systems

MDPI

*Article*

# Evaluating the Quality of AI-Generated Digital Educational Resources for University Teaching and Learning

**Qian Huang [1], Chunlan Lv [2], Li Lu [3] and Shuang Tu [4,*]**

[1] College of Management Science, Chengdu University of Technology, Chengdu 610059, China; hqian@cdut.edu.cn
[2] School of Economics and Management, Yibin University, Yibin 644007, China; lvchunlan@yibinu.edu.cn
[3] Tourism School, Sichuan University, Chengdu 610065, China; lulirudy@scu.edu.cn
[4] School of Health Management, Xihua University, Chengdu 610039, China
* Correspondence: 0720150019@mail.xhu.edu.cn

**Abstract:** With the proliferation of artificial intelligence in education, AI-generated digital educational resources are increasingly being employed as supplements for university teaching and learning. However, this raises concerns about the quality of the content produced. To conduct a comprehensive quality assessment, this paper presents an evaluation index system for AI-generated digital educational resources by combining the Delphi method and the Analytic Hierarchy Process. The initial quality indicators across the dimensions of content, expression, and user and technical aspects are identified through a systematic literature review of the recent research. Then, the Delphi method is utilized to modify the quality indicators according to experts' opinions through two rounds of questionnaire surveys. Subsequently, the weight coefficients of the quality indicators are calculated using the Analytic Hierarchy Process. Finally, a quality indicator system for evaluating AI-generated digital educational resources is developed, which comprises four dimensions and twenty indicators. The findings reveal that content characteristics are of critical importance in assessing the quality of AI-generated educational resources, followed by expression characteristics as the second most significant factor, with user and technical characteristics also being recognized. Among the second-level indicators, "authenticity", "accuracy", "legitimacy", and "relevance" are accorded greater importance relative to other indicators. The proposed system equips relevant stakeholders with a framework for selecting high-quality AIGDERs and steering AI tools in line with educational standards. Finally, some implications are provided to support the selection of high-quality AI-generated resources and guidance on aligning these resources with educational standards.

**Keywords:** AI-generated content; digital educational resources; quality; evaluation index system; AHP; Delphi method

check for updates

Academic Editor: Timothy Ferris

Received: 1 February 2025
Revised: 27 February 2025
Accepted: 2 March 2025
Published: 3 March 2025

**Citation:** Huang, Q.; Lv, C.; Lu, L.; Tu, S. Evaluating the Quality of AI-Generated Digital Educational Resources for University Teaching and Learning. *Systems* **2025**, *13*, 174. https://doi.org/10.3390/systems13030174

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

In recent years, artificial intelligence (AI) has achieved significant advancements, leading to transformations across various industries and facets of human life [1]. One major area where AI has had a profound impact is content creation. With advancements and breakthroughs in generative language models, artificial-intelligence-generated content (AIGC) that supplements conventional content creation methods like user-generated content (UGC) and professional-generated content (PGC) has become a key focus worldwide [2,3]. AIGC refers to a novel approach to the creation of content, such as text, images, videos, and other forms of data, through artificial intelligence algorithms [4]. AIGC is achieved by extracting

intent information from instructions provided by humans and generating the content according to its knowledge and the intent information [5]. This streamlined content creation process enables faster and more efficient generation of high-quality content [6]. Prominent research organizations and companies in artificial intelligence, such as DeepSeek, OpenAI, and DeepMind, have prioritized the development of AIGC [7]. ChatGPT-3.5, a generative AI chatbot developed by OpenAI and launched in 2022, reached 1 million users within a week of its launch [8]. DeepSeek R1, a surprisingly efficient and powerful Chinese AI model, has drawn significant attention for its outstanding price–quality ratio and fully open-source nature.

AIGC has brought profound transformations to various societal contexts, with university education demonstrating substantial influence [9]. Numerous studies have demonstrated the important role of AIGC in university teaching and learning, such as providing wide varieties of customized educational resources, improving learning experiences, and enhancing educational effectiveness [10–12]. Digital educational resources, one of the pivotal components in modern university education, have increasingly been shaped by the influence of AIGC. Educational resources generated by AI technology and operated on varieties of digital platforms are called artificial-intelligence-generated digital educational resources (AIGDERs). They have the potential to promote inclusive and efficient education in universities. AIGDERs enable personalized learning by analyzing undergraduates' abilities and learning styles to provide personalized feedback and recommendations for continuous improvement [13]. AIGDERs are also a boon for rapidly evolving fields as they can quickly incorporate the latest research and knowledge, keeping undergraduates well informed. Additionally, AIGDERs can create engaging learning experiences through interactive elements like simulations and virtual reality, fostering the innovative consciousness and creative abilities of the undergraduates [14]. Although AIGDERs provide many benefits to university education, they concurrently pose potential risks and challenges.

Cognitive biases represent a significant concern in the application of AIGDERs [14,15]. AIGDERs utilize probabilistic models to identify patterns within textual data and automatically generate content based on syntactic rules. However, since AIGDERs fail to comprehend language semantics, they may readily generate content that is fraught with inaccuracy, bias, or unfairness [16]. A new term, "hallucination", is used to describe the occurrence of disinformation in AIGDERs [17]. According to data from GitHub, the ChatGPT series of platforms exhibit a 3% to 3.5% error rate, the LLaMA series of platforms have an error rate of 5.1% to 5.9%, and Google's PaLM platform has a notably higher error rate of 12.1% [18]. Consequently, AIGDERs can inevitably lead to cognitive biases [19]. Scholars have emphasized the necessity of cautious engagement with AIGC platforms. In addition, certain researchers have devised mathematical models to comprehensively assess both their hallucinatory tendencies and creative capabilities. However, the absence of standardized criteria and methodologies for evaluating AIGDER systems has led to inconsistent practices and operational inefficiencies. Therefore, there is an immediate need to establish a comprehensive evaluation system for AIGDERs.

To ensure the quality of AIGDERs in the context of the emerging digital world, this study proposes a framework for evaluating their quality by integrating the Delphi method with the Analytic Hierarchy Process (AHP). While Delphi seeks a convergence of opinions among a group of experts, the AHP helps to solve complex decision problems using a hierarchical structure. The integration of these two methods assists in systematically identifying the evaluation criteria and then setting priorities among them. The procedures are as follows. Through a comprehensive literature review, various dimensions and indicators consisting of content, expression, and user and technical aspects are identified. Subsequently, a two-round questionnaire is disseminated among a panel of experts to affirm the

validity of each indicator via the Delphi method. Then, the AHP is employed to determine the weights of the identified quality indicators, resulting in an evaluation index system that includes four dimensions and twenty indicators. The objective of this paper is to select key indicators that can inform the development of strategies for stakeholders involved in AIGDERs.

The remainder of this paper is organized as follows. Section 2 provides a comprehensive summary of the related literature. Section 3 describes and details the methodology. Section 4 presents the research results. Section 5 presents a discussion and some implications. Section 6 provides the main conclusions.

## 2. Literature Review

Many previous studies have analyzed the quality of AIGDERs and the evaluation criteria of digital educational resources.

### 2.1. The Quality of AIGDERs

As AIGC is increasingly widely used in the field of education, the quality of AIGDERs has attracted more research interest. For example, Denny et al. suggested that AI-generated resources could serve as viable supplementary materials in certain contexts [20]. Morales-Chan et al. demonstrated AIGC's capacity to substantially enhance the creation and customization of digital educational materials [21]. Darko et al. compared two AI-powered learning resources with traditionally human-created materials and concluded that the integration of AIGDERs can significantly enhance the instructional processes traditionally managed by humans [22]. However, Lund et al. stated that concerns have been raised regarding the quality and accuracy of AIGDERs [22]. Salovich and Rapp claimed that inaccurate information could significantly impact academic research and teaching [23]. These studies have made progress in the recognition that the quality of AIGDERs is multifaceted, but they fail to provide a comprehensive framework for quality evaluation. The absence of a quality evaluation index system may result in inefficient education practices. Therefore, to ensure the utilization of high-quality AIGDERs, it is imperative to develop a comprehensive framework for quality evaluation.

### 2.2. Evaluation Criteria of Digital Educational Resources

As the amount of digital educational resources available is rapidly increasing, many previous studies have focused on the evaluation criteria of digital educational resources. For example, Estrada-Molina et al. integrated general criteria from ISO/IEC 9241-11 and ISO/IEC 9126-1 and pedagogical usability criteria when evaluating the usability of digital educational resources [24]. Xie et al. thought that the evaluation criteria for educational digital content should include four aspects: content relevance, pedagogical effectiveness, technical support, and curriculum alignment [25]. Rozo and Real identified pedagogical, technological, social, communicational, organizational, energizing, and ethical elements that must be considered when designing and creating adaptive digital educational resources [26]. Lee and Han summarized that the five decision-making criteria for selecting and evaluating educational digital resources were the primary benefit holder, function of resources, opportunity for improving skills, affordances, and constraints [27]. Almendro and Silveira described a framework including pedagogical, technical, and content dimensions to support the quality assurance and validation processes of open educational resources [28]. Although previous studies on the evaluation criteria and quality framework of digital educational resources have made significant contributions, few have analyzed them from the perspective of AIGDERs. Therefore, this study addresses this gap by proposing a comprehensive framework for evaluating the quality of AIGDERs.

## 3. Research Methods

The overall framework of this study is mainly divided into three steps: extraction of quality indicators, confirmation of quality indicator sets, and establishment of the quality evaluation index system, as illustrated in Figure 1.

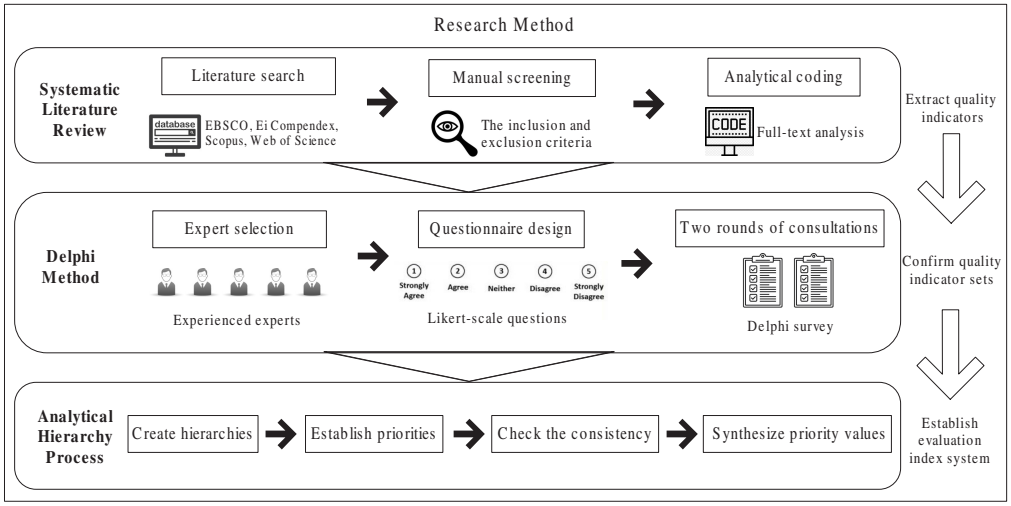



**Figure 1.** The framework of the evaluation study.

### *3.1. Systematic Literature Review*

To evaluate the quality of AIGDERs, the factors that affect the quality are identified using systematic literature review (SLR). SLR is a methodology that employs a standardized analytical approach to summarize the current state and development tendency within a research field or topic [29]. SLR follows a structured process proposed by Watson [30], which includes predefined search criteria and systematic screening processes [31]. Compared to traditional literature review, SLR has higher reliability because it addresses uncontrollable issues such as subjectivity and bias [32].

#### 3.1.1. Literature Search

According to the PRISMA guidelines, SLR typically involves searching three or more databases to retrieve literature [33]. To comprehensively acquire the necessary literature data for evaluating the quality of AIGDERs, four literature databases are selected: EBSCO, Ei Compendex, Scopus, and Web of Science. These databases extensively cover leading educational journals, and their inclusion criteria ensure that the literature within them largely meets the quality requirements of SLR [34].

This study identified several relevant keywords from existing literature on AIGDERs and selected appropriate terms for research focus. For example, the thematic search query for the Web of Science is structured as follows: TS = ((“Artificial intelligence generated content” or “AIGC” or “ChatGPT” or “Large language model*” or “Generative artificial intelligence” or “Generative AI”) AND (“Education*”) AND (“Resource*” OR “Material*”) AND (“Quality” or “Quality evaluation” or “Evaluation index system” or “Evaluation criteria”)). As ChatGPT was released by OpenAI in November 2022, the search timeframe is set from November 2022 to October 2024. For four chosen databases, appropriate filtering tools are employed to refine the research results and avoid the occurrence of duplication [35]. Following the literature search, a total of 187 papers are retrieved from four databases after deduplication.

#### 3.1.2. Manual Screening

The literature obtained in the initial search often contains papers that initially seem relevant but may not be. For efficient analytical coding, manual screening is performed to guarantee the quality and relevance of the selected literature. The review process carefully follows the inclusion and exclusion criteria specified in Table 1. The manual screening is not just a redundant step but a crucial one to rectify the potential flaws in the initial automated filtering. For example, some databases, despite their sophisticated algorithms, do not always accurately index all aspects of a publication. By manually checking, it is possible to guarantee that the literature included is relevant to the research focus and maintains a certain level of quality and consistency.

**Table 1.** Selection criteria.

| Criteria | Inclusion | Exclusion |
| --- | --- | --- |
| Type of publication | Scholarly articles, books, book chapters | Conference proceedings, reports, PhD theses, technical papers, tutorials |
| Peer-reviewed | Peer-reviewed | Non-peer-reviewed |
| Content | Full-text and complete information | Without in-depth discussion |
| Language | English | Non-English |
| Publication year | From November 2022 to October 2024 | Before November 2022 |
| Primary topic | AI-generated educational resources and quality evaluation | Neither AI-generated educational resources nor quality evaluation |

The manual screening process is conducted in two phases [35]. Initially, exclude completely irrelevant literature after reading titles and abstracts. Subsequently, read the full content of the selected articles to accurately identify the required literature. During this in-depth screening, 54 papers from the databases were ultimately retained.

#### 3.1.3. Analytical Coding

Upon completing the manual screening, a full-text analysis of the chosen literature is conducted. From a comprehensive review of existing literature, information is collected across multiple dimensions, and specific coding content involves creating codes that accurately represent the information.

### *3.2. Delphi Method*

The Delphi method, named after a Greek oracle who was famed for prophetic abilities, was initiated by the Rand Corporation in the 1950s to forecast the effect of technology on warfare [36]. The Delphi method is a systematic and qualitative method used to gather opinions from panels of selected experts to achieve consensus on outcomes and responses [37]. It is featured with anonymity and confidentiality. This method has the capability to brainstorm from experts' opinions without face-to-face contact, which avoids groupthink [38]. Moreover, location and time differences are not permitted to influence the investigation [39]. The Delphi method has been widely used to identify and evaluate indicators in education field. For example, Khan et al. established a consensus on the curriculum viability indicators using the modified Delphi technique to provide a framework for evaluating curriculum viability [40]. Seo et al. employed the modified Delphi method

to refine key indicators for the course subject "Environment" in the secondary school curriculum in South Korea [41]. Encouraged by these studies, the Delphi method is utilized in this study to determine the indicators for the quality of AIGDERs.

#### 3.2.1. Expert Selection

In the Delphi method, expert selection is critical for obtaining critical and reliable results [42]. A group of 10 to 50 participants was originally suggested as appropriate for Delphi studies [43]. The ideal size for a panel of experts is considered between 10 and 20 participants, which enables meaningful statistical analysis [44]. Delphi studies that involve fewer panelists tend to have a very specific focus or seek outcomes that require the participation of highly specialized experts [45]. Conducting a Delphi study with a larger number of experts can be logistically challenging and may lead to increased response fatigue, potentially compromising the quality of feedback [46]. Based on these principle, 15 experts are selected in this study. The experts are expected to have up-to-date knowledge and extensive work experience in the relevant field, and they are willing to participate in the Delphi process [47]. Each participant is chosen based on their extensive experience and recognized expertise in AI and education. Moreover, these experts are teachers of different disciplines from Sichuan University, Chengdu University of Technology, Xihua University, and Yibin University. Their collective knowledge ensures a comprehensive understanding of AIGDERs.

#### 3.2.2. Delphi Survey

According to Delphi method, two rounds of questionnaires are sent to 15 selected experts. The first round is mainly to consult the rationality of each indicator. The selected experts are invited to score and put forward suggestions for modification. According to the experts' scores and opinions from the first round, the index system is modified accordingly.

The Delphi questionnaire involves two sections: one part collects the backgrounds of the experts, while the other one includes quantitative Likert-scale and qualitative semi-open questions. The Likert-scale questions comprise one of the essential rating scales used as a measurement tool in the importance of the recommended indicators, with options ranging from 1 (very important) to 5 (very unimportant) [48]. Then, the degree of consensus among the experts can be assessed. The open-ended questions are intended to collect possible recommendations if the experts think the indicators are not appropriate.

Expert reliability is represented by coefficient of expert authority, which is indicated by $Cr$. The range of $Cr$ is between 0 and 1; the higher $Cr$ means the the higher reliability of the expert consultation. In general, $Cr < 0.70$ is regarded as acceptable degree. $Cr$ can be obtained using Equation (1).

$$Cr = \frac{(Ca + Cs)}{2} \tag{1}$$

where $Ca$ represents reasons for expert's judgment and $Cs$ represents the expert's familiarity with the indicators. To collect these data, experts are required to provide a 5-level score on the familiarity of the indicators and a 3-level score on the reasons for judgment (3 means large impact, 2 means medium impact, and 1 means small impact) during consultation. The aspects of evaluation include practical experience, logical reasoning, understanding of international and domestic conditions, and intuition.

### *3.3. The Analytical Hierarchy Process*

The AHP, first developed by Thomas L. Saaty in the 1980s, is one of the most widely used multiple-criteria decision-making tools [49]. It provides a structured procedure for assessing the importance of multiple criteria and assigning priorities to alternatives on a predefined scale for the given alternatives [50]. In this technique, depending on the

qualitative judgments of experts and users, the priorities of criteria could be established through the pairwise comparison matrix [51]. The AHP method is the most commonly used for multi-criteria weighting. For example, Yildiz et al. used the AHP method to evoke experts' perspectives regarding the relative importance of the criteria for GIS-based offshore wind farm site selection [52]. Chen et al. used the AHP to generate the weights of indicators for hazard, vulnerability, and restorability, and then established a multi-method evaluation approach for urban flood risk assessment [53]. Inspired by these studies, the same panel of 15 Delphi experts provide the pairwise comparisons for the AHP method though the second round of questionnaires, and the AHP method is used to determine the weights of the quality indicators for AIGDERs in the following four steps during the decision-making process.

Step 1: Create hierarchies. Decisions are broken down into a three-level hierarchical structure model, which consists of the objective level, the criterion level, and the solution level. The quality of AIGDERs is the objective level. Content characteristics, expression characteristics, user characteristics, and technical characteristics comprise the criterion level.

Step 2: Establish priorities. The relative priorities are assigned by the participants based on pairwise comparisons at each hierarchy level for each criterion compared with the pairwise criterion. The participating experts are invited to construct a judgment matrix by comparing the criteria pairwise using Saaty's 1–9 scale method of importance degree evaluation. Table 2 shows Saaty's 1–9 scale of pairwise comparisons and the meaning. The relevant importance of two criteria are determined by conducting pairwise comparisons, as shown in Table 3.

**Table 2.** Saaty's 1–9 scale of pairwise comparisons.

| Scale | Meaning |
|---|---|
| 1 | i factor is as important as j factor |
| 3 | i factor is moderately more important than j factor |
| 5 | i factor is strongly more important than j factor |
| 7 | i factor is very strongly more important than j factor |
| 9 | i factor is extremely more important than j factor |
| 2, 4, 6, 8 | The relative importance of i factor and j factor is between adjacent levels |

**Table 3.** The pairwise comparison judgment matrix.

| | $C_1$ | $C_2$ | $\cdots$ | $C_N$ |
|---|---|---|---|---|
| $C_1$ | $C_1/C_1$ | $C_1/C_2$ | $\cdots$ | $C_1/C_N$ |
| $C_2$ | $C_2/C_1$ | $C_2/C_2$ | $\cdots$ | $C_2/C_N$ |
| $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ |
| $C_N$ | $C_N/C_1$ | $C_N/C_2$ | $\cdots$ | $C_N/C_N$ |

Note: $C_i/C_j$ is just shorthand for the relative importance of $C_i$ to $C_j$.

Step 3: Check the consistency. $\lambda_{\max}$ represents the largest eigenvalue of the matrix $A$ and can be calculated as

$$\lambda_{\max} = \frac{1}{n}\sum_{i=1}^{n}\frac{(AW)_i}{W_i} \quad (2)$$

where $n$ is the matrix order, $W$ is the eigenvector, and $W_i$ is the eigenvalue of criterion I. $\lambda_{\max}$ is employed as a critical validating parameter to quantify the consistency, which can be calculated as

$$CI = \frac{\lambda_{\max} - n}{n-1} \quad (3)$$

where $CI$ stands for consistency index.

To assess the consistency of the response, Saaty created random matrices for various matrix dimensions and calculated their random index ($RI$), which is a random index determined by the number ($n$) of involved criteria, as shown in Table 4.

**Table 4.** Random index ($RI$) for different values of $n$.

| $n$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $RI$ | 0 | 0 | 0.52 | 0.89 | 1.11 | 1.25 | 1.35 | 1.4 | 1.45 | 1.49 |

The consistency ratio ($CR$) was calculated to assess the consistency of a pairwise comparison matrix considering its random index ($RI$) and consistency index ($CI$), shown as

$$CR = \frac{CI}{RI} \tag{4}$$

When $CR < 0.1$, the consistency of the judgment matrix is acceptable [54,55] unless the judgment matrix needs to be revised until the consistency requirement is satisfied.

Step 4: Synthesize these priority values. Each priority value is multiplied by the weight for that criterion, and they are aggregated to obtain a set of overall priorities for each criterion in the hierarchy.

## 4. Results

### *4.1. The Initial Identified Quality Indicators*

Using the SLR, Table 5 illustrates the four dimensions for coding and offers a description of the quality indicators coded within each dimension. Information is collected across the four dimensions for coding: content characteristics, expression characteristics, user characteristics, and technical characteristics. Specifically, content characteristics are derived from a comprehensive review of the existing literature on high-quality educational content, such as [56,57]. Expression characteristics are identified through research on the expression of AIGC in educational settings, such as [58]. User characteristics are developed based on user-centered design principles and research on how users interact with AIGC, such as [31,59,60]. Technical characteristics are determined by considering the technological requirements of modern educational tools, such as [17,61]. For each indicator in the four dimensions, the frequency of its occurrence in the literature is counted. The codes are ranked based on their frequency, providing a clear picture of the relative importance of each theme. These codes are then used as input indicators for the Delphi and AHP methods in the subsequent methodological steps.

**Table 5.** List of codes.

| **Dimension** | **Specific Coding Content** |
|---|---|
| Content characteristics | Authenticity, accuracy, specifications, relevance, novelty, diversity, timeliness |
| Expression characteristics | Legitimacy, knowledgeability, logicality, comprehensible |
| User characteristics | Achievement, acquisition, compatibility, friendliness |
| Technical characteristics | Conciseness, stability, human-analogy, security, big data |

### *4.2. Analysis of the Delphi Survey*

Through the first round of questionnaires from the selected experts, the expert consultation results are analyzed based on the mean value. Those indicators with mean values greater than 3 are retained. Moreover, the feedback from the experts regarding the refinement of the evaluation criteria has been incorporated. The descriptions of some indicators

have been revised, and and the newly proposed indicators put forward by the experts have been integrated into the framework. The "diversity" indicator is excluded due to its conceptual overlap with other metrics and insufficient specificity for targeted evaluation. Similarly, the "human-analogy" indicator is eliminated as it may not guarantee the fairness of the quality judgment. To enhance terminological precision, the "knowledgeability" indicator is revised to the "normativity" indicator to emphasize adherence to established standards, while the "specifications" indicator is redefined as the "disciplinarity" indicator to clarify alignment with domain-specific rigor. The "computation power" indicator is introduced to evaluate generative AI's functional efficiency in processing and synthesizing data. The "experience feeling" indicator is incorporated to underscore the pedagogical significance of evaluating learning processes alongside outcomes. According to Equation (1), the *Cr* value representing the expert authority coefficient of the first round is between 0.7 and 1.0, with an average value of 0.885. Consequently, a new set of twenty indicators have been established based on the initial round of responses from a panel of selected experts through a questionnaire survey. They are authenticity, accuracy, disciplinarity, relevance, novelty, timeliness, legitimacy, normativity, logicality, comprehensible, achievement, acquisition, compatibility, friendliness, experience feeling, conciseness, stability, security, computational power, and big data.

In the second round of questionnaires, the revisions collected from the first round are incorporated. Experts are consulted to evaluate the importance of the indicators with Likert-scale questions. Based on the calculations, the Cr value of the second round varies from 0.7 to 1.0, with the average being 0.85. Given that the Cr values for the two rounds of questionnaires do not exceed 0.7, the results of the expert consultations have high authority and credibility. Following two rounds of expert questionnaires, the evaluation index system of AIGDERs has been established, including four first-level dimensions and twenty second-level indicators.

### *4.3. Determination of the Weight for Identified Quality Factors Using AHP*

The AHP method is used to determine the weights for the quality of AIGDERs. Figure 2 shows the constructed hierarchical model used in this study.

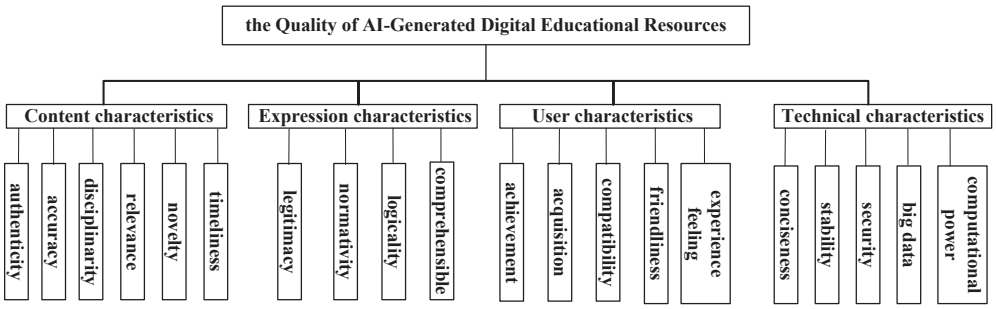



**Figure 2.** The constructed hierarchical model used in this study.

Through pairwise comparisons of the criteria, a collection of preference relations are constructed. The priority weights of alternatives are obtained by analyzing the given comparison matrices. Taking the questionnaire results of expert 1 as an example, judgment matrices of the evaluation indexes are shown in Tables 6–10. The questionnaire responses from other experts have also been utilized to construct the similar judgement matrixes of different indexes.

**Table 6.** Judgment matrix of the first-level evaluation indexes.

| | Content Characteristics | Expression Characteristics | User Characteristics | Technical Characteristics |
|---|---|---|---|---|
| Content characteristics | 1 | 3 | 2 | 3 |
| Expression characteristics | 1/3 | 1 | 1/2 | 1 |
| User characteristics | 1/2 | 2 | 1 | 2 |
| Technical characteristics | 1/3 | 1 | 1/2 | 1 |

**Table 7.** Judgment matrix of the second-level evaluation indexes in content characteristics.

| | Authenticity | Accuracy | Disciplinarity | Relevance | Novelty | Timeliness |
|---|---|---|---|---|---|---|
| Authenticity | 1 | 1 | 5 | 3 | 3 | 4 |
| Accuracy | 1 | 1 | 5 | 3 | 3 | 4 |
| Disciplinarity | 1/5 | 1/5 | 1 | 1/3 | 1/3 | 1/2 |
| Relevance | 1/3 | 1/3 | 3 | 1 | 2 | 3 |
| Novelty | 1/3 | 1/3 | 3 | 1/2 | 1 | 1 |
| Timeliness | 1/4 | 1/4 | 2 | 1/3 | 1 | 1 |

**Table 8.** Judgment matrix of the second-level evaluation indexes in expression characteristics.

| | Legitimacy | Normativity | Logicality | Comprehensible |
|---|---|---|---|---|
| Legitimacy | 1 | 2 | 1/3 | 1/2 |
| Normativity | 1/2 | 1 | 1/5 | 1/4 |
| Logicality | 3 | 5 | 1 | 2 |
| Comprehensible | 2 | 4 | 1/2 | 1 |

**Table 9.** Judgment matrix of the second-level evaluation indexes in user characteristics.

| | Achievement | Acquisition | Compatibility | Friendliness | Experience Feeling |
|---|---|---|---|---|---|
| Achievement | 1 | 3 | 5 | 2 | 2 |
| Acquisition | 1/3 | 1 | 2 | 1 | 3 |
| Compatibility | 1/5 | 1/2 | 1 | 1/3 | 1/2 |
| Friendliness | 1/2 | 1 | 3 | 1 | 1 |
| Experience feeling | 1/2 | 1/3 | 2 | 1 | 1 |

**Table 10.** Judgment matrix of the second-level evaluation indexes in technical characteristics.

| | Conciseness | Stability | Security | Computational Power | Big Data |
|---|---|---|---|---|---|
| Conciseness | 1 | 3 | 2 | 5 | 3 |
| Stability | 1/3 | 1 | 2 | 4 | 3 |
| Security | 1/2 | 1/2 | 1 | 3 | 2 |
| Computational power | 1/5 | 1/4 | 1/3 | 1 | 1/2 |
| Big data | 1/3 | 1/3 | 1/2 | 2 | 1 |

The consistency results of the five mathematical judgment matrices are 0.0039, 0.0248, 0.0079, 0.0495, and 0.0352, respectively. Since all these values are below the threshold of 0.10, the matrices are considered to be consistent. Similarly, the consistency of the judgment matrices derived from the questionnaire results of other experts has also been verified. After synthesizing these priority values, the quality evaluation index system of AIGDERs is shown in Table 11.

**Table 11.** The quality evaluation index system of AIGDERs.

| First-Level Indicator | First-Level Weight | Second-Level Indicator | Second-Level Weight | Rank |
|---|---|---|---|---|
| Content characteristics | 0.503 | Authenticity | 0.1851 | 1 |
| | | Accuracy | 0.1070 | 2 |
| | | Disciplinarity | 0.0529 | 8 |
| | | Relevance | 0.0672 | 4 |
| | | Novelty | 0.0347 | 12 |
| | | Timeliness | 0.0560 | 6 |
| Expression characteristics | 0.251 | Legitimacy | 0.1040 | 3 |
| | | Normativity | 0.0411 | 11 |
| | | Logicality | 0.0506 | 9 |
| | | Comprehensible | 0.0554 | 7 |
| User characteristics | 0.137 | Achievement | 0.0566 | 5 |
| | | Acquisition | 0.0286 | 13 |
| | | Compatibility | 0.0151 | 19 |
| | | Friendliness | 0.0179 | 16 |
| | | Experience feeling | 0.0187 | 14 |
| Technical characteristics | 0.109 | Conciseness | 0.0455 | 10 |
| | | Stability | 0.0181 | 15 |
| | | Security | 0.0157 | 17 |
| | | Computational power | 0.0142 | 20 |
| | | Big data | 0.0154 | 18 |

## 5. Discussion and Implications

### *5.1. Discussion*

The proposed method simplifies the complex problem of evaluating the AIGDER quality into simple dimensions and various indicators. The evaluation index system includes four dimensions (content, expression, and user and technical characteristics) and contains a total of twenty indicators. The weights for the four dimensions and the various indicators are illustrated in Figures 3 and 4.

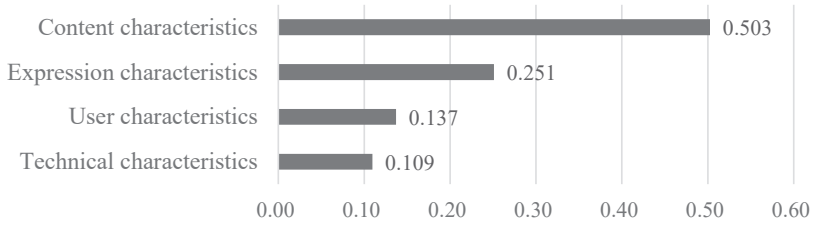



**Figure 3.** Analysis of the weights for the four evaluation dimensions.

Content characteristics have the highest weight, with the value being 0.503, indicating their great importance to the quality of AIGDERs. This finding aligns with the prevailing viewpoint that content plays a vital role in the application of AIGDERs within higher-education institutions [13]. Among all the indicators, the panel of experts assigns the top ranks to "authenticity" (Rank 1) and "accuracy" (Rank 2). The sum of their weights is 0.282, which occupies a considerable part. AIGDERs have posed challenges regarding the authenticity and accuracy of information in the era of digital information explosion [62]. Distorted information has appeared within artificial intelligence systems and has been a topic of significant interest [17]. Therefore, authenticity and accuracy are very important considerations. Although "novelty" (Rank 6) is considered as the least important indicator within the content characteristics, it is also essential to ensure that the content is appealing

and offers a fresh perspective. Other notable indicators include "relevance" (Rank 3), "timeliness" (Rank 4), and "disciplinarity" (Rank 5).

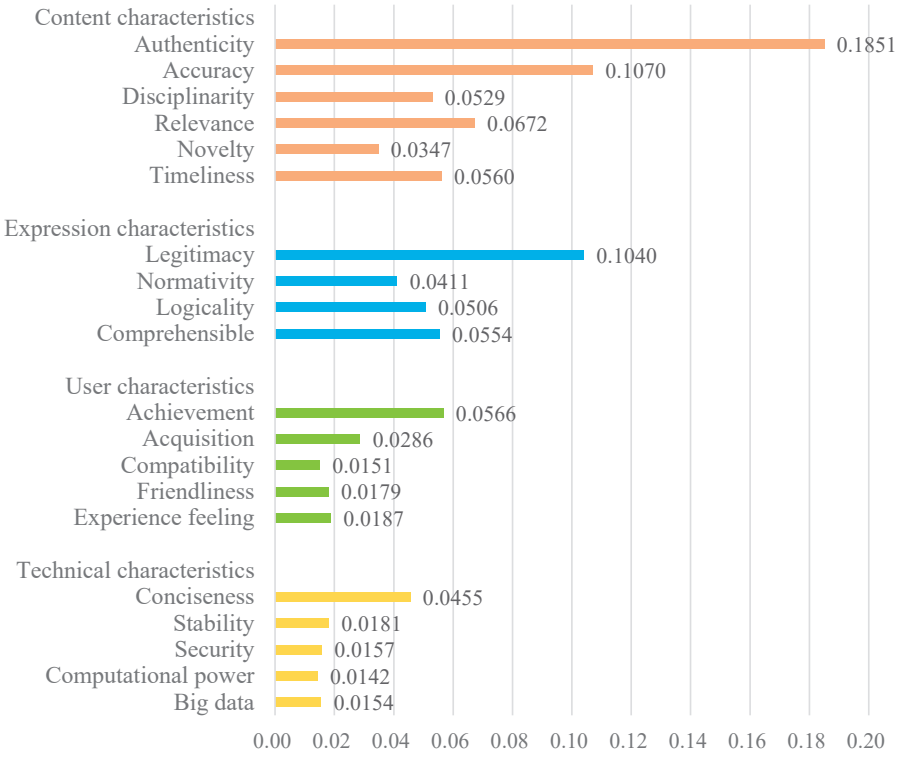



**Figure 4.** Analysis of the weights for the twenty evaluation indicators.

Expression characteristics are assigned the weight of 0.257 and placed second to content characteristics. Expression characteristics refer to the expression of AIGDERs in language and form. AIGDERs adapt to rapidly changing expression through achieving efficient processing and automatically extracting information [63]. "Legitimacy" (Rank 1) is the most influential factor in expression characteristics. It indicates that AIGDERs demonstrate respect for the diverse values and beliefs of stakeholders, maintain impartiality in their operations, and ensure equitable treatment of opposing viewpoints and interests [64]. Furthermore, "comprehensible" (Rank 2) demonstrates significant importance as AIGDERs should be easily comprehensible by users with varying literacy levels [65]. "Logicality" (Rank 3) means that AIGDERs are supposed to demonstrate self-organizing and deductive logical reasoning skills. In particular, when encountering certain highly specialized problems, expression logic is of great concern [59]. Other noteworthy indicators include "normativity" (Rank 4).

The weight assigned to user characteristics is 0.137, which is marginally higher than that of technical characteristics. The primary purpose of AIGDERs is to meet users' requirements and sequentially to improve the learning outcome and learning quality. For educators, effective interaction with AI is crucial as it enhances efficiency and produces the most relevant information [66]. For students, these AIGDERs can transform obscure concepts into more concrete and comprehensible forms, which significantly enhances students' understanding of content and classroom engagement [67]. In this dimension, the panel acknowledges the paramount importance of "achievement" (Rank 1) as a key indicator of users' aspects. Achievement denotes the extent to which a learner has attained or accomplished the short- or long-term educational goals. Another important indicator is "acquisition" (Rank 2), which indicates whether access is convenient. "Experience feeling" (Rank 3) is considered to be important as AIGDERs have the advantages of personalized

learning, efficiency improvement, and time-saving, which can enhance learning experiences for both students and teachers [61]. Furthermore, "friendliness" (Rank 4) is identified as a significant influencing indicator. As researched by Maheshwari, users' inclination to adopt AI-generated content is influenced by their perception of its user-friendliness [60]. Although compatibility positively influences the adoption intention of AI-generated content [68], it is worth noting that "compatibility" (Rank 5) is considered to be the least influential factor in user characteristics.

Technical characteristics have the lowest weight of 0.109, but they cannot be ignored in AIGDERs' application. It is also evident from the previous study that university teachers acknowledge technical aspects when effectively leveraging AIGC tools [61], and the technical efficiency of AIGDERs in education has been extensively researched and emphasized [69,70]. Among the indicators in technical characteristics, experts emphasize the significance of "conciseness" (Rank 1) in improving the quality of AIGDERs. "Conciseness" represents the user-friendly and intuitive nature of the AIGC interactive interface and is a criterion for readability [71]. Other important indicators are "stability" (Rank 2) and "security" (Rank 3). AIGDERs rely on complex algorithms and models, and this raises concerns about the stability of AIGC technology [72]. Meanwhile, the current AIGC tools come in both online and locally deployed versions, and the computational and generative processes are often viewed as black boxes, leading to concerns about the security of AIGDERs [73]. "Big data" (Rank 4) and "computational power" (Rank 5) are afforded the lowest importance. "Big data" does not only refer to a large amount of data but, more importantly, contains multiple types of data that require processing and analysis via complex algorithms [74]. Based on the above data and the user's behavioral information, generative AI can continuously enhance the quality of the generated content. "Computational power" is crucial to support the large-scale deployment of AIGC technology in education [75,76]. Especially in the case of high concurrency and high traffic, computing power can ensure that the AIGC platforms respond to user requests in time and provide high-quality services.

Previous studies had a more general focus on digital educational resources. For example, Atenas and Havemann evaluated the effectiveness of repositories for open educational resources (ROER) regarding the following 10 quality indicators: featured resources, user evaluation tools, peer review, authorship of the resources, keywords of the resources, use of standardized metadata, multilingualism of the repositories, inclusion of social media tools, specification of the creative commons license, and availability of the source code or original files [77]. Mhouti et al. presented the four dimensions of digital learning resource quality—academic quality, pedagogical quality, didactic quality, and technical quality—and described how each dimension could be interpreted to evaluate digital learning resources [78]. In contrast, this study specifically targets AI-generated digital educational resources. This specificity enables exploring the unique characteristics and challenges associated with resources created by artificial intelligence, such as the influence of AI algorithms on content quality and user–AI interaction aspects. The proposed evaluation framework presents the final set of overall priorities for alternative indicators, thus helping stakeholders to make decisions. For example, it can help educators and educational institutions to select the best-quality resources. If a particular AIGDER has high overall priority values based on the AHP analysis, it is more likely to be a high-quality resource and a better choice for use in university teaching and learning.

### *5.2. Managerial Implications*

The theoretical implications of this study are multifaceted. Firstly, this paper provides a comprehensive framework for assessing the quality of AIGDERs and enriches educational evaluation theory. The quality evaluation index system includes four dimensions (content,

expression, and user and technical characteristics) and twenty indicators, contributing to the literature by simplifying the complex problem of quality evaluation into manageable dimensions. This is a response to researchers' calls that educators should critically evaluate any AI-generated content and adapt it to their specific teaching contexts [79]. Secondly, the study's findings highlight the importance of content characteristics as the most significant factor in the quality assessment of AIGDERs. This aligns with the growing recognition among scholars that content is central to effective learning [80].

On the practical side, the evaluation index system established in this study offers educators and educational technologists an instrument to assess and select high-quality AIGDERs for university teaching and learning. Recognizing the relative importance of each dimension and indicator, educators are empowered to make informed decisions about which resources are more likely to enhance teaching quality and learning outcomes. Moreover, the insights from our Delphi survey and the AHP analysis can guide the development of AI algorithms to better align with educational standards and users' needs. For instance, the emphasis on user characteristics suggests that AI tools should be designed not only for effectiveness but also for user-friendliness, promoting a positive learning experience. The study's findings have implications for educational policy as well. Policymakers can use the evaluation index system to develop guidelines for the integration of AIGDERs, which leads to more effective use of technology in education and helps to bridge the gap between traditional teaching methods and innovative technological advancements.

## 6. Conclusions and Future Study

The increasing incorporation of AIGDERs within university settings has become a notable trend. Concurrently, this development has sparked concerns regarding the quality of AIGDERs. To ensure the utilization of high-quality AIGDERs, this paper developed a comprehensive evaluation framework for AIGDERs by combining the Delphi method and the AHP. A systematic literature review of the recent literature on AIGDERs was conducted to identify criterion-level indicators, encompassing the dimensions of content, expression, and user and technical characteristics. Then, two rounds of questionnaires were sent to a panel of knowledgeable experts to confirm the rationality of each indicator using the Delphi method. Subsequently, the weights for the identified quality indicators were determined through the AHP, and the evaluation index system including four dimensions and twenty indicators was obtained. The findings underscored the paramount importance of content characteristics in the quality assessment of AIGDERs. Moreover, expression characteristics were considered to be the second most important factor, and the importance of user and technical characteristics was properly recognized. Among the second-level indicators, the experts regarded "authenticity", "accuracy", "legitimacy", and "relevance" more important than other indicators. The proposed system provided relevant stakeholders with a framework to select high-quality AIGDERs and guide AI tools towards educational standards.

However, this study has two main limitations. Firstly, the quality evaluation indicators of AIGDERs may not be comprehensive enough. This study employed the Delphi method and the AHP to construct the evaluation index system, which included four first-level dimensions and twenty second-level indicators. However, these methodologies are heavily dependent on the subjective assessments of experts, potentially resulting in unstable and biased results. In future research, multiple samples will be used and the number of participants will be increased to provide diverse inputs to the study. It will incorporate a broader range of professionals, including educators from different educational levels, instructional designers with varying specializations, and AI researchers with a focus on educational applications. Secondly, the quality evaluation index system of AIGDERs proposed in

this study has not yet been tested for reliability, validity, and empirical application. In future research, to make the evaluation index system more convincing, different levels of universities and various types of courses will be selected for empirical analysis. Thirdly, as large-group decision-making methods have been applied to solve complex decision-making problems [81], incorporating these methods is considered in future research. For example, large-group decision-making methods could be applied to the weighting process of different factors in our evaluation model.

The proposed evaluation framework is a starting point aimed at evaluating AIGDERs. Future research will conduct case-specific applications to test and verify the validity of the framework. In the short term, a pilot test of the framework will be carried out in a select few educational institutions. Then, the implementation will be expanded to a larger scale. In the long term, the framework will be continuously refined and updated based on feedback from the field.

**Author Contributions:** Conceptualization, Q.H. and S.T.; methodology, Q.H. and C.L.; software, C.L.; validation, S.T.; formal analysis, L.L.; investigation, Q.H. and C.L.; resources, L.L.; writing—original draft preparation, Q.H. and C.L.; writing—review and editing, L.L. and S.T.; visualization, L.L.; funding acquisition, S.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Chinese Society of Educational Development Strategy (Grant No. ZLB20240831) and Xihua University (Grant No. xjjg2023038).

**Data Availability Statement:** The data are available from the corresponding author on reasonable request.

**Acknowledgments:** The authors would like to thank the editors and anonymous referees for their valuable comments, which have significantly improved this paper.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Wang, W.; Siau, K. Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: A review and research agenda. *J. Database Manag.* **2019**, *30*, 61–79. [CrossRef]
2. Lin, H.; Jiang, X.; Deng, X.; Bian, Z.; Fang, C.; Zhu, Y. Comparing AIGC and traditional idea generation methods: Evaluating their impact on creativity in the product design ideation phase. *Think. Ski. Creat.* **2024**, *54*, 101649. [CrossRef]
3. Xu, M.; Du, H.; Niyato, D.; Kang, J.; Xiong, Z.; Mao, S.; Han, Z.; Jamalipour, A.; Kim, D.I.; Shen, X.; et al. Unleashing the power of edge-cloud generative AI in mobile networks: A survey of AIGC services. *IEEE Commun. Tutorials* **2024**, *26*, 1127–1170. [CrossRef]
4. Du, H.; Zhang, R.; Liu, Y.; Wang, J.; Lin, Y.; Li, Z.; Niyato, D.; Kang, J.; Xiong, Z.; Cui, S.; et al. Enhancing deep reinforcement learning: A tutorial on generative diffusion models in network optimization. *IEEE Commun. Tutorials* **2024**, *26*, 2611–2646. [CrossRef]
5. Cao, Y.; Li, S.; Liu, Y.; Yan, Z.; Dai, Y.; Yu, P.S.; Sun, L. A comprehensive survey of ai-generated content (AIGC): A history of generative AI from gan to chatgpt. *arXiv* **2023**, arXiv:2303.04226.
6. Lee, D.; Yeo, S. Developing an AI-based chatbot for practicing responsive teaching in mathematics. *Comput. Educ.* **2022**, *191*, 104646. [CrossRef]
7. Guo, D.; Chen, H.; Wu, R.; Wang, Y. AIGC challenges and opportunities related to public safety: A case study of ChatGPT. *J. Saf. Sci. Resil.* **2023**, *4*, 329–339. [CrossRef]
8. Tao, W.; Gao, S.; Yuan, Y. Boundary crossing: An experimental study of individual perceptions toward AIGC. *Front. Psychol.* **2023**, *14*, 1185880. [CrossRef]
9. Rudolph, J.; Tan, S.; Tan, S. ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *J. Appl. Learn. Teach.* **2023**, *6*, 342–363.
10. Kasneci, E.; Seßler, K.; Küchemann, S.; Bannert, M.; Dementieva, D.; Fischer, F.; Gasser, U.; Groh, G.; Günnemann, S.; Hüllermeier, E.; et al. Chatgpt for good? on opportunities and challenges of large language models for education. *Learn. Individ. Differ.* **2023**, *103*, 102274. [CrossRef]
11. Jalil, S.; Rafi, S.; LaToza, T.D.; Moran, K.; Lam, W. Chatgpt and software testing education: Promises & perils. In Proceedings of the 2023 IEEE International Conference on Software Testing, Verification and Validation Workshops, Dublin, Ireland, 16–20 April 2023; pp. 4130–4137.

12. Liang, Y.; Zou, D.; Xie, H.; Wang, F.L. Exploring the potential of using ChatGPT in physics education. *Smart Learn. Environ.* **2023**, *10*, 52. [CrossRef]
13. Yang, S.; Yang, S.; Tong, C. In-depth application of artificial intelligence-generated content AIGC large model in higher education. *Adult High. Educ.* **2023**, *5*, 9–16.
14. Wu, T.-T.; Lee, H.-Y.; Li, P.-H.; Huang, C.-N.; Huang, Y.-M. Promoting self-regulation progress and knowledge construction in blended learning via ChatGPT-based learning aid. *J. Educ. Comput. Res.* **2024**, *61*, 3–31. [CrossRef]
15. Lambert, J.; Stevens, M. ChatGPT and generative AI technology: A mixed bag of concerns and new opportunities. *Comput. Sch.* **2024**, *41*, 559–583. [CrossRef]
16. Kortemeyer, G. Could an artificial-intelligence agent pass an introductory physics course? *Phys. Rev. Phys. Educ. Res.* **2023**, *19*, 010132. [CrossRef]
17. Sun, Y.; Sheng, D.; Zhou, Z.; Wu, Y. AI hallucination: Towards a comprehensive classification of distorted information in artificial intelligence-generated content. *Humanit. Soc. Sci. Commun.* **2024**, *11*, 1278. [CrossRef]
18. Li, J.; Zhu, S.; Yang, H.H.; Xu, J. What does artificial intelligence generated content bring to teaching and learning? A literature review on AIGC in education. In Proceedings of the 2024 International Symposium on Educational Technology, Macao, China, 29 July–1 August 2024; pp. 18–23.
19. Watts, F.M.; Dood, A.J.; Shultz, G.V.; Rodriguez, J.-M.G. Comparing student and generative artificial intelligence chatbot responses to organic chemistry writing-to-learn assignments. *J. Chem. Educ.* **2023**, *100*, 3806–3817. [CrossRef]
20. Denny, P.; Khosravi, H.; Hellas, A.; Leinonen, J.; Sarsa, S. Can we trust AI-generated educational content? Comparative analysis of human and ai-generated learning resources. *arXiv* **2023**, arXiv:2306.10509.
21. Morales-Chan, M.; Amado-Salvatierra, H.R.; Hernandez-Rizzardini, R. AI-driven content creation: Revolutionizing educational materials. In Proceedings of the Eleