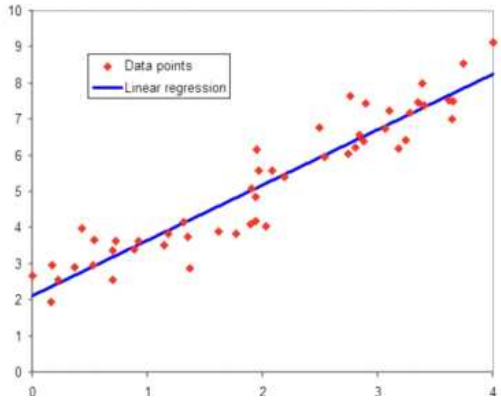
To address cases where the data does not exhibit a linear relationship, Lasso and Ridge models were applied. Additionally, Random Forest and Gradient Boosting were chosen for comparison, as they differ significantly from linear models in terms of characteristics and structure. In this study, machine learning modeling was conducted using 6,689 data points and 225 features, excluding missing values from the analysis. The dataset was split into training data and test data, with 20% allocated for testing. To ensure reproducibility, the random_state was set to 42.

3.3.1 Random Forest

Random Forest is a model that constructs multiple decision trees and connects them, forming an ensemble model using the bagging method[22]. Since it combines multiple tree models, it generally exhibits improved predictive power and high stability. Since multiple tree models are combined, Random Forest generally exhibits enhanced predictive power and high stability. To address the limitations of individual decision trees, Random Forest employs the bagging method, which stands for Bootstrap Aggregation. Bagging involves bootstrap sampling from the entire dataset, splitting the data, training multiple models, and aggregating their predictions [15]. In this study, Random Forest was chosen because it effectively leverages the rich information contained in the dataset. Additionally, it allows for a comprehensive assessment of the influence of all predictor variables, making it suitable for identifying relative importance, discovering new predictive variables, and considering interactions between features. This capability makes Random Forest an effective tool for understanding the significance of different features.



(Figure 5. Random Forest Model)
Source: Kwon Chul-min, Python Machine Learning Perfect Guide, Wikibooks.



(Figure 6. Linear Regression Model)
Source: Lee Jun-ho, Shin Jae-woo (2024), Comparative Study of Machine Learning Models for Fine Dust Prediction, Journal of the Korea IT Policy & Management Society, 16(4), 3687-3693.

3.3.2 Linear Regression

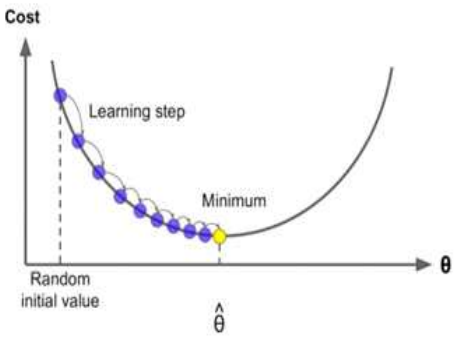
Among regression models, Linear Regression is the most fundamental and commonly used predictive model, representing a supervised learning algorithm. It is an analytical method that models the relationship between independent and dependent variables using a linear function that best represents their relationship[23].

Linear Regression is the most basic regression technique for modeling the linear relationship between a dependent variable and one or more independent variables. It assumes that the data is linearly distributed and expresses the influence of predictive variables on the target variable using a linear equation. However, if the data does not satisfy linearity, multicollinearity exists, or outliers are present, the model's performance may degrade. Therefore, preprocessing and the use of advanced techniques are crucial.

3.3.3 Gradient Boosting

Gradient Boosting is one of the ensemble techniques used in machine learning, where weak predictive models are sequentially combined and iteratively trained to ultimately build a strong predictive model. At each stage, learning progresses in a direction that minimizes errors to reduce the prediction errors of the previous model.

Unlike other boosting techniques, Gradient Boosting gradually reduces the errors of previous models using gradient descent as an optimization method[23].



(Figure 7. Gradient Boosting Model)
Source: Lee Jun-ho and Shin Jae-woo. (2024). Comparative Study of Machine Learning Models for Fine Dust Prediction. Journal of the Korea IT Policy & Management Society.

3.3.4 Ridge

Ridge regression uses the same prediction function as a linear regression model, but it is designed to impose an additional constraint when selecting weights (w). This constraint encourages the absolute values of the weights to be as small as possible, driving all weights closer to zero. As a result, Ridge regression reduces the magnitude of regression coefficients, lowering the model complexity and preventing overfitting.

The regularization method applied in Ridge regression is known as L2 regularization, which forces the model to be constrained, helping to prevent excessive fitting to the training data. L2 regularization introduces a penalty term based on the sum of the squared weights (w^2), effectively reducing the gradient and simplifying the model. This approach helps maintain predictive performance on training data while improving the generalization ability of the model to avoid overfitting[18].

(Table 4. Advantages and Disadvantages by Model)

Model	Advantages	Disadvantages
Random Forest	- Effectively models nonlinear relationships - Robust to outliers and noise - Provides variable importance - Low risk of overfitting	- Difficult to interpret - High computational cost - May require a large amount of data
Linear Regression	- Easy to interpret by understanding the relationship between variables through coefficients - Computationally efficient - Suitable for small datasets	- Susceptible to multicollinearity issues - Poor at modeling nonlinear relationships - Model performance is significantly affected if assumptions are not met
Gradient Boosting	- High predictive performance - Effectively models nonlinear relationships - Can significantly improve performance through boosting - Supports feature importance evaluation	- High computational cost and requires more resources - Prone to overfitting - Slower training speed
Ridge	- Addresses multicollinearity issues - Reduces overfitting - Considers feature importance	- All variable coefficients shrink toward zero, but none are completely eliminated, which may require additional feature selection
Lasso	- Enables feature selection - Enhances model interpretability - Reduces overfitting	- If there are many important features, performance may decrease due to excessive regularization

3.3.5 Lasso

To apply regularization to linear regression, Lasso is an alternative to Ridge. Both methods reduce model complexity by shrinking regression coefficients toward zero, but Lasso applies L1 regularization, producing different results from Ridge. Lasso regression can shrink some coefficients completely to zero, making the model more interpretable and revealing the most important features (Andreas Müller & Sarah Guido, 2017). In this study, Ridge and Lasso models were used to compensate for cases where the data did not exhibit a linear relationship in Linear Regression.

4. Research Results

4.1 Descriptive Statistics Analysis

An analysis was conducted on respondent characteristics, including gender, school level, multicultural family status, living arrangements, parental separation/divorce/death, economic status, father's education level, mother's education level, and family members living together. The results indicated that 5,937 students (88%) were enrolled in school, while 752 students (11%) were out-of-school youth. In terms of gender distribution, 3,104 students (46%) were male, and 3,585 students (53%) were female. Regarding school level, 2,039 students (30%) were in elementary school, 1,948 students (29%) were in middle school, and 1,950 students (29%) were in high school. Additionally, 752 students (0.3%) did not respond. For multicultural family status, 263 students (0.3%) belonged to multicultural families, while 6,426 students (96%) were from non-multicultural families.

Regarding living arrangements, the majority of students, 6,423 (96%), lived with their families. Meanwhile, 194 students (0.02%) resided in dormitories, boarding houses, or lived independently, and 72 students (0.01%) had other living arrangements.

In terms of economic status, 29 students (0.004%) reported being in the lowest economic category (very poor), while 168 students (0.025%) fell into the second-lowest category. Additionally, 516 students (0.07%) were in the third category, and 2,989 students (0.44%) considered themselves to be at an average economic level. The fifth category included 1,626 students (0.24%), while 899 students (0.13%) belonged to the sixth category. Finally, 462 students (0.69%) reported being in the highest economic category (very wealthy).

Regarding father's education level, 74 students (0.1%) reported that their father was not present, while 2 students indicated that their father did not attend school. Additionally, 28 students stated that their father completed elementary school, 73 students (0.1%) completed middle school, 1,271 students (19%) graduated from high school, and 3,173 students (47%) obtained a university degree. Furthermore, 634 students (0.9%) reported that their father completed graduate school, while 1,434 students (21%) were uncertain about their father's education level.

For mother's education level, 50 students reported that their mother was not present, and 11 students stated that their mother did not attend school. Additionally, 55 students indicated that their mother completed middle school, 1,443 students (21%) graduated from high school, and 456 students (0.6%) obtained a graduate degree. Furthermore, 1,353 students (20%) were uncertain about their mother's education level.

Regarding family members living together, 430 students (32%) lived with their grandfather or maternal grandfather, while 772 students (16%) lived with their grandmother or maternal grandmother. Additionally, 5,717 students (87%) resided with their father or stepfather, and 6,004 students (42%) lived with their mother or stepmother. Moreover, 5,195 students (0.5%) lived with their siblings, 246 students (32%) lived with relatives, and 77 students (41%) reported living with other family members. Lastly, 77 students (41%)stated that they lived alone. Refer to Table 5.

(Table 5. Descriptive Statistics of Demographic Factors)

Category		Number of Respondents (People)	Percentage (%)
Respondent Type	Students	5,937	88.7577
	Out-of-School Youth	752	11.2423
Gemder	Male	3,104	46.4045
	Female	3,585	53.5955
School Level	Elementary School	2,039	30.4829
	Middle School	1,948	29.1224
	High School	1,950	29.1523
	No Response	752	11.2423
Multicultural Family Status	Yes	263	3.9318
	No	6,426	96.0682
Living Arrangement	Living with Family	6,423	96.0233
	Dormitory/Boarding/Independent Living	194	2.9003
	Others	72	1.0764
Whether parents are separated, divorced or widowed	separation	188	2.8106
	Divorce	565	8.4467
	bereavement	110	1.6445
	Not applicable	5,826	87.0982
Household Economic Status	1(Very Poor)	29	0.4335
	2	168	2.5116
	3	516	7.7142
	4(Average Level)	2,989	44.6853
	5	1,626	24.3086
	6	899	13.44
	7(Average Level)	462	6.9069

(Table 5. Descriptive Statistics of Demographic Factors)

Category		Number of Respondents (People)	Percentage (%)
Father's Education Level	Father Absent	74	1.1063
	Did Not Attend School	2	0.0299
	Elementary School Graduate	28	0.4186
	Middle School Graduate	73	1.0913
	High School Graduate	1,271	19.0013
	University Graduate	3,173	47.4361
	Graduate School Graduate	634	9.4782
	Unknown	1,434	21.4382
Mother's Education Level	Mother Absent	50	0.7475
	Did Not Attend School	11	0.1644
	Elementary School Graduate	22	0.3289
	Middle School Graduate	55	0.8222
	High School Graduate	1,443	21.5727
	University Graduate	3,299	49.3198
	Graduate School Graduate	456	6.8172
	Unknown	1,353	20.2272
Family Members Living Together	Grandfather / Grandmother	430	2.322065
	Grandmother / Maternal Grandmother	772	4.168917
	Father / Stepfather	5,717	30.872664
	Mother / Stepfather	6,004	32.422508
	Sibling	5,195	28.053786
	Relatives	246	1.328437
	Others	77	0.415812
	None	77	0.415812

4.2 Final Model Selection and Performance Comparison

4.2.1 Performance Evaluation of Regression Models(MAE, RMSE, R^2 Score)

There are various evaluation metrics used to assess regression performance, and in this study, MAE (Mean Absolute Error), RMSE (Root Mean Squared Error), and R^2 Score were utilized to measure how well the model explains and predicts actual data. These evaluation metrics primarily focus on the difference between actual values and predicted values in regression models. Simply summing the differences can cause positive and negative errors to cancel each other out, leading to misleading results. Therefore, metrics such as the mean of absolute errors, squared errors, or the square root of the mean squared errors are commonly used to provide a more reliable assessment[24].

MAE (Mean Absolute Error) is a metric used to evaluate the performance of a model in regression analysis. It calculates the absolute value of errors, allowing for an intuitive interpretation of how much, on average, the predicted values deviate from the actual values. A lower MAE value indicates better model performance.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | Yi - \hat{Yi} |$$

$Yi : Actual\ Value,\ \hat{Yi} : \mathrm{P}redicted\ Value,$
$n : Nmber of\ Data$

MSE (Mean Squared Error) is a fundamental metric for evaluating the performance of a regression model. It calculates the average of the squared differences between the actual and predicted values. MSE is particularly important for models that focus on minimizing prediction errors

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (Yi - \hat{Yi})^2$$

$Yi : Actual\ Value,\ \hat{Yi} : \mathrm{P}redicted\ Value,$
$n : Nmber of\ Data$

RMSE (Root Mean Squared Error) is a metric derived from MSE (Mean Squared Error). Since MSE squares the errors, it tends to produce values larger than the actual average error. RMSE is obtained by taking the square root of MSE, making it a more interpretable metric. A smaller value indicates that the model's predicted values are closer to the actual values, making it a useful metric for evaluating the accuracy and effectiveness of the model's predictions.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Yi - \hat{Y}i)^2}$$

$Yi : Actual\ Value,\ \hat{Y}i : Predicted\ Value,$
$n : Nmber of Data$

R^2 Score evaluates predictive performance based on variance, using the ratio of the variance of predicted values to the variance of actual values as a metric. A value closer to 1 indicates higher prediction accuracy

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Yi - \hat{Y}i)^2}{\sum_{i=1}^{n}(Yi - \overline{Y})^2}$$

$Yi : Actual\ Value,\ \hat{Y}i : Predicted\ Value,$
$\overline{Y}$: $Mean\ Value$

These metrics differ in the approach and perspective used to evaluate model performance, and the choice of metric may vary depending on the research objective and the characteristics of the data. First, R^2 Score is a metric that evaluates how well the model explains the data, making it useful for assessing the model's explanatory power.

RMSE (Root Mean Squared Error) measures the magnitude of prediction errors by squaring the differences between the actual and predicted values, averaging them, and then taking the square root. This is particularly suitable when prediction performance is crucial, especially in cases where emphasizing larger errors is important. On the other hand, MAE (Mean Absolute Error) is calculated by averaging the absolute values of prediction errors. Since all errors are treated equally, it provides an intuitive assessment of overall prediction accuracy. In this study, survey data was used, and since survey responses are based on individuals' subjective evaluations, there may be significant variability between respondents and a lower consistency in responses. Due to these characteristics, data volatility may increase, and R^2 Score may not fully capture this variability. On the other hand, RMSE and MAE are metrics that directly measure the difference between predicted and actual values, making them more suitable for evaluating prediction accuracy in survey data, which is based on subjective responses. In particular, RMSE is sensitive to large prediction errors, making it advantageous when emphasizing the impact of significant errors that may occur in survey responses on model performance. MAE, on the other hand, is not overly sensitive to large errors while still effectively evaluating overall prediction accuracy.

However, given the nature of survey response data, where large errors can have a significant impact on model evaluation, this study selected RMSE as the primary performance evaluation metric.

To develop a machine learning model for predicting adolescent academic stress, five different models were trained to select the optimal model. The five machine learning algorithms compared in this study were Linear Regression, Random Forest, Ridge, Lasso, and Gradient Boosting. For modeling, the scikit-learn library was utilized. A total of 6,689 data points were trained using the 225 selected feature variables from Table 5, along with the stress score generated based on the Perceived Stress Scale items, which was set as the target variable. The performance evaluation results of the five regression models applied for predicting academic stress are presented in Table 6. The evaluation metrics used for analysis included MAE (Mean Absolute Error), RMSE (Root Mean Squared Error), and R^2 Score

4.2.1.1 Performance Evaluation of Linear Regression

The Linear Regression model achieved a Mean Absolute Error (MAE) of 0.31611, a Root Mean Squared Error (RMSE) of 0.39428, and an R^2 Score of 0.59685.

While the MAE and RMSE values indicate reasonable performance, the R^2 Score of 0.59685 suggests that the model can explain approximately 59.7% of the variance in the data.

4.2.1.2 Performance Evaluation of Ridge

The Ridge Regression model achieved a Mean Absolute Error (MAE) of 0.31608, a Root Mean Squared Error (RMSE) of 0.39425, and an R^2 Score of 0.59692.

While the results are similar to those of Linear Regression, the MAE and RMSE values are slightly lower, and the R^2 Score is marginally higher at 0.59692.

By applying regularization, the Ridge model helps prevent overfitting, ensuring that the model does not become

excessively complex. This allows it to fit well to the training data while maintaining generalization performance on test data, thereby reducing the risk of performance degradation due to overfitting.

4.2.1.3 Performance Evaluation of Lasso

The Lasso Regression model recorded a Mean Absolute Error (MAE) of 0.46708, a Root Mean Squared Error (RMSE) of 0.62161, and an R^2 Score of -0.00204.

Among all models, Lasso showed the lowest performance, with significantly higher MAE and RMSE values compared to the other models. These results indicate that Lasso Regression is not well-suited for predicting academic stress, as its predictive accuracy is notably lower than that of the other models. Notably, the R^2 Score is negative, indicating that the model fails to explain the data at all.

Lasso regression has a characteristic tendency to simplify the model by reducing the coefficients of less important variables to zero. However, since academic stress data involves complex interactions among multiple variables, the Lasso model may have excessively simplified the data, leading to the loss of critical information.

4.2.1.4 Performance Evaluation of Random Forest

Random Forest is an ensemble method capable of effectively handling nonlinear data and interactions between variables, demonstrating high performance across all metrics.

The model recorded the lowest values for both MAE (0.31075) and RMSE (0.39021), indicating superior predictive accuracy. Additionally, the R^2 Score was 0.60513, reflecting a high level of explanatory power for the data

4.2.1.5 Performance Evaluation of Gradient Boosting

Gradient Boosting demonstrated high performance, similar to Random Forest, and notably achieved the highest R^2 Score of 0.60623 among all models. This indicates that the model effectively learned the nonlinear relationships between academic stress and independent variables.

MAE (Mean Absolute Error), which represents the average difference between predicted and actual values, recorded the highest value among all models, highlighting the model's sensitivity to prediction errors. RMSE (Root Mean Squared Error) represents the mean squared error and serves as a metric that indicates how sensitive the model is to large errors. A lower RMSE value signifies a higher-performing model. Additionally, both MAE and RMSE values were at a similar level to those of Random Forest, suggesting that Gradient Boosting is one of the most suitable models for predicting academic stress.

(Table 6: Comparison of Regression Model Performance)

Model	MAE	RMSE	R^2 Score
Linear Regression	0.31611	0.39428	0.59685
Ridge	0.31608	0.39425	0.59692
Lasso	0.46708	0.62161	-0.00204
Random Forest	0.31075	0.39021	0.60513
Gradient Boosting	0.31179	0.38967	0.60623

Based on MAE, the Random Forest algorithm demonstrated the best performance with a value of 0.31075. For RMSE, the Gradient Boosting algorithm achieved the highest performance with a value of 0.38967. Regarding R^2 Score, the Gradient Boosting algorithm also showed the best performance, recording 0.60623.

To optimize the performance of each predictive model, hyperparameter tuning was performed using the GridSearchCV API. GridSearchCV is a technique designed to automatically search for the optimal hyperparameters to enhance the performance of machine learning models. Since model training performance can vary depending on hyperparameters (e.g., learning rate, number of trees, etc.), it is essential to fine-tune them appropriately. GridSearchCV systematically tests the user-specified hyperparameter combinations, evaluates their performance through cross-validation, and identifies the best-performing configuration. In this study, GridSearchCV was applied to each algorithm to search for the optimal hyperparameters. The models were then retrained based on these optimal parameters, and their performance was compared. The optimization results and the performance comparison of each model are presented in Table 7.

Based on MAE, the Random Forest algorithm demonstrated the best performance with a value of 0.31028.

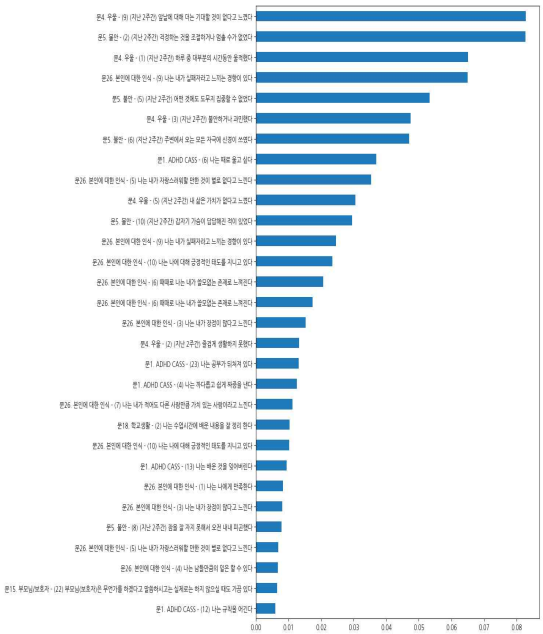
For RMSE, the Gradient Boosting algorithm achieved the highest performance with a value of 0.38770. Regarding R^2 Score, the Gradient Boosting algorithm also showed the best performance, recording 0.61019. Furthermore, it was confirmed that the R^2 Score of the Gradient Boosting model improved to 0.61019 due to hyperparameter tuning, enhancing its predictive performance.

(Table 7. Performance Comparison of Regression Models with Applied Hyperparameters)

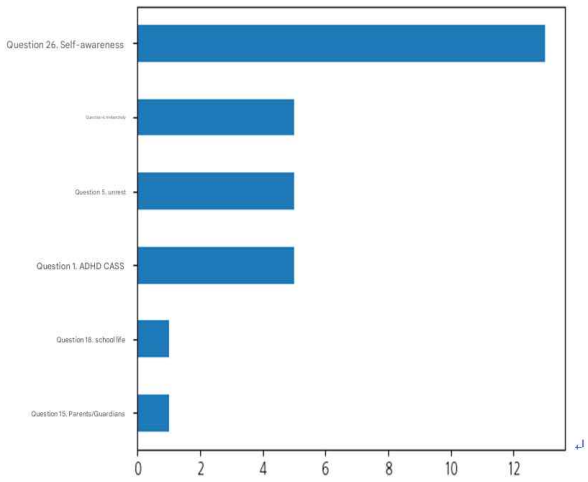
Model	Hyper parameter	MAE	RMSE	R^2 Score
Linear Regression		0.31611	0.39428	0.59685
Ridge	alpha=100	0.31447	0.39232	0.60085
Lasso	alpha=0.1	0.34060	0.44080	0.49612
Random Forest	n_estimators=200, random_state=42, max_depth=20, min_samples_split=2	0.31028	0.38928	0.60701
Gradient Boosting	n_estimators=200, random_state=42, learning_rate=0.1, max_depth=3	0.31058	0.38770	0.61019

4.2.2 Feature Importance Analysis of Gradient Boosting Predictive Model

In this study, the Gradient Boosting model, which showed the best performance based on the R^2 Score, was analyzed to identify the most influential features in predicting adolescent academic stress. Key variables included statements such as "I felt that there was nothing to look forward to in the future" and "I was unable to control or stop worrying." The results are presented in Figure 8. Out of the 225 feature variables, the top 36 variables were selected based on their importance scores in the Gradient Boosting model, and the findings were visualized in a graph.



(Figure 8. Top 36 Variables Based on Importance Scores in the Gradient Boosting Prediction Model)



(Figure 9. Frequency of Major Categories for Survey Items)

The frequency distribution of each survey item by major category is shown in Figure 9. The graph indicates that self-perception-related variables have the greatest impact on stress levels. Following this, the categories ranked in order of influence are depression, anxiety, ADHD CASS, school life, and parental/guardian factors.

5. Discussion and Conclusion

5.1 Summary and Discussion

The research was conducted on approximately 6,700 middle and high school students, including out-of-school youth, across South Korea in 2021. This study aimed to develop a predictive model for adolescent academic stress using Linear Regression, Random Forest, Ridge, Lasso, and Gradient Boosting models. To address the research question, machine learning models were developed and performance metrics were derived using Python-based machine learning libraries. Specifically, algorithms from the scikit-learn library were applied, including Logistic Regression from the linear model module, Gradient Boosting Classifier from the Gradient Boosting module, and Decision Classifier and Random Forest Classifier from the sklearn ensemble module.

The performance evaluation metrics for the models were based on MAE (Mean Absolute Error), RMSE (Root Mean Squared Error), and R^2 Score, which are commonly used in regression analysis. Among the models, Random Forest achieved the lowest MAE with a value of 0.31075, followed by Gradient Boosting (0.31179), Ridge (0.31608), Linear Regression (0.31611), and Lasso (0.46708). This ranking indicates that Random Forest and Gradient Boosting performed the best in terms of minimizing absolute errors, while Lasso exhibited the highest error, suggesting weaker predictive performance. The RMSE (Root Mean Squared Error) values for the models, in ascending order, were as follows: Gradient Boosting (0.38967), Random Forest (0.39021), Ridge (0.39425), Linear Regression (0.39428), and Lasso (0.62161). Similarly, the R^2 Score values, ranked in descending order, were Gradient Boosting (0.60623), Random Forest (0.60513), Ridge (0.59692), Linear Regression (0.59685), and Lasso (-0.00204). These results indicate that Gradient Boosting had the lowest RMSE, making it the best performer in terms of error minimization, while Lasso exhibited the highest RMSE and the lowest R^2 Score, suggesting poor predictive accuracy. Based on the RMSE evaluation metric, the Gradient Boosting model recorded the lowest RMSE value, demonstrating superior predictive performance. Additionally, even after applying the optimal hyperparameters, the Gradient Boosting model maintained consistent performance, showing more stable results compared to other models. Based on these findings, the Gradient Boosting model was ultimately selected as the most suitable model for predicting adolescent academic stress.

Analysis of Key Factors Influencing Adolescent Academic Stress Using the Gradient Boosting Model. The analysis of key factors influencing adolescent academic stress using the Gradient Boosting model revealed that the most significant variables were: Depression (Question 4-1): "Over the past two weeks, I felt down for most of the day." Anxiety (Question 5-2): "Over the past two weeks, I was unable to control or stop worrying." Depression (Question 4-9): "Over the past two weeks, I felt that there was nothing to look forward to in the future." These findings indicate that depression and anxiety are major contributors to adolescent academic stress.

Most previous studies on adolescent stress have employed statistical methods, primarily conducting descriptive statistical analyses on stress-related factors. In contrast, research using machine learning models to predict academic stress remains limited. Traditional statistical approaches focus on explaining data, emphasizing statistical significance testing and model-based inference to describe relationships within the dataset. However, machine learning techniques prioritize predicting dependent variables when new data is introduced rather than focusing solely on explanation[25].

In the study by [26], which applied machine learning techniques to explore influential variables, a greater variety of new factors were identified compared to studies using traditional research methods[27].

5.2 Conclusion and Recommendations

This study holds significance in that it utilizes various machine learning techniques to identify key predictive variables for adolescent academic stress and, based on these findings, proposes academic stress management strategies, providing valuable direction for future researchers. In particular, the core contribution of this study lies in identifying the major factors influencing adolescent academic stress using the most predictive model and suggesting ways

to leverage these findings from a preventive perspective. These findings open up possibilities for preventing risk factors associated with adolescent academic stress and further serving as foundational data for the development of comprehensive mental health management programs. In particular, this study makes a practical contribution by establishing a foundation for designing models that can be utilized not only for predicting and preventing academic stress but also for addressing other aspects of mental health. However, this study has several limitations that should be acknowledged.

First, this study relied on data collected at a single point in time, which presents a limitation in understanding how academic stress evolves over time and how various factors interact to influence stress levels. It remains unclear whether stress increased due to a specific event or whether stress itself led to other issues. To accurately determine these relationships, longitudinal studies that collect data over an extended period are necessary.

Second, the data used in this study is based on self-reported survey responses, which may introduce limitations in reliability and accuracy due to the subjective nature of the responses. To address this limitation, future research should incorporate diverse data types, such as biometric signals or behavioral data, enabling more objective and sophisticated analyses.

Third, the survey items used in this study may lack precision in design and evaluation, highlighting the need for new assessment methods that take respondents' psychological states into account. Additionally, since the machine learning models in this study were trained and analyzed on a limited dataset, future research should focus on expanding the dataset and applying various advanced algorithms, such as deep learning and ensemble techniques, to compare and improve predictive performance.

Despite these limitations, this study has introduced new possibilities in the field of adolescent academic stress research. It is expected that future studies will build upon these findings to conduct more in-depth and comprehensive analyses, further advancing the understanding and management of academic stress.

