

## Fuzzy AHP-based Weight Analysis of Selection Attributes for ChatGPT and Gemini

Yujin Kim\*, Hyung-Seok Lee\*\*

\*Ph. D., Dept. of Business Administration, Chungbuk National University, Chungbuk, Korea

\*\*Professor, School of Business, Chungbuk National University, Chungbuk, Korea

### [Abstract]

Extending prior research on the selection attributes of generative AI services, this study develops a two-level hierarchical model comprising four upper-level dimensions-functional attributes, experiential attributes, adaptive attributes, and brand image-and eight lower-level criteria, namely information accuracy, explainability, enjoyment, anthropomorphism, personalization, creativity, corporate reputation, and corporate innovativeness, to better understand the complexity of users' choice environments arising from the increasing diversity of generative AI services. To account for the imprecision inherent in human judgment, Fuzzy AHP was employed to derive the priority weights, and a conventional AHP analysis on the same data yielded nearly identical weights, thereby confirming the robustness of the Fuzzy AHP results. The analysis confirms that information accuracy emerges as the highest-priority attribute of generative AI services, consistent with prior findings. An evaluation of the two alternatives, ChatGPT and Gemini, showed that Gemini held a slight advantage over ChatGPT in terms of overall score. These findings provide empirical guidelines for service differentiation and strategic resource allocation in the design of future AI services.

▶ **Key words:** Generative AI service, Analytic Hierarchy Process, Fuzzy AHP, Priority weights

### [요 약]

본 연구는 생성형 AI 서비스 선택속성에 관한 선행연구를 확장하여 생성형 AI 서비스의 다양성 증가에 따라 복잡해지는 사용자 선택 환경을 보다 체계적으로 이해하기 위해, 기능적 속성, 경험적 속성, 적응적 속성, 브랜드 이미지의 4개 상위 차원과 정보 정확성, 설명 가능성, 즐거움, 의인화, 개인화, 창의성, 기업 명성, 기업 혁신성의 8개 하위 기준으로 구성된 계층적 의사결정 모델을 확장하여 분석하였다. 인간 판단에 내재된 모호성을 반영하여 분석의 견고성을 높이기 위해 퍼지 AHP를 적용하였다. 중요도 가중치 분석 결과, 정보 정확성이 생성형 AI 서비스의 핵심 가치로 확인되었다. ChatGPT와 Gemini 두 대안에 대한 평가에서는 종합 점수 기준으로 Gemini가 ChatGPT에 비해 근소한 우위를 보였다. 이러한 결과는 향후 AI 서비스 설계에 있어 서비스 차별화와 전략적 자원 배분을 위한 실증적 지침을 제공한다.

▶ **주제어:** 생성형 AI 서비스, 계층분석과정, 퍼지 AHP, 우선순위 가중치

• First Author: Yujin Kim, Corresponding Author: Hyung-Seok Lee

\*Yujin Kim (yjk824@naver.com), Dept. of Business Administration, Chungbuk National University

\*\*Hyung-Seok Lee (hyunglee@chungbuk.ac.kr), School of Business, Chungbuk National University

• Received: 2026. 08. 21, Revised: 2026. 09. 20, Accepted: 2026. 10. 06.

## I. Introduction

The contemporary generative AI service market is expanding at an unprecedented pace. Since ChatGPT amassed 100 million users worldwide within two months of its launch, competing services—including Google's Gemini, Anthropic's Claude, and xAI's Grok—have emerged in rapid succession, compelling users to choose among multiple high-performing services rather than a single option [1, 2, 3]. This shift indicates that the fundamental nature of decision-making surrounding generative AI has moved from a technology-adoption question of whether to use AI to a service-selection question of which AI to choose [4]. It further suggests that this selection process—encompassing not only functional characteristics such as information accuracy and response speed but also heterogeneous attributes such as creative output, anthropomorphic interaction, personalized experience, and brand trust—is inherently a multi-criteria decision-making (MCDM) problem [4, 5].

However, prior studies have yet to adequately address this issue. Much of the existing literature has drawn on technology acceptance frameworks such as the Technology Acceptance Model (TAM) and the Unified Theory of Acceptance and Use of Technology (UTAUT), focusing primarily on functional attributes such as perceived usefulness and ease of use, while experiential characteristics such as enjoyment and anthropomorphism have received comparatively little attention. In addition, structural equation modeling (SEM), the prevailing methodology in this literature, is well suited to identifying causal relationships but is structurally limited in quantitatively addressing the relative priority among multiple competing attributes. Although recent studies such as Kim and Lee [4] derived the relative weights of generative AI selection attributes using AHP and Fuzzy AHP, the attributes were treated as a single-level set, and the relative strengths of competing services were

not examined directly. Consequently, evidence on hierarchically structured attributes and service-level comparison remains limited.

To address this gap, the present study builds on the Fuzzy Analytic Hierarchy Process (Fuzzy AHP) approach of Kim and Lee [4] as its analytical framework. Conventional AHP systematically derives relative weights through pairwise comparisons and enables quantitative verification of judgment consistency, but because it constrains judgments to single crisp values, it cannot fully capture the linguistic ambiguity and uncertainty inherent in human judgment [6, 7]. Fuzzy AHP addresses this limitation by mapping pairwise comparison values onto triangular fuzzy numbers, making it particularly well suited to analyzing generative AI service selection, where user perceptions tend to be subjective and context-dependent.

Prior studies, such as Kim and Lee [4], identified six selection attributes for generative AI services—anthropomorphism, personalization, information accuracy, enjoyment, creativity, and response uncertainty—and applied AHP and Fuzzy AHP, establishing information accuracy as the highest-priority attribute. Building upon this foundational work, this study extends the previous research by addressing two directions suggested for future research: (1) expanding the hierarchical structure of selection attributes and (2) conducting direct alternative comparisons among individual services, and corroborates its key finding. Specifically, this study reorganizes the attributes of Kim and Lee [4] into four upper-level dimensions—functional, experiential, and adaptive attributes, along with brand image—and eight lower-level criteria. Five of the original attributes were retained (information accuracy, enjoyment, anthropomorphism, personalization, and creativity). Response uncertainty was excluded because it is a cost-type attribute, for which higher values denote lower preference, and is therefore incompatible with the benefit-criteria structure adopted here.

Explainability was added because the transparency of the reasoning underlying outputs is a core dimension of the functional attributes. Brand image, with its two criteria of corporate reputation and corporate innovativeness, was introduced because it has been established as a determinant of consumer response independent of the functional attributes. The derived weights are then applied to two real-world alternatives, ChatGPT and Gemini, to empirically investigate their comparative advantages—an aspect not addressed by Kim and Lee [4]. The key differences between the study by Kim and Lee [4] and the present study in terms of model, data, analysis, and results are summarized in Table 6.

Through this research design, the present study offers three contributions. Theoretically, it extends the single-level attribute set of Kim and Lee [4] into a two-level classification (four dimensions, eight criteria) tailored to the generative AI service selection context, and examines whether the top-priority status of information accuracy reported by Kim and Lee [4] holds under this extended structure. Methodologically, it extends Kim and Lee's [4] application of AHP and Fuzzy AHP by evaluating two real-world services as alternatives through criterion-level pairwise comparisons, and by adding a sensitivity analysis and a conventional AHP comparison to assess the effect of fuzzification. Practically, it supports rational decision-making by providing individual users and organizational decision-makers with concrete evaluative criteria grounded in attribute-level priorities, along with a comparative assessment of ChatGPT and Gemini as alternatives.

## II. Theoretical Background

### 1. Literature Review on Generative AI

Recently, as generative AI services such as ChatGPT and Gemini have rapidly permeated everyday life, scholarly interest has grown in

understanding the criteria on which users base their choice among competing services [1]. Scholars have found that this selection is not determined by technical performance alone, but constitutes a multilayered judgment process encompassing the reliability of information, the affective experience elicited during use, the adaptive characteristics of the service, and perceptions of the service provider [2, 4, 8].

Functional attributes have long been established as the most fundamental evaluative criterion. Holbrook [9] cross-classified consumer value along three dichotomies—extrinsic/intrinsic, self-oriented/other-oriented, and active/reactive—with extrinsic values such as efficiency and excellence underpinning the theoretical foundation for functional attributes centered on information accuracy, while Sweeney and Soutar [10] systematized this quality dimension through the PERVAL model. Shin and Park [11], from an algorithmic affordance perspective, decomposed transparency into understandability, explainability, and observability, demonstrating that explainability operates directly as a core dimension of transparency and thus qualifies as a sub-criterion of functional attributes. Sun et al. [12] similarly identified content reliability and *ex ante* transparency as key tasks underpinning information accuracy and explainability, respectively.

With respect to experiential attributes, researchers have shown that affective and psychological responses elicited during use exert an independent influence on service selection. Holbrook's [9] intrinsic values of play and aesthetics, together with the emotional value dimension in Sweeney and Soutar's [10] PERVAL model, provide the theoretical and measurement basis for enjoyment. Solakis et al. [13] identified hedonic motivation and anthropomorphism as key customer factors in AI-driven value co-creation, operating distinctly from functional dimensions, while Hollebeek et al. [14], reviewing 89 studies on

AI-based consumer engagement, identified anthropomorphism and personalization as technological antecedents and hedonic motivation as a key recurring variable. Doshi and Hauser [15] further provided empirical grounding by employing enjoyment as a measurement item in examining generative AI's effect on creative writing.

Adaptive attributes represent a dimension unique to generative AI, primarily discussed in terms of personalization and creativity. Vargo et al. [16], drawing on service-dominant logic, reconceptualized AI as an emergent property of adaptive service ecosystems and proposed the generative duality of adaptation and intelligence, providing the theoretical basis for classifying personalization and creativity as adaptive attributes. Mehmood et al. [17] showed that AI-based personalization simultaneously expands and constrains users' sense of self, underscoring personalization's complexity as an adaptive attribute, while Doshi and Hauser [15] demonstrated that generative AI enhances individual creativity but reduces collective diversity—a social dilemma further supported by Sun et al.'s [12] identification of creativity stimulation and personalized prompting as core elements of generative AI use.

With regard to brand image, corporate reputation and corporate innovativeness have emerged as independent selection criteria. Keller's [18] brand equity model established brand image as a core antecedent of consumer response, independent of functional attributes, and Walsh et al. [19] provided a five-dimensional measurement basis for corporate reputation. Henard and Dacin [20] found that a reputation for product innovation increases loyalty and tolerance for failure yet paradoxically reduces willingness to pay a premium, while Pappu and Quester [21] showed that brand innovativeness affects loyalty entirely through perceived quality—together justifying innovativeness as a sub-dimension of brand image. In a related study, Kim and Lee [4] applied AHP and

Fuzzy AHP to generative AI selection attributes and identified information accuracy as the highest-priority criterion.

Taken together, these studies consistently indicate that generative AI service selection involves a complex structure in which four dimensions—functional, experiential, adaptive, and brand image—operate jointly.

## 2. Analytic Hierarchy Process

The Analytic Hierarchy Process (hereafter AHP) is a methodology developed by Saaty [6] to address multi-criteria decision-making problems that entail the simultaneous consideration of multiple—and often conflicting—evaluation criteria. AHP decomposes a decision problem into a hierarchical structure consisting of a goal, criteria, sub-criteria, and alternatives, and systematically quantifies the relative importance of elements within the same hierarchical level through pairwise comparisons [6, 22]. Its principal strengths lie in its capacity to incorporate both qualitative judgments that resist numerical expression and quantitative data, as well as its ability to internally verify the logical consistency of judgments via the consistency ratio [23]. Owing to these characteristics, AHP has been widely applied across diverse managerial decision-making contexts, including service quality evaluation, technology selection, and supplier selection, making it a suitable methodology for analyzing the problem of generative AI service selection, in which multiple attributes must be weighed simultaneously [24].

The analytical procedure adopted in this study comprises five steps, as summarized in Table 1. First, the decision problem of generative AI service selection is defined, and the selection attributes derived from the literature review are organized into an upper- and lower-level hierarchy. Second, data are collected by eliciting the relative importance of factors within the same hierarchical level through pairwise comparisons using Saaty's nine-point scale [22]. Third, the consistency of

each respondent's pairwise comparison judgments is assessed by computing the Consistency Ratio (CR), and only responses satisfying a CR of 0.1 or below are retained for subsequent analysis. Fourth, the pairwise comparison matrices that satisfy the consistency requirement are mapped onto triangular fuzzy numbers and aggregated via the geometric mean method to derive the Fuzzy AHP weights, which explicitly account for linguistic ambiguity and uncertainty inherent in respondents' judgments, thereby overcoming the limitations of conventional AHP [25, 26]. Fifth, the derived attribute weights are applied to the two alternatives—ChatGPT and Gemini—to compute the overall score for each service and determine the most preferred alternative.

Table 1. Steps of the AHP and Fuzzy AHP

| Step   | Description                                                                                                                                                                                         |
|--------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Step 1 | Establish the research objective and decompose the generative AI service selection problem into a structured hierarchy comprising upper-level dimensions and lower-level criteria.                  |
| Step 2 | Conduct pairwise comparisons for all criteria at each hierarchical level to quantify their relative importance using Saaty's nine-point scale.                                                      |
| Step 3 | Assess the logical consistency of respondents' judgments by computing the Consistency Ratio (CR); responses with a CR exceeding 0.1 are excluded from subsequent analysis.                          |
| Step 4 | Convert the pairwise comparison matrices into triangular fuzzy numbers (TFNs) and aggregate them using the geometric mean method to derive the fuzzy AHP weights.                                   |
| Step 5 | Apply the derived fuzzy priority weights to evaluate and rank the two target alternatives (ChatGPT and Gemini), thereby identifying the relative preference between the two generative AI services. |

### III. Research Design

#### 1. Design of Generative AI Service Selection Factors

Selecting among competing services in the generative AI market constitutes a complex decision-making situation in which multiple heterogeneous attributes operate simultaneously rather than a single criterion. To systematically analyze this multi-criteria decision-making structure, this study employed Fuzzy AHP. Fuzzy AHP enables both the reorganization of the selection attributes of generative AI services—treated as a single-level set by Kim and Lee [4] or addressed in a fragmented manner in prior research—into a two-level hierarchical structure and the quantitative comparison of their relative importance.

The selection attributes were derived from the literature review presented above, and the classification scheme was refined and finalized through an iterative review process involving researchers and practitioners in generative AI-related fields to ensure content validity. Consequently, the selection attributes of generative AI services were classified into four upper-level dimensions—functional attributes, experiential attributes, adaptive attributes, and brand image—with eight lower-level criteria established as their respective sub-criteria, as presented in Figure 1. Functional attributes comprise information accuracy—the degree to which generated responses are factually correct and free from errors or

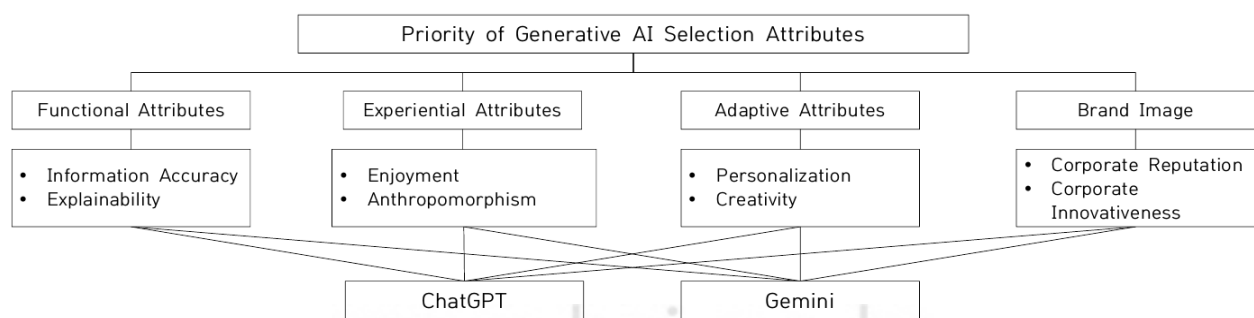


Fig. 1. Hierarchical Model of Generative AI Selection Attributes

hallucinations-and explainability-the extent to which the reasoning behind outputs is transparent and comprehensible to users. Experiential attributes comprise enjoyment-the intrinsic pleasure and positive affect experienced during use-and anthropomorphism-the tendency to attribute human-like characteristics to the service. Adaptive attributes comprise personalization-the capability to tailor responses to individual users' preferences and needs-and creativity-the ability to produce novel and valuable outputs beyond conventional information retrieval. Brand image comprises corporate reputation-users' overall evaluation of the provider's trustworthiness and reliability-and corporate innovativeness-users' perception of the provider's continuous technological advancement. The hierarchical structure in this study comprises two levels, encompassing upper-level dimensions and lower-level criteria, with each lower-level criterion formulated to directly evaluate the two alternatives-ChatGPT and Gemini. This design reflects the actual decision-making context encountered by users in generative AI service selection-namely, the process of comprehensively judging which of the two competing services performs better along each attribute dimension. Furthermore, all attributes were operationalized as benefit criteria, for which higher values denote greater preference, and a two-level hierarchy without additional sub-layers was maintained to minimize respondents' cognitive burden and to ensure clarity in comparative judgments.

## 2. Data Collection and Sample Characteristics

This study engaged an online panel research firm to administer a survey targeting users who had prior experience with generative AI services. The survey was administered over a period of approximately one week, from July 6 to July 13, 2026. Of the 200 responses collected, 141 were retained for the final analysis after excluding responses that did not satisfy the consistency

requirement for pairwise comparisons ( $CR > 0.1$ ).

The demographic characteristics and usage patterns of the respondents included in the analysis are presented in Table 2. Regarding gender, the sample comprised 74 women (52.5%) and 67 men (47.5%). With respect to age, respondents in their 20s accounted for the largest proportion (61.7%), followed by those in their 30s (14.2%), 50s or older (13.5%), and 40s (10.6%). Regarding occupation, the sample comprised employees (50.4%) and university students (49.6%). All respondents were required to have prior experience with both ChatGPT and Gemini, the two alternatives evaluated in this study.

Table 2. Sample Characteristics

| Category   |                    | n   | %     |
|------------|--------------------|-----|-------|
| Gender     | Male               | 67  | 47.5  |
|            | Female             | 74  | 52.5  |
| Age        | 20 ~ 29            | 87  | 61.7  |
|            | 30 ~ 39            | 20  | 14.2  |
|            | 40 ~ 49            | 15  | 10.6  |
|            | ≥ 50               | 19  | 13.5  |
| Occupation | University Student | 70  | 49.6  |
|            | Employee           | 71  | 50.4  |
| Total      |                    | 141 | 100.0 |

## 3. Priority Determination

This study employed Fuzzy AHP, which integrates fuzzy set theory into the AHP framework proposed by Saaty [6], to identify the relative importance of the factors influencing the selection of generative AI services.

To model the inherent vagueness and subjective uncertainty involved in this judgment process, fuzzy set theory introduced by Zadeh [27] is employed. Among various fuzzy numbers, Triangular Fuzzy Numbers (TFNs) are widely adopted to convert qualitative judgments into precise numerical representations. A TFN is conventionally specified by a triplet  $A = (l, m, u)$ , where  $l$ ,  $m$ , and  $u$  denote the lower bound, the most plausible value, and the upper bound of the fuzzy assessment, respectively, and the triangular membership function  $\mu_A(x)$  is formulated as Eq. (1):

$$\mu_A(x) = \begin{cases} 0 & x < l \\ \frac{x-l}{m-l} & l \leq x \leq m \\ \frac{u-x}{u-m} & m \leq x \leq u \\ 0 & x > u \end{cases} \quad (1)$$

In this study, each Saaty scale value  $k$  ( $k = 1, \dots, 9$ ) used in the pairwise comparisons was converted into a triangular fuzzy number  $(l, m, u) = (k - 1, k, k + 1)$ , with  $k = 1$  mapped onto  $(1, 1, 1)$ ; reciprocal comparisons were mapped onto  $(1/u, 1/m, 1/l)$ .

Based on this formulation, a nine-point scale was employed for the pairwise comparisons among evaluation criteria, and the judgment values elicited from individual respondents were mapped onto triangular fuzzy numbers and aggregated across respondents via the geometric mean method proposed by Buckley [28] to construct a single, integrated fuzzy judgment matrix reflecting the group's collective judgment. This study follows Buckley's procedure: the row geometric mean of the group fuzzy matrix is normalized through fuzzy arithmetic to obtain fuzzy priority weights, which are then converted into crisp weights via the defuzzification procedure detailed below.

To obtain a non-fuzzy, representative scalar value for further decision-making, the aggregated fuzzy numbers were converted into crisp values. This study adopted the Center of Gravity defuzzification technique, which calculates the geometric centroid of the fuzzy membership area, as expressed in Eq. (2) [29]:

$$A = \frac{l + m + u}{3} \quad (2)$$

The relative weight of each evaluation criterion was derived from the aggregated fuzzy judgment matrix through the defuzzification process. The consistency ratio (CR) was verified in two stages. First, the CR was calculated for the individual judgment matrix of each respondent, and responses with a CR exceeding 0.1 were excluded to ensure logical consistency. Second, the retained responses were aggregated via the geometric mean method to construct a single, integrated group

judgment matrix, and the CR of this matrix, computed from its defuzzified values, was also verified. Because each lower-level comparison matrix involves only two criteria, it is a  $2 \times 2$  reciprocal matrix whose CR is mathematically equal to zero regardless of the judgment values; the substantive consistency check therefore applies to the comparison matrix of the four upper-level dimensions. The Fuzzy AHP weight computation was performed using Microsoft Excel.

For the alternative evaluation, respondents performed pairwise comparisons of ChatGPT and Gemini for each of the eight lower-level criteria, using the same nine-point scale applied in the criteria comparisons. The collected judgments were converted into the local weights of the two services for each criterion through the same procedure: conversion into triangular fuzzy numbers, aggregation via the geometric mean method, and defuzzification via the center of gravity method. The overall score of each service was then computed by summing, across the eight criteria, the product of each criterion's overall weight and the service's local weight for that criterion.

## IV. Results

### 1. Analysis of Weights and Priorities

The results of the relative importance analysis of the selection attributes of generative AI services are presented in Tables 3 and 4. With respect to the upper-level attributes, functional attributes received the highest weight (0.525), emerging as the most important factor among the four attribute dimensions, followed by experiential attributes, adaptive attributes, and brand image. The weight of functional attributes was approximately 2.5 times that of experiential attributes, the second-ranked dimension, indicating that the utilitarian value directly linked to task performance—such as the accuracy and reliability of information—is given substantially greater priority than any other factor

Table 3. Fuzzy AHP Weights and Ranking of Higher Criteria

| Criteria          | Functional Attributes                             | Experiential Attributes | Adaptive Attributes | Brand Image |
|-------------------|---------------------------------------------------|-------------------------|---------------------|-------------|
| Fuzzy AHP Weights | 0.525                                             | 0.212                   | 0.172               | 0.091       |
| Rank              | 1                                                 | 2                       | 3                   | 4           |
| Consistency       | $\lambda_{\max} = 4.033$ , CI = 0.011, CR = 0.012 |                         |                     |             |

Table 4. Fuzzy AHP Weights and Ranking of Sub-Criteria for Generative AI Service Selection

| Higher Criteria            | Lower Criteria           | Fuzzy Weights | Overall Weight | Overall Rank | ChatGPT (Local Weight) | Gemini (Local Weight) | ChatGPT (Contribution Score) | Gemini (Contribution Score) |
|----------------------------|--------------------------|---------------|----------------|--------------|------------------------|-----------------------|------------------------------|-----------------------------|
| Functional Attributes      | Information Accuracy     | 0.809         | 0.425          | 1            | 0.397                  | 0.603                 | 0.1688                       | 0.2560                      |
|                            | Explainability           | 0.191         | 0.101          | 4            | 0.443                  | 0.557                 | 0.0445                       | 0.0561                      |
| Experiential Attributes    | Enjoyment                | 0.531         | 0.112          | 2            | 0.655                  | 0.345                 | 0.0737                       | 0.0388                      |
|                            | Anthropomorphism         | 0.469         | 0.099          | 5            | 0.595                  | 0.405                 | 0.0591                       | 0.0401                      |
| Adaptive Attributes        | Personalization          | 0.587         | 0.101          | 3            | 0.502                  | 0.498                 | 0.0508                       | 0.0504                      |
|                            | Creativity               | 0.413         | 0.071          | 6            | 0.479                  | 0.521                 | 0.0341                       | 0.0371                      |
| Brand Image                | Corporate Reputation     | 0.328         | 0.030          | 8            | 0.503                  | 0.497                 | 0.0150                       | 0.0148                      |
|                            | Corporate Innovativeness | 0.672         | 0.061          | 7            | 0.467                  | 0.533                 | 0.0285                       | 0.0325                      |
| Evaluation of Alternatives |                          |               |                |              |                        |                       | 0.4742                       | 0.5258                      |

Table 5. Sensitivity Analysis of Alternative Evaluation Scores ( $\pm 20\%$  Variation in Upper-Level Dimension Weights)

| Scenario                     | Varied Dimension        | Change | ChatGPT Score | Gemini Score | Gap (Gemini–ChatGPT) | Preferred Alternative |
|------------------------------|-------------------------|--------|---------------|--------------|----------------------|-----------------------|
| Baseline (no change)         | -                       | -      | 0.474         | 0.526        | 0.052                | Gemini                |
| Functional Attributes –20%   | Functional Attributes   | –20%   | 0.489         | 0.511        | 0.022                |                       |
| Functional Attributes +20%   |                         | +20%   | 0.459         | 0.541        | 0.082                |                       |
| Experiential Attributes –20% | Experiential Attributes | –20%   | 0.466         | 0.534        | 0.068                |                       |
| Experiential Attributes +20% |                         | +20%   | 0.482         | 0.518        | 0.036                |                       |
| Adaptive Attributes –20%     | Adaptive Attributes     | –20%   | 0.473         | 0.527        | 0.054                |                       |
| Adaptive Attributes +20%     |                         | +20%   | 0.475         | 0.525        | 0.050                |                       |
| Brand Image –20%             | Brand Image             | –20%   | 0.474         | 0.526        | 0.052                |                       |
| Brand Image +20%             |                         | +20%   | 0.474         | 0.526        | 0.052                |                       |

when selecting a generative AI service. This suggests that users still perceive generative AI services primarily as reliable task-performing tools rather than as sources of emotional experience or symbolic value. Conversely, the fact that adaptive attributes—which capture a dimension unique to generative AI—were rated lower than experiential attributes suggests that personalization and creativity function as secondary differentiating factors that become salient only after the fundamental trust requirement of accuracy has been met. The lowest ranking of brand image indicates that, because both ChatGPT and Gemini

are services provided by well-established major technology firms, actual performance differences, rather than corporate reputation, serve as the primary determinant of selection. The consistency of the judgments was confirmed at both stages. Of the 200 respondents, 141 satisfied  $CR \leq 0.1$  and were retained for analysis. For the integrated group matrix of the four upper-level dimensions (Table 3),  $\lambda_{\max} = 4.033$ , CI = 0.011, and CR = 0.012, well below the 0.1 threshold. The lower-level comparison matrices yielded CR = 0.000, as expected for  $2 \times 2$  matrices.

In terms of the overall weights of the lower-level

Table 6. Comparison of Kim and Lee [4] and the Present Study

| Dimension | Kim and Lee [4]                                                                                                                                                                                          | Present Study                                                                                                                                                                                                                                                                          |
|-----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Model     | Six single-level attributes (anthropomorphism, personalization, information accuracy, enjoyment, creativity, response uncertainty)                                                                       | Four upper-level dimensions and eight lower-level criteria in a two-level hierarchy (functional, experiential, adaptive, brand image → information accuracy, explainability, enjoyment, anthropomorphism, personalization, creativity, corporate reputation, corporate innovativeness) |
| Data      | 310 responses collected, 77 valid ( $CR \leq 0.1$ ); January 25-31, 2026; online panel                                                                                                                   | 200 responses collected, 141 valid ( $CR \leq 0.1$ ); July 6-13, 2026; online panel                                                                                                                                                                                                    |
| Analysis  | AHP and Fuzzy AHP (geometric mean method); Mann-Whitney U test for weight differences by service type and user group                                                                                     | Fuzzy AHP (Buckley's geometric mean method) with conventional AHP conducted in parallel; alternative evaluation via pairwise comparison of ChatGPT and Gemini; sensitivity analysis ( $\pm 20\%$ weight variation with renormalization)                                                |
| Results   | Information accuracy ranked first (AHP: 0.339; Fuzzy AHP: 0.291); attribute weight structures differed significantly by service type (ChatGPT vs. Gemini users) and by user group (students vs. workers) | Information accuracy ranked first (overall weight: 0.425); Gemini slightly preferred over ChatGPT (0.526 vs. 0.474); Gemini's advantage remained robust across all sensitivity scenarios                                                                                               |

criteria, information accuracy received a weight of 0.425, ranking first among the eight criteria by a substantial margin, followed by enjoyment, personalization, explainability, anthropomorphism, creativity, corporate innovativeness, and corporate reputation. The fact that information accuracy was rated considerably higher than explainability, which belongs to the same functional attributes dimension, indicates that users prioritize the reliability of the response itself—that is, whether the output is correct—over an explanation of why a given output was generated.

Moreover, the fact that personalization (ranked third overall) was slightly higher than explainability (ranked fourth overall) indicates that personalization functions as a potent differentiating factor at the individual level. Within brand image, corporate innovativeness was rated higher than corporate reputation, suggesting that, in the rapidly evolving generative AI industry, innovation capability carries greater weight than reputation in building user trust.

Meanwhile, to verify the practical effect of fuzzification on the analytical results, a conventional AHP analysis was conducted in parallel using the same 141 responses. The results showed that the weights for the upper-level dimensions were 0.527 for functional attributes, 0.211 for experiential attributes, 0.172 for adaptive

attributes, and 0.090 for brand image. These rankings and point estimates are nearly identical to the Fuzzy AHP results (0.525 / 0.212 / 0.172 / 0.091), yielding an identical consistency ratio (CR) of 0.012.

This pattern was similarly confirmed for both the eight lower-level criteria and the final alternative evaluation results. The conventional AHP global weights for the eight lower-level criteria were 0.426 for information accuracy, 0.112 for enjoyment, 0.101 for personalization, 0.100 for explainability, 0.099 for anthropomorphism, 0.071 for creativity, 0.060 for corporate innovativeness, and 0.030 for corporate reputation. These rankings match those of the Fuzzy AHP results (0.425 / 0.112 / 0.101 / 0.101 / 0.099 / 0.071 / 0.061 / 0.030), with point estimates differing by at most 0.001. Furthermore, the overall contribution scores of ChatGPT and Gemini, the central finding of this study, were also nearly identical under conventional AHP (0.474 vs. 0.526) and Fuzzy AHP (0.474 vs. 0.526), identical to the third decimal place. This confirms that Gemini's modest advantage observed in this study is a robust finding regardless of whether fuzzification is applied.

This similarity arises because the most plausible value ( $m$ ) of each triangular fuzzy number equals the original Saaty judgment. Consequently, the

aggregated  $m$  values coincide with the geometric mean of the conventional judgments, and the defuzzified value obtained via the center of gravity method lies close to  $m$ . Because the triangular fuzzy numbers were generated with a fixed width of  $\pm 1$  around each judgment, fuzzification did not alter the point estimates or the ranking of the weights in this study. The contribution of fuzzification is to express each weight as an interval that reflects the imprecision of the nine-point scale. Accordingly, the conclusions of this study do not depend on fuzzification, and Fuzzy AHP serves here as a robustness check on the conventional AHP result rather than as a source of different conclusions.

## 2. Analysis of Alternative Evaluation

The results of the overall evaluation, derived by applying the attribute weights to the two alternatives, ChatGPT and Gemini, are presented in Table 4. The analysis showed that Gemini obtained an overall score of 0.526 and ChatGPT obtained an overall score of 0.474, indicating a slight advantage for Gemini over ChatGPT.

The difference in overall scores between the two services was relatively small (0.052). This suggests that this gap does not reflect absolute superiority of one service over the other, but rather indicates that each service competes in the market by leveraging a distinct set of strengths.

To ensure transparency and reproducibility in the evaluation process, this study presents the local weights and contribution scores of ChatGPT and Gemini for each of the eight lower-level criteria in Table 4. The results reveal domain-specific strengths for each service. Within the functional attributes, Gemini holds a clear advantage in information accuracy (0.603 vs. ChatGPT's 0.397) and explainability (0.557 vs. 0.443). Conversely, within the experiential attributes, ChatGPT shows a marked advantage in enjoyment (0.655 vs. Gemini's 0.345) and anthropomorphism (0.595 vs. 0.405). For the adaptive attributes (personalization, creativity) and

the brand-image criteria, the local weights of the two services are largely comparable, ranging from 0.47 to 0.53. Because information accuracy carries the largest overall weight by a substantial margin (0.425), Gemini's advantage in this single criterion yields a contribution score gap of 0.087 (0.2560 vs. 0.1688). This lead is partly offset by ChatGPT's advantage in the experiential attributes, which amounts to 0.054 (enjoyment: 0.035; anthropomorphism: 0.019), while Gemini's smaller advantages in explainability, creativity, and corporate innovativeness add 0.019 in total. As a result, the net gap in the overall score is 0.052.

To assess the robustness of the alternative evaluation results, a sensitivity analysis was conducted by varying the weights of the four upper-level dimensions by  $\pm 20\%$  individually and renormalizing the remaining three dimensions to maintain their original relative proportions. As presented in Table 5, Gemini consistently maintains a higher composite score than ChatGPT across all eight extreme scenarios (4 dimensions  $\times$   $\pm 20\%$ ). The score gap between the two services ranges from a minimum of 0.022 to a maximum of 0.082, both occurring under variation in the functional attributes dimension—the dimension with the largest baseline weight. These results indicate that Gemini's observed superiority is not an accidental outcome driven by estimation errors in specific attribute weights, but rather a robust result that remains stable within a reasonable range of weight structure fluctuations. To clarify the contributions of this study relative to the most closely related prior work, Table 6 provides a side-by-side comparison of the study by Kim and Lee [4] and the present study.

## V. Conclusions

### 1. Summary and Implications

Extending the work of Kim and Lee [4], this study employed Fuzzy AHP in a two-level hierarchy to identify the relative importance of the attributes

that users consider when selecting a generative AI service, and to compare ChatGPT and Gemini directly.

First, among the upper-level dimensions, functional attributes showed the highest weight at 0.525, identifying this dimension as the most important selection criterion among the four attribute dimensions. This is closely related to the fact that generative AI, being based on the probabilistic generation mechanism of large language models, inherently entails the risk of hallucination, and that users are often unable to independently verify the accuracy of the generated output; they are thus positioned to first assess whether the output can be trusted before applying it to actual tasks or decisions.

Second, among the lower-level criteria, information accuracy received an overall weight of 0.425, ranking first by a substantial margin. The substantially higher rating of information accuracy than of explainability—both criteria within the functional attributes dimension—reflects users' persistent concern regarding hallucination. The low ranking of anthropomorphism can be interpreted in the same light, suggesting that users place greater importance on the trustworthiness of the output than on emotional attachment to the AI. This corroborates the primacy of information accuracy reported in Kim and Lee [4], under a different attribute structure and with a different sample.

Third, the comparative evaluation showed that Gemini obtained an overall score of 0.526 against ChatGPT's 0.474, indicating a slight advantage for Gemini. This narrow gap suggests that neither service is uniformly superior to the other. Taken together, these findings demonstrate that the selection of generative AI services is not governed by a single criterion but constitutes a complex structure in which experiential and adaptive values are hierarchically layered atop the basic trust requirement of accuracy. This underscores the need for phased differentiation strategies.

These findings offer important academic and

practical implications. Academically, prior research on generative AI services has largely relied on causal frameworks (e.g., TAM, UTAUT) to test whether specific attributes such as perceived usefulness influence usage intention, leaving limited evidence on how the relative weights of attributes are structured hierarchically and how competing services compare on those attributes. Although Kim and Lee [4] derived attribute weights using AHP and Fuzzy AHP, the overall evidence on hierarchical weight structures and cross-service comparisons remains limited. By structuring the attributes into a two-level hierarchy of four upper-level dimensions and eight lower-level criteria, and comparing two services directly, this study extends the AHP-based work of Kim and Lee [4]. The results corroborate the centrality of information accuracy, as reported by Kim and Lee [4], and show that generative AI selection attributes form a layered structure rather than a flat, parallel set. Although the point estimates in this study coincided with those of conventional AHP, Fuzzy AHP additionally provides uncertainty bounds.

First, from a practical standpoint, given that information accuracy was identified as the decisive criterion, service providers should prioritize enhancing models' fact-verification capabilities and providing supporting evidence for outputs—an investment in reducing user uncertainty rather than merely improving response quality. Second, given the narrow overall score gap between the two services, each provider should reinforce a positioning strategy centered on its relative strengths while gradually improving weaker attributes. Finally, the attribute-level priorities and overall evaluation results identified here can serve as a practical basis for users to rationally select the service best suited to their intended purposes.

## 2. Limitations and Future Research

Despite these academic and practical contributions, this study has several limitations.

First, this study analyzed the selection attributes

of generative AI services in an integrated manner, irrespective of users' specific purposes—such as learning, work assistance, creative work, or decision support—even though attribute importance may vary by usage context. Future research should segment samples by purpose of use to examine how attribute priorities differ across usage contexts.

Second, this study is limited in that it set only ChatGPT and Gemini as alternatives, excluding competing services such as Anthropic's Claude, which has secured a substantial user base and offers agentic capabilities beyond simple response generation—limiting the generalizability of the alternative evaluation results. Given the rapid pace of change in this technology, service performance and features may also have shifted since the survey. Future research should include Claude and other emerging agentic services as comparison targets and specify version details at the time of the survey.

Third, this study has a limitation in that it did not individually verify or require respondents to specify the exact model versions of ChatGPT and Gemini used during the survey. Because both services operate by automatically applying the latest versions to general users, it can reasonably be inferred that respondents utilized the versions available as of the survey period in July 2026 (ChatGPT: GPT-5.5, subsequently updated to GPT-5.6 during the survey; Gemini: Gemini 3.5); however, this inference was not stringently validated at the individual respondent level. In particular, given that ChatGPT's model version was updated during the one-week survey period, it cannot be ruled out that evaluations of different versions were confounded depending on when respondents submitted their responses. Consequently, future research should address this methodological limitation by incorporating procedures that require respondents to explicitly state the exact model version used at the time of the survey.

Fourth, due to the response burden inherent in AHP-based methods, a total of 59 out of 200 responses failed to satisfy the consistency ratio criterion and were excluded, resulting in a final valid sample size of 141. While such exclusions ensure the logical consistency and reliability of the Fuzzy AHP analysis, future research should explore ways to optimize response consistency, such as refining the scale structure or improving survey administration techniques.

Fifth, this study did not account for the effects of individual characteristics—such as users' proficiency or experience with generative AI—on the priority structure; incorporating such variables would allow more in-depth interpretation.

Finally, the sample of this study is limited in its ability to fully represent the broader population of generative AI users, as individuals in their 20s account for the majority, at 61.7%, and the occupational distribution is concentrated primarily among employees and university students. Because the evaluation attributes users prioritize when selecting services may vary depending on age and occupational characteristics, future research should verify the generalizability of these findings using more balanced samples that include a broader range of age groups and occupations.

Notwithstanding these limitations, this study quantitatively identifies the complex selection attributes of generative AI services using Fuzzy AHP. By determining both attribute priorities and alternative performances, it provides a solid foundation for service design, positioning strategies, and future research on user behavior.

## REFERENCES

- [1] J. Peng, C. Jiang, X. Zhang, W. Guan, and Y.-H. Lin, "How the artificial intelligence industry is growing: A comparison of the current development of ChatGPT's Chinese and American 'fast followers' Ernie Bot and Google Gemini," *Finance & Economics*, Vol. 1, pp. 1-8, July 2025. DOI: 10.61173/4b553551

- [2] J. Yeon, Y. Jung, Y. Baek, D. Lee, J. Shin, and W. Y. Chung, "User preferences on a generative AI user interface through a choice experiment," *International Journal of Human-Computer Interaction*, Vol. 41, No. 12, pp. 7626-7637, September 2024. DOI: 10.1080/10447318.2024.2400379
- [3] S. Gupta, A. Arora, S. Singh, and J. Jain, "AI feel millennials: Prioritizing the intentions towards adoption of AI-enabled chatbots using fuzzy-AHP approach," *Journal of Science and Technology Policy Management*, Vol. ahead-of-print, No. ahead-of-print, February 2025. DOI: 10.1108/JSTPM-09-2023-0159
- [4] Y. Kim and H.-S. Lee, "Analyzing the Priority of Selection Attributes for Generative AI Services: An Application of AHP and Fuzzy AHP," *Journal of the Korea Society of Computer and Information*, Vol. 31, No. 5, pp. 251-261, May 2026. DOI: 10.9708/jksci.2026.31.05.251
- [5] Y. Kim and H.-S. Lee, "Factors Affecting the Continuous Intention to Use ChatGPT: Evidence from Korea and USA," *The Journal of Korean Institute of Communications and Information Sciences*, Vol. 51, No. 1, pp. 146-168, January 2026. DOI: 10.7840/kics.2026.51.1.146
- [6] T. L. Saaty, "*The Analytic Hierarchy Process*," McGraw-Hill, pp. 52-64, 1980.
- [7] D.-Y. Chang, "Applications of the Extent Analysis Method on Fuzzy AHP," *European Journal of Operational Research*, Vol. 95, No. 3, pp. 649-655, December 1996. DOI: 10.1016/0377-2217(95)00300-2
- [8] Y. Kim and H.-S. Lee, "Effects of Generative AI Characteristics on User Satisfaction and Trust: AI Literacy as a Moderator," *Journal of the Korea Society of Computer and Information*, Vol. 30, No. 12, pp. 253-263, December 2025. DOI: 10.9708/jksci.2025.30.12.253
- [9] M. B. Holbrook, "*Consumer Value: A Framework for Analysis and Research*," Routledge, pp. 91-97, 1999.
- [10] J. C. Sweeney and G. N. Soutar, "Consumer Perceived Value: The Development of a Multiple Item Scale," *Journal of Retailing*, Vol. 77, No. 2, pp. 203-220, June 2001. DOI: 10.1016/S0022-4359(01)00041-0
- [11] D. Shin and Y. J. Park, "Role of Fairness, Accountability, and Transparency in Algorithmic Affordance," *Computers in Human Behavior*, Vol. 98, pp. 277-284, September 2019. DOI: 10.1016/j.chb.2019.04.019
- [12] Y. Sun, E. Jang, F. Ma, and T. Wang, "Generative AI in the wild: Prospects, challenges, and strategies," *Proc. 2024 CHI Conf. Human Factors in Computing Systems*, pp. 1-16, Honolulu, HI, USA, May 2024. DOI: 10.1145/3613904.3642160
- [13] K. Solakis, V. Katsoni, A. B. Mahmoud, and N. Grigoriou, "Factors Affecting Value Co-creation through Artificial Intelligence in Tourism: A General Literature Review," *Journal of Tourism Futures*, Vol. 10, No. 1, pp. 116-130, May 2022. DOI: 10.1108/JTF-06-2021-0157
- [14] L. D. Hollebeek, C. Menidjel, M. Sarstedt, J. Jansson, and S. Urbonavicius, "Engaging Consumers through Artificially Intelligent Technologies: Systematic Review, Conceptual Model, and Further Research," *Psychology & Marketing*, Vol. 41, No. 4, pp. 880-898, April 2024. DOI: 10.1002/mar.21957
- [15] A. R. Doshi and O. P. Hauser, "Generative AI enhances individual creativity but reduces the collective diversity of novel content," *Science Advances*, Vol. 10, No. 28, p. eadn5290, July 2024. DOI: 10.1126/sciadv.adn5290
- [16] S. L. Vargo, H. Wieland, and M. A. Akaka, "Service-dominant logic in the AI era," *Journal of Service Research*, Vol. 29, No. 3, pp. 360-376, August 2026. DOI: 10.1177/10946705261430147
- [17] K. Mehmood, K. Verleye, A. De Keyser, and B. Lariviere, "The Transformative Potential of AI-enabled Personalization across Cultures," *Journal of Services Marketing*, Vol. 38, No. 6, pp. 711-730, May 2024. DOI: 10.1108/JSM-08-2023-0286
- [18] K. L. Keller, "Conceptualizing, Measuring, and Managing Customer-Based Brand Equity," *Journal of Marketing*, Vol. 57, No. 1, pp. 1-22, January 1993. DOI: 10.1177/002224299305700101
- [19] G. Walsh, V. W. Mitchell, P. R. Jackson, and S. E. Beatty, "Examining the Antecedents and Consequences of Corporate Reputation: A Customer Perspective," *British Journal of Management*, Vol. 20, No. 2, pp. 187-203, June 2009. DOI: 10.1111/j.1467-8551.2007.00557.x
- [20] D. H. Henard and P. A. Dacin, "Reputation for Product Innovation: Its Impact on Consumers," *Journal of Product Innovation Management*, Vol. 27, No. 3, pp. 321-335, May 2010. DOI: 10.1111/j.1540-5885.2010.00719.x
- [21] R. Pappu and P. G. Quester, "How Does Brand Innovativeness Affect Brand Loyalty?," *European Journal of Marketing*, Vol. 50, No. 1/2, pp. 2-28, February 2016. DOI: 10.1108/EJM-01-2014-0020
- [22] T. L. Saaty, "A Scaling Method for Priorities in Hierarchical Structures," *Journal of Mathematical Psychology*, Vol. 15, No. 3, pp. 234-281, June 1977. DOI: 10.1016/0022-2496(77)90033-5
- [23] H.-S. Lee, "A Study on the Priority of Family Restaurant Selection Determinants," *Journal of Foodservice Management*, Vol. 9, No. 2, pp. 7-26, June 2006.
- [24] M. Tavana, M. Soltanifar, and F. J. Santos-Arteaga, "Analytical Hierarchy Process: Revolution and Evolution," *Annals of Operations Research*, Vol. 326, No. 2, pp. 879-907, July 2023. DOI: 10.1007/s10479-021-04432-2
- [25] Y. Liu, C. M. Eckert, and C. Earl, "A review of fuzzy AHP methods for decision-making with subjective judgements," *Expert Systems with Applications*, Vol. 161, Article 113738, December 2020. DOI: 10.1016/j.eswa.2020.113738
- [26] S. E. Helmy, G. H. Eladl, and M. Eisa, "Fuzzy Analytical

- Hierarchy Process (FAHP) Using Geometric Mean Method to Select Best Processing Framework Adequate to Big Data," *Journal of Theoretical and Applied Information Technology*, Vol. 99, No. 1, pp. 207-226, January 2021.
- [27] L. A. Zadeh, "Fuzzy sets," *Information and Control*, Vol. 8, No. 3, pp. 338-353, 1965. DOI: 10.1016/S0019-9958(65)90241-X
- [28] J. J. Buckley, "Fuzzy Hierarchical Analysis," *Fuzzy Sets and Systems*, Vol. 17, No. 3, pp. 233-247, December 1985. DOI: 10.1016/0165-0114(85)90090-9
- [29] G. J. Klir and B. Yuan, "*Fuzzy Sets and Fuzzy Logic: Theory and Applications*," Prentice Hall, 1995.

## Authors



Yujin Kim received the B.B.A. degree in Management Information Systems and the Ph.D. degree in Operations Management from Chungbuk National University, Korea, in 2018 and 2026, respectively.

Dr. Kim is currently a Lecturer of the School of Business at Chungbuk National University, Korea. Her research interests include service operations management, digital service innovation, and information technology policy, and the application of quantitative methodologies to decision-making problems.



Hyung-Seok Lee received the B.E. degree from Kwangwoon University in 1996 and the M.S. and Ph.D. degrees in Operations Management from Korea University, Korea, in 2000 and 2003, respectively.

Dr. Lee joined the faculty of the School of Business at Chungbuk National University, Korea, in 2011, where he is currently a Professor. His research interests include service operations management, digital service innovation, service quality, and the application of quantitative methodologies to decision-making problems.