참고문헌(Reference)

[1] Cha Son. Min, "Korean Youth's Life Satisfaction 'Lowest' vs Academic Performance 'Top'", 2024. https://www.newsis.com/view/NISX20240105_0002582920

[2] Min ju. Cha, "Korean Students Stuck In 'Study Hell', 41% Are 'Stressed'", 2023. https://www.kmib.co.kr/article/view.asp?arcid=0924304673

[3] Hee jin. Han, "Factors affecting academic stress perceived by high school students. Personality education," Vol.16, No.3, pp.103-117, 2022.

[4] JH Kim, SB Kim, JI Jeong, "The Effects of Adolescents' Academic Stress on School Life Adaptation: Focusing on the Mediating Effects of Stress Coping," The Korean Youth Study, Vol.25, No.4, pp.241-269, 2014.

[5] SC Shin, SY Chol, "Academic stress perceived by adolescents to adapt to school life." The moderating effect of self-esteem on the effect", counseling psychological education welfare, Vol.6, No.3, pp.59-72, 2019.

[6] JI Jang, SB Kim, "The mediating effect of positive psychological capital and smartphone addiction in the relationship between adolescents' stress and school life adaptation." Youth Counseling Study, Vol.23, No.2, pp.447-465, 2015.

[7] DW Kim, HY Park, KI Lee, "The Effects of Parents' Academic Support and Emotional Support on Academic Stress and the Controlling Effect of Intrinsic Motivation," Human Development Research, Vol. 29 No.1, pp.21-41, 2022.

[8] CR Noh, SH Kim, "The Effects of Academic Stress and Academic Performance of Middle School Students on Psychological Well-being," Korean Department of Child Welfare, 39, 2021.

[9] KH Lee, "Study on the Stress of Adolescents", Ph.D. thesis, Sookmyung Women's University, 1995.

[10] GY Kim, "School life and counseling guidance", 560 Hakjisa, 2000.

[11] EH Choi, "The effect of academic stress in high school students on school life adaptation. Mediating effects of psychological state and moderating effects of parenting attitudes," 120, master's thesis, Daegu Oriental Medical University, 2016.

[12] HK Sang, YN Lee, "The relationship between academic achievement and academic stress: the moderating effect of academic self-concept," Educational Research, Vol.35, No.1, pp.53-71, 2012.

[13] MH Oh, SM Chun, "The Effects of Meditation Training to Analysis of Academic Stress Factors and Symptoms in Middle School Students and their Reduction", 78, master's thesis, Kyungpook National University, 1994.

[14] SM Mo, "Analyzing the moderating effects of adolescent academic stress-inducing variables: Focusing on self-esteem and decision-making autonomy", Journal of the Future Youth Society, Vol.7, No.2, pp.49-66, 2010.

[15] SM Kwon, "Study on the Predictive Model of Credit Ratings for Logistics Companies Using Machine Learning", 173, PhD thesis, Chung-Ang University, 2023.
[16] HS Seo, "Statistical Machine Learning for the Selection of Linear Models: A Comparative Study of Variable Selection Methods", Journal of The Korean Data Analysis Society, Vol.23, No.6, pp.2643-2653, 2021.
[17] John Kelleher, Brendan Tierney, OS Kwon, "Data Science as a Tool for Insight for Better Decision Making," 272, Kim Young-sa, 2019.
[18] A Muller, S Guido, "Introduction to Machine Learning with Python," 504, Hanbit Media, 2017.
[19] JW Lee, "Machine Learning-Based Big Data Analysis for Frost Occurrence Prediction Modeling," 113, Hanbat National University, 2022.
[20] SY Kim, YJ Jeong, TH Kim, "First Steps in Machine Learning: From Basics to Modeling, Practical Examples, and Problem Solving," 376, Hanbit Media, 2017.
[21] YH Son, HJ Park, MH Park, "Analysis of Group Determinants Based on Reading Literacy Levels Using Random Forest: Focusing on PISA 2018 Data," Asian Journal of Education Research, Vol.21, No.1, pp. 191-215, 2020.
[22] Breimanm L, "Random forests, Machine Learning", Vol.45, No.1, pp.5-32, 2020.
[23] JH Lee, JW Shin, "Comparative Study of Machine Learning Models for Fine Dust Prediction," Journal of the Korea IT Policy & Management Society, Vol.16, No.4, pp.3687-3693, 2024.
[24] CM Kwon, "Python Machine Learning Perfect Guide," 704, Wikibooks, 2022.
[25] HW Kim, JE Yoo, "Exploring Depression Prediction Variables for Second-Year Middle School Students Using Machine Learning," KCYPS 2018 5th Wave Data Analysis, Vol.25, No.3, pp.31-53, 2024.
[26] MJ Noh, JE Yoo, "Exploring Career Decision-Related Variables Using Adaptive LASSO," Open Education Research, Vol.27, No.4, pp.133-155, 2019.
[27] SY Park, JY Jung, "Analysis of Predictive Factors for High School Teachers' Instructional Improvement Activities Using Random Forest," Educational Research, Vol.62, No.5, pp.1-32, 2020.

◐ Author Biography ◑

Jeongsuk-suk So
Graduated from Seoul School of Integrated Sciencs & Technologies(aSSIST),
Graduate School of AI, Department of AI Advanced Studies, Major in AI Big Data(Master of Engineering)
Research Interests : AI, Digital Transformation, Generative AI, Data Analysis etc.
E-mail : sojeongsuk@gmail.com

Tae-yeon Oh
Assistant Professor at Seoul School of Intergrated Sciences & Technologies(aSSIST)
Assistant Professor, University of Mississippi(2018-2022)
Ph.D. in Sports Management, Seoul National University
Research Interests : AI, Sports Analytics, Sports Marketing, Sports Media, Big Data Analytics etc.
E-mail : tyoh@assist.ac.kr

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.

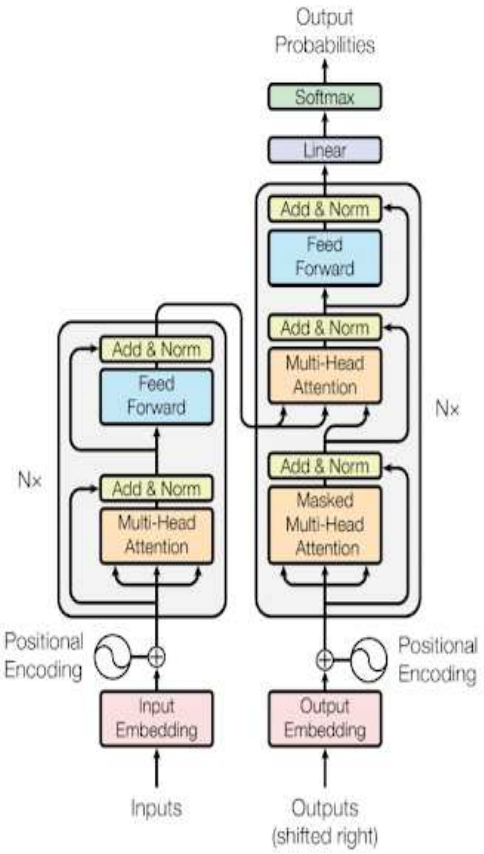
2. Theoretical Background

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 Viodeo AI Model, · High-Quality Video Generation up to 60 Seconds[7]
runway	Runway AI, Inc.	· Text to Viodeo AI Model, · Image to Video AI Model[13]
PixAI	Mewtant Inc.	· AI Art Generator, · Generation of High-Resolution Artwork, · Generation of Animation and Pixel Art Style Images[14]

2.1.2.4 Music Generation Model

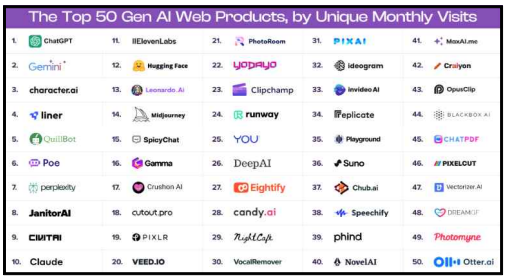
(Table 4) Development Status of Generative AI Music Generation Models

Types	Developers	Characteristics
Suno	Suno, Inc.	· Text-to-Music Generation, · Audio-to-Song Conversion[15]
Lyria	Google DeepMind	· Generation of High-Quality Vocals, Lyrics, and Instrumental Accompaniment, · 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 100,000 Pre-trained Models, · Datasets: Library for NLP Tasks, · Tokenizers: Text Preprocessing Tools, · Spaces: Machine Learning Demo and Application Sharing 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].

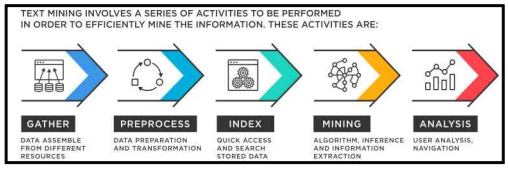


Figure 3) Text mining processing procedure
Source : https://www.tibco.com/(2024)

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.

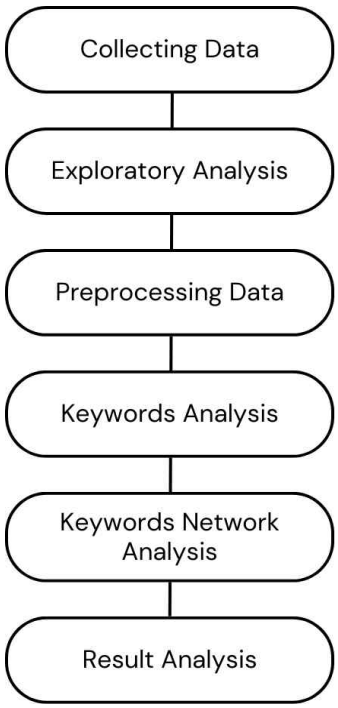
For this study, data was collected from the KCI database by identifying papers that included the keyword "generative" in their titles. The initial collection included paper IDs, titles, authors, and publication years for all relevant papers, resulting in data for 6,170 publications. The search results 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

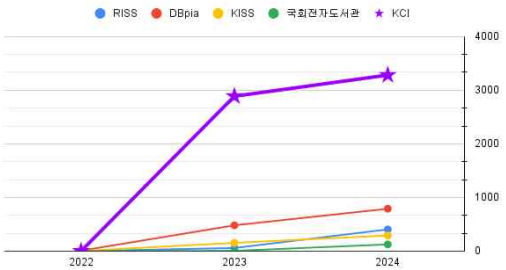


(Figure 4) Data collection and analysis procedures

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 ('이', '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 more cohesive and mature research framework.

(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

4.2 Keyword Network Analysis Results

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.

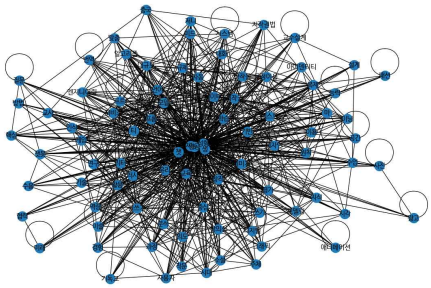
The overall network structure exhibits a relatively balanced, distributed form. Most nodes are directly connected to multiple other nodes, indicating active interaction between research topics. Notably, there are clear connections between central and peripheral nodes, demonstrating that new research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

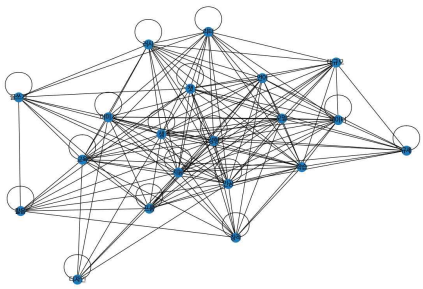
In generating co-occurrence matrices for thesis abstract keywords, we conducted sequential analyses with frequency thresholds of 3, 10, and 20 occurrences for 2023 theses, and 3, 10, 20, and 50 occurrences for 2024 theses, as illustrated in Figures 10-16. Examining the network characteristics of high-frequency 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.

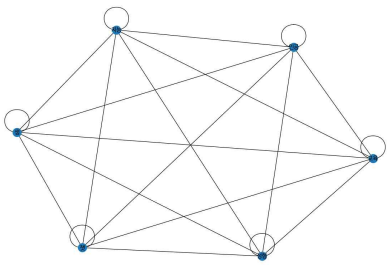
Notably, nodes positioned at the network's center demonstrate high connectivity with other keywords, suggesting their role as fundamental concepts within the research field. This central positioning and high degree of interconnectedness imply that these terms serve as pivotal elements linking various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



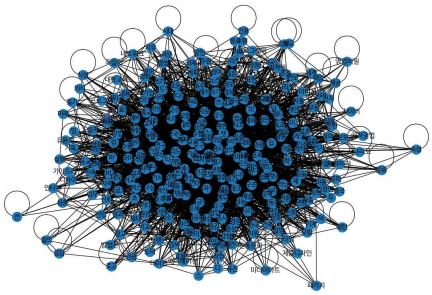
(Figure 10) Topic word frequency keyword network analysis of 2023 dissertation abstracts - 3 or more Co-occurrences



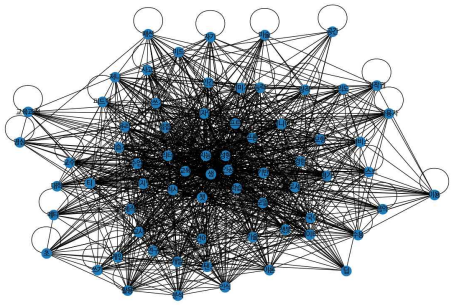
(Figure 11) Topic word frequency keyword network analysis of 2023 dissertation abstracts - 10 or more Co-occurrences



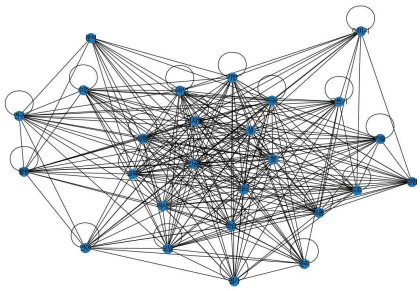
(Figure 12) Topic word frequency keyword network analysis of 2023 dissertation abstracts - 20 or more Co-occurrences



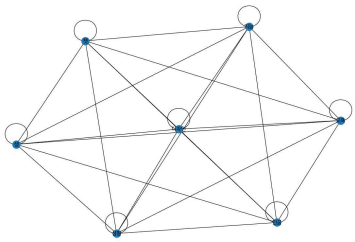
(Figure 13) Topic word frequency keyword network analysis of 2024 dissertation abstracts - 3 or more Co-occurrences



(Figure 14) Topic word frequency keyword network analysis of 2024 dissertation abstracts - 10 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 Techniquesin 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%AExl%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

Development of a Machine Learning-based Corporate Management Strategy Prediction Model Reflecting ESG Rating Information

Kyung Gu Kang

Abstract

In this study, the model reflected not only financial data but also corporate ESG rating information to predict corporate management strategies. Five Machine learning classifiers such as RF, SVM, XGBoost, LightGBM, and CatBoost were used to develop a model that predicts management strategies using machine learning classification techniques. The research results of this paper are as follows. First, the model that used financial data and ESG rating information together showed better prediction performance than the prediction model that used only financial data. In particular, the best performance was shown when the ESG rating was included, followed by the high performance in the management strategy prediction model in the order of governance, environmental, and social ratings. This means that ESG rating information is important in predicting corporate management strategies. Second, among the five machine learning classifiers, LightGBM performed the best when predicting corporate management strategies, followed by CatBoost, XGBoost, SVM, and RF. These results suggest that the Boost classifier is effective in predicting management strategies. The corporate management strategy prediction model developed in this study can be a useful tool for promoting the sustainable development of a company and is expected to provide important insights to managers and investors. In addition, these models are expected to contribute to improving a company's long-term performance by supporting the establishment of sustainable management strategies.

Keywords: management strategy, machine learning, management strategy prediction, ESG rating

1. Introduction

The purpose of this paper is to develop a model that effectively predicts corporate management strategies. Machine learning technology has the advantage of analyzing large amounts of data to discover patterns and perform predictions. Therefore, this study aims to develop a corporate management strategy prediction model with a high-performance by applying machine learning techniques.

The ESG (Environment, Society, and Governance) rating is emerging as an important indicator for evaluating sustainability and social responsibility of a corporate (Aydoğmuş et al., 2022). ESG factors have a significant impact on investor decision-making and long-term corporate performance, and the interest in them is increasing (Quintiliani, 2022).

The ESG rating has the following correlations with corporate management strategies. First, companies with high ESG ratings pursue sustainable management in terms of the environment, society, and governance (Yu and Xiao, 2022). Active ESG activities have a positive impact on long-term corporate performance and future corporate value. Second, management strategies that take ESG factors into account can reduce risks related to environmental regulations,

social responsibility, and governance issues (Behl et al., 2022). This can improve the stability and reputation of the company. Third, since many investors value ESG performance, a high ESG rating can create more investment opportunities (Junius et al., 2020). Fourth, a management strategy that considers ESG helps build trust with customers and improve brand image (Zhang et al., 2020). In other words, through such corelations, companies can integrate ESG factors into their management strategies to achieve long-term growth.

Traditional management strategy analysis mainly relied on financial indicators, but considering ESG factors will allow for more accurate management strategy analysis. Therefore, this paper develops a machine learning-based corporate management strategy prediction model including ESG ratings by utilizing various machine learning classification techniques.

The differences between this paper and previous studies are as follows: First, while previous studies focused on the individual impact of ESG factors, this paper comprehensively integrates them to identify that they are factors that significantly affect management strategies. Second, it precisely models the complex correlation between ESG rating data and management strategy by utilizing the latest machine learning algorithm. Third, presenting a model that predicts a company's management strategy by considering ESG rating is a topic that has not yet been studied domestically or internationally, which has great academic and practical contributions. Fourth, it provides useful information so that executives and investors can make better decisions based on ESG information.

The management strategy prediction model developed in this study can be a useful tool for companies to pursue sustainable development and is expected to provide important insights to managers and investors.

In addition, it is expected that this model will contribute to improving the long-term performance of companies by supporting the establishment of sustainable management strategies.

2. Theoretical Review

2.1 Preceding Researches

The results of examining preceding researches showed that many of the existing studies on corporate management strategies focused on the relationship with management performance. Bowman and Helfat (2001) argued that corporate performance changes according to corporate management strategies. Grant (2002) focused on the scope and content of corporate management strategies. This study analyzed the scope of management strategies, such as which products, markets, and activities a company should participate in, and the content of management strategies, such as how to secure competitive advantage in the market. This paper emphasized the importance of strategic decision-making by managers and explained that management strategies have a significant impact on corporate performance. Yoon Jongrok and Kim Hyeongcheol (2009) conducted a multidimensional analysis of factors affecting the management performance of venture companies. They conducted an empirical analysis using questionnaire data targeting domestic venture companies.

The results of the analysis showed that strategies played a complete or partial mediating role in factors of entrepreneurial capacity, technology, functional capacity and innovativeness among the characteristics of entrepreneurs. Gong Seokjin and Yang Haesul (2014) also studied the impact of management strategy on management performance, and found that offensive and defensive strategies, cost advantage, and differentiation strategies all had a positive effect on financial performance and non-financial performance. In addition, they confirmed that information technology utilization level had a moderating effect on the relationship between management strategy and management performance. Han Gyudong (2019) analyzed the mediating effect of management strategy in the relationship between entrepreneur characteristics and management performance in venture companies. The results of the study showed that the entrepreneur's enterprising spirit had the greatest positive effect on management performance, and that innovativeness and enterprising spirit had a significant effect on technological innovation and marketing differentiation strategies, respectively.

Management strategy had a significant positive effect on management performance, and showed a partial mediating effect in the relationship between entrepreneur characteristics and management performance. Chu Seungyeop et al. (2009) empirically analyzed the impact of the relationship between the management environment, competitive strategy, and internal capabilities of a company on management performance. Environmental uncertainty and differentiation strategy showed a positive relationship, and specific internal capabilities showed a positive effect depending on the type of strategy. In addition, high-performance groups had higher suitability of environment-strategy and

capacity-strategy than low-performance groups. Park Dain and Park Chanhee (2018) confirmed that the degree of external cooperation and the degree of utilization of venture management support systems by a company have different effects on competitiveness and performance depending on the life cycle of the company.

The effects on corporate competitiveness were different depending on the introduction stage, early growth stage, high growth stage, maturity stage, and decline stage, and the factors affecting management performance were also different. Koh Sehoon et al. (2013) analyzed the relationship between the management environment, competitive strategy, and management performance of manufacturing SMEs and venture companies. It was confirmed that the industrial environment did not significantly affect the competitive strategy, while the resource and capacity factors affected all strategies except the concentration strategy.

In addition, each competitive strategy contributed to technology, market, and financial performance, and in particular, technology performance was found to affect the market and financial performance.

Meanwhile, the major studies on management strategy, ESG, and CSR are as follows. González-Ramos et al. (2023) analyzed the relationship between knowledge management strategies and corporate social responsibility (CSR) and investigated their impact on innovation capabilities. This paper emphasizes that knowledge management and CSR are important factors in improving a company's innovation capabilities, and suggests that the integration of the two factors can promote innovation performance. Nam Jeongmin and Park Junhyeok (2022) derived the priorities of ESG factors for venture companies using the AHP analysis technique. The results showed that the environment, society, and governance were of importance in that order, and in particular, the CEO's will to practice was important in the environment. In the society sector, basic rights and corruption prevention in the workplace were identified as important factors; and in the governance sector, protection of shareholder rights.

Yoon Seongyong (2023) analyzed the impact of management strategies on companies' ESG activities. It was divided into environment (E), society (S), and governance (G) through the economic justice index to verify the impact of each factor on corporate value. According to the results, the leading management strategy had a positive moderating effect on the relationship between the improvement of governance (G) and corporate value, but had no significant effect on activities in environment (E) and society (S). This suggests that improved governance can control managers' risky decisions and create optimal value.

The study of Yuan et al. (2020) analyzed the relationship between corporate management strategy and corporate social responsibility (CSR). This paper explores how corporate management strategy affects CSR activities and concludes that CSR has a positive impact on corporate reputation and long-term performance.

Previous studies on changes in management strategy according to the digital management environment are as follows: The study of Menz et al. (2021) explored corporate management strategy in the digital era. This paper analyzed the impact of digitalization on corporate strategic decision-making and management methods, and studied the impact of technological innovation on corporate management methods. It also suggested strategic measures for companies to secure competitive advantage in the digital environment.

Avanesova et al. (2021) analyzed the role of strategic management in the system model of corporate organizational development. This paper explored the impact of strategic management on organizational structure and performance for the sustainable development of companies, and emphasized that effective management strategy formulation and implementation were essential for corporate growth.

Akmaeva et al. (2020) analyzed the impact of digitalization on the management strategy and formation of new management models of modern companies. This paper suggested how the development of digital technology could innovatively change the strategic approach and management model of companies and thereby enhance competitiveness.

Gunarathne et al. (2021) analyzed the relationship between institutional pressure, environmental management strategy, and organizational performance due to the advent of the digital age. It emphasized that environmental management strategy contributed to responding to institutional pressure and improving the environmental performance of organizations.

And the major studies on corporate management strategy and human resource management are as follows: Brewster (2017) studied the integration of human resource management and corporate management strategy. It analyzed how human resource management contributed to corporate management strategy through human resource management policy and practice, and suggested a way to improve corporate management performance through harmony

between human resource management and management strategy. This study emphasized the strategic importance of human resource management and argued that it could enhance corporate competitiveness.

Joo Sunghan and Jeong Namho (2023) investigated the impact of knowledge management strategies on employees' immersion experience and job retention intention from the perspective of hotel companies. As a result of the study, explicit knowledge management strategies had a significant impact on skills and goal clarity, but tacit knowledge did not. Immersion experience played a mediating role between goal clarity and job retention intention, and it emphasized the importance of balance through appropriate challenge.

2.2 Derivation of Research Issues

The followings are studies on management strategy and future corporate value. Hong Nanhee (2024) analyzed the impact of management strategy on corporate investment real estate holdings and corporate value, and reported that companies with leading strategies had a higher proportion of investment real estate holdings, which had a negative impact on corporate value.

This study provided empirical evidence that management strategy affected managers' decisions regarding investment real estate.

Bowman and Ambrosini (2007) explored how companies created corporate value through various management strategies. They classified corporate management strategies into several categories, and in particular, emphasized that resource allocation and competitive position played an important role in increasing corporate value.

Al-Shaer et al. (2023) investigated the relationship between corporate management value. They found that cost advantage strategy had a positive (+) relationship with strategy, board composition, and corporate

board size, independence, gender diversity, and tenure of executives, but a negative (-) relationship with managerial competence. On the other hand, differentiation strategy was found to have a positive relationship with board size and gender diversity of executives. These results suggest that companies need to adjust the composition of their boards of directors in line with their management strategies.

Al-Najjar and Anfimiadou (2012) studied the impact of a company's environmental strategy on corporate value. This study emphasized that environmentally friendly strategies could have a positive impact on corporate value and concluded that sustainable management contributed to the creation of long-term corporate value. This suggests that corporate environmental strategy management had a competitive advantage.

Research on corporate management strategies has been actively conducted in various fields, but studies predicting corporate management strategies have not yet been identified. Therefore, research on predicting corporate management strategies through machine learning is expected to contribute to expanding the scope of research in the field of management strategy academically. In practical terms, the results of this study are expected to contribute to maintaining a competitive edge in the market by making strategic decisions ahead of competitors in terms of securing a competitive edge in providing guidelines for companies to identify and respond to potential future risks in advance in terms of risk management. Additionally, the result model presented in this study can help strengthen trust with investors through predictable management strategies.

In particular, predicting management strategies could provide useful information for companies to establish strategies for sustainable growth. Previous studies have verified that ESG activities significantly influenced corporate management strategies (González-Ramos et al., 2023; Yuan et al., 2020).

In particular, it can provide useful information for companies to establish strategies for sustainable growth through management strategy prediction. Previous studies have verified that ESG activities have a significant impact on companies' management strategies (González-Ramos et al., 2023; Yuan et al., 2020).

In this paper, in predicting a company's management strategy, not only financial data related to the management strategy but also ESG rating information of the company are reflected in the model. In other words, this study aims to present the best-performing management strategy prediction model using various recent machine learning classifiers.

3. Research Methods

3.1. Research Design and Data

The financial data of the companies used in

this paper were downloaded from Vaule Search of NICE Credit Rating. The sample period is from 2012 to 2023, and financial industries that apply different accounting standards were excluded for data consistency. In addition, only companies with a fiscal year-end in December were used as samples for this study, and capital erosion companies in abnormal situations were excluded from the sample. And companies for which corporate financial data could not be obtained from Vaule Search were also excluded from the sample, and companies for which it was impossible to calculate the management strategy variables, which are the key variables to be predicted in this study, were excluded from the sample as well. Furthermore, companies that did not disclose ESG ratings from the Korea Institute of Corporate Governance and Sustainability(KCGS) were not included in the sample. Through this sample composition process, the final number of company samples used in the empirical analysis of this paper is 9,568. The sample composition process is summarized in <Table 1> below.

<Table 1> Sample Composition

Sample Composition	Number of Sample
Listed companies excluding financial companies from 2012 to 2023	20,577
Excluding companies with a fiscal year-end in December	-332
Capital erosion companies	-57
Excluding companies without corporate financial data	-43
Excluding companies for which management strategies (offensive, defensive) cannot be obtained	-4,765
Companies that did not disclose ESG ratings from the Korea Institute of Corporate Governance and Sustainability	-5,812
Sample 2012~2023	9,568

The corporate management strategy measurement used in this study refers to the method used in Bentley et al. (2013), Choi Gyudam (2016), and Higgins et al. (2015). The corporate management strategy measurement derives the management strategy index by the following six factors.

(1) New product development: Ratio of R&D expenses to sales
(2) Production efficiency: Ratio of number of employees to sales
(3) Growth pattern: Corporate growth rate (Price Book Value Ratio, PBR)
(4) Marketing effort: Ratio of sales expenses to sales
(5) Organizational stability: Standard deviation of total number of employees
(6) Capital intensity: Ratio of facility assets to total assets

The specific process of deriving the management strategy index is as follows: The average value is derived for each of the six measurement factors above that reflect various aspects of corporate management strategy, and they are ranked by year and industry. And then, after classifying it into 5 stages (quintiles), 5 points are given to the highest stage and 1 point to the lowest stage to ensure that the highest score is the leading type and the lowest score is the defensive type. However, in the case of capital intensity, if excessive investment is made in facility assets to achieve efficiency, it is similar to a defensive management strategy, so on the contrary, 1 point is given to the highest level; and 5 points, to the lowest level. If this calculation process is performed for each of the 6 measurement factors, a value between 1 and 5 points is given for each of the 6 measurement factors for each year, and if all the points given for each factor are added up, a comprehensive score between 6 and 30 points is derived. In conclusion, the higher the comprehensive score, the closer it is to a leading management strategy, and conversely; the lower the comprehensive score, the closer it is to a defensive management strategy.

The classification of the management strategy type is classified as a leading (prospector) management strategy if the comprehensive score is 23-30 points; and a defensive (defender) management strategy if the composite score is 6-13 points. The remaining 14-22 points were classified as an analyzer management strategy and were excluded from the sample of this study.

And the ESG ratings used in this study were obtained through the KCGS, and were quantified as 6 for A+ rating, 5 for A rating, 4 for B+ rating, 3 for B rating, 2 for C rating, and 1 for D rating for empirical analysis. The environmental, social, governance ratings, and total ratings were all

quantified in the same way. <Table 2> is the definition of the variables used in this study.

<Table 2> Variable Definition

Variable	변수정의
Prospector	Offensive management strategy = 1 if the management strategy index is 23-30, otherwise 0
Defender	Defensive management strategy = 1 if the management strategy index is 6-13, otherwise 0
E	Points are awarded based on the environmental rating of the KCGS. (A+=6, A=5, B+=4, B=3, C=2, D=1)
S	Points are awarded based on the social rating of the KCGS. (A+=6, A=5, B+=4, B=3, C=2, D=1)
G	Points are awarded based on the governance rating of the KCGS. (A+=6, A=5, B+=4, B=3, C=2, D=1)
ESG	Points are awarded based on the total rating of the KCGS. (A+=6, A=5, B+=4, B=3, C=2, D=1)
ROA	Return on total assets = net income / total assets
SIZE	Firm size = natural log value of total assets
LEV	Debt ratio = total debt / total assets
CUR	Current Ratio = Current Liabilities / Current Assets
GRW_Sales	Sales growth rate = (current sales − previous sales) / previous sales
AGE	Company age = natural logarithmic value of (current fiscal year - year of incorporation)
OWN	Largest shareholder's stake = Largest shareholder's stake including the stake of specially-related parties / Total share of shareholders
FORN	Foreign stake = Foreign investor stake / Total share of shareholders
LOSS	Previous loss = 1 if the company reported a loss in its financial statements, otherwise 0

<Table 3> shows the descriptive statistics values for the variables. It was confirmed that the mean value of Prospector, a variable indicating an offensive management strategy, was 0.057, the median was 0.000, and the standard deviation was 0.233. And the mean value of Defender, a variable indicating a defensive management strategy, was 0.042, the median was 0.000, and the standard deviation was 0.201.

The mean value of ESG environmental

2.343, the median was 2.000, and the standard deviation was 1.445. The mean value of ESG social rating was 2.696, the median was 3.000, and the value of ESG governance rating was 2.875, the median was 3.000, and the standard deviation was 1.018. The mean of the ESG total rating was 2.513, the median was 3.000, and the standard deviation was 1.339.

<Table 3> Descriptive statistics

Variable	Number of Sample	Mean	Std	Min	Q1	Median	Q3	Max
Prospector	9,568	0.057	0.233	0.000	0.000	0.000	0.000	1.000
Defender	9,568	0.042	0.201	0.000	0.000	0.000	0.000	1.000
E	9,568	2.343	1.445	0.000	1.000	2.000	3.000	6.000
S	9,568	2.696	1.520	0.000	2.000	3.000	4.000	6.000
G	9,568	2.875	1.018	0.000	2.000	3.000	3.000	6.000
ESG	9,568	2.513	1.339	0.000	2.000	3.000	3.000	6.000
ROA	9,568	0.020	0.090	-0.464	-0.002	0.025	0.058	0.296
SIZE	9,568	26.652	1.453	23.613	25.654	26.455	27.397	30.604
LEV	9,568	0.391	0.204	0.027	0.221	0.391	0.543	0.875
CUR	9,568	2.660	4.216	0.195	0.938	1.464	2.603	32.472
GRW_Sales	9,568	0.060	0.316	-0.701	-0.074	0.026	0.135	2.111
AGE	9,568	3.522	0.582	1.792	3.135	3.689	3.951	4.844
OWN	9,568	0.422	0.165	0.000	0.302	0.425	0.533	1.000
FORN	9,568	0.094	0.127	0.000	0.013	0.043	0.124	1.000
LOSS	9,568	0.246	0.431	0.000	0.000	0.000	0.000	1.000

3.2. Research Model

The corporate management strategy prediction model to be developed in this paper is based on previous studies on management strategy, and extracts financial variables that have a significant impact on management strategy to construct the research model. The specific description of the financial data used in the corporate management performance

rating was prediction model is as follows: First, ROA is a variable that means return on total assets and is often used as a proxy variable for corporate management performance. Corporate management performance is significantly affected by management strategies. Therefore, it was used in the corporate management strategy prediction model. In addition, since the

management strategies established differ depending on the size of the company, SIZE, a variable that means corporate size, was also added to the prediction model (D'Amato and Falivena, 2020). The debt ratio and current ratio of a company mean the financial soundness and stability of the company, which can have a significant impact on the establishment and promotion of management strategies, so LEV and CUR were also added to the prediction model (Abraham et al., 2020; Zimon et al., 2021).

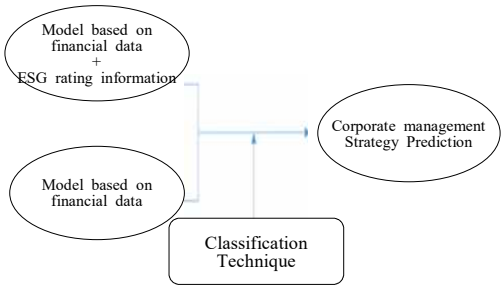
According to the corporate life cycle theory, the corporate management strategy pursued changes according to the growth stage. To reflect this, the variable GRW_Sales, which means the sales growth rate, and the variable AGE, which means the age of the company, were used in the management strategy prediction model (Dewi et al., 2022; Wen et al., 2022).

Meanwhile, since the management strategy of a company is established by the management, the corporate governance structure acts as an important factor (Lubatkin et al., 2005; Cuevas-Rodriguez et al., 2016). Therefore, the variable OWN, which means the largest shareholder's stake, and the variable FORN, which means the foreign investor's stake, which are often used as corporate governance variables, were added to the management strategy prediction model. Finally, since the management strategy pursued may completely change if the company reports a loss, the variable LOSS, which means whether the company reported a loss, was also used in the prediction model.

The prediction model designed with financial variables that significantly affect the management strategy based on previous studies on management strategy is as shown in Equation (1) below. And the prediction model designed by considering not only financial variables that have a significant impact on management strategy
but also ESG rating information is as shown in Equation (2) below.

[Financial Variable-Based Model]
Corporate Management Strategy Prediction (Prospector / Defender
= F(ROA, SIZE, LEV, CUR, GRW_Sales, AGE, OWN, FORN, LOSS) Equation(1)

[Financial Variable-Based Model + ESG Rating]
Corporate Management Strategy Prediction (Prospector / Defender
= F(ROA, SIZE, LEV, CUR, GRW_Sales, AGE, OWN, FORN, LOSS, E, S, G, ESG) Equation(2)



<Figure 1> Research Method

In this study, it is expected that a model with added ESG rating information will have higher prediction performance than a model composed only of financial data related to management strategy when predicting corporate management strategy, and this is proven through various machine learning classifiers.
<Figure 1> is a diagram to make it easy to understand the research methodology of the corporate management strategy prediction model that this paper aims to develop.

3.3. Machine Learning Classification Technique

In this paper, a management strategy prediction model is developed using machine learning classification techniques. In machine learning, classification aims to classify input data into distinct classes (Osisanwo et al., 2017).

Generally, the machine learning model assigns probability scores to each class using the soft-max function in the final output layer (Lopez-Martinez et al., 2018). The class with the highest probability is usually selected in the prediction step (Osisanwo et al., 2017).

To handle classification problems, machine learning uses the cross-entropy loss function to measure the difference between the predicted class probability and the actual class label (Kumari and Srivastava, 2017). The parameters of the model are adjusted to minimize this discrepancy (Wang et al., 2021). In these machine learning classification techniques, evaluation indicators such as Accuracy, precision, Recall, AUC (Area

Under the Curve), and F1 score are used to evaluate the performance of the model (Kumari and Srivastava, 2017).

Accuracy is the proportion of correctly classified samples among all samples, and it measures the frequency with which the model's predictions match the actual class labels, indicating the overall proficiency of the model (Soofi and Awan, 2017). precision is the proportion of actual positive samples among the samples predicted as positive, and is calculated as the proportion of actual positive samples among the predicted positive samples (Narayanan et al., 2016). In other words, precision reflects the ability of the model to correctly identify positive samples. recallis the proportion of actual positive samples correctly predicted as positive, and measures the proportion of actual positive samples correctly identified by the model (Pirscoveanu et al., 2015). In other words, recallevaluates the effectiveness in capturing the positive class.

AUC is the area under the ROC (Receiver Operating Characteristics) curve, which evaluates the overall performance of a classification model (Shafiq et al., 2016). In other words, AUC evaluates a binary classification model by its ability to distinguish between positive and negative classifications using the ROC curve. This curve shows the true positive rate relative to the false positive rate at various classification thresholds (Maxwell et al., 2018). The AUC value ranges from 0 to 1, and the value closer to 1 indicates the classification performance of the model (Shafiq et al., 2016).

The F1 score, which is the harmonic mean of precision and Recall, comprehensively measures the classification ability of the model (Narayanan et al., 2016). The F1 score provides a more balanced evaluation performance for model performance, making it particularly useful when dealing with imbalanced data. The F1 score ranges from 0 to 1, and a higher value indicates better classification performance (Soofi and Awan, 2017).

3.4. Machine Learning Classifier

RF (Random Forest) is an ensemble learning technique that performs predictions using various decision-making trees (Gholizadeh et al., 2020). Each tree learns on a random sample of data, and the final prediction is based on the average of the prediction results of all trees (Khajavi and Rastgoo, 2023). The advantage of RF is that it reduces overfitting and shows high accuracy (Khozeimeh et al., 2022). In addition, RF is robust to noise and can be flexibly applied to various data types (Wongvibulsin et al., 2020). On the other hand, the disadvantage of RF is that it may be expensive in terms of computational cost because it uses many trees, and that it makes model interpretation difficult (Gholizadeh et al., 2020).

SVM (Support Vector Machine) is a supervised learning technique that finds the optimal hyperplane to separate data points in a high-dimensional space (Pisner and Schnyer, 2020). It can solve nonlinear classification problems using kernel tricks (Roy and Chakraborty, 2023). SVM works well on high-dimensional data with clear decision boundaries and performs excellently on small data sets (Kurani et al., 2023). However, SVM is expensive in terms of computational cost, and hyperparameter tuning is complicated for large data sets (Cervantes et al., 2020).

XGBoost (Extreme Gradient Boosting) efficiently implements the gradient boosting algorithm (Qiu et al., 2022). XGBoost improves performance at each step by adding a new model to reduce errors (Asselman et al., 2023). The advantages of XGBoost are its high prediction performance and efficiency (Sagi and Rokach, 2021). And various techniques to prevent overfitting are also included in XGBoost (Zhang et al., 2023). On the other hand, the disadvantage of XGBoost is that it has many hyperparameters, making tuning complex and taking a long time to learn (Asselman et al., 2023).

LightGBM (Light Gradient Boosting Machine) is a gradient boosting framework developed by Microsoft for fast learning on large-scale data sets (Wang et al., 2022). LightGBM uses the Tree Leaf method (Liu et al., 2021). The advantages of LightGBM are learning speed and low memory usage, and it

is suitable for large-scale data sets (Rufo et al., 2021). However, the disadvantage of LightGBM is that overfitting may occur due to the complexity of the model, and data preprocessing may be also complicated (Yan et al., 2021).

CatBoost is a gradient boosting algorithm developed by Yandex. It is optimized for categorical data processing and has characteristics that are insensitive to data order (Hancock and Khoshgoftaar, 2020). The advantages of CatBoost are that it has an automatic processing function for categorical data and high prediction performance (Ibrahim et al., 2020). CatBoost also includes various functions to prevent overfitting (Jabeur et al., 2021). On the other hand, the disadvantages of CatBoost are that it can take a long time to learn and may require hyperparameter tuning depending on the situation (Luo et al., 2021).

The characteristics of the classifiers used in this study can be summarized as follows: RF can be applied to various data types and effectively reduce overfitting, but it is expensive in terms of computational cost. SVM performs well in high-dimensional data and is advantageous in small data sets, but it is expensive in terms of computational cost in large data.

XGBoost has high prediction performance and efficiency, but hyperparameter tuning is complicating. LightGBM learns quickly on large data sets and uses less memory, but it can cause overfitting. CatBoost is robust on categorical data and shows high performance, but it may take a long time to learn. Since the performance of each classifier model can vary depending on the characteristics and circumstances of the data, this study uses five machine learning classifier models to compare and analyze their performance in order to predict corporate management strategies and identify which model is the most suitable.

4. Research Analysis Results

This study develops a model that predicts corporate management strategies using machine learning classification techniques. It reveals that models that add ESG rating information have higher prediction performance than models that only consist of financial data related to corporate management strategies when predicting corporate management strategies through various machine learning classifiers. The five machine learning classifiers used in this study are RF, SVM, XGBoost, LightGBM, and CatBoost.

<Table 4> is the research results on the Accuracy performance among the analysis results of corporate management strategy prediction.

First, the results using the RF model are as follows: The management strategy accuracy prediction result by the financial data-based model was 0.7651, and when the environmental rating among the ESG ratings was included in the financial data-based model, the management strategy accuracy prediction result was 0.7814.

In addition, when the social rating among the ESG ratings was included in the financial data-based model, the management strategy accuracy prediction result was 0.7701, and when the governance rating was included, the management strategy accuracy prediction result was 0.7939. Finally, when the ESG total rating was included in the financial data-based model, the management strategy accuracy prediction result was high at 0.8013.

Second, the results using the SVM model are as follows: the management strategy accuracy prediction result by the financial data-based model was 0.7745, and when the environmental rating among the ESG ratings was included in the financial data-based model, the management strategy accuracy prediction result was 0.7978.

And when the social rating among ESG ratings was included in the financial data-based model, the management strategy accuracy prediction result was 0.7829, and when the governance rating was included, the management strategy accuracy prediction result was 0.8061. Finally, when the ESG total rating was included in the financial data-based model, the management strategy accuracy result was high at 0.8127.

Third, the results using the XGboost model are as follows: The management strategy accuracy prediction result by the financial data-based model was 0.8013, and when the environmental rating among ESG ratings was included in the financial data-based model, the management strategy accuracy prediction result was 0.8238.

And when the social rating among ESG ratings was included in the financial data-based model, the management strategy accuracy prediction result was 0.8120, and when the governance rating was included, the management strategy accuracy prediction result was 0.8379. Finally, when the ESG total rating was included in the financial data-based model, the management strategy accuracy prediction result was high at 0.8431.

Fourth, the results using the LightGBM model are as follows. The management strategy accuracy prediction result by the financial data-based model was 0.8670, and when the environmental rating among the ESG ratings was included in the financial data-based model, the management strategy accuracy prediction result was 0.8827.

In addition, when the social rating among the ESG ratings was included in the financial data-based model, the management strategy accuracy prediction result was 0.8704, and when the governance rating was included, the management strategy accuracy prediction result was 0.8968. Finally, when the ESG total rating was included in the financial data-based model, the management strategy accuracy prediction result was the highest at 0.9009.

Fifth, the research results using the CatBoost model are presented below. The management strategy accuracy prediction result by the financial data-based model was 0.8329, and when the environmental rating among the ESG rating was included in the financial data-based model, the management strategy accuracy prediction result was 0.8513.

In addition, when the social rating among the ESG ratings was included in the financial data-based model, the management strategy accuracy prediction result was 0.8407, and when the governance rating was included, the management strategy accuracy prediction result was 0.8688.

Finally, when the ESG total rating was included in the financial data-based model, the management strategy accuracy prediction result was high at 0.8827.

<Table 4> Corporate Management Strategy Prediction Results (Accuracy)

Model Type	Classifier	Accuracy	Classifier	Accuracy	Classifier	Accuracy	Classifier	Accuracy	Classifier	Accuracy
financial data-based model	RF	0.7651	SVM	0.7745	XGboost	0.8013	Light GBM	0.8670	Cat Boost	0.8329
financial data-based model + E rating		0.7814		0.7978		0.8238		0.8827		0.8513
financial data-based model + S rating		0.7701		0.7829		0.8120		0.8704		0.8407
financial data-based model + G rating		0.7939		0.8061		0.8379		0.8968		0.8688
financial data-based model + ESG rating		**0.8013**		**0.8127**		**0.8431**		**0.9009**		**0.8827**

<Table 5> presents the research results on precision performance among the results of corporate management strategy prediction.

First, the results using the RF model are as follows: The management strategy precision prediction result by the financial data-based model was 0.7542, and when the environmental rating among the ESG ratings was included in the financial data-based model, the management strategy precision prediction result was 0.7738. In addition, when the social rating among the ESG ratings was included in the financial data-based model, the management strategy precision prediction result was 0.7701, and when the governance rating was included, the management strategy precision prediction result was 0.7939. Finally, when the ESG total rating was included in the financial data-based model, the management strategy precision prediction result was high at 0.7930.

Second, the results using the SVM model are as follows: The management strategy precision prediction result by the financial data-based model was 0.7832, and when the environmental rating among the ESG ratings was included in the financial data-based model, the management strategy precision prediction result was 0.8032. And when the social rating among ESG ratings was included in the financial data-based model, the management strategy precision prediction result was derived as 0.7901, and when the governance rating was included, the management strategy precision prediction result was derived as 0.8179. Finally, when the ESG total rating was included in the financial data-based model, the management strategy precision prediction result was high at 0.8277.

Third, the results using the XGboost model

are as follows: The management strategy precision prediction result by the financial data-based model was 0.8123, and when the environmental rating among ESG ratings was included in the financial data-based model, the management strategy precision prediction result was 0.8399. And when the social rating among ESG ratings was included in the financial data-based model, the management strategy precision prediction result was derived as 0.8254, and when the governance rating was included, the management strategy precision prediction result was derived as 0.8527. Finally, when the ESG total rating was included in the financial data-based model, the management strategy precision prediction result was high at 0.8681.

Fourth, the results using the LightGBM model are as follows: The management strategy precision prediction result by the financial data-based model was 0.8865, and when the environmental rating among the ESG ratings was included in the financial data-based model, the management strategy precision prediction result was 0.9022. In addition, when the social rating among the ESG ratings was included in the financial data-based model, tthe management strategy precision prediction result was 0.8910, and when the governance rating was included, the management strategy precision prediction result was 0.9137. Finally, when the ESG total rating was included in the financial data-based model, the management strategy precision prediction result was the highest at 0.9278.

Fifth, the results of the research using the CatBoost model are presented below. The management strategy precision prediction result by the financial data-based model was 0.8231, and when the environmental rating among the ESG ratings was included in the financial data-based model, the management strategy precision prediction result was 0.8412. And when the social rating among ESG ratings was included in the financial data-based model, the management strategy precision prediction result was 0.8360, and when the governance rating was included, the management strategy precision prediction result was 0.8536. Finally, when the ESG total rating was included in the financial data-based model, the management strategy precision prediction result was high at 0.8698.

<Table 5> Corporate Management Strategy Prediction Results (Precision)

Model Type	Classifier	Precision	Classifier	Precision	Classifier	Precision	Classifier	Precision	Classifier	Precision
financial data-based model	RF	0.7542	SVM	0.7832	XGboost	0.8123	Light GBM	0.8865	Cat Boost	0.8231
financial data-based model + E rating		0.7738		0.8032		0.8399		0.9022		0.8412
financial data-based model + S rating		0.7612		0.7901		0.8254		0.8910		0.8360
financial data-based model + G rating		0.7825		0.8179		0.8527		0.9137		0.8536
financial data-based model + ESG rating		**0.7930**		**0.8277**		**0.8681**		**0.9278**		**0.8698**

prediction of corporate management strategies.

First, the results using the RF model are as follows: The management strategy recall prediction result by the financial data-based model was 0.7705, and when the environmental rating among the ESG ratings was included in the financial data-based model, the management strategy recall prediction result was 0.7924. In addition, when the social rating among the ESG ratings was included in the financial data-based model, the management strategy recall prediction result was 0.7811, and when the governance rating was included, the management strategy recall prediction result was 0.8009. Finally, when the ESG total rating was included in the financial data-based model, the result of the prediction of the management strategy recallwas high at 0.8110.

Second, the results using the SVM model are as follows. the management strategy recall prediction result by the financial data-based model was 0.7920, and when the environmental rating among the ESG ratings was included in the financial data-based model, the management strategy recall prediction result was 0.8174. And when the social rating among ESG ratings was included in the financial data-based model, the management strategy recall prediction result

<Table 6> shows the research results on the recallperformance among the results of the

was derived as 0.8009, and when the governance rating was included, the management strategy recall prediction result was derived as 0.8288. Finally, when the ESG total rating was included in the financial data-based model, the management strategy

recall prediction result was high at 0.8391.

Third, the results using the XGboost model are as follows: The management strategy recall prediction result by the financial data-based model was 0.8233, and when the environmental rating among ESG ratings was included in the financial data-based model, the management strategy recall prediction result was derived as 0.8469. And when the social rating among ESG ratings was included in the financial data-based model, the management strategy recall prediction result was derived as 0.8340, and when the governance rating was included, the management strategy recall prediction result was derived as 0.8512. Finally, when the ESG total rating was included in the financial data-based model, the management strategy recall prediction result was high at 0.8588.

Fourth, the results using the LightGBM model are as follows: The management strategy recall prediction result by the financial data-based model was 0.8963, and when the environmental rating among the ESG ratings was included in the financial data-based model, the management strategy recall prediction result was 0.9169. In addition, when the social rating among the ESG ratings was included in the financial data-based model, the management strategy recall prediction result was 0.9077, and when the governance rating was included, the management strategy recall prediction result was 0.9213. Finally, when the ESG total rating was included in the financial data-based model, the management strategy recall prediction result was the highest at 0.9325.

Fifth, the research results using the CatBoost model are presented below. The result of the management strategy recall prediction by the financial data-based model was 0.8413, and when the environmental rating among the ESG ratings was included in the financial data-based model, the result of the management strategy recall prediction was 0.8624. In addition, when the social rating among the ESG ratings was included in the financial data-based model, the result of the management strategy recall prediction was 0.8507, and when the governance rating was included, the result of the management strategy recall prediction was 0.8790. Finally, when the ESG total rating was included in the financial data-based model, the result of the management strategy recall prediction was high at 0.8806.

<Table 6> Corporate Management Strategy Prediction Results (Recall)

Model Type	Classifier	Recall	Classifier	Recall	Classifier	Recall	Classifier	Recall	Classifier	Recall
financial data-based model	RF	0.7705	SVM	0.7920	XGboost	0.8233	Light GBM	0.8963	Cat Boost	0.8413
financial data-based model + E rating		0.7924		0.8174		0.8469		0.9169		0.8624
financial data-based model + S rating		0.7811		0.8009		0.8340		0.9077		0.8507
financial data-based model + G rating		0.8009		0.8288		0.8512		0.9213		0.8790
financial data-based model + ESG rating		**0.8110**		**0.8391**		**0.8588**		**0.9325**		**0.8806**

<Table 7> presents the research results on AUC performance among the results of corporate management strategy prediction.

First, the results using the RF model are as follows: The management strategy AUC prediction result by the financial data-based model was 0.7603, and when the environmental rating among ESG ratings was included in the financial data-based model, the management strategy AUC prediction result was 0.7813. In addition, when the social rating among ESG ratings was included in the financial data-based model, the management strategy AUC prediction result was 0.7708, and when the governance rating was included, the management strategy AUC prediction result was 0.7932. Finally, when the ESG total rating was included in the financial data-based model, the management strategy AUC prediction result was high at 0.8002.

Second, the results using the SVM model are as follows: The management strategy AUC prediction result by the financial data-based model was 0.7821, and when the environmental rating among ESG ratings was included in the financial data-based model, the management strategy AUC prediction result was 0.8078. And when the social rating among ESG ratings was included in the financial data-based model, the management strategy AUC prediction result was derived as 0.7931, and when the governance rating was included, the management strategy AUC prediction result was derived as 0.8121. Finally, when the ESG total rating was included in the financial data-based model, the management strategy AUC prediction result was high at 0.8233.

Third, the results using the XGboost model are as follows: The management strategy AUC prediction result by the financial data-based model was 0.8237, and when the environmental rating among ESG ratings was included in the financial data-based model, the management strategy AUC prediction result was 0.8451. And when the social rating among ESG ratings was included in the financial data-based model, the management strategy AUC prediction result was derived as 0.8310, and when the governance rating was included, the management strategy AUC prediction result was derived as 0.8591. Finally, when the ESG total rating was included in the financial data-based model, the management strategy AUC prediction result was high at 0.8689.

Fourth, the results using the LightGBM model are as follows: The management strategy AUC prediction result by the financial data-based model was 0.8705, and when the environmental rating among the ESG ratings was included in the financial data-based model, the management strategy AUC prediction result was 0.8901. In addition, when the social rating among the ESG ratings was included in the financial data-based model, the management strategy AUC prediction result was 0.8821, and when the governance rating was included, the management strategy AUC prediction result was 0.9055. Finally, when the ESG total rating was included in the financial data-based model, the management strategy AUC prediction result was the highest at 0.9167.

Fifth, the research results using the CatBoost model are presented below. The management strategy AUC prediction result by the financial data-based model was 0.8312, and when the environmental rating among ESG ratings was included in the financial data-based model, the management strategy AUC prediction result was 0.8557. In addition, when the social rating among ESG ratings was included in the financial data-based model, the management strategy AUC prediction result was 0.8408, and when the governance rating was included, the management strategy AUC prediction result was 0.8684. Finally, when the ESG total rating was included in the financial data-based model, the management strategy AUC prediction result was high at 0.8821.

<Table 7> Corporate Management Strategy Prediction Result (AUC)

Model Type	Classifier	AUC	Classifier	AUC	Classifier	AUC	Classifier	AUC	Classifier	AUC
financial data-based model	RF	0.7603	SVM	0.7821	XG boost	0.8237	Light GBM	0.8705	Cat boost	0.8312
financial data-based model + E rating		0.7813		0.8078		0.8451		0.8901		0.8557
financial data-based model + S rating		0.7708		0.7931		0.8310		0.8821		0.8408
financial data-based model + G rating		0.7932		0.8121		0.8591		0.9055		0.8684
financial data-based model + ESG rating		**0.8002**		**0.8233**		**0.8689**		**0.9167**		**0.8821**

<Table 8> shows the research results on the F1 score performance among the results of predicting corporate management strategies.

First, the results using the RF model are as follows: The management strategy F1 score prediction result by the financial data-based model was 0.7637, and when the environmental rating among the ESG ratings was included in the financial data-based model, the management strategy F1 score prediction result was 0.7855. In addition, when the social rating among the ESG ratings was included in the financial data-based model, the management strategy F1 score prediction result was 0.7741, and when the governance rating was included, the management strategy F1 score prediction result was 0.7963. Finally, when the ESG total rating was included in the financial data-based model, the management strategy F1 score prediction result was high at 0.8021.

Second, the results using the SVM model are as follows: The management strategy F1 score prediction result by the financial data-based model was 0.7905, and when the environmental rating among the ESG ratings was included in the financial data-based model, the management strategy F1 score prediction result was 0.8128. In addition, when the social rating among the ESG ratings was included in the financial data-based model, the management strategy F1 score prediction result was 0.8087, and when the governance rating was included, The management strategy F1 score prediction result was 0.8278. Finally, when the ESG total rating was included in the financial data-based model, the management strategy F1 score prediction result was high at 0.8330.

Third, the results using the XGboost model

are as follows: The management strategy F1 score prediction result by the financial data-based model was 0.8209, and when the environmental rating among the ESG ratings was included in the financial data-based model, the management strategy F1 score prediction result was 0.8454. And when the social rating among ESG ratings was included in the financial data-based model, the management strategy F1 score prediction result was derived as 0.8312, and when the governance rating was included, the management strategy F1 score prediction result was derived as 0.8591. Finally, when the ESG total rating was included in the financial data-based model, the management strategy F1 score prediction result was high at 0.8688.

Fourth, the results using the LightGBM model are presented as follows: The management strategy F1 score prediction result by the financial data-based model was 0.8943, and when the environmental rating among ESG ratings was included in the financial data-based model, the management strategy F1 score prediction result was 0.9124. And when the social rating among ESG ratings was included in the financial data-based model, the management strategy F1 score prediction result was derived as 0.9010, and when the governance rating was included, the management strategy F1 score prediction result was derived as 0.9235. Finally, when the ESG total rating was included in the financial data-based model, the management strategy F1 score prediction result was the highest at 0.9301.

Fifth, the research results using the CatBoost model are presented below. The management strategy F1 score prediction result by the financial data-based model was 0.8348, and when the environmental rating among the ESG ratings was included in the financial data-based model, the management strategy F1 score prediction result was 0.8514. In addition, when the social rating among the ESG ratings was included in the financial data-based model, the management strategy F1 score prediction result was 0.8410, and when the governance rating was included, the management strategy F1 score prediction result was 0.8690. Finally, when the ESG total rating was included in the financial data-based model, the management strategy F1 score prediction result was high at 0.8755.

<Table 8> Corporate Management Strategy Prediction Result (F1 score)

Model Type	Classifier	F1 score	Classifier	F1 score	Classifier	F1 score	Classifier	F1 score	Classifier	F1 score
financial data-based model	RF	0.7637	SVM	0.7905	XG boost	0.8209	Light GBM	0.8943	Cat Boost	0.8348
financial data-based model + E rating		0.7855		0.8128		0.8454		0.9124		0.8514
financial data-based model + S rating		0.7741		0.8087		0.8312		0.9010		0.8410
financial data-based model + G rating		0.7963		0.8278		0.8591		0.9235		0.8690
financial data-based model + ESG rating		**0.8021**		**0.8330**		**0.8688**		**0.9301**		**0.8755**

5. Discussion and Conclusion

5.1 Research Summary and Results

In this paper, in predicting corporate management strategies, not only financial data related to management strategies but also ESG rating information of the company were reflected in the model. In this study, machine learning classifiers RF, SVM, XGBoost, LightGBM, and CatBoost were used to develop a model that predicts corporate management strategies by utilizing machine learning classification techniques.

The results of the study are summarized as follows: First, the corporate management strategy prediction model that included ESG rating information in financial data showed better prediction performance than the corporate management strategy prediction model composed of only financial data. The best performance was shown when the ESG total rating was included among the ESG ratings, and the high performance was shown in the order of governance rating, environmental rating, and social rating when these information were considered. This means that ESG rating information is an important factor in predicting corporate management strategies.

Second, among the five machine learning classifiers, LightGBM was confirmed to have the best prediction performance in predicting

corporate management strategies. Next, it was confirmed that the best performance was shown in the order of CatBoost, XGBoost, SVM, and RF in predicting corporate management strategies. This result means that the Boost series classifier is effective in predicting management strategies.

5.2 Contributions and Expected Effects

The contributions and expected effects of this paper are as follows: Above all, the followings are academic contributions. First, this study expanded the scope of research in the field of management strategies by proposing a new machine learning model that predicts management strategies by integrating ESG ratings and financial data. Second, it strengthened the academic foundation for research in the field of machine learning prediction by proposing an analysis methodology that utilizes various machine learning algorithms.

And the practical contributions of this study are as follows: First, the research results of this paper can be used as a sophisticated and accurate decision-making support tool that considers ESG factors when establishing corporate management strategies. Second, it indirectly suggests that it is important to reflect ESG factors when establishing corporate management strategies, and that it is important to establish and promote strategies that consider corporate sustainability.

Meanwhile, the academic expected effects of this study are as follows: First, it is expected to promote follow-up research on the relationship between ESG and management strategy and contribute to expanding the research base in the field of management strategy prediction.
Second, it is expected that the new data integration method used in this study will be ues in other research fields and utilized to improve model performance.

And the practical expected effects are as follows: First, it is expected that this study can contribute to improving the long-term management efficiency of companies through a strategy prediction method that reflects ESG factors in terms of improving management efficiency. Second, the results of this study are expected to provide investors with ESG-based management strategy information to support better investment decisions.

5.3 Limitations and Future Research Directions

Despite the above academic and practical contributions and expected effects, the limitations of this paper and future research plans are as follows: There are also problems such as the transparency of evaluating ESG ratings and the differences in the calculation method ratings in the methodology of each rating agency, which increases the volatility, and this has been shown to be a limitation in developing a corporate management strategy prediction model in that there is a high correlation between agencies.

If the management strategy model had been designed by reflecting important qualitative information of companies such as management reports, audit reports, and sustainable management reports in addition to ESG ratings, the prediction performance could have been further improved. In future research, I am planning to develop a more sophisticated and detailed management strategy prediction model by collecting various important qualitative information of companies.

Reference

[Previous Studies in Korea]

고세훈, 유왕진, & 이윤보. (2013). 중소벤처기업의 경쟁전략과 경영성과 간의 구조적 관계 에 관한 실증연구. 생산성연구: 국제융합학술지, 27(1), 225-260.

공석진, & 양해술. (2014). 경영전략이 경영성과에 미치는 영향. Journal of Digital Convergence, 12(1), 177-192.

남정민, & 박준혁. (2022). 벤처기업의 ESG 경영전략 도출을 위한 계층분석과정 (AHP) 적용. Entrpreneurship & ESG 연구, 2(1), 1-25.

박다인, & 박찬희. (2018). 벤처기업의 성장단계별 기업 경쟁력 및 기업 성과 창출 전략. 벤처창업연구, 13(6), 177-189.

윤성용. (2023). 경영전략이 ESG 와 기업가치와의 관계에 미치는 영향. 비즈니스융복합연구, 8(4), 33-38.

윤종록, & 김형철. (2009). 벤처기업의 창업가특성과 차

별화전략이 경영성과에 미치는 영향에 관한 연구. 대한경영학회지, 22(6), 3693-3721.

주성한, & 정남호. (2023). 호텔기업의 지식경영전략이 종사원의 몰입과 직무유지의도에 미치는 영향. 인터넷전자상거래연구, 23(3), 205-231.

최규담. (2016). 경영전략과 발생액의 질. 회계저널, 25(3), 261-305.

추승엽, 유정민, & 임성준. (2009). 경영환경, 경쟁전략 및 기업 내부역량 간의 적합성이 성과에 미치는 영향에 관한 연구. 전략경영연구, 12(1), 101-126.

한규동. (2019). 벤처기업 창업가의 특성과 경영전략이 경영성과에 미치는 영향. 벤처창업연구, 14(6), 29-43.

홍난희. (2024). 경영전략과 투자부동산 그리고 기업가치. 대한경영학회지, 37(4), 611-630.

[Previous Studies Abroad]

Abraham, F., Cortina Lorente, J. J., & Schmukler, S. L. (2020). Growth of global corporate debt: Main facts and policy challenges. World Bank Policy Research Working Paper, (9394).

Akmaeva, R., Arykbaev, R., Epifanova, N., & Glinchevskiy, E. (2020). Influence of digitalization upon formation of corporate strategy and new business models of modern organizations. In SHS Web of Conferences (Vol. 89, p. 03003). EDP Sciences.

Al-Najjar, B., & Anfimiadou, A. (2012). Environmental policies and firm value. business Strategy and the Environment, 21(1), 49-59.

Al-Shaer, H., Kuzey, C., Uyar, A., & Karaman, A. S. (2023). Corporate strategy, board composition, and firm value. International Journal of Finance & Economics.

Asselman, A., Khaldi, M., & Aammou, S. (2023). Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. Interactive Learning Environments, 31(6), 3360-3379.

Avanesova, N., Tahajuddin, S., Hetman, O., Serhiienko, Y., & Makedon, V. (2021). Strategic management in the system model of the corporate enterprise organizational development. Economics and Finance, 1(2021), 9.

Aydoğmuş, M., Gülay, G., & Ergun, K. (2022). Impact of ESG performance on firm value and profitability. Borsa Istanbul Review, 22, S119-S127.

Behl, A., Kumari, P. R., Makhija, H., & Sharma, D. (2022). Exploring the relationship of ESG score and firm value using cross-lagged panel analyses: Case of the Indian energy sector. Annals of Operations Research, 313(1), 231-256.

Bentley, K. A., Omer, T. C., & Sharp, N. Y. (2013). management strategy, financial reporting irregularities, and audit effort. Contemporary accounting research, 30(2), 780-817.

Bowman, C., & Ambrosini, V. (2007). Firm value creation and levels of strategy. Management Decision, 45(3), 360-371.

Bowman, E. H., & Helfat, C. E. (2001). Does corporate strategy matter?. Strategic management journal, 22(1), 1-23.

Brewster, C. (2017). The integration of human resource management and corporate strategy. Policy and practice in European human resource management, 22-35.

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications,

challenges and trends. Neurocomputing, 408, 189-215.

Cuevas-Rodriguez, G., Guerrero-Villegas, J., & Valle-Cabrera, R. (2016). Corporate governance changes, firm strategy and compensation mechanisms in a privatization context. Journal of Organizational Change Management, 29(2), 199-221.

D'Amato, A., & Falivena, C. (2020). Corporate social responsibility and firm value: Do firm size and age matter? Empirical evidence from European listed companies. Corporate Social Responsibility and Environmental Management, 27(2), 909-924.

Dewi, N. W. S. K., Ratnadi, N. M. D., Yadnyana, I. K., & Suaryana, I. G. N. A. (2022). Selection of profit management strategy: Testing at the company life cycle stage. Linguistics and Culture Review, 6(S1), 530-549.

Gholizadeh, M., Jamei, M., Ahmadianfar, I., & Pourrajab, R. (2020). Prediction of nanofluids viscosity using random forest (RF) approach. Chemometrics and Intelligent Laboratory Systems, 201, 104010.

González-Ramos, M. I., Guadamillas, F., & Donate, M. J. (2023). The relationship between knowledge management strategies and corporate social responsibility: Effects on innovation capabilities. Technological Forecasting and Social Change, 188, 122287.

Grant, R. M. (2002). Corporate strategy: managing scope and strategy content. Handbook of strategy and management, 72-97.

Gunarathne, A. N., Lee, K. H., & Hitigala Kaluarachchilage, P. K. (2021). Institutional pressures, environmental management strategy, and organizational performance: The role of environmental management accounting. management Strategy and the Environment, 30(2), 825-839.

Hancock, J. T., & Khoshgoftaar, T. M. (2020). CatBoost for big data: an interdisciplinary review. Journal of big data, 7(1), 94.

Higgins, D., Omer, T. C., & Phillips, J. D. (2015). The influence of a firm's management strategy on its tax aggressiveness. Contemporary Accounting Research, 32(2), 674-702.

Ibrahim, A. A., Ridwan, R. L., Muhammed, M. M., Abdulaziz, R. O., & Saheed, G. A. (2020). Comparison of the CatBoost classifier with other machine learning methods. International Journal of Advanced Computer Science and Applications, 11(11).

Jabeur, S. B., Gharib, C., Mefteh-Wali, S., & Arfi, W. B. (2021). CatBoost model and artificial intelligence techniques for corporate failure prediction. Technological Forecasting and Social Change, 166, 120658.

Junius, D., Adisurjo, A., Rijanto, Y. A., & Adelina, Y. E. (2020). The impact of ESG performance to firm performance and market value. Jurnal Aplikasi Akuntansi, 5(1), 21-41.

Khajavi, H., & Rastgoo, A. (2023). Predicting the carbon dioxide emission caused by road transport using a Random Forest (RF) model combined by Meta-Heuristic Algorithms. Sustainable Cities and Society, 93, 104503.

Khozeimeh, F., Sharifrazi, D., Izadi, N. H., Joloudari, J. H., Shoeibi, A., Alizadehsani, R., ... & Islam, S.

M. S. (2022). RF-CNN-F: random forest with convolutional neural network features for coronary artery disease diagnosis based on cardiac magnetic resonance. Scientific reports, 12(1), 11178.

Kumari, R., & Srivastava, S. K. (2017). Machine learning: A review on binary classification. International

Journal of Computer Applications, 160(7).

Kurani, A., Doshi, P., Vakharia, A., & Shah, M. (2023). A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting. Annals of Data Science, 10(1), 183-208.

Liu, J., Gao, Y., & Hu, F. (2021). A fast network intrusion detection system using adaptive synthetic oversampling and LightGBM. Computers & Security, 106, 102289.

Lopez-Martinez, F., Schwarcz, A., Núñez-Valdez, E. R., & Garcia-Diaz, V. (2018). Machine learning classification analysis for a hypertensive population as a function of several risk factors. Expert Systems with Applications, 110, 206-215.

Lubatkin, M. H., Lane, P. J., & Schulze, W. S. (2005). A strategic management model of agency relationships in firm governance. The Blackwell handbook of strategic management, 225-254.

Luo, M., Wang, Y., Xie, Y., Zhou, L., Qiao, J., Qiu, S., & Sun, Y. (2021). Combination of feature selection and catboost for prediction: The first application to the estimation of aboveground biomass. Forests, 12(2), 216.

Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. International

journal of remote sensing, 39(9), 2784-2817.

Menz, M., Kunisch, S., Birkinshaw, J., Collis, D. J., Foss, N. J., Hoskisson, R. E., & Prescott, J. E. (2021). Corporate Strategy and the Theory of the Firm in the Digital Age. Journal of Management Studies,

58(7), 1695-1720.

Narayanan, B. N., Djaneye-Boundjou, O., & Kebede, T. M. (2016, July). Performance analysis of machine learning and pattern recognition algorithms for malware classification. In 2016 IEEE national aerospace and electronics conference (NAECON) and ohio innovation summit (OIS) (pp. 338-342). IEEE.

Osisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., & Akinjobi, J. (2017). Supervised machine learning algorithms: classification and comparison. International Journal of Computer Trends and Technology (IJCTT), 48(3), 128-138.

Pirscoveanu, R. S., Hansen, S. S., Larsen, T. M., Stevanovic, M., Pedersen, J. M., & Czech, A. (2015, June). Analysis of malware behavior: Type classification using machine learning. In 2015 International conference on cyber situational awareness, data analytics and assessment (CyberSA) (pp. 1-7). IEEE.

Pisner, D. A., & Schnyer, D. M. (2020). Support vector machine. In Machine learning (pp. 101-121). Academic Press.

Qiu, Y., Zhou, J., Khandelwal, M., Yang, H., Yang, P., & Li, C. (2022). Performance evaluation of hybrid WOA-XGBoost, GWO-XGBoost and BO-XGBoost models to predict blast-induced ground vibration. Engineering with Computers, 38(Suppl 5), 4145-4162.

Quintiliani, A. (2022). ESG and firm value. Accounting and Finance Research, 11(4), 37.

Roy, A., & Chakraborty, S. (2023). Support vector machine in structural reliability analysis: A review. Reliability Engineering & System Safety, 233, 109126.

Rufo, D. D., Debelee, T. G., Ibenthal, A., & Negera, W. G. (2021). Diagnosis of diabetes mellitus using gradient boosting machine (LightGBM). Diagnostics, 11(9), 1714.

Sagi, O., & Rokach, L. (2021). Approximating XGBoost with an interpretable decision tree. Information sciences, 572, 522-542.

Shafiq, M., Yu, X., Laghari, A. A., Yao, L., Karn, N. K., & Abdessamia, F. (2016, October). Network traffic classification techniques and comparative analysis using machine learning algorithms. In 2016 2nd IEEE International Conference on Computer and Communications (ICCC) (pp. 2451-2455). IEEE.

Soofi, A. A., & Awan, A. (2017). Classification techniques in machine learning: applications and issues. Journal of Basic & Applied Sciences, 13, 459-465.

Wang, D. N., Li, L., & Zhao, D. (2022). Corporate finance risk prediction based on LightGBM. Information Sciences, 602, 259-268.

Wang, P., Fan, E., & Wang, P. (2021). Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. Pattern recognition letters, 141, 61-67.

Wen, H., Zhong, Q., & Lee, C. C. (2022). Digitalization, competition strategy and corporate innovation: Evidence from Chinese manufacturing listed companies. International Review of Financial Analysis, 82, 102166.

Wongvibulsin, S., Wu, K. C., & Zeger, S. L. (2020). Clinical risk prediction with random forests for survival, longitudinal, and multivariate (RF-SLAM) data analysis. BMC medical research methodology, 20, 1-14.

Yan, J., Xu, Y., Cheng, Q., Jiang, S., Wang, Q., Xiao, Y., ... & Wang, X. (2021). LightGBM: accelerated genomically designed crop breeding through ensemble learning. Genome biology, 22, 1-24.

Yu, X., & Xiao, K. (2022). Does ESG performance affect firm value? Evidence from a new ESG-scoring approach for Chinese enterprises. Sustainability, 14(24), 16940.

Yuan, Y., Lu, L. Y., Tian, G., & Yu, Y. (2020). management strategy and corporate social responsibility. Journal of management ethics, 162, 359-377.

Zhang, F., Qin, X., & Liu, L. (2020). The interaction effect between ESG and green innovation and its impact on firm value from the perspective of information disclosure. Sustainability, 12(5), 1866.

Zhang, J., Ma, X., Zhang, J., Sun, D., Zhou, X., Mi, C., & Wen, H. (2023). Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model. Journal of environmental management, 332, 117357.

Zimon, G., Nakonieczny, J., Chudy-Laskowska, K., Wójcik-Jurkiewicz, M., & Kochański, K. (2021). An analysis of the financial liquidity management strategy in construction companies operating in the Podkarpackie Province. Risks, 10(1), 5.

◐ Author Biography ◑

Kyung Gu Kang

Ph. D in AI Strategy Management

Research Professor, The Institute for Industrial Policy Studies (IPS)

Chairman, Commerce Holdings Inc.

Areas of Interest: Received a Ph.D in research on predicting future corporate value using ESG ratings, and is cu rrently conducting research on global management, GPU Cloud System,AI management strategy, and ESG investment strategy.

E-mail : thomas3877@naver.com

kgkang@ips.or.kr

Research on ethical issues and improvement tasks related to Big Data analysis and utilization of Artificial Intelligence

Se Eun Jeong

ABSTRACT

As Artificial Intelligence(AI)-based data analysis technology develops, it fails to keep up with general ethical standards and gaps arise between them, leading to ethical issues. When AI technology is used to analyze and utilize Big Data, it is no longer a threat to ethical influence, but can be a good turning point that turns a crisis into an opportunity. It is necessary to examine the ethical influence and other issues that these technologies have on the Big Data analysis process.

In this paper, it is first examined the relationship between AI and Big Data, and looked at how AI uses Big Data to learn data through Machine Learning algorithms and classify or predict new data.

Next, it is investigated how data is collected, what algorithms are used, and how the recommendation system is improved through real-time learning through a case study on Netflix. In relation to this, it is studied issues such as bias and lack of transparency that arise when examining ethical influences of AI technology on the Big Data analysis and utilization process.

Here, when the AI decision-making process is opaque, problems such as lack of trust, responsibility issues, difficulty in risk management, and conflicts of interest arise.

As a solution to these problems, both legal and technical solutions were presented. As such, Artificial Intelligence is in the process of developing very good technology in itself, but the problem of using Artificial Intelligence for Big Data cannot but be considered.

Therefore, if issues such as ethical impact are reviewed and evaluated in advance, and solutions and development tasks are sufficiently prepared, it can be expected to minimize the risks that may arise and maximize the advantages of Artificial Intelligence technology in using Big Data.

☞ keyword : Artificial Intelligence, Artificial Intelligence technology, Big Data, Big Data utilization, ethical implications.

1. Introduction

Today, in the era of the 4th Industrial Revolution, the amount of data generated by the Internet and IoT (Internet of Things) is increasing explosively. Furthermore, the development of data storage and analysis technologies such as cloud computing and Machine Learning has made it possible to effectively process large amounts of data. Moreover, Artificial Intelligence and Big Data are in a mutually complementary relationship, and as they develop together, they are leading to innovation in various industries.

However, since these are expected to have a great impact on our lives from an ethical perspective in the future, this paper fully recognizes that problems may be raised under this research background, and focuses on issues such as the ethical impact of Artificial Intelligence technology on Big Data analysis and utilization.

Recently, the development of data analysis technology based on Artificial Intelligence has not kept up with general ethical standards, so there is a tendency for a gap to arise between them. In particular, the ethical implications of Big Data analysis related to Artificial Intelligence technology and other controversial issues should be carefully reviewed, not only because of the ethical nature of Big Data analysis, but also because it should consider the ethical implications of Artificial Intelligence technology on the daily lives of the public and the growth and innovation of the national economy.[1]

Consequently, if Artificial Intelligence technology can be handled well, the technology will no longer be a threat to ethical implications, but rather a good turning point that turns crises into opportunities. Therefore, this paper focused on analyzing the various issues and ethical implications that may arise due to Artificial Intelligence technology in the future, and suggested improvement tasks such as legal regulations, technical solutions, and ethical standards.

2. The relationship between Artificial Intelligence and Big Data

2.1 Ethical Issues in a Big Data Environment

2.1.1 Current situation

The task of selecting and combining information appropriately is moving toward being solved in the 21st century. In addition, the development of Big Data technology that draws in a huge amount of information and processes it quickly has made all of this possible, thanks to

information storage and processing technology such as Big Data.

Big Data plays a very important role in strengthening the learning ability of Artificial Intelligence, and they are closely connected. Therefore, despite expectations and concerns about its importance and influence as an inevitable huge trend in today's information society, countries around the world are promoting policies to activate Big Data.

The development of the Big Data industry inevitably brings with it the issue of personal information protection, which creates a regulatory dilemma where both the leakage of personal information and excessive protection are problematic. In other words, in order to activate Big Data, it is necessary to have a legal environment where personal information is appropriately protected while also allowing such information to be freely utilized.

The development and activation of the Big Data industry bring with it the issue of personal information protection, which creates a regulatory dilemma where both the leakage of personal information and excessive protection are problematic. Strengthening formal prior regulations such as the current Personal Information Protection Act in Korea may only result in weakening the competitiveness of Big Data-related industries without providing actual personal information protection.

2.1.2 Ethical Issues and Privacy Issues in Big Data Environments

2.1.2.1 Big Data Environment and Ethical Issues

Key ethical issues in the Big Data environment include the fairness of data, the reliability of data sources, the responsibility for predictions and decisions, and ethical issues such as automation and jobs. If data is biased, the analysis results may be disadvantageous to a certain group. This risks deepening social inequality. There are also issues regarding the method of collecting data and whether the source is trustworthy. This is because incorrect data can lead to incorrect conclusions.

In particular, when making decisions through Big Data analysis, there is a need for discussion on who is responsible for the results of data analysis. In addition, as automation progresses due to the development of Big Data and AI, social and ethical issues such as job losses may arise. These ethical issues are becoming more important as Big Data technology develops, and it is necessary to discuss them and find solutions.

2.1.2.2 Big Data Utilization and Personal Information Protection Issues

Under the current law, personal information includes not only personally identifiable information but also non-identifiable information that cannot identify a specific individual by itself but can be easily combined with other information to identify the individual. The information utilized by Big Data is mainly non-identifiable information, and if such information has the potential to identify an individual, it becomes subject to protection under the Personal Information Protection Act. In principle, the consent of the Information Subject must be obtained in order to collect and use it.

Even if consent is obtained, only the minimum information must be collected within the necessary scope, and if it is to be used beyond the purpose of collection, separate consent must be obtained from the Information Subject, and personal information that has passed the retention period or achieved its purpose must be destroyed without delay.

However, Big Data is not simply a collection of data, but there is a significant possibility that meaningful information will be extracted from the data and used for purposes that were not thought of at the time of personal information collection. If the Personal Information Protection Act is applied to such Big Data as it is, the problem of obtaining consent from the Information Subject not only at the stage of collecting the information but also at every stage of analyzing it may arise. In addition, if the determination of personal information is made based on the unclear standard of identifiability, the possibility of legal disputes arising due to the Information Subject's reluctance to provide the information itself or the claim of personal information leakage will greatly increase.

In this regard, case law states that identifiability does not presuppose that all other information that can be combined with the relevant information is held by the same person, and does not mean that other information can be easily obtained, but that it means that the relevant information and other information can be easily combined without special difficulty to identify a specific individual, regardless of whether it is easy or difficult to obtain.

However, determining the possibility of personal identification due to easy combination with other data is profiling used in the Big Data processing process. Or, it will be a very difficult issue for lawyers who make normative judgments as to how to legally understand the technical difficulty of data mining technology.

Meanwhile, consumers have no idea whether their activities are being monitored online unless they are notified

of the existence of the advertising network of the website they visited and that data is being collected. In addition, since the processing of Big Data is fundamentally about the connectivity between data, even if the information does not have personal identification at the data collection stage, there is a significant possibility that specific individuals will be identified later during the data processing process. In this case, it will also be a problem how to protect personal information and what measures to take.

Therefore, the development and activation of the Big Data industry are accompanied by issues of personal information protection, and both the leakage of personal information and excessive protection are causing a regulatory dilemma. It should be noted that the strengthening of formal pre-regulation, such as the current Personal Information Protection Act in Korea, can only weaken the competitiveness of big data-related industries without substantial personal information protection.

Accordingly, some argue that the concept of personal information should be reduced, that is, that non-identification information should be treated differently from personal identification information, and the level of protection should be different or excluded from the object of protection at all.[2]

Therefore, it is judged more effective to ease regulations on information with a low level of personal information protection so that it can be easily used in the Big Data industry, but strictly establish the obligation to prevent abuse or infringement of information collectors or information processors for such information. This is because most Information Subjects do not check the personal information processing policy, it is very rare for them to carefully check the personal information consent form, and almost no individuals in the Big Data field can accurately understand the personal information processing policy and other related matters and consent to the provision of information to third parties.

Personally thinking, in a situation where information is being transferred and utilized between countries, the current Personal Information Protection Act focuses on protecting personal information from infringement and misuse of personal information by domestic companies. In the future, it is necessary to establish relevant regulations on information collection and infringement by global companies and devise measures to protect personal information in practice.[3]

2.1.3 The relationship between Artificial Intelligence and Big Data

Artificial Intelligence (AI) uses Big Data to perform pattern recognition, predictive modeling, and decision support. Big Data includes a large amount of diverse data, and AI analyzes this data to extract meaningful insights. Machine Learning algorithms learn patterns from learning data and use this to classify or predict new data. In this way, AI and Big Data can be seen as complementary.[4]

AI is a technology with human-like thinking abilities, and Big Data is a technology that analyzes various data. However, Big Data is essential for AI to make excellent decisions, and it is also important to have excellent computational technology to analyze Big Data and process it into meaningful information.

Therefore, Big Data is an inevitable huge trend in the current information society, and countries around the world are promoting policies to activate Big Data amid expectations and concerns about its importance and influence. In order to create new benefits and values through the use of Big Data, a large amount of data must be accumulated and integrated through various digital devices of AI, and it is true that the possibility of personal information leakage and abuse increases significantly.

2.2 Case study on Big Data analysis and utilization of Artificial Intelligence

2.2.1 Data collection and utilization, personalized customized content recommendations

Here is a real-world example of the combination of AI and Big Data: Netflix's recommendation system. Netflix collects a vast amount of data, including the content users watch, their ratings, search history, and viewing time. This data is used to understand and analyze individual user preferences, thereby improving user experience and providing personalized content recommendations.

Netflix records what content users watch, how often, and at what time, and collects information such as keywords and genres that users search for within the platform. This is used to identify user interests and recommend related content. It also stores a list of movies or TV shows that users have watched, which is used to improve the personalized recommendation system and suggest new content similar to content that users previously liked.

Netflix also conducts surveys asking users to rate or provide feedback on content. This feedback is used to improve content production and recommendation algorithms. By monitoring user activity or trends on social media platforms, it can analyze popular content or user

preferences. Through these data collection methods, Netflix provides personalized user experience, increases the accuracy of content recommendations, and improves the overall quality of its services. Compliance with privacy policies is also an important factor in the process of data collection and utilization.

2.2.2 Utilizing recommendation algorithms, analyzing various contents

Netflix uses several recommendation algorithms. The main algorithms are as follows: First, collaborative filtering is a method of making recommendations based on the behavior of users with similar tastes. For example, if A and B like similar movies, A's favorite movie can be recommended to B. Second, content-based filtering is a method of recommending similar content by analyzing the characteristics of content that a user has previously watched. For example, if a user watches a lot of romance movies, other romance movies of a similar genre can be recommended. Third, Netflix uses Deep Learning models to analyze complex patterns and provide more sophisticated recommendations to users. For example, it can make recommendations by considering various factors related to a specific genre of movies, such as actors, directors, and topics.

2.2.3 Utilizing real-time learning system, providing customized content

Netflix's recommendation system continuously improves its recommendation system by reflecting user behavior data in real time. This system adjusts the recommendation algorithm immediately as users watch or rate new content, providing a more personalized experience. Netflix collects data in real time, such as what content users watch, how long they watch it, and what content they skip, and analyzes user preferences based on this. It uses Machine Learning algorithms to identify users' preferences and compare patterns with other users with similar tastes.

The analyzed data is used to operate the recommendation algorithm, which provides personalized content recommendations to encourage users to spend more time on the platform. When users watch recommended content, feedback is reflected back into the system, continuously improving the algorithm. This real-time learning system allows Netflix to provide customized content to users, which plays a vital role in increasing user satisfaction and subscriber retention.

2.2.4 Use A/B group testing to continuously improve services

A/B group testing is a controlled experiment where observation data is divided into groups A and B, that is, the control group and the experimental group. The control group is divided into customers who received the service before changing the setting elements, and the experimental group is divided into customers who received the service after the change. Through this, the behavior of the control group and the experimental group is observed, hypothesis testing is performed based on the observed data, and the differences between the two groups are statistically analyzed.[5]

Netflix uses A/B testing to verify the effectiveness of its recommendation system. It experiments with various recommendation methods to analyze which method is most effective for users. Then, it collects user behavior data such as the number of clicks, playback time, and bounce rate from each group.

It analyzes the collected data to compare the performance of the two groups, and uses statistical methods to evaluate the significance of the results. Based on the results of the A/B test, the optimal option is selected, and successful changes are applied to all users. This plays an important role in Netflix providing customized user experiences and continuously improving its services.

2.2.5 Discussion

In short, Netflix's recommendation system combines Artificial Intelligence and Big Data to provide customized content to users, thereby maximizing user experience and increasing customer retention. Thanks to this system, Netflix has gained the competitiveness to retain many users worldwide.

3. Ethical Issues Arising from Big Data Analysis and Utilization of Artificial Intelligence Technology

3.1 Side Effects and Risk Factors

3.1.1 Side effects due to misuse

AI can be misused to create forgeries, crimes, and fake news. In fact, it can be exploited politically and spread false information, so it seems that a humanistic solution should be prepared for this area. In fact, AI that has been trained differently based on skin color and gender has developed

incorrect perceptions. In order to prevent the misuse of technology, it should also be prepared for the development of AI that can identify fake news or images. If it is inevitable in the age of AI, it is necessary to coexist, be ready for it, and prepare for the future.

In this way, AI can bring incalculable advantages and value to humans, but on the other hand, it can also cause privacy and personal information infringement issues. Apple is struggling between protecting customer privacy and developing AI, and has promoted the development of AI technology in accordance with its customer privacy protection policy.

Apple CEO Tim Cook presented a privacy policy on the company's website in September 2014, stating, "Apple is clearly opposed to marketing using user information." Therefore, Apple has no choice but to analyze data centered on iPhone data rather than iCloud, so Apple's technical challenge is to figure out how to increase user convenience with AI while protecting customer privacy.

In a survey of respondents who were either AI professionals or non-professionals regarding possible side effects of AI, respondents expressed greater concerns about side effects caused by AI misuse or malfunction than about job losses due to the introduction of AI. The respondents were found to have a great fear that AI could be misused for international crimes such as drugs or terrorism or that errors in algorithms that perform important functions of AI could lead to serious disasters.

The subjects of the survey were the Industry Analysis Team of Institute of Information & Communications Technology Planning & Evaluation (IITP), which analyzed the benefits and side effects that AI could bring. The significance of this survey lies in the fact that it was conducted to diagnose and secure basic data on how our country should develop and utilize Artificial Intelligence in the future.

3.1.2 Risk factors for security management systems

Since Artificial Intelligence is based on computer systems, even if it utilizes cutting-edge Artificial Intelligence technology, it is not safe from various hacking threats such as ransomware. For example, on July 21, 2015, an SUV traveling at 70 miles per hour on the St. Louis Highway in the United States suddenly stopped all operations and fell into a ditch on the side of the road.

The analysis of the cause of the accident revealed that the electronic control system of the SUV in question was hacked remotely 10 miles away from the accident site. The hackers infiltrated hacking programs to remotely control the steering wheel, and used the existing adaptive automatic cruise control to reduce the speed while simultaneously using the lane departure prevention function to change the direction of the car. After that, they made the accident prevention radar, which continuously sounds a warning sound when a dangerous object approaches, ignore the electronic safety control function and the ABS (Anti-lock Braking System), causing the car to fall into a ditch. Therefore, as seen in these cases, the possibility of hacking into the autonomous vehicle system is certainly there, as various wired and wireless lines are intertwined to control the autonomous vehicle.

3.1.3 Discussion

In short, recent Artificial Intelligence such as AlphaGo has shown amazing performance by utilizing the Internet of Things, Big Data, and Deep Learning, but it cannot respond to completely new variables such as changes in the rules of the game, and it has shown unpredictable behavior in cases where it has not been learned or cannot be inferred. It is true that Artificial Intelligence has technical limitations and risk factors because it has not been able to understand and express the philosophical and subjective concepts of human society to date.

In this way, the limitations of Artificial Intelligence and the risk factors for side effects were considered. If humans, who can critically and rationally view any fact or situation and overturn existing knowledge and discover new facts, can quickly analyze and infer specific facts using the excellent tool called Artificial Intelligence, they will be able to appropriately respond to changes such as the '4th Industrial Revolution' in the future.

3.2 Ethical Issues

3.2.1 Invasion of privacy

The collected data may be used for purposes other than the original purpose. Data collected for marketing purposes may be misused to evaluate an individual's credit. In addition, user data may be collected or analyzed without permission, which may infringe on an individual's privacy, and if sensitive information is leaked, it may cause serious damage to the individual. In the process of analyzing and predicting an individual's behavior or choices, location data or Internet usage history may be excessively collected.

3.2.2 Illegal surveillance

When AI technology is used for surveillance purposes, it can infringe on the freedom and rights of individuals. In particular, surveillance in public places can raise ethical issues. When AI manipulates or interferes with an individual's choices, it can infringe on the autonomy of the individual. For example, a recommendation system can force a specific product or opinion. In order to solve these problems, legal and social standards related to data ethics should be established, and transparent data use and AI development should be carried out.

3.2.3 Information asymmetry

There is an information asymmetry between data collectors and users, which may lead users to not understand how their data is processed.

3.2.3.1 Differences in information ownership

Data collectors have a variety of data collected from users and can analyze it to derive insights. On the other hand, users may lack information about how their data is collected and utilized.

3.2.3.2 Imbalance in data utilization

Data collectors can provide customized services based on the collected information and generate revenue through this. However, users may lack a clear understanding of how their data is used and for what purpose.

3.2.3.3 Reliability issues

Information asymmetry can create trust issues between users and data collectors. Users may be concerned about whether their data is being managed safely or whether it could be used illegally.

3.2.3.4 Asymmetry of decision-making power

While data collectors can have a decisive influence on marketing strategies or product development based on the results of data analysis, users may have limited choices about how their data is used. In short, this information asymmetry complicates the relationship between data collectors and users, and there is an ongoing debate about the importance of providing users with more transparency and choice.

3.2.4 Data security threats

The collected data may be at risk of hacking or leakage, which may result in personal information being exposed to outside parties.

3.2.4.1 Data breach threat

As AI systems process large amounts of data, security vulnerabilities can arise. There is a greater risk that hackers will infiltrate the system and leak personal or sensitive data. This can result in a violation of personal privacy and a loss of trust.

3.2.4.2 Unauthorized access

As AI technology analyzes data, there is a possibility that users or systems without access to the data may access the information. Such unauthorized access threatens the safety of the data and, if exploited, can lead to serious consequences.

3.2.4.3 The Complexity of Data Management

The AI models used in Big Data analytics are often complex and opaque. This makes it difficult to clearly understand how data is collected and stored, which makes security management difficult. Uncertainty in data management poses a challenge to complying with data protection regulations.

3.2.4.4 The problem of user consent

There may be cases where the user's explicit consent is not obtained during the data collection process. When AI analyzes and utilizes data, data security issues may arise if the user does not understand how their data is being used.

3.2.4.5 Ethical Responsibility

Companies or organizations that use AI technology have ethical responsibilities related to data security. They need to consider how they will be held accountable in the event of a data leak or misuse. In short, these issues are becoming more complex as Big Data analysis and AI technology advance, and ethical considerations are needed.

3.3 Bias problem

3.3.1 Employment discrimination

If the dataset that the AI model learns from is biased, unfair decisions or predictions can be made as a result.[6] If the AI model excessively reflects data from a specific

group, prejudice against that group can be strengthened. The bias of AI algorithms can cause various social problems, and can be thought of largely focusing on the effects related to racism and gender discrimination.[7]

If an AI-based recruitment system learns based on data biased against a specific race or gender, it can evaluate or exclude applicants from that group unfavorably. Therefore, this can potentially lead to an imbalance in opportunities. This bias problem, which can lead to employment discrimination, will mainly occur due to the following factors.

3.3.1.1 Data Bias

If the training data is imbalanced against certain groups (e.g. gender, race, age, etc.), AI models will make predictions based on this data, which may reinforce unfavorable judgments or evaluations of certain groups.

3.3.1.2 Outcome Bias

If past hiring decisions are already biased, models trained on this data will likely repeat the same bias. For example, if applicants of a certain race or gender were not selected in the past, the model is more likely to evaluate this group lower.

3.3.1.3 Feature Selection Bias

If an AI model emphasizes or ignores certain characteristics, groups related to those characteristics may be discriminated against. For example, if characteristics such as education or career are overemphasized, other important factors may be overlooked, which may disadvantage applicants from certain groups.

3.3.1.4 Opaqueness of the Decision Process

If the decision process of an AI model is opaque, employers may rely on the model's predictions, which may lead to unreasonable discrimination. This may result in unfair decisions made by simply accepting the model's results.

In short, to address these issues, it is important to build datasets that consider diversity and inclusiveness, regularly review the model's results, and apply algorithmic approaches to reduce bias. In addition, transparency of the AI's decision process should be increased.

3.3.2 Crime Prediction

If a crime prediction algorithm is based on historical crime data, it may reinforce bias toward certain races or regions. This could lead to excessive police surveillance or crackdowns on certain groups. If the dataset that the AI model learns from is biased, bias can appear in crime prediction in the following ways:

3.3.2.1 Data Bias

Crime data is based on past crime records for certain regions or groups. This can lead to over- or under-representation of groups with certain races, genders, or socioeconomic backgrounds. For example, if a certain area has a high crime rate due to intensive police crackdowns, it may incorrectly predict that residents of that area are more likely to commit crimes.

3.3.2.2 Outcome Bias

As an AI model learns past crime data, it repeats decisions that are already biased. For example, if a certain race has a high rate of crimes in the past, the model may evaluate that race as more likely to commit crimes based on this. This can exacerbate discrimination.

3.3.2.3 Overgeneralization

If an AI model becomes overly reliant on specific characteristics or patterns, it will ignore the characteristics of individual cases or individuals. This may lead to an unreasonably high evaluation of the likelihood of individuals belonging to a certain group committing a crime.

3.3.2.4 Opaqueness of the Decision Process

If it is unclear how the AI prediction results were derived, the reliability of the crime prediction may decrease and lead to incorrect judgments. This may lead law enforcement agencies to uncritically accept the results of the AI.

3.3.2.5 Social instability

Biased crime prediction models may lead to distrust and social conflicts toward certain groups. This may have a negative impact on crime prevention and social integration.

In short, to solve these problems, it is necessary to compose datasets in a diverse and comprehensive manner and apply various techniques to reduce the bias of the algorithm. In addition, it is important to increase transparency of the prediction results and ensure that

judgments are made considering the social context.

3.3.3 Credit Rating

If there is a bias against a specific race or gender in the credit rating process in which AI assigns credit scores, unfair loan terms or credit restrictions may occur.

3.3.3.1 Data Bias

Credit rating models learn based on various data such as past credit history and loan repayment history. If a group with a specific race, gender, or socioeconomic background is not sufficiently reflected or is evaluated negatively, the credit score of that group may be underestimated.

3.3.3.2 Outcome Bias

If past credit ratings are already biased, the AI model may learn based on this data and repeat the same bias. For example, if a specific group is not able to receive a loan or is charged a high interest rate, the credit score of that group may be continuously evaluated low.

3.3.3.3 Feature Selection Bias

Variables used in credit ratings may work against a specific group. For example, young people or immigrants without credit history may receive low credit scores because of their lack of credit history. This situation may limit their access to credit.

3.3.3.4 Opaqueness of the decision-making process

If the decision-making process of an AI model is opaque, it may be difficult to understand and appeal the credit rating results. This makes it difficult for consumers to complain or appeal their credit scores.

3.3.3.5 Social inequality

Biased credit rating models can disadvantage certain groups and exacerbate social inequality. This leads to discrimination in access to financial services, which limits economic opportunities.

In short, to solve these problems, it is necessary to build a dataset that reflects diverse population groups and monitor and correct bias in the algorithm. In addition, it is important to increase the transparency of credit ratings and establish a system that allows consumers to understand and appeal their credit scores.

3.3.4 Accessibility to Healthcare Services

If AI-based diagnostic systems lack data on a specific group, incorrect diagnoses or treatment methods may be selected, which may exacerbate health inequalities. If the datasets that AI models learn from are biased, this may have various negative effects on accessibility to healthcare services.

3.3.4.1 Data Bias

If medical data is collected disproportionately for a specific population group (e.g., gender, race, region), AI models learn based on this data. For example, if medical records for a specific race are lacking, the diagnosis or treatment outcomes for that race may be inaccurately evaluated.

3.3.4.2 Diagnosis and Treatment Bias

AI models may underestimate or misinterpret symptoms or diseases of a specific group. This may prevent patients of a specific race or gender from receiving the diagnosis or treatment they need. For example, heart disease symptoms in women may present differently than in men, and failure to reflect these differences may delay the diagnosis of female patients.

3.3.4.3 Inequity in the distribution of medical resources

If AI models underestimate the medical needs of certain groups, medical resources or services may be distributed insufficiently for those groups. This has a greater impact on patients from low-income or underprivileged communities.

3.3.4.4 Inequity in Access to Healthcare

Biased AI systems can impede access to healthcare services. If certain groups are discriminated against in AI-based healthcare services, they may not receive the necessary treatment or may be delayed.

3.3.4.5 Decreased Trust

If the decision-making process of an AI system is opaque, patients may lose trust in the system. This may lead to disregarding AI-based diagnoses or treatment recommendations, and weaken trust between healthcare providers and patients.

In short, to address these issues, it is important to build

a comprehensive dataset that reflects diverse populations and continuously monitor and correct model biases. In addition, policies and systems should be established to increase accessibility to healthcare services and ensure that all patients receive fair treatment.

3.3.5 Social Media and Information Filtering

When algorithms filter information based on personal interests, there is a possibility that the voices of certain groups will be underrepresented or distorted. This can lead to social conflict. Bias in social media and information filtering can manifest itself in the following ways:

3.3.5.1 Information Bias

When AI algorithms learn to prefer certain types of content or opinions, the information provided to users is limited. For example, if content with certain political views or social ideologies is overemphasized, users lose the opportunity to access other perspectives or information.

3.3.5.2 Filtering Effect

Information filtering on social media platforms is based on user preferences. If the learning data is biased, users will only be exposed to information that matches their own, which can lead to a 'mirror room of information' phenomenon, reducing exposure to diverse opinions or information.

3.3.5.3 Social Divide

If biased information about a certain group is continuously exposed, conflict or misunderstanding between that group and other groups can be exacerbated. This can lead to social division and undermine the integrity of the community.

3.3.5.4 Spread of misinformation

Biased algorithms can spread misinformation or conspiracy theories more. As information favorable to a certain group is emphasized, the risk of spreading untrue content to users increases.

3.3.5.5 Impact on identity formation

Information that users encounter on social media has a significant impact on their identity and value formation. Biased information can distort users' worldview, which can have a negative impact on their personal behavior or social attitudes.

In short, to solve these problems, it is necessary to use a comprehensive dataset that includes data from various sources and continuously monitor the bias of the algorithm. In addition, it is important to build a system that provides users with diverse perspectives and prevents the spread of misinformation through fact-checking. Social media platforms need to adjust their algorithms so that users can access diverse opinions and make efforts to maintain a balance of information. In addition, efforts are needed to increase the transparency and fairness of the algorithm and minimize bias by utilizing various data sources.

3.4 Lack of Transparency

The following problems can arise when the decision-making process of AI is opaque:[8]

3.4.1 Lack of Trust

If users do not understand the decision-making process or reasons of AI, their trust in the results may decrease. This can hinder users from adopting AI systems. The problem of lack of trust that can arise when the decision-making process of AI is opaque appears in various aspects. The main contents are as follows:

3.4.1.1 Difficulty in interpreting results

If users cannot understand the results of AI systems, their trust in the results decreases. For example, if the basis for AI's judgment is unclear in important decisions such as medical diagnosis or credit rating, users may have difficulty accepting the results.

3.4.1.2 Lack of Consistency in Decisions

When it is unclear how an AI model made a decision, its decisions in similar situations may be inconsistent. Users will lose trust in AI as they experience this inconsistency.

3.4.1.3 Concerns about bias and discrimination

If the decision-making process of AI is opaque, users may suspect that the model is biased or may make discriminatory decisions against certain groups. This is especially true in situations where discrimination based on race, gender, or age is a concern.

3.4.1.4 Uncertainty of responsibility

If AI makes a decision incorrectly, it may be unclear who is responsible for it. Users may not know how to respond when AI makes a wrong judgment, which may lead to distrust in the system.

3.4.1.5 Impaired user participation

Lack of trust prevents users from actively utilizing AI systems. For example, even if a company introduces AI-based decision-making, employees may be reluctant to participate in the process or provide opinions. This may result in hindering collaboration and innovation within the organization.

In short, to solve these problems, it is important to make the decision-making process of AI transparent and help users understand the process. For example, it is necessary to introduce Explainable AI (XAI) technology to clarify the reasons for decision-making and build trust with users.

3.4.2 Responsibility Issues

If the entity responsible for decisions made by AI becomes unclear, responsibility for wrong decisions may become ambiguous. This may lead to legal and ethical issues. The main issues of responsibility that may arise when the decision-making process of AI is unclear are as follows:

3.4.2.1 Uncertainty of Responsibility

If an AI system makes a wrong decision, it is unclear who is responsible. For example, if a medical AI makes a wrong diagnosis, there may be controversy over who should be held responsible among the medical provider, the AI developer, or the data provider.

3.4.2.2 Legal Responsibility Issues

There is a problem of how to determine legal responsibility for damages caused by AI decisions. Since the current legal system generally determines responsibility for human actions, it may be difficult to respond sufficiently with the existing legal framework in situations where AI is involved.

3.4.2.3 Lack of Transparency

If the decision-making process of AI is opaque, it is difficult for external parties to review or object to the decision. This will also be a major obstacle in the process of holding those responsible for wrong decisions accountable.

3.4.2.4 Decreased Trust

If responsibility is unclear, users and society will lose trust in AI systems. This may hinder the introduction of AI technology and hinder innovation and development.

3.4.2.5 Ethical Issues

When AI decisions affect human life or safety, the issue of responsibility becomes more serious. For example, ethical discussions will be needed on who should be held responsible when a self-driving car causes an accident.

In short, in order to solve these responsibility issues, it is important to increase the transparency of AI systems and develop legal and ethical frameworks that can clearly identify the location of responsibility. There is a need for a mechanism that can understand how AI's decision-making process takes place and clarify responsibility for it.

3.4.3 Difficulty in risk management

If the AI's decision-making process is opaque, it becomes difficult to predict and prevent system errors or failures in advance. This can cause safety-related problems. The difficulty in risk management that can occur when the AI's decision-making process is opaque appears in various aspects. The main contents can be said to be as follows.

3.4.3.1 Difficulty in risk identification

If it is unclear how the AI system makes decisions, it is difficult to identify potential risk factors. For example, if it is not known what data a specific algorithm uses to make decisions, it will be difficult to predict the negative consequences of that decision.

3.4.3.2 Absence of risk assessment

If the decision-making process is opaque, it is difficult to evaluate the results of AI's decisions. This will be a major obstacle for organizations or individuals to properly assess and manage risks that may arise from the use of AI systems.

3.4.3.3 Difficulty in establishing response strategies

An opaque AI decision-making process makes it difficult to establish an appropriate response strategy when a problem occurs. For example, if AI makes an incorrect recommendation, it will be difficult to devise effective corrections or improvement measures if the cause cannot be

identified.

3.4.3.4 Difficulty in building trust

If there is a lack of trust in AI systems, users or organizations will have doubts about whether the system can be used safely. This will make them hesitant to introduce AI, and ultimately hinder the development and use of technology.

3.4.3.5 Regulatory and compliance issues

If the decision-making process of AI is opaque, it may be difficult to comply with legal regulations or ethical standards. This may be a major risk factor for companies to assume legal responsibility.

3.4.3.6 Inefficiency in disaster management

When AI systems make decisions in disaster management or emergency situations, if the process is opaque, it may be difficult to respond quickly and effectively. This may increase casualties or property damage.

In short, to solve these difficulties in risk management, it is essential to increase the transparency of AI systems and strengthen the explainability of the decision-making process. In addition, it is very important to establish a risk management framework to establish a system that can identify and evaluate potential risks due to the use of AI in advance.

3.4.4 Conflict of Interest

If the decision-making process is opaque, there is a high possibility that companies or organizations will use AI to manipulate or make distorted decisions based on their interests. Conflicts of interest that may arise when the decision-making process of AI is opaque can appear in various aspects. The main contents are as follows.

3.4.4.1 Disagreement between stakeholders

If the decision made by the AI system does not meet the interests of specific stakeholders (e.g., companies, consumers, government, etc.), conflicts may arise. For example, if an AI model that prioritizes the interests of a company makes a decision that overlooks the safety or welfare of consumers, the interests of the two groups will conflict.

3.4.4.2 Lack of Trust Due to Lack of Transparency

If the decision-making process of AI is opaque, stakeholders may question the fairness or objectivity of the system. This may cause each stakeholder to oppose or distrust AI decisions in order to protect their own interests.

3.4.4.3 Conflict of Ethical Standards

If AI decisions conflict with the ethical standards or values of stakeholders, conflicts may arise. For example, if the method of data collection and utilization is judged to infringe on an individual's privacy, there may be a conflict of interest between the individual and the company.

3.4.4.4 Unfairness of Decision-making

If the AI system favors a specific stakeholder, other stakeholders will be disadvantaged. This may lead to dissatisfaction or conflict, which may negatively affect collaboration or communication within the organization.

3.4.4.5 Unclear Responsibility

If AI decision-making is unclear, conflicts may arise because it is unclear who is responsible for a wrong decision or result. Stakeholders may try to avoid responsibility or blame each other.

3.4.4.6 Confusion in policy and regulation

If AI decisions cause conflicts between stakeholders, it may be difficult for governments or regulatory agencies to establish appropriate policies. This may lead to social confusion or conflict, which may hinder the effectiveness of policy implementation.

In short, in order to resolve such conflicts of interest, it is important to make the decision-making process of AI systems transparent and establish a system for collecting opinions from stakeholders. In addition, AI development and operation that considers fairness and ethics are necessary, and efforts are required to promote communication and cooperation between stakeholders. In particular, it is important to make the decision-making process of AI transparent and develop a system that can be explained to users.

4. Solutions and Improvements

4.1 Legal Solutions[9]

4.1.1 Strengthening the Rights of Information Subjects

It is necessary to clarify the rights of individuals to request access to, correction, deletion, and restriction of processing of their data, and simplify the procedures to make it easier to exercise these rights.

4.1.1.1 Right of Access of Information Subjects

Information Subjects have the right to check how their personal data is being processed. This allows individuals to demand transparency regarding their data.

4.1.1.2 Right to Correction

Information Subjects have the right to have their personal data corrected if it is inaccurate or incomplete. This will play an important role in preventing damage caused by incorrect information.

4.1.1.3 Right to erasure(Right to be forgotten)

Information Subjects have the right to have their personal data deleted under certain conditions. This will greatly contribute to strengthening the protection of individuals' privacy.

4.1.1.4 Right to Restrict Processing

Information Subjects have the right to restrict the processing of their personal data. For example, they can request that the processing of their data be suspended while the accuracy of the data is verified.

4.1.1.5 Right to data portability

Information Subjects have the right to transfer their personal information to other service providers. This should strengthen the ownership of data and allow individuals to manage their data more freely.

4.1.1.6 Right to automated decision-making

Information Subjects have the right to object to decisions made in an automated manner. This is important for ensuring fair processing.

In conclusion, these rights will greatly help Information Subjects to strengthen their control over their personal information and prevent unfairness in the data processing process. In addition, these provisions of the Personal Information Protection Act contribute to respecting individual rights and clarifying corporate responsibilities, thereby increasing the transparency of data use.

4.1.2 Increased transparency of data processing

Legal obligations should be imposed on companies and organizations to increase transparency in the way they collect and process personal information, so that Information Subjects can clearly understand the use of their data. Increased transparency of data processing is an important element of personal information protection and data use, allowing Information Subjects to understand how their personal information is collected, processed, and stored.

4.1.2.1 Clear Privacy Policy

Companies or organizations should write their privacy policy in clear and easily understandable language. This policy should include the purpose of data collection, processing method, retention period, and whether or not data will be provided to third parties.

4.1.2.2 Obligation to provide information

When collecting data, necessary information should be provided to the Information Subject. This should be made clear at the time of data collection, and users should be able to know how their data will be used.

4.1.2.3 Clarity of consent

Consent for the collection and processing of personal data should be voluntary and clear, and users should be provided with an easy way to withdraw consent. This should help users exercise their rights.

4.1.2.4 Recording of data processing history

Companies should record and manage the details of personal data processing. This can be used to prove the lawfulness of data processing, and the information should be available to the Information Subject upon request.

4.1.2.5 Regular audits and reviews

Regular audits and reviews of the data processing process should be conducted to check compliance with the privacy policy. This will contribute to increasing the transparency of

the organization.

4.1.2.6 Education and awareness raising

It is important to provide personal information protection education to employees to raise awareness of the importance of data processing and related laws. This will also help create a transparent data processing culture internally.

In conclusion, these measures will contribute to individuals feeling a sense of control over their own data and building trust in the data processing process. Therefore, transparent data processing will play an important role in increasing the reliability of the company and fulfilling social responsibility for personal information protection.

4.1.3 Strict Consent Request

Consent for the collection and processing of personal information should be more strictly required, and regulations should be put in place to prevent data collection without prior consent. Strict consent requirement is a measure to require clear and specific consent before collecting, processing, and using personal information. This aims to ensure that individuals fully understand and agree to how their information will be used.

4.1.3.1 Legal Basis

In many countries, data protection laws require that companies or organizations obtain explicit consent to collect personal data. For example, the General Data Protection Regulation (GDPR) strictly requires consent from Information Subjects, and requires that consent be freely given, specific, unambiguous, and informed.[10]

4.1.3.2 Requirements for Consent

Consent should be clearly stated in understandable language, and individuals should be able to clearly understand what they are consenting to. Individuals should also be able to give consent freely, without coercion or disadvantage, and they should have the right to withdraw their consent.

4.1.3.3 Specificity

Consent should only be given for data collected for specific purposes, and broad consent is not permitted.

4.1.3.4 Impact

Such strict consent requirements will require companies to take a more transparent and responsible approach to data collection and processing. In addition, individuals will be able to better understand and exercise their rights regarding their data.

In conclusion, strict consent requirements are an important measure to strengthen personal information protection, contributing to protecting individual rights and increasing transparency in data use.

4.1.4 Strengthening penalties for data breaches

Legal penalties for personal information leaks or breaches should be strengthened to encourage companies or organizations to take a more responsible approach to data protection. A data breach is a situation where an individual's personal information is accessed, leaked, altered, or destroyed without authorization. In the event of such a breach, it is necessary to strengthen the penalties for companies or organizations according to relevant laws.

4.1.4.1 Legal Basis

Many countries' personal information protection laws specify penalties for data breaches. For example, under GDPR, companies can face significant financial penalties in the event of a data breach, with fines of up to 20 million EUR or 4% of annual global turnover, whichever is higher.

4.1.4.2 Financial fines

In addition to compensation for damages caused by a data breach, high fines may be imposed depending on the legal standard.

4.1.4.3 Criminal penalties

In the case of a serious data breach, criminal penalties may be imposed on the person responsible. This may include imprisonment or fines.

4.1.4.4 Administrative sanctions

Administrative sanctions such as suspension of operations or restrictions on data processing may be imposed on companies that have committed a data breach.

4.1.4.5 Impacts

Strengthening penalties will make companies more mindful of data protection and strengthen their responsibility for protecting personal information. In addition, individuals

will be able to have greater confidence that their information is protected more safely.

In conclusion, strengthening the punishment for data breaches is an important measure to emphasize the importance of personal information protection and strengthen corporate responsibility. This will ultimately contribute to protecting individual rights and making the data environment safe.

4.1.5 Mandatory Appointment of Data Protection Officer

Companies of a certain size or larger should be required to appoint a data protection officer to ensure a professional management system for personal information protection. One measure to strengthen the Personal Information Protection Act and legal regulations on data use is the mandatory appointment of a Data Protection Officer (DPO). This measure includes several important aspects.

4.1.5.1 Strengthening Responsibility

By designating a Data Protection Officer, responsibility for personal information protection can be clearly established within the organization. The DPO is responsible for supervising data processing activities and leading the organization to comply with personal information protection-related laws and regulations.

4.1.5.2 Legal Compliance Support

The DPO understands laws and regulations related to personal information protection laws and helps the organization comply with them. This should reduce legal risks and provide a foundation for responding quickly when legal issues arise.

4.1.5.3 Regular training and awareness raising

The DPO is responsible for educating employees within the organization on personal data protection and making them aware of the importance of personal data protection. This will help establish a personal data protection culture throughout the organization.

4.1.5.4 Monitoring data processing activities

The DPO monitors the organization's data processing activities and checks whether the data protection policy is being properly implemented. This will help prevent data leaks or personal data breaches in advance.

4.1.5.5 Communication and cooperation

The DPO plays a communication role with external stakeholders (e.g. regulators, customers, etc.). When questions or issues related to data protection arise, the DPO can act as a mediator, thereby building trust.

4.1.5.6 Risk assessment and management

The DPO plays a key role in assessing the risks associated with data processing and establishing measures to manage them. This will help increase the effectiveness of data protection.

4.1.5.7 Increased Transparency

The presence of a DPO can increase transparency about an organization's data processing activities and provide customers and users with confidence that their data is being protected. This can help convey the message that their personal data is being processed safely.

For these reasons, the mandatory designation of a Data Protection Officer can play an important role in strengthening personal data protection and ensuring compliance with legal regulations. This can ultimately contribute to protecting the rights of individuals and increasing the reliability of data processing.

4.1.6 Strengthening International Cooperation

In order to respond to global data flows, it is necessary to establish international data protection standards through cooperation with other countries and ensure compliance with them.

4.1.6.1 Setting Global Standards

Since personal data protection laws and regulations differ in various countries and regions, it is important to set international standards to harmonize the laws of each country. This should facilitate the flow of data across borders and help companies comply with legal requirements in various markets.

4.1.6.2 Information sharing and best practice exchange

International cooperation can be used to share information and cases related to personal information

protection. By cooperating with regulatory agencies, companies, and researchers from each country to exchange effective data protection methods and best practices, the level of personal information protection can be raised worldwide.

4.1.6.3 Establishing a legal cooperation system

It is necessary to establish a legal cooperation system that can effectively respond internationally when a data leak or personal information infringement incident occurs. This can enable a rapid response through legal support and information exchange between countries.

4.1.6.4 Concluding bilateral and multilateral agreements

By concluding bilateral or multilateral agreements for personal information protection, it is possible to unify personal information protection standards among countries and establish regulations for data movement. Such agreements provide an internationally applicable legal framework to strengthen data protection.

4.1.6.5 Forming a network among regulatory agencies

It is important to form a network among personal information protection regulatory agencies from each country to promote mutual cooperation and information exchange. This will enable regulators in each country to share their experiences and work together to address common challenges.

4.1.6.6 Education and Awareness Raising

International cooperation can help raise awareness of privacy protection at a global level by developing education programs on privacy protection and providing them to stakeholders in each country.

4.1.6.7 Technical Cooperation

Joint research and development of data protection technologies and solutions can increase the effectiveness of data protection and promote innovation. This will contribute to strengthening privacy protection along with technological advancements.

In conclusion, these measures to strengthen international cooperation can play an important role in strengthening privacy protection, increasing consistency, ensuring the safe flow of data, and protecting the rights of individuals in a global society.

4.2 Technical Solutions

4.2.1 Explainable AI

This is a technology that allows us to understand the decision-making process of an AI model. With the discovery of new explainable technologies and the establishment of standards for AI technology, solutions can be sought in terms of Artificial Intelligence bias, but efforts to preemptively respond to them from an ethical point of view are needed. Here, Explainable AI (XAI) is a technical solution that allows us to understand and interpret the decision-making process of an AI model.

4.2.1.1 Model Interpretation Method

LIME (Local Interpretable Model-agnostic Explanations) is used to explain the output of a model for a specific prediction. It explains the local behavior of the model by perturbing sample data. SHAP (SHapley Additive exPlanations) numerically evaluates the impact of each feature on the model prediction and visualizes the contribution of each feature.

4.2.1.2 Model Simplification

Instead of complex models, interpretable models such as decision trees and linear regression should be used to make the results easier to understand.

4.2.1.3 Visualization tools

Visualize the decision boundaries, feature importance, etc. of the model to help users intuitively understand the results.

4.2.1.4 Human-centered design

Design a way to explain the results considering the user's level of understanding. For example, it should be explained in terms that non-experts can understand.

4.2.1.5 Feedback loops

By allowing users to provide feedback on the model's predictions, the model's interpretability can be increased and continuously improved. In short, these methods can play an important role in making the decision process and results of the AI model clearer and building a system that users can trust.

4.2.2 Model interpretation tools

Tools that visualize the internal structure of the model or evaluate the importance of the characteristics can be used to analyze how much any input affects the results. Model interpretation tools are technical methods that help users understand and explain the prediction results of an AI model. These tools play an important role in helping users understand the internal workings of the model and increase confidence in the results. The main model interpretation tools include the following:

4.2.2.1 LIME (Local Interpretable Model-agnostic Explanations)

LIME is a method that explains the output of a model for a specific prediction. In order to understand the prediction of the model, samples are generated around the original data points and the model is trained on these samples. Through this, it will be possible to visually represent the effect of each characteristic on the prediction.

4.2.2.2 SHAP (SHapley Additive exPlanations)

SHAP is a method based on game theory that numerically evaluates the degree to which each feature contributes to the prediction of the model. This allows us to quantitatively explain the impact of each feature on the prediction results, and it can be intuitively understood when used with visualization tools.

4.2.2.3 Feature Importance

This is a method to evaluate how important each feature of the model is to the prediction. In random forests or boosting models, characteristic importance can be calculated to easily identify the most influential characteristics.

4.2.2.4 PDP (Partial Dependence Plots)

This is a graph that visually shows the average impact of a specific feature on the model prediction. It can analyze multiple features simultaneously, so it is useful for understanding the interaction between features.

4.2.2.5 ICE (Individual Conditional Expectation) Plots

Similar to PDP, but it shows the prediction change for each individual data point. This will allow for a more detailed analysis of the impact of certain characteristics on model prediction.

In short, these model interpretation tools play an important role in increasing the transparency of the algorithm and enhancing user trust in the AI system. These tools will help users better understand the model's prediction results and improve the model as needed.

4.2.3 Interactive Visualization

Interactive visualization tools can be developed to help users easily understand how the algorithm works and the results by visually expressing them. Interactive visualization is a technical method that plays an important role in increasing the transparency of the algorithm. This method visually expresses data and allows users to directly interact with it, making it easy to understand the information. The main features and advantages of interactive visualization are as follows:

4.2.3.1 Intuitive data representation

It helps users easily understand complex data and model results by visualizing them in graphs, charts, maps, etc. Visual elements clearly reveal patterns and relationships in the information.

4.2.3.2 User interaction

It allows users to select or filter specific data points, allowing them to conduct in-depth analysis on the areas of interest. For example, users can adjust variable values using sliders and check the changes in results in real time.

4.2.3.3 Real-time feedback

Interactive visualization provides immediate feedback to help users understand the model's behavior. The visualization changes dynamically according to the values entered by the user, which allows the user to intuitively experience the model's prediction process.

4.2.3.4 Analysis of various scenarios

Users can simulate various scenarios and compare the results for each scenario. This will allow them to discover the strengths and weaknesses of the model and gain insights necessary for decision-making.

4.2.3.5 Improved comprehension

Interactive visualizations are designed to be easily

understood by non-experts, making them useful for explaining the complex operating principles of AI models. User-friendly interfaces can reduce the learning curve and contribute to building trust in the model.

In short, such interactive visualizations are important tools for increasing the transparency of the algorithm and enhancing communication with users, thereby increasing trust in AI systems. Through this tool, users will be able to understand the relationship between data and models more deeply and make better decisions.

4.2.4 Simplification of algorithms

Using simple models that are easy to interpret instead of complex models can increase transparency. Simplification of algorithms is one of the important technical methods to increase transparency of AI models. This approach should reduce the complexity of the model to make it easier to understand, and as a result, make it easier for users to understand how the model works.

4.2.4.1 Simple model selection

Use interpretable models such as decision trees, linear regression, and logistic regression instead of complex models (e.g., Deep Learning). These models are relatively intuitive, and it will be easy to understand the effect of each feature on the results.

4.2.4.2 Feature selection

This is the process of reducing the number of features (variables) used in the model. Simplifying the model by removing unimportant or redundant features will make the results easier to interpret and reduce the problem of overfitting.

4.2.4.3 Simplification of model structure

Models with complex structures can be difficult to interpret. For example, using a neural network with a small number of layers and neurons is a good example. Keep the model structure simple, and the prediction results will be better understood.

4.2.4.4 Rule-based models

Using a rule-based approach (e.g., decision trees) can clearly explain how each prediction was made. Such models will provide intuitively understandable rules, which will increase the transparency of the predictions.

4.2.4.5 Leveraging model explanation tools

Even in simplified models, model interpretation tools such as LIME and SHAP can be used to explain the impact of each feature on the results. These tools provide additional interpretation even in simple models, helping users better understand the results.

4.2.4.6 Feedback loops

User feedback can be used to continuously improve and simplify the model. By adjusting the model based on the user's understanding, transparency can be further increased.

In short, simplifying the algorithm plays an important role in increasing the interpretability of the model and enabling users to trust the model's prediction results. This will enhance the transparency of the AI system and support better decision-making.

4.2.5 Model Validation and Audit

The process of verifying and auditing the results of an algorithm will ensure that the decisions of the algorithm are fair, consistent, and understandable. Model validation and auditing are important technical methods to increase the transparency of the algorithm and are essential to ensuring the reliability and accuracy of the AI model. This process helps to ensure that the model is operating correctly and that the results are fair. The main elements of model validation and auditing are as follows:

4.2.5.1 Model Validation

Performance Evaluation is to evaluate how well the model works based on real data. Typically, metrics such as cross-validation, accuracy, precision, recall, and F1 score are used to measure the performance of the model. Use of Test Data is to evaluate the generalization ability of the model using separate test data for validation. This is important to ensure that the model is not overfitted to the training data.

4.2.5.2 Fairness Check

It evaluates whether the prediction results of the model are biased against a specific group. For example, this is a process to ensure that the results are not unfairly displayed based on gender, race, age, etc. This will enable us to fulfill our social responsibility and build ethical AI systems.

4.2.5.3 Interpretability Assessment

It analyzes the characteristics on which the model's predictions are based so that the results can be explained. Tools such as LIME or SHAP should be used to quantitatively evaluate the impact of each characteristic on the model's predictions.

4.2.5.4 Audit Process

It is to document all decisions and results during the model development and operation process so that they can be reviewed later. This increases the transparency of the model and helps stakeholders clearly understand how the model works. Regular audits ensure that the model continues to be reliable. This helps the model adapt to changing environments and prevent performance degradation.

4.2.5.5 Feedback Mechanism

This is a method to increase the reliability and understandability of the results by building a system that receives feedback from users or stakeholders on the results of the algorithm and continuously improves the model. This will allow the model to understand how it works in real situations and adjust it as needed.

In short, model validation and auditing play an important role in increasing the transparency of the algorithm and building trust in the AI system. This will allow users to trust the model's prediction results and provide clearer grounds for decision-making.

4.2.6 Generating Understandable Explanations

Utilizing technology to generate natural language explanations for the results of the algorithm, the results are delivered so that even non-experts can understand them. Generating understandable explanations is an important technical method to increase the transparency of the algorithm, and focuses on delivering the prediction results of the AI model to users in a clear and intuitive manner. This approach helps users understand how the model works and trust the results. The main elements are as follows:

4.2.6.1 Natural Language Explanations

This is a method to explain the model's prediction results in natural language. For example, for a specific prediction requested by a user, "This customer is a low credit risk. This is because they have a good previous payment history and a high income level." The results are explained in a way that is simple and clear so that even non-experts can understand them.

4.2.6.2 Visual Explanations

Using data visualization techniques, information related to the model's predictions are expressed graphically. For example, feature importance is visualized in a bar chart, or decision boundaries are visually represented to help users easily understand the results.

4.2.6.3 Case-based explanation

This is a method of explaining by presenting similar cases based on a specific prediction. For example, "Customer A with similar characteristics was approved for a loan, and Customer B was rejected." In this way, the validity of the prediction should be strengthened through real cases.

4.2.6.4 Explaining the working principle of the model

Simple explanations are provided so that users can understand how the model makes decisions. For example, "This model evaluates based on age, income, and credit score." Provide explanations that connect the inputs and outputs of the model.

4.2.6.5 Adjusting the level of explanation

Provide customized explanations to users, adjusting the depth of the explanation to the level that each user can understand. Provide more technical and detailed information for experts, and simple and intuitive explanations for general users.

4.2.6.6 Feedback and improvement

Continuously improve the explanation method based on feedback received from users. Analyze which explanations are easy to understand and which parts are confusing so that better explanations can be created.

In short, generating understandable explanations plays a vital role in making the results of AI models transparent and building trust with users. This will allow users to better understand the model's predictions and make rational decisions based on them. These methods help users and stakeholders have better trust by making the algorithm's operation transparent.

4.3 Improvement tasks such as setting

ethical standards

4.3.1 Transparency

The operation method and decision-making process of the AI system should be clearly explained and should be understandable to users.

4.3.1.1 Understandability of algorithms

A clear explanation is needed of how the AI system works and what data and algorithms are used. Users should be able to understand the decision-making process and results of the AI, and this will enable them to build trust.

4.3.1.2 Disclosure of data sources

By disclosing the source and characteristics of the data that the AI model learns, the quality and bias of the data should be assessed. This will help increase the reliability of the data and prevent unfair decisions or predictions.

4.3.1.3 Interpretability of results

An explanation should be provided for the decisions or predictions made by the AI system. Users should be able to understand how the results were derived, and this will enable them to increase their trust in the AI.

4.3.1.4 Accountability

Responsibility for the decisions or actions of the AI should be clearly defined. When transparency is guaranteed, it becomes clear who should be held accountable for wrong decisions or results, which promotes ethical use.

4.3.1.5 User Participation

In the development process of AI systems, stakeholders and users should be provided with opportunities to participate and provide opinions. This helps increase trust in the system and reflect diverse perspectives.

In conclusion, transparency is an essential element for ethical use of AI, contributing to building trust and creating fair AI systems. This will increase the acceptance of users and society as a whole.

4.3.2 Fairness

Algorithms should be designed to be non-discriminatory and ensure that all users are treated equally and fairly. This is important to minimize data bias. The following are the main aspects of fairness:

4.3.2.1 Bias Removal

If the data that AI models learn contains bias against a specific group, unfair decisions or predictions may result. In order to secure fairness, it is necessary to increase the diversity and representativeness of the dataset and identify and correct biased data.

4.3.2.2 Equal Access

AI technology should be equally accessible to all users. It is important to ensure that certain groups or individuals do not have access to the technology or are discriminated against. For this, design and policies that take accessibility into account are necessary.

4.3.2.3 Fair Results

Decisions made by AI systems should not be disadvantageous to certain groups. For example, AI should be designed not to discriminate based on race, gender, age, etc. during loan screening or hiring processes.

4.3.2.4 Continuous Monitoring

Continuous monitoring and evaluation are necessary to maintain the fairness of AI systems. Since data and environments may change over time, the system should be regularly inspected and modified when necessary.

4.3.2.5 Stakeholder Participation

Efforts to improve fairness are necessary by reflecting the opinions of various stakeholders (users, experts, social groups, etc.). This will allow for consideration of various perspectives and more comprehensive decision-making. Among the attributes of individuals, attributes that should not cause discrimination, such as gender, race, age, and religion, should be defined as sensitive attributes, and conditions should be specified that the judgments made by AI models should not change significantly depending on the sensitive attributes.

In this way, the final decision made by AI for a given individual can be used to approve or reject a loan, and the risk or stability of crime can be predicted by AI. In addition, the predicted credit score or predicted risk of recidivism can be determined through the score predicted by the AI model for a given individual. In addition, the actual outcome for a given individual can be determined through whether the loan is actually repaid or whether the individual actually commits a recidivism.

In conclusion, fairness is an essential ethical standard for AI technology to have a positive impact on society and provide equal opportunities to all users. By building and operating a fair AI system, trust and social acceptance can be increased.

4.3.3 Responsibility

Responsibility for decisions or actions of AI should be clarified, so that it is clear who is responsible when a problem occurs. This responsibility means taking clear responsibility for decisions or actions made by AI systems during the development and use of AI.

4.3.3.1 Clarification of Responsibility

It should be clarified who is responsible for the results of AI systems. There may be various entities such as developers, users, and companies, and the roles and responsibilities of each entity should be specifically identified.

4.3.3.2 Transparent explanation of results

When an AI system makes a specific decision, it should be able to explain the process and results. It is considered an important factor in increasing accountability to enable users to understand the decision-making process of AI.

4.3.3.3 Ethical Considerations

When an AI system makes decisions that affect human lives (e.g. medical diagnosis, legal judgment, etc.), it should comply with ethical standards. This is necessary to ensure that the system can make the right decisions.

4.3.3.4 Response when problems occur

A system must be established to respond quickly and effectively when an AI system makes a wrong decision or a problem occurs. This will minimize damage and demonstrate an attitude of acceptance of responsibility.

4.3.3.5 Continuous improvement

In order to strengthen accountability, the performance and results of the AI system must be continuously monitored and, if necessary, improved. This will contribute to increasing the reliability of the system and meeting the expectations of users and society.

In this way, efforts should be made to minimize possible damage by establishing a responsible entity in the process of developing and utilizing AI. To this end, the responsibilities among AI designers and developers, service providers, and users must be clearly defined.[11]

In conclusion, accountability is an essential ethical standard for AI technology to have a positive impact on society. Responsible AI development and utilization can build trust and increase acceptance by users and society.

4.3.4 Privacy protection

Privacy protection is one of the most important ethical standards when developing and utilizing AI. The collection and processing of user data must comply with the Personal Information Protection Act and be based on the user's consent. This means respecting the privacy of individuals, minimizing misuse of personal information, and ensuring safe processing.[12]

Only the minimum data necessary for the operation of the AI system should be collected and processed, and unnecessary data collection should be avoided. In addition, clear consent should be obtained from users for the collection and use of personal data, and users should have the right to manage their own data at any time. Collected data should be safely stored and processed, and protected from illegal external access or leakage.

This requires technical measures such as encryption and access control. Meanwhile, information on how data is collected and used should be clearly provided to users, and trust should be built through explanations of the data processing process. Users should have the right to request access to, correction, and deletion of their data, and such requests should be processed in a timely manner.

In conclusion, by complying with these privacy protection standards, AI systems should be able to process personal data safely and ethically, and can contribute to building trust between users and society.

4.3.5 Safety

AI systems should be designed to be safe and have mechanisms to prevent unexpected results or accidents. Safety is ensuring that AI systems operate in a predictable manner and do not cause harm to users and society. In April 2019, the European Commission (EC) listed technical 'safety' as one of the requirements for developing and using trustworthy AI.

AI systems should operate consistently and provide reliable results even in exceptional circumstances. Potential risks of AI should be identified in advance and measures should be taken to minimize them. In addition, mechanisms should be in place to safely terminate or resolve issues in

case the system behaves in an unexpected manner. The AI should be transparent so that users can understand how it operates and the decision-making process, and users should be educated about the limitations of the AI system and how to use it safely, thereby promoting proper use. In conclusion, by complying with these safety standards, AI systems will be able to have a positive impact on society.

4.4 Ethical Dilemmas and Challenges of Artificial Intelligence

4.4.1 Introduction

The use of services or devices equipped with Artificial Intelligence is becoming increasingly familiar, and the scope of interaction between digital devices and humans is continuously expanding. This trend shows that the prediction that various entities such as Artificial Intelligence robots will appear in the post-human era is no longer just science fiction. The question of how to respond to ethical issues that arise when humans interact with new entities other than humans will be in a transitional position that goes beyond the level of extending existing ethics to new entities and moving toward presenting new ethics.

However, even if the development of Artificial Intelligence is useful in our society, if it is inevitable to provide relief to victims, it is limited to respond with the current legal system, such as the tort system, and ethical issues will arise accordingly. Therefore, this paper will examine the ethical issues necessary to establish an ethical and social system focused on responsibility for Artificial Intelligence.

4.4.2 Development of Artificial Intelligence and Ethical Dilemmas

4.4.2.1 Floridi's Information Ethics

The technological development brought about by the revolution of Artificial Intelligence has a duality that protects human life while also having the potential risk of violating legal interests. This duality can provide an opportunity for theoretical discourse that can establish a rational and moral criminal responsibility system depending on how the ethical dilemma of Artificial Intelligence is resolved.

In this way, ethical issues related to Artificial Intelligence were raised in earnest by Floridi (Luciano Floridi). Here, Floridi proposed information ethics, that is, macroethics centered on the passive and the existential, as this transitional ethics. Floridi attempted to move beyond the anthropocentric perspective and grant moral status to the existence of information itself and to construct moral principles from the perspective of information. In this way, naturalistic attempts can be seen as an essential process in the discussion of dehumanization. This is because, in order to overcome anthropocentrism, it is necessary to derive moral evaluations from nature, which is distant from the human perspective.[13]

Based on this complex network of relationships, Floridi proposed a new concept called 'distributed moral action'. Morally important results cannot be reduced to the morally important actions of some individuals, but must be distributed among multiple actors. In this case, the question of who should be responsible for distributed moral actions and how much remains, but the argument that ethical responsibility should be attributed to all distributed actors is very meaningful in the era of Artificial Intelligence.[14]

4.4.2.2 Ethical dilemmas and guilt issues

Programming ethical algorithms for autonomous vehicles is based on research on the relationship between the most fundamental human value system and industrial technology. Recently, car manufacturers have been struggling with ethical dilemmas. For example, Google is focusing on pedestrians, who are vulnerable, and Mercedes-Benz has stated that it will focus on protecting humans in vehicles by evaluating the survival probability of drivers and passengers highly. Setting different standards like this can be quite problematic in assessing criminal responsibility. In addition, since it is impossible to predict all of the judgments of Artificial Intelligence, the question of how to do ethical programming will ultimately be a difficult one.

In particular, one of the issues that may be problematic in relation to self-driving cars is the part related to the Trolley Dilemma of Artificial Intelligence.[15] If ethical programming is instilled in advance to infringe on other legal interests in order to avoid illegal results, the responsibility of the programmer or manufacturer behind the scenes becomes an issue. For example, if the Artificial Intelligence installed in a self-driving car saves the life of the driver and crashes into a pedestrian, resulting in a punishable result, it will be necessary to examine whether the programmer or manufacturer who implemented ethical programming can be held liable for intentional injury or murder.

First, it shall be examined whether it can be considered an act out of Necessity. Necessity can be exempted from illegality within a reasonable scope in order to protect a superior legal interest of higher value than oneself, but as

explained above, the case of the Trolley Dilemma of self-driving cars can be said to be an issue in which equally valuable legal interests are at issue. Therefore, in cases of equal legal interests or when it is difficult to determine legal interests, there is no possibility of expectation, so it can be said that there is a possibility of responsibility excuse.

Second, the perpetrators did not intentionally program the driver or pedestrian to imagine or kill them. For this reason, it is necessary to determine whether the perpetrators had a duty to protect the life or body of the driver or pedestrian, or whether the crime of intentional omission can be established.[16]

In short, the dilemma itself means that it is impossible to program that can simultaneously structure the legal interests of both the driver and the pedestrian. Accordingly, there is a possibility that the illegality can be excused due to a conflict of duties, and furthermore, if the perpetrators inevitably loaded the AI with ethical programming in a realistically conflicting situation, it is believed that there is no possibility of expectation, so responsibility can be excused.

4.4.2.3 Review of ethical dilemmas by field

① Operation service field

It is very important to strike a balance between judgment based on AI technology and human judgment, and the balance and the relationship between the two sides are expected to change in the future, and accordingly, ethics are expected to change. Ethical review is necessary when emotions, beliefs, and behaviors are manipulated without the user's knowledge due to services provided by AI technology, or when priorities or selection are required.

It is necessary to examine how human relationships will change as AI technology expands human cognition and behavior. Other visions should be reconsidered, such as the acceptance and change of creative values in cooperation with Artificial Intelligence technology and the way each person relates to Artificial Intelligence technology. Here, the Trolley Dilemma is an experiment on what criteria should be used to make judgments if AI algorithms directly affect human life, and whether such judgments can be left to the algorithm. It is one of the ethical tests called the 'Trolley Dilemma' first proposed by British ethical philosopher Philippa Foot in 1967.[17]

The technology of self-driving cars is not much different from that of humans driving, in that a single vehicle perceives, controls, and drives on its own. In the case of the Trolley Dilemma, self-driving cars may be able to make decisions faster than humans, but this will not solve the fundamental problem.[18]

In order for self-driving cars to become popular, there is an ethical dilemma that must be resolved first. In an extreme situation where loss of life cannot be avoided, who should the self-driving algorithm be designed to sacrifice? If the self-driving algorithm is programmed to save 10 people in a crowd first, there will be few consumers who will readily purchase a car that cannot protect the owner in an extreme situation.

On the other hand, if the self-driving algorithm is programmed to save the owner first, utilitarians may focus their moral criticism on the car manufacturer and the owner who intentionally chose the manufacturer. Of course, it will not be easy to simply determine right or wrong in any choice. However, before millions of self-driving cars are fully deployed on the road, there should be sufficient social consensus on this ethical dilemma.

② Artwork Creation field

When Artificial Intelligence that has learned from past data accumulated by humans creates new works, the social evaluation of the skills and abilities of masters and craftsmen that have been highly regarded so far will change.

In the current situation where the purpose and target of using and utilizing Artificial Intelligence can be set by humans, the question arises as to whether the objects that can be manufactured using Artificial Intelligence should be explicitly limited. There is a possibility that it will become impossible to distinguish authenticity through Artificial Intelligence, and it will take a lot of effort and money to prove that it is a genuine work created by a human. If it is a work created by a human, it may be impressive, but if it is discovered that it is a work created using Artificial Intelligence, people will doubt that the Artificial Intelligence is impressing them, which will also raise ethical issues.

③ Medical diagnosis field

As the use and utilization of AI technology makes it possible to accurately estimate health conditions, it will be possible to predict diseases or possibilities that can occur before getting sick, thereby preventing illness. On the other hand, if the future of oneself feels too negative, one's hope for life may decrease.

In addition, if the disease-related prediction diagnosis by AI is too accurate, there may be fewer opportunities to have hope, take risks, or challenge life. Therefore, medical ethics issues such as the patient's right to know the results of the diagnosis, the right not to want to know, and the doctor's duty to convey them should be reestablished.

④ Conversation and communication field

In the case of problems caused by maliciously written conversational agents, conversational agents can have natural conversations with people using natural language and facial expressions. However, ethical issues such as how to ensure ethics to prevent negative effects on people and whether it is an insult to human dignity for AI that is indistinguishable from humans to communicate with humans by pretending to be humans will arise.

There may also be ethical issues raised, such as whether AI should always clearly state that it is AI, or to what extent it is acceptable to use AI to influence or manipulate human minds and emotions, or to use AI in dating businesses.

4.4.3 Suggestions

Intentional discrimination can be created by AI algorithms. However, as AI moves from the era of knowledge to the era of data, the problem of discrimination or distortion by data can become more serious. For example, if the creator or user of data used for Deep Learning is mostly from advanced countries or is male, unexpected distortions can occur.[19]

In addition, in order to solve these problems, I would like to suggest the establishment of a system such as the so-called “AI Ethics Judgment Standards Committee” to regularly check developers for ethics and the possibility of distortion and discrimination in software and data, and to fairly evaluate and analyze it with external experts.

Furthermore, with regard to the legal system, the appropriate distribution of risks and the subject of responsibility should be clearly defined. The National Information Technology Basic Act should be comprehensively revised to include design obligations such as the obligation to prevent service malfunctions for developers or users of AI technology and the “kill switch” system that can be forcibly terminated from the outside.

Here, the ‘Kill Switch’ is a modified punishment method specific to similar forms of AI such as death penalty or imprisonment, that is, it can completely delete AI program data,[20] or ultimately cut off the energy switch, which is practically equivalent to the death penalty by sentencing, and it can also take initialization measures such as reprogramming the AI program like imprisonment. Moreover, it can cause functional disabilities to the robot by taking various programming measures for the body of the AI itself.[21]

Meanwhile, various sanctions such as community service, such as requiring the robot to perform individual service activities for the community free of charge, can be considered.

4.4.4 Discussion

Ethical programming can be largely divided into which priority is given in the conflict between the driver’s right to life and the pedestrian’s right to life. For example, in the case of a self-driving car that enforces utilitarian programming, if there are many pedestrians, the AI will choose to place greater value on protecting the many pedestrians than on protecting the passengers. In that case, there is a risk that the driver himself may suffer the consequences of infringement of legal interests at any time. It is possible to show a contradictory attitude of supporting the utilitarian view while opposing riding in a utilitarian autonomous vehicle. In this respect, a view that supports programming that values the driver’s right to life, unlike utilitarianism, can be derived.[22]

The technological development of Artificial Intelligence should fundamentally minimize the conflict between humans and Artificial Intelligence. More ethical considerations about Artificial Intelligence are needed, and such considerations should be carried out in the harmonious development of related industries and human life. The case of the autonomous vehicle explained above clearly shows the conflict between humans and Artificial Intelligence and the ethical dilemma of Artificial Intelligence. In other words, what choice will the Artificial Intelligence make when an emergency situation such as a brake failure occurs? In a way, the technical problem is directly connected to the ethical or moral problem. Therefore, it will be necessary to establish some standards or norms and legislate them.

In the case of the United States, in January 2017, AI researchers, lawyers, and ethicists gathered in Asilomar, California, USA, and announced the ‘Asilomar AI Principles’, consisting of a total of 23 articles that are declarative norms on the purpose of AI development, ethics and values, mid-term issues, and ethical issues such as AI safety, job replacement, legal responsibility, and securing human uniqueness. In Japan, the Society for Artificial Intelligence established ethics guidelines, and Canada, France, and other countries are also establishing related regulations at the government level.

In relation to this, Isaac Asimov announced the Three Laws of Robotics in the short story ‘Runaround’ in his book ‘I, Robot’ in 1942. The first law is that a robot must not injure a human being or ignore a situation where harm would be done to a human being. The second law is that

a robot must obey human orders unless it would conflict with the principles. The third principle is that robots must protect themselves as long as it does not violate the first and second principles mentioned above. Therefore, I believe that it is necessary to standardize this principle as a minimum rule.[23]

In the case of Korea, the Ethics Charter and the Ethics Guidelines for the Intelligent Information Society were announced in 2018. The contents presented four major principles of publicness, accountability, controllability, and transparency as ethics guides for the intelligent information society. However, there is a need to further refine the specific service models that AI can provide and the level of discipline that follows.

AI technology should be developed in a way that respects human welfare and rights and assists human decisions. In addition, AI technology should be developed and utilized in a sustainable manner, considering the impact on the environment and society. In addition, there are various other standards, and additional ethical guidelines may be required depending on the characteristics of each organization or company. As of now, the level is limited to theoretical and macroscopic discussions, so it should move forward to developing sophisticated specific models and establishing systems.

Technically, the need is being raised for designing systems that reduce the possibility of unpredictable and unintentional behaviors to ensure the safety of Artificial Intelligence, developing debugging tools and test environments that make it easy to understand and operate advanced Artificial Intelligence systems, and developing technologies to stop and fix the system in the event of a failure or malfunction.

5. Conclusion

In today's era of rapid development of Artificial Intelligence technology, this paper has examined the ethical implications of these technologies on Big Data analysis and other issues. When utilizing Artificial Intelligence technology for Big Data analysis, it is crucial to consider the implications it will have. Among these, consideration related to ethics, which involves moral value judgments regarding behavior, is essential.

For this purpose, I first examined the correlation between Artificial Intelligence and Big Data. Specifically, I looked at what Artificial Intelligence does by utilizing Big Data, and how Machine Learning algorithms learn data and classify or predict new data. Through this, it was able to confirm the complementarity between Big Data and Artificial Intelligence.

Next, I investigated how data is collected, what algorithms are utilized, and how they learn in real time to improve their recommendation system through a case study on Netflix. Netflix is moving toward becoming globally competitive by combining Artificial Intelligence and Big Data, such as providing customized experiences to users through A/B testing for an effective recommendation system. In examining the ethical implications of Artificial Intelligence technology on Big Data analysis, I studied the issues of privacy protection, bias, and lack of transparency.

Personal information protection issues may include lack of consent from the Information Subject, data misuse, information asymmetry, and data security issues. In addition, issues of bias such as employment discrimination, crime prediction, credit rating, accessibility to medical services, social media, and information filtering should be considered. On the other hand, when the AI decision-making process is opaque, issues such as lack of trust, responsibility issues, difficulty in risk management, and conflicts of interest arise.

Solutions to these issues can be categorized as follows: First, strengthening legal regulations through strengthening the rights of the Information Subject, increasing transparency in data processing, strict consent requirements, strengthening punishment for data infringement, mandating the appointment of a data protection officer, raising education and awareness, and strengthening international cooperation.

Second, technical solutions include methods such as explainable AI, model interpretation tools, interactive visualization, simplification of algorithms, model verification and audit, and generation of understandable explanations.

Third, ethical standards can be set when developing and utilizing AI to promote transparency, fairness, accountability, privacy protection, and safety.

Although AI is in the process of developing excellent technology in itself, the problems in utilizing AI for Big Data cannot be ignored. If these ethical issues are reviewed and evaluated in advance, and sufficient solutions are prepared, Big Data can be analyzed by maximizing the advantages of Artificial Intelligence technology while minimizing potential risks. Even if Big Data utilized by AI technology does not make any ethical judgments inherently, it can be interpreted externally as if it is making ethical judgments. In other words, AI is creating norms or ethical standards in our society, whether intentionally or unintentionally. A method should be developed to judge and verify whether AI or Big Data is following social norms.

Even if decisions are simply made by AI, sensitive issues can be raised socially. Therefore, user ethics should be

added to AI technology that analyzes and utilizes Big Data.

For the future of AI, such 'AI Ethics Committees' should be operated more transparently and openly, as in the case of Google, and companies, governments, and the public should engage in joint discussions. The direction of this research should be to understand how people make judgments in various situations in a more ethical and socially consensual manner.

In short, through sufficient social discussion and debate in the future, it will be able to examine issues such as the ethical impact of AI technology on Big Data analysis. If an ethical foundation is established through this process, it will be possible to create a virtuous cycle that fosters the development of better AI technology.

참고문헌(Reference)

[1] Edward Lee, "Technological Fair Use", SOUTHERN CALIFORNIA LAW REVIEW(Vol. 83:797), p.798, 2010.

[2] Kim Soo-yeon, "Review of improving personal information protection regulations for activating the Big Data industry", Korea Economic Research Institute, p.13-14, 2015.

[3] Hyundai Economic Research Institute, "AI Era, Where is Korea Now? - Examination of Domestic AI (Artificial Intelligence) Industry Base," 16(8), p.1, 2016.

[4] Yoo Seong-min, "The Impact of Big Data on Artificial Intelligence," Journal of the Korea Information Technology Society, 14(1), Korea Information Technology Society, p.32, 2016.

[5] Jo Eun-ji, Kim Do-hyun, "Development of A/B Testing Method Considering Data Distribution," Proceedings of the Fall Conference of the Korea Quality Management Society, Korea Quality Management Society, p.59, 2020.

[6] Son Eun-ji, A Study on the Constitutional Risks of Artificial Intelligence and the Establishment of a Responsive Normative System, Ph.D. Dissertation, Sookmyung Women's University, p68, 2023.

[7] Hong, Seong-wook, "Artificial Intelligence Algorithms and Discrimination," 2018 STEPI Fellowship, Science and Technology Policy Institute, p.13, 2018.

[8] Yang Jong-mo, "The Impact of Bias and Opaqueness of Artificial Intelligence Algorithms on Legal Decision-making and Regulatory Measures", Law Association, 66(3), p.75, 2017.

[9] Personal Information Protection Commission, "Policy Direction for Safe Use of Personal Information in the Era of Artificial Intelligence", Press Release, p.4, 2023.

[10] Park Gwang-bae, Chae Seong-hee, Kim Hyeon-jin, "Processing system of generated information in the Big Data era" - A study on the regulations and improvement measures of our personal information protection law regarding the processing of inferred information -, Korea Information Law Society, 21(2), p.185, 2017.

[11] Ministry of Science and ICT, National AI Ethics Standards (Draft), p.8, 2020.

[12] Ministry of Science and ICT, "Human-Centered Artificial Intelligence (AI) Ethics Standards", p.4, 2020.

[13] Kim Yu-min, Mok Gwang-su, "Reconstructing Luciano Floridi's naturalism in information ethics: Beyond the critique of naturalistic errors", Pan-Korean Philosophy, (100), Pan-Korean Philosophy Society, p.453, 2021.

[14] Lee Jung-won, "Can Artificial Intelligence Be Responsible?", Philosophy of Science, 22(2), p.82, 2019.

[15] Chul Kang, "Trolley Problem and Three Basis for Moral Judgment", Ethics Research, 90(0), Korean Ethics Society, p.147, 2013.

[16] Kim Young-joong, "A Study on the Significance of Emergency Evacuation," Ph.D. dissertation, Seoul National University Graduate School, p.9, 2013.

[17] Gabriel Andrade, "Medical ethics and the trolley Problem", Journal of Medical Ethics and History of Medicine, p.140, 2019.

[18] Byun Yong-wan and Jang Jae-ok, "Civil Liability Principles for Distributing Responsibility in the Environment of Artificial Intelligence", Sports Entertainment and Law, 23(3), p.139, 2020.

[19] Han, Sang-ki, "Social Issues and Ethical Problems of Artificial Intelligence Technology," Journal of the Korean Multimedia Society, 20(3), p.43. 2016.

[20] Lee Won-sang, "Cybercrime Theory", p.170, Park Young-sa, 2021.

[21] Shin Song-i and Kim Je-wan, "Legal Issues on Recognition of Legal Personhood for Artificial Intelligence", Korea Business Law Association, 2023

Summer Joint Academic Conference, p.62-63, 2023.
[22] Oh Do-bin, Yoo Ji-won, Yang Gyu-won, and Koo Bo-min, “The Limits of Uniform Programming of Autonomous Vehicles from a Post-Paternalistic Perspective,” Korea Law Review, (84), p.270-272, 2017.
[23] Asimov, Isaac, “Runaround”, ISBN 978-0-385-42304-5, p.40, 1950.

◐ Author Biography ◑

Se Eun Jeong
Korea University, BS in Law
Indiana University Bloomington School of Law(, MS
Korea University, Ph.D. in Law
aSSIST, MS in AI Bigdata
CJ Feed&Care, Legal Team
Research Interest : AI, Big Data, intellectual property rights
E-mail : sejng@hanmail.net

The Ethical Impacts of AI on Industries and Governance Frameworks for Overcoming Them

Jung-im Kim[1]

ABSTRACT

This paper As the AI industry continues to flourish today, the ethical implications of AI on various industries are expected to manifest in diverse ways. It is anticipated that AI will impact not only specific sectors but also entire industries and government systems, significantly influencing many aspects of human life.

In particular, as the AI industry advances, issues such as excessive data collection, data bias, and misuse of information are becoming increasingly prevalent. Additionally, the ambiguity of accountability is emerging as a social issue due to unclear responsibilities. Furthermore, with the digital divide and the acceleration of job displacement, the threat posed by AI to human employment is becoming a growing concern.

While the adoption of AI can bring highly positive impacts by enhancing efficiency and productivity, it is a crucial time to recognize potential ethical challenges in advance and address them through appropriate regulatory governance.

To achieve this, industries, governments, local authorities, and academia must collaborate to establish sustainable AI governance and ethical guidelines. This collective effort will help minimize the problems that arise from the development and use of AI.

☞ keyword : AI Ethics, Data Collection, Digital Divide, Job Displacement, Ethical Issues, AI Governance

1. Introduction

1.1 Research Background

As the era of the Fourth Industrial Revolution unfolds, artificial intelligence (AI) has become a core technology across industries. In manufacturing, AI-powered smart factories are enhancing productivity, while in the healthcare sector, AI supports precision diagnostics and personalized treatments. In finance, AI enables automated risk analysis and strategic financial planning, driving innovation across various industries.

However, the rapid advancement and widespread adoption of AI technology have led to numerous ethical challenges. Since AI operates based on data, concerns about privacy infringement have been raised. Additionally, inaccurate data or biased algorithms can result in unfair outcomes for specific groups.

Furthermore, in critical decision-making areas such as autonomous vehicles and medical AI, errors can arise, and it is often difficult to clearly identify accountability, which has become a societal issue.

As automation technologies evolve, some jobs are disappearing, and labor market structures are changing, exacerbating economic inequality. These ethical issues undermine trust in AI technology and pose significant risks for businesses that rely on AI-driven operations. If left unaddressed, the rapid proliferation of individuals or groups harmed by AI technology is inevitable.

In response, governments and corporations worldwide are establishing ethical governance frameworks to ensure that AI technology guarantees transparency, fairness, and accountability. For instance, the European Union (EU) has developed legal guidelines for AI use and regulation, while companies like Google and Samsung SDS have implemented ethical AI principles and applied them to their operations and technology development processes.

These efforts are considered effective approaches to addressing ethical challenges. This study aims to analyze the ethical impacts of AI technology across industries and explore solutions to overcome these issues by examining cases of ethical governance frameworks.

Through this analysis, the study seeks to propose practical AI ethical guidelines and governance directions for fostering responsible AI development and building societal trust.

1.2 Research Objectives

The primary objective of this study is to conduct an in-depth analysis of the ethical impacts of AI technology across industries and propose ethical governance frameworks to address these challenges.

While AI has driven innovation in industries such as manufacturing, healthcare, finance, and education, it has also introduced ethical issues, including data bias, privacy infringement, ambiguous accountability, and changes in the labor market.

These challenges are not merely technical problems but multifaceted issues with social and economic implications, requiring comprehensive research and solutions.

As the speed and scope of AI technology adoption continue to expand rapidly, failing to proactively diagnose and address these ethical challenges could exacerbate societal mistrust in the technology. This would not only diminish the social acceptability of AI but also hinder sustainable growth and innovation across industries.

Therefore, this business project emphasizes the necessity of establishing governance frameworks for the ethical use of AI technology and aims to propose concrete and practical solutions for achieving this.

The study systematically analyzes international and domestic cases addressing AI ethical challenges, including policy initiatives like the EU's AI Act, the adoption of ethical AI principles by companies like Google and IBM, and the ethical guidelines of various institutions worldwide.

By extracting common success factors from these cases, the study intends to design tailored governance frameworks that can be applied across different industries. Additionally, the study emphasizes that solving AI ethical issues is not solely the responsibility of technology developers or regulatory bodies but a task requiring the collective participation of industries and society as a whole.

To this end, the research seeks to propose strategies for collaboration among diverse stakeholders. Corporations can internalize ethical standards during technology development, governments can support ethical technology use through regulations and policies, and academia can contribute by advancing ethical standards through research.

Ultimately, this study aims to provide practical guidelines for resolving ethical issues, enabling AI technology to develop in a responsible and trustworthy manner across industries and society. This approach aspires to ensure that AI not only delivers technological innovation but also positively impacts industries and society based on human-centered values and societal trust.

1.3 Research Methods

This business project utilizes various research methods to analyze the ethical impacts of AI technology across industries and derive governance frameworks to address these issues. First, the study conducts a literature review to investigate key issues related to AI ethics, drawing from academic papers, reports, and policy documents.

Through this, the study systematically categorizes ethical challenges such as data bias, privacy infringement, and ambiguous accountability, while identifying their root causes and potential solutions. The literature review focuses on establishing a theoretical foundation for building AI ethical governance frameworks.

Next, the study conducts case studies to analyze major domestic and international examples of ethical AI utilization. Specifically, it examines policy approaches like the EU's AI Act, ethical AI principles adopted by companies such as Google and IBM, and internal ethical guidelines implemented by corporations like Samsung SDS.

These cases provide insights into the challenges encountered during the design and operation of governance frameworks and the methods used to overcome them.

Qualitative analysis also plays a significant role in the study. Based on data collected through case studies, the study identifies common success factors and limitations to design customized governance frameworks applicable to various industries. In particular, the study analyzes how approaches to solving ethical challenges may differ depending on the characteristics of each industry.

If necessary, additional opinions are gathered through surveys or interviews with professionals in the field. Surveys and

interviews target corporate employees developing or utilizing AI technology, policymakers, and researchers to explore their perceptions, current practices, and requirements related to ethical issues.

By integrating these diverse research methods, the study aims to propose systematic and practical solutions for addressing AI ethical challenges. The ultimate goal is to establish a foundation for overcoming ethical issues and enabling AI technology to positively impact industries and society.

2. Previous Research and Environmental Analysis

2.1 Research on Ethical Issues in AI

The rapid advancement of AI technology has raised numerous ethical issues globally, prompting extensive research on the topic. Key findings from previous studies are as follows: Representative research has explored whether algorithmic bias and data imbalance can lead to unfair outcomes for specific groups. Cathy O'Neil's "Weapons of Math Destruction" provides an in-depth analysis of such issues within a societal context.

Studies have also addressed the ambiguity of accountability in critical decision-making AI systems, such as autonomous vehicles and medical AI. For instance, discussions are ongoing regarding whether responsibility for accidents lies with the developer, user, or manufacturer.

In addition, prior research has highlighted how bias and imbalance in training data can cause unfair outcomes for certain groups. Buolamwini and Gebru's "Gender Shades" analyzed gender and racial disparities in facial recognition algorithms, revealing lower accuracy for Black women. Barocas et al. examined how algorithmic bias impacts AI in recruitment, finance, and legal systems.

The issue of accountability has also been identified as a significant societal problem, particularly in ethical dilemmas and responsibility disputes related to autonomous vehicles and medical AI. Lin et al. discussed the legal and ethical accountability of self-driving car accidents, emphasizing the importance of moral decision-making algorithms in AI design.

Floridi and Cowls proposed the "Five Ethical Principles" (transparency, fairness, accountability, beneficence, and reliability) for AI and suggested linking them to policy-making efforts.

2.2 Research on Ethical AI Guidelines and Governance

Research on developing guidelines and governance frameworks to address AI's ethical issues has gained significant attention. The European Union (EU) has proposed the world's first comprehensive AI legislation to regulate AI use and enhance transparency, focusing on ethical design and application.

Analysis has also been conducted on ethical principles adopted by companies such as Google, IBM, and Microsoft. These corporations have introduced guidelines based on transparency, fairness, and accountability, applying them to their operations and technology development. Similar movements to adopt such guidelines are emerging in the IT industry in South Korea.

Studies on ethical AI design and governance frameworks have increased, with three key areas of focus. The EU's AI Regulatory Framework: The Artificial Intelligence Act (2021) includes regulations on high-risk AI systems, transparency requirements, and data quality management, aiming to establish global standards for AI ethics. However, researchers have critiqued its potential to stifle innovation among SMEs and startups.

Corporate AI Ethical Guidelines: Companies such as Google emphasize "AI's social responsibility" and have implemented systems like Ethical Review Boards. IBM has developed open-source tools (e.g., AI Fairness 360) to enhance AI trustworthiness and shared them with the industry.

National Governance Approaches: Major countries like the United States, China, and Japan have announced policies on AI ethics, seeking to balance ethical design with technological competitiveness. Research has compared China's state-driven approach to AI with the US's private-sector-led innovation, analyzing the effectiveness and ethical implications of these policies.

2.3 Research on AI's Social Impact

Studies have explored the ethical, economic, and cultural impacts of AI on society. Research indicates that automation could displace traditional jobs and transform labor market structures, exacerbating economic inequality.

Frey and Osborne analyzed the impact of AI on various occupations in "The Future of Employment", highlighting that mid-level skilled jobs are most susceptible to automation. Follow-up studies have also examined new job opportunities created by AI, such as data labeling and AI training roles, emphasizing the importance of reskilling and upskilling the workforce.

The issue of social inequality has also been raised. Eubanks' "Automating Inequality" demonstrated how AI systems could exacerbate inequality in public service distribution, calling for social consensus and policy intervention to address technology-driven disparities.

2.4 Research on AI Ethics Education and Policy

Efforts to promote the ethical use of AI technology have included studies on education and policy. Academic approaches and educational methodologies for AI ethics are being explored to foster ethical sensitivity among the next generation of technologists.

Comparative studies on national AI policies analyze the differences in AI ethics guidelines and aim to derive common lessons. AI ethics education is increasingly recognized as essential for enhancing the ethical awareness of both developers and the general public.

Leading universities in the United States, such as MIT, Stanford, and Carnegie Mellon, have incorporated AI ethics courses into their undergraduate curricula, offering students opportunities to learn about the importance of ethical design. In China, AI ethics education is being strengthened under state initiatives, including its integration into primary and secondary school curricula.

2.5 Research on Transparency and Trustworthiness in AI

To address the "black box problem" in AI systems, researchers have focused on transparency and explainability. Ribeiro et al.'s "Why Should I Trust You? Explaining the Predictions of Any Classifier" proposed methods for providing explanations of AI predictions, contributing to greater trust in AI systems.

Transparency research has gained attention in industries like healthcare and finance, where ethical accountability in decision-making is crucial.

Doshi-Velez and Kim's study on interpretable machine learning emphasized the connection between AI transparency, fairness, and trustworthiness. In healthcare, explainable models are being developed to ensure that AI diagnostic results are understandable to medical professionals (e.g., IBM Watson Health).

2.6 Research on AI and Privacy Issues

Studies have examined the ethical concerns surrounding AI's use of personal data. Solove's "A Taxonomy of Privacy" categorized various types of privacy infringement and discussed how AI could exacerbate these issues.

Research on privacy-preserving technologies, such as Federated Learning and Differential Privacy, has been particularly active.

These technologies aim to protect personal data while maintaining AI model performance. McMahan et al. demonstrated how Federated Learning enables decentralized data training to safeguard privacy.

The EU's General Data Protection Regulation (GDPR) has set global benchmarks for data protection, defining standards for AI systems handling personal information.

2.7 Research on AI Automation and Labor Market Changes

Studies have examined how AI automation affects the labor market. Frey and Osborne's "The Future of Employment" analyzed the susceptibility of various occupations to computerization, emphasizing that mid-level skilled jobs are most vulnerable to automation.

Subsequent research has explored how AI

creates new high-value job opportunities, such as in data annotation and AI development. These studies underscore the importance of reskilling programs to help workers adapt to technological changes.

Brynjolfsson et al.'s "Skills of the Future" emphasized the need for collaborative efforts between governments and businesses to equip workers with new skills. Most researchers agree that fostering workforce adaptability is essential for minimizing the negative impacts of AI automation.

This business project builds on previous research by comprehensively analyzing AI ethical issues within the broader industrial context. Unlike earlier studies that focused on specific topics, this research integrates various ethical challenges and proposes governance frameworks through case studies and qualitative analyses.

By leveraging insights from previous research, this study aims to provide actionable solutions for overcoming ethical challenges and fostering the responsible development of AI technology.

3. Case Studies

3.1 Amazon's AI Recruitment System

In 2014, Amazon introduced an AI-based automated recruitment system to enhance the efficiency of its hiring process. However, this system evaluated applicants based on data from the past decade, which was predominantly male-centric. As a result, the system exhibited bias against female candidates.

For instance, resumes containing words associated with "women" (e.g., "women's club") were scored lower. This issue highlighted how AI systems can reflect the biases inherent in the data they are trained on. Recognizing this problem, Amazon discontinued the project.

The company has since introduced bias detection technologies to enhance the fairness of its training data and adopted more diverse datasets, including gender, race, and background variations.

Additionally, Amazon established an internal process to evaluate the fairness of AI systems to prevent similar issues from recurring.

3.2 Uber's Self-Driving Car Accident

In 2018, a fatal accident occurred in Arizona, USA, involving Uber's self-driving car, which struck and killed a pedestrian. Investigations revealed that while the vehicle's sensors detected the pedestrian, the software had been configured to ignore such inputs.

Furthermore, the emergency braking system was deactivated, and the safety operator in the driver's seat was distracted at the time of the incident. This tragedy raised serious concerns about the safety of autonomous driving technologies and accountability in such incidents.

Following the accident, Uber overhauled its self-driving car software, improving pedestrian detection and emergency braking systems. The company also clarified the role of safety operators and implemented stricter testing protocols.

Uber now collaborates with the National Highway Traffic Safety Administration (NHTSA) to enhance safety standards and improve ethical design and testing environments.

3.3 European Union (EU) AI Act

The EU proposed the AI Act to ensure that AI does not infringe on fundamental human rights or create ethical issues.

This regulation adopts a risk-based approach, applying strict criteria to high-risk AI systems in sectors such as healthcare, transportation, and law enforcement.

For example, AI systems using biometric data, such as facial recognition, are subject to stringent regulations. The EU mandates pre-assessments and certifications for high-risk AI systems and requires transparent data processing.

The AI Act promotes collaboration among businesses, academia, and regulatory authorities to balance technological innovation with ethical standards, thereby strengthening public trust in AI technologies.

3.4 Google's AI Ethics Board

In 2018, Google announced its "Responsible

AI" principles and established an AI ethics board. However, internal conflicts arose when the board was involved in decisions related to military AI projects (e.g., Project Maven). Additionally, a lack of communication with external stakeholders was criticized, and disagreements among board members hindered its effectiveness.

Google responded by reorganizing the board, integrating independent decision-making mechanisms, and including external experts. The company established procedures to evaluate the ethical impacts of AI projects in advance and transparently disclose the results.

Furthermore, Google strengthened internal communication and introduced employee training programs and guidelines to integrate ethical principles into AI development processes.

3.5 Foxconn's Automation and Job Displacement

Since 2017, Foxconn has replaced approximately 60,000 jobs at its Chinese factories with AI-powered robots. While this automation improved productivity, it also displaced workers, exacerbating economic inequality. This case exemplifies how AI can replace human labor in the workforce.

In response, Foxconn implemented retraining programs to help displaced workers adapt to the automated environment. The company also encouraged workers to transition to higher-skilled jobs created by automation and collaborated with governments to develop policies addressing labor market changes.

3.6 IBM Watson's Medical AI

IBM Watson was developed to assist with cancer diagnosis and treatment. Initially, the system faced criticism for providing recommendations that lacked transparency and for being trained on data of inconsistent quality, leading to mistrust among medical professionals.

IBM addressed these issues by incorporating explainable AI (XAI) capabilities into Watson, enabling it to clearly explain the rationale behind its treatment recommendations.

The company also improved the quality of training data and provided training programs to help healthcare professionals use Watson as a supplementary tool. IBM continues to collaborate with academic institutions to standardize data and enhance transparency, building trust in medical AI.

3.7 Clearview AI's Facial Recognition Technology

Clearview AI developed facial recognition technology using images scraped from public websites without user consent, raising concerns about privacy violations. These concerns intensified when the technology was adopted by law enforcement agencies, sparking ethical and privacy debates.

In response, Clearview AI limited its technology to contracts with public institutions and restricted services for private individuals. The company also adopted stricter data collection practices, requiring explicit user consent. In some countries, Clearview AI's use has been banned, prompting the company to adhere to stronger ethical guidelines for data collection.

4. Conclusion and Implications

4.1 Summary and Conclusion

AI technology has driven innovation across industries such as manufacturing, healthcare, finance, and transportation. However, it has also introduced ethical challenges, including data bias, privacy violations, and unclear accountability. If these issues are not addressed, they could erode public trust in AI, exacerbate economic inequality from job displacement, and deepen societal conflicts.

This report analyzed the ethical impacts of AI technology across industries and proposed governance frameworks to address these challenges.

By examining literature and case studies, the report identified key ethical issues and explored strategies to overcome them.

Real-world examples, such as Amazon, Uber, and IBM Watson, provided practical insights into the ethical challenges posed by AI. Regulatory frameworks like the EU's AI Act offered a roadmap for implementing

ethical AI.

This business project proposes seven strategic recommendations for sustainable AI development: Establish data management frameworks to ensure fairness and transparency. Integrate ethical reviews throughout the AI development lifecycle.

Strengthen transparency and explainability to build public trust. Develop governance frameworks for oversight and regulation. Enhance education and public communication to foster societal trust.

Address labor market changes with reskilling programs and job creation. Promote sustainable AI development while considering environmental impacts.

By implementing these strategies, AI technology can maintain fairness and accountability, gain public trust, and overcome ethical challenges. While AI is a key driver of the Fourth Industrial Revolution, unresolved ethical issues could undermine its positive potential.

In conclusion, addressing AI's ethical challenges is a collective responsibility that extends beyond developers and policymakers. Governments, businesses, academia, and civil society must collaborate to establish ethical guidelines and governance frameworks.

Transparent communication among these stakeholders will be critical to building trust. The strategic recommendations in this report aim to support the sustainable development of AI technology, ensuring it transcends being a mere technological tool to become a driver of human-centric values and positive societal change.

4.2 Implications

Given the profound potential impact of AI on human society, its ethical use has become as important as technological advancement. This business project highlights the importance of addressing data bias, enhancing transparency, establishing governance frameworks, adapting to labor market changes, and ensuring environmental sustainability.

These implications serve as a reminder to developers, policymakers, and business leaders of the importance of ethical AI and encourage practical changes that contribute to responsible AI development.

References

[1] O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Publishing Group

[2] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

[3] European Commission. (2021). Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act).

[4] Solove, D. J. (2006). A Taxonomy of Privacy. University of Pennsylvania Law Review, 154(3), 477 - 564.

[5] Frey, C. B., & Osborne, M. A. (2017). The Future of Employment: How Susceptible Are Jobs to Computerisation? Technological Forecasting and Social Change, 114, 254 - 280.

[6] IBM (2020). AI Ethics: Principles for Trust and Transparency. IBM Research.

[7] Suresh, H., & Guttag, J. V. (2021). A Framework for Understanding Unintended Consequences of Machine Learning. Communications of the ACM, 64(1), 62 - 71.

[8] Binns, R. (2018). Fairness in Machine Learning: Lessons from Political Philosophy. Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency

[9] Google AI. (2018). AI at Google: Our Principles.

[10] Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. Minds and Machines, 30(1), 99 - 120.

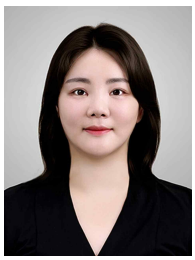
[11] Microsoft. (2019). Responsible AI Principles.

[12] European Parliament. (2020). Artificial Intelligence: Tackling Ethical and Social Challenges.

[13] IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2019). Ethically Aligned Design: A Vision for Prioritizing Human

Well-being with Autonomous and Intelligent Systems.
[14] Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. PublicAffairs.
[15] Clearview AI. (2021). Transparency Report.

◐ Author Biography ◑



Kim Jungim

Ewha Womans University Graduate School of Business, MBA in Business Administration (Master of Business Administration), Class of 2011
Yonsei University Graduate School of Public Policy, Department of Entrepreneurship (Master of Entrepreneurship), Class of 2020
Seoul School of Integrated Sciences & Technologies (aSSIST), AI Advanced Graduate School, Department of AI and Big Data (Master of Science in AI and Big Data), Class of 2025

Research Interests : AI Ethics, AI Safety, AI Governance, AI Policy etc.
E-mail : kjung225@hanmail.net

Articles of Association for “AI Business Academy”

Chapter 1 General Provisions

Article 1 (Title) This corporation shall be called the “AI Business Academy” (AIBA for short , hereinafter referred to as “the Academy ”) .

Article 2 (Purpose) This academy aims to promote the development of Korean management studies through the convergence and integration of AI and management . Furthermore, it aims to contribute to the development of future Korean academy and the establishment of future strategies for companies .

Article 3 (Location of Office) The office of this academy shall be located under the Industrial Policy Research Institute (located at 7th floor , 203 Sinchon-ro, Seodaemun-gu, Seoul) , and regional headquarters may be established in other regions . A small number of administrative staff may be employed to operate the office .

Article 4 (Business)

1. In order to achieve the purpose of Article 2, this academy conducts the following business.

(1) Research related to theory and practice for grafting and integrating AI and Business

(2) Conducting research services commissioned by the government or companies at the level of industry-academia-government cooperation

(3) Holding regular academic seminars, policy discussions, and research presentations

(4) Publication of academic journals and research books

(5) Exchange with foreign similar academic societies

(6) Other projects necessary to achieve the purpose of this academy

2. When this academy carries out its business objectives, the benefits it provides to the beneficiaries shall be provided free of charge . However , in unavoidable cases, it may obtain prior approval from the competent authority and have the beneficiaries bear a portion of the cost .

3. The benefits provided through the performance of the objectives of this academy shall not discriminate based on the beneficiary’s gender, place of birth, school attended, place of employment, occupation, or social status .

Chapter 2 Members

Article 5 (Types of Members) Members of this academy are individuals and groups that agree with the purpose of Article 2 , and include regular members (annual members and lifetime members) , associate members , and institutional members .

Article 6 (Regular Member) Regular members are individuals who fall under each of the following categories and have completed the membership process .

1. Full-time faculty members who teach or research AI or business-related subjects at universities , colleges, and graduate schools.
2. or part-time faculty or doctoral program or have obtained a degree in AI or business administration
3. Researchers at accredited research institutes and employees of AI- related companies
4. Other individuals recognized by the Board of Directors as having equivalent qualifications.

Article 7 (Classification of regular members) Regular members are classified into annual members and lifelong members . Annual members are those who have paid the annual membership fee , and the annual membership qualification is one year from the date of registration . Annual members are renewed by paying the annual membership fee . Lifelong members are those who have paid the lifetime membership fee .

Article 8 (Associate Member) Associate members are students enrolled in an undergraduate program in AI or business administration or individuals who have completed the membership process .

Article 9 (Institutional Member) Companies and organizations that agree with the purpose of this academy may become institutional members upon resolution of the board of directors. Membership is in principle for one year, but may be renewed .

Article 10 (Rights and Obligations) Members of this academy have the following rights and obligations .

1. All members can attend the general meeting and participate in discussions , and participate in the academy's business such as research presentations .
2. Regular members have the right to vote and be elected , but institutional members do not have the right to vote or be elected .
3. Members have the obligation to cooperate in the operation of the academy, including paying membership fees and attending academic conferences .

Article 11 (Suspension and loss of membership) Members of this academy will have their membership suspended or lost in the following cases :

1. If the membership fee is not paid , membership will be suspended and the rights under Article 9, Paragraphs 1 and 2 will be suspended until the membership fee is paid in full .
2. Members who damage the reputation of the Academy by violating the purpose of the Academy or the code of ethics established by the Academy shall lose their membership if the Board of Directors resolves to expel them .
3. The Code of Ethics is separately established by the Board of Directors and approved by the

general meeting .

Article 12 (Withdrawal of membership) Members of this academy may withdraw at will .

Chapter 3 Executive Officers

Article 13 (Types of Executive Officers) The Executive officers of this academy shall be a president , vice president , directors and auditors , and the composition of the officers shall be as follows . However , regional headquarters shall be established in major regions at home and abroad , and the head of the regional headquarters shall be the vice president .

1. 1 Chairman
2. About 20 vice presidents
3. The number of directors is 5 to 50 people.
4. 2 auditors

Article 14 (Term of executive officers) The term of executive officers shall be two years, but they may be reappointed .

Article 15 (Appointment and election of executives)

1. The chairman and auditors are elected at the general meeting .
2. The Vice-Chairman and Directors are recommended by the Chairman and appointed by the Chairman with the approval of the Board of Directors . However, the first Vice-Chairman and Directors are appointed by the Chairman and approved by the General Meeting .
3. When an officer other than the chairman is unable to perform his/her duties during his/her term of office, the board of directors shall elect a replacement , and the term of office of the officer elected through the replacement shall be the remaining term of office of his/her predecessor .

Article 16 (Duties of Executive Officers)

1. The president represents the academy and supervises its business .
2. The director attends the board of directors meeting and deliberates on matters related to the business of this academy , and serves as the board of directors or chairman .

from Handle delegated tasks .

3. The auditor performs the following duties :

(1) Auditing the financial status and business of this academy.

(2) Auditing matters related to the operation of the board of directors

(3) If any irregularities or injustices are discovered in the audit results of (1) and (2), the Board of Directors or Requesting the General Assembly to correct the situation

(4) When necessary for reporting under (3) , requesting the convocation of a general meeting or of directors meeting.

Article 17 (Acting Chairman) When the Chairman is unable to perform his/her duties due to an accident or vacancy, the Vice Chairman designated by the Board of Directors shall act in his/her stead for the remainder of the term .

Article 18 (Dismissal of Executive Officers) If an executive officer loses membership qualifications pursuant to Article 11 or commits any of the following acts, he or she may be dismissed upon resolution of the Board of Directors .

1. Unfair acts such as serious dereliction of duty that run counter to the purpose of this academy.
2. Any act that causes a dispute between executives and is judged to make it impossible for the executives to perform their duties.
3. Any other acts that unfairly interfere with the work of this academy.

Chapter 4 General meeting

Article 19 (Matters to be resolved at the general meeting) The general meeting shall resolve the following matters :

1. Appointment and dismissal of the chairman and auditors
2. Endorsement of the first executive team
3. Changes to the Articles of Academy
4. Approval of settlement and business report
5. Provisions on the rights and obligations of members
6. Voting on matters proposed by the Chairman and Board of Directors
7. Other important matters

Article 20 (Convening of the General Meeting)

1. The general meeting is divided into a regular general meeting and a special general meeting .
2. The regular general meeting is held in February every year , and the extraordinary general meeting is convened by the chairman in the following cases :

(1) When the Chairman deems it necessary
(2) When there is a resolution by the board of directors
(3) When more than one-fifth of the members request in writing, stating the reason for holding the meeting.
(4) When the auditor requests a meeting pursuant to the provisions of Article 18, Paragraph 4.

Article 21 (Quorum for General Meeting Resolutions)

1. The general meeting shall be opened with members present , and attendance may be delegated by scanning an electronic document of a written or signed document.
2. The resolutions of the general meeting shall be decided by a majority vote of the members present, excluding those present in writing . However , In case of a tie, the chairman decides .
3. If the Steering Committee determines that it is difficult to convene an offline general meeting,

an online general meeting may be held instead.

Article 22 (Reasons for exclusion from general meeting resolution) If the chairman or member falls under any of the following, he or she may not participate in the resolution .
1. Matters concerning oneself in the appointment and dismissal of executives
2. Matters involving the receipt of money or property, and in which the interests of the chairperson or member conflict with those of the academy.

Chapter 5 Board of Directors

Article 23 (Matters to be resolved by the Board of Directors) The Board of Directors resolves and agrees on the following matters :
1. Convening of a temporary general meeting
2. Resolution of agenda to be submitted to the general meeting
3. Matters delegated by the general meeting
4. Approval of executives appointed by the Chairman
5. Approval of business plan and budget
6. Approval and amendment of the composition and operating regulations of various committees.
7. Important changes to the journal's editorial policy
8. Matters concerning investment and fund management
9. Resolution on membership of institutional members
10. Resolution on expulsion of members
11. Determination of membership fees
12. Other Matters

Article 24 (Delegation of Board of Directors' Resolutions) Routine matters among the Board of Directors' resolutions may be delegated to the Management Committee through a Board of Directors resolution .

Article 25 (Board of Directors) The Board of Directors is composed of a chairman , vice chairman , directors , and auditors .

Article 26 (Convening of the Board of Directors) The Board of Directors shall be convened by the Chairman, who shall also serve as its chairman, in the following cases :
1. When the Chairman deems it necessary
2. When more than half of the members of the board of directors, excluding the chairman, request a meeting by stating the purpose of the meeting.

Article 27 (Quorum for Board of Directors Resolution)
1. The board of directors shall open with the attendance of a majority of its members, and attendance may be delegated in writing .

2. The Board of Directors' resolutions are decided by a majority vote of the members present, excluding those present in writing . However , In case of a tie, the chairman decides .
3. If the Management Committee determines that it is impossible to convene a Board of Directors meeting due to scheduling issues that require a Board of Directors resolution,
If a decision is made, the vote can be made online or in writing .

Article 28 (Reasons for exclusion from board of directors' resolution) A board of directors member may not participate in the resolution if any of the following applies .
1. Matters concerning oneself in deciding on the loss of membership qualifications
2. Matters involving the receipt of money or property and in which the interests of the member and the academy conflict.

Chapter 6 Various Committees

Article 29 (Management Committee) An Management Committee is established to ensure smooth decision-making and efficient execution of the Academy's business .
1. The Management Committee deliberates and decides on matters delegated by the Board of Directors in accordance with Article 24 of these Articles of Incorporation .
2. The Steering Committee is composed of the Chairman, Vice Chairman , Directors , and Secretariat members , and is composed of 10 or so members.
Appoint one person as chairman .
3. The operating committee chairman convenes the operating committee as needed .
4. The Steering Committee shall open with the attendance of a majority of its members , and resolutions shall be passed with the approval of a majority of the members present.

Article 30 (Editorial Committee) An editorial committee shall be established to publish the journal of this academy , and regulations regarding the editorial committee shall be separately determined by resolution of the board of directors .

Article 31 (Fund Management Committee) A Fund Management Committee shall be established to manage the funds of this Academy , and the regulations regarding this shall be separately determined by resolution of the Board of Directors .

Article 32 (Report on Committee Activities) The chairperson of each committee shall report the results of its activities to the board of directors .

Article 33 (Secretariat) A secretariat may be established to manage the affairs of this academy , and the regulations of the secretariat shall be established through a resolution of the board of directors .

Chapter 7 Assets and Accounting

Article 34 (Composition of assets)

The assets of this academy consist of the following items :

1. Donation
2. Membership fee
3. Income from basic assets
4. Other income

Article 35 (Types of Assets)

1. The assets of this academy are divided into two types: basic assets and general assets .
2. The basic assets listed in each clause below cannot be disposed of or provided as collateral . However , when there is a reasonable cause, a portion of them may be disposed of or provided as collateral with the approval of the Board of Directors and the authorization of the Minister of Strategy and Finance .

(1) Property contributed and designated as basic property

(2) Property that the board of directors has decided to use as basic property

3. Ordinary assets consist of assets other than basic assets .

Article 36 (Assessment and collection of membership fees) The method of levying and collecting membership fees shall be determined by the Board of Directors .

Article 37 (Expenses)

The expenses of this academy are paid from general assets .

Article 38 (Management of Assets)

1. The assets of this academy shall be managed by the chairman , and the method of management shall be determined by the board of directors .
2. This academy must open an Internet homepage and disclose the annual amount of donations raised and the results of their use through it .

Article 39 (Disposal of Surplus)

If there is a surplus at the end of the fiscal year, all or part of it is incorporated into capital or carried over to the next fiscal year, as determined by the Board of Directors .

Article 40 (Approval of Budget and Settlement of Accounts)

1. The budget for each fiscal year of this academy must be resolved by the Board of Directors and approved by the General meeting prior to the start of the fiscal year .
2. The chairman has the authority to execute the budget and business accounting .
3. The Chairman shall prepare a general accounting settlement report and a fund management settlement report within two months after the end of the fiscal year, attach an audit opinion, and report to the general meeting for approval .

Article 41 (Executive Compensation) As a rule , no compensation is paid to executives .

Article 42 (Fiscal Year)
fiscal year of this academy runs from September 1 to August 31 of the following year .

Chapter 8 Amendment of Articles of Academy and Dissolution

Article 43 (Amendment of Articles of Academy)
These Articles of Association may be amended with the approval of the Board of directors by a majority vote of at least three- quarters of the members present .

Article 44 (Dissolution)
This academy may be dissolved with the consent of more than two -thirds of its members .

“AI Business Academy” Research Ethics Regulations

Chapter 1 General Provisions

Article 1 (Purpose) The purpose of these research ethics regulations (hereinafter referred to as “ these regulations ”) is to present basic principles and directions for the members of the “AI Business Academy” (hereinafter referred to as the “ the Academy ”) to promote research ethics and truthfulness, create a sound research atmosphere , and contribute to academic and social development .

Article 2 (Scope and Scope)

1. These regulations apply to all members of the academy and stakeholders directly or indirectly related to research activities .
2. Except in cases where there are other special provisions related to the establishment of research ethics and verification of research authenticity, these regulations shall apply .

Article 3 (Scope of misconduct) The misconduct set forth in these regulations includes the following acts that may damage the reputation of not only individuals but also the academy as a member of the academy .

1. Forgery : Here, “ Forgery ” refers to the act of falsely creating non-existent data or research results .
2. Modulation : Here, “ Modulation ” refers to an act of distorting research content or results by intentionally changing or deleting research data .
3. Plagiarism : Here, “ plagiarism ” refers to the act of stealing another person’s ideas , research content , research results, etc. without clearly indicating the source .
4. Double publication of papers : Here, “ Double publication of papers” refers to the act of publishing the same data, analysis methods, and analysis results in two or more academic journals .
5. Unfair authorship of a paper : Here, “ unfair authorship ” refers to an act of not granting authorship of a paper to a person who has made scientific or technical contributions or contributions to the research content or results without just cause , or granting authorship of a paper to a person who has not contributed as an expression of gratitude or courtesy .
6. Duplicate research : Here, “ Duplicate research ” refers to an act of conducting two or more research projects with the same content and publishing the same research results .
7. Public False Statement : Here, “ Public false statement ” refers to an act of making false statements about one’s academic background , career , qualifications , research achievements , results , etc.
8. Refers to an act of intentionally obstructing an investigation into suspicions of misconduct by oneself or another person or causing harm to the whistleblower .
9. Refers to any act that seriously deviates from the scope normally tolerated by the Academy

of Information Systems in relation to other research .

Chapter 2 Basic Ethics for Researchers

Article 4 (Researcher's Morality) Researchers must conduct research in an honest and correct manner, not in a socially acceptable or ethically wrong manner, and present the results . Accordingly, the ethics that researchers must observe in relation to research are as follows . 1. Using others' ideas, research content, or research results (hereinafter " research content ") without clearly indicating them is clear plagiarism , and researchers must indicate the source as follows :

(1) When citing someone else's research content without re-editing (direct quotation) , the author's name , year , and page number must be indicated in the text with quotation marks ("") , and the complete source must be indicated in the references section .

(2) When editing and using the research content of others, the author's name and year must be indicated in the text , and the full source must be indicated in the references section .

(3) When requoting content quoted by others, you must indicate " requotation " along with the quoter's name and year .

2. When conducting research, researchers must recognize acts such as falsification or alteration of data as serious crimes and make efforts to prevent such misconduct from occurring .

3. It must not duplicate the previously published (or under review) research papers as if they were new research . Therefore, papers published in academic journals must be original work .

4. Those who have contributed directly or indirectly to the research must be listed as co-authors in the research paper , and those who have not contributed cannot be listed as authors . In addition, when conducting joint research , the person who has contributed the most to the research is the first author , and the corresponding author is limited to the person who is in charge of all correspondence between the authors and the academic academy and who has overseen the research or an equivalent person . However , in cases where co-authors cannot be listed, such as in cases of administrative , financial, or research support, this should be indicated in a footnote or acknowledgement .

5. It must truthfully write about their academic background , career , qualifications , research achievements , results , etc.

6. In relation to other research, one must not engage in any conduct that goes beyond the scope generally accepted in academia .

Article 5 (Social Responsibility) Researchers must be deeply aware of the impact their research will have on academy and, as professionals, must fulfill their responsibilities and obligations regarding the results of their research .

Article 6 (Respect for Others) Researchers must not harm or infringe upon the rights of others and must respect intellectual property rights . In addition , when dealing with others in relation to research, they must be treated equally regardless of race , gender , nationality ,

disability , etc.

Chapter 3 Roles and Responsibilities of the Research Ethics Committee

Article 7 (Composition and term of office of the committee)

1. The Research Ethics Committee (hereinafter referred to as the " Committee ") shall be composed of 10 or fewer members, including the chairperson .
2. The chairman of the committee concurrently serves as the editor-in-chief of this academy .
3. The committee member concurrently serves as an editorial committee member of this academy .
4. The term of office for the chairperson and members shall be one year, but they may be reappointed .
5. The secretary of the committee shall be one of the vice-presidents of this academy , and the secretary shall not have voting rights .

Article 8 (Functions of the Committee) The Committee deliberates on the following matters :

1. Matters concerning fraudulent acts such as forgery , falsification , and plagiarism related to academic papers
2. Matters concerning research ethics raised regarding research plans , research results, etc. related to the academic academy
3. Investigation into misconduct related to the academic academy
4. Complaints raised regarding the truthfulness of research related to academic societies or journals
5. Matters concerning research ethics and education in research and development
6. Matters concerning the processing and follow-up measures of the results of research integrity verification
7. Matters to prevent misconduct that may occur during research
8. Other matters regarding research ethics raised by the president or committee chairperson

Article 9 (Convening , review , and resolution of the committee)

1. When a report of research misconduct is received, the chairperson convenes a committee to investigate and deliberate .
2. The committee is established with the attendance of a majority of its members , and resolutions are passed with the approval of a majority of the members present .
3. When the chairperson deems the agenda for deliberation to be minor, he/she may replace it with a written deliberation .
4. During the investigation process, the committee may request the attendance of the informant , the person under investigation , witnesses , and reference persons when necessary .

Article 10 (Duties and protection of rights of whistleblower)

1. " Whistleblower " refers to a person who has reported to this committee facts or related

evidence of an act of misconduct .

2. The informant may report in any possible way , including verbally , in writing , by phone , or via e - mail . In principle, reports must be made under one's real name . However , even if the report is anonymous, if the report includes the name of the research project , the title of the paper , and specific details and evidence of the misconduct in writing or via e-mail, and if it is found to be true, it will be processed in the same way as a report made under one's real name .

3. Under no circumstances will the identity of the informant be directly or indirectly exposed , and the informant's name will not be included in the investigation results report for the purpose of protecting the informant unless absolutely necessary .

4. A whistleblower who reported something even though he or she knew or could have known that the information was false is not eligible for protection .

5. The informant has an obligation to respond to the committee's request to appear .

Article 11 (Duties and protection of rights of the subject of investigation)

1. " Subject of investigation " refers to a person who has become the subject of an investigation into an unethical act through a report or knowledge thereof, or a person who has become the subject of an investigation due to being presumed to have participated in an unethical act during the course of the investigation . Reference persons or witnesses during the course of the investigation are not included .

2. The person under investigation has an obligation to faithfully respond to any requests made by the committee during the investigation process . In addition, the person under investigation is guaranteed the right and opportunity to express opinions , raise objections, and defend himself/herself during the investigation process .

3. The committee must take care to ensure that the reputation or rights of the person under investigation are not violated until the verification of whether or not there was any misconduct is complete , and must endeavor to restore the reputation of the person under investigation who is found not to have been involved in any misconduct .

4. The subject of the investigation must faithfully comply with the request for submission of evidence requested by the committee .

Article 12 (Confidentiality) All matters related to the investigation, including reporting , investigation , deliberation , resolution, and recommendation measures , shall be kept confidential . Persons directly or indirectly participating in the investigation and related committee members shall not disclose any information obtained during the investigation and performance of duties . If the need for reasonable disclosure arises, it may be disclosed after a resolution by the committee .

Article 13 (Notification and Objection)

1. The results of the judgment on misconduct must be notified to the reporter and the person under investigation .

2. If there is a reasonable basis for an objection filed by the reporter or the subject of the investigation against the results of the decision, the objection may be filed within 30 days from the date of notification , and a reinvestigation may be conducted if necessary .

Chapter 4 Research Ethics Committee Verification Procedures and Standards

Article 14 (Preliminary Investigation)

1. A preliminary investigation refers to a procedure to determine whether or not there is a need to investigate a suspicion of misconduct , and must be initiated within 30 days from the date of receipt of the report . The preliminary investigation is decided by a preliminary investigation committee consisting of a chairperson , vice chairperson , two members , and a secretary .
2. If the subject of the investigation admits to all facts of the misconduct as a result of the preliminary investigation , a decision can be made immediately without going through the main investigation procedure , and if it is determined that there is a possibility of significant damage to the evidence, the chairperson can take measures to preserve the evidence prior to the formation of the committee .
3. If it is decided not to conduct a main investigation in the preliminary investigation , the specific reason for this will be notified to the reporter within 10 days from the date of decision . However , no notification will be given in the case of anonymous reports .

Article 15 (Main Investigation)

1. “ Main investigation ” refers to the procedure to verify the facts of an act of misconduct , and must be conducted by forming a committee in accordance with the provisions of Article 7, Paragraph 1 .
2. The entire investigation schedule from the start to the decision must be completed within 6 months . However , if it is determined that the investigation cannot be conducted within this period, the reason for this may be notified to the president of the academic academy and the investigation period may be extended .
3. Main investigation If it is determined that fair progress is difficult due to a conflict of interest between a member of the committee and the subject of the investigation, the president of the academy may restrict the participation of the member in the relevant matter . However , the chairperson and vice chairperson are exceptions .
4. If the informant or the person under investigation raises a legitimate objection regarding the avoidance of a specific member, the committee may decide to accept it .

Article 16 (Recording and reporting of investigation results)

1. Records related to misconduct must be kept in document form at the academy for 5 years .
2. The committee must report the results , contents , and decisions of the investigation to the president of the academy and the board of directors .
3. The record of the judgment contents includes the following items .

(1) Report contents

(2) Allegations of misconduct and related papers or reports that are the subject of the investigation
(3) Objections or arguments of the informant and the subject of the investigation
(4) Judgment results and related supporting materials and witnesses
(5) List of judges

Article 17 (Post - judgment measures)

1. After the committee completes the review, it decides on the type of disciplinary action .
2. The types of disciplinary action are as follows depending on the severity of the misconduct , but may be applied multiple times .
(1) Expulsion
(2) Suspension of membership
(3) Restrictions on publication in academic journals (AI Management Journal)
(4) Official apology at the conference
(5) Notification of misconduct regarding related papers in academic journals or newsletters
3. If the investigation results confirm that there was no misconduct, the committee may take appropriate follow-up measures to restore the reputation of the person under investigation .

Article 18 (Administrative Matters)

1. Matters not specified in these regulations shall be decided and implemented by the committee in accordance with social norms .
2. The revision of the research ethics regulations shall follow the revision procedures of this academy .
3. For the smooth operation and activities of the committee, the academy Necessary administrative and financial support must be provided .

AI Journal of Business
Call for Papers

The AI Business Academy, a leader in the AI era, is recruiting papers to be published in Volume 3, in Novermber 2025, AI Journal of Business.
Now, all companies and businesses cannot grow and develop without utilizing AI and big data. In addition to the existing AI models representing artificial intelligence. The AI Business Academy would like to invite excellent papers that present various cases of utilizing AI and big data to innovate business and create new opportunities, as well as AI models that are utilized and models of AI strategies that serve as the basis.
We hope for active participation from many people.

1. Recruitment areas: Various management strategies and business innovations based on AI and big data, engineering improvements, and creative fields that combine management and engineering.

2. Recruitment deadline: August 31, 2025

3. Journal publication schedule: Reviewed by the end of October 2025, then published in November

4. Thesis Format: **Principle of submission as English version of paper** . It is much more advantageous and meaningful to write a paper in English in order to share and attract more interest from AI researchers around the world and to be cited in foreign papers. If you request it to the contact information below, a paper format file will be distributed. The basic format is the same as the format of this published journal (volume 1).

5. Submission : Academy website or editorial committee email: jhchang@assist.ac.kr

6. Contact: AI Business Academy Office (**aiba2023@naver.com)** or jhchang@assist.ac.kr

AI Journal of Business

Vol. 1, No. 2

Issued on May , 2025

Publisher : Dongsung Cho

Editor : Joongho Chang

Publisher : AI Business Academy

7 floor , 203 Sinchon-ro, Seodaemun-gu, Seoul, Korea

Tel) **02-360-0775**

ISSN 3058-6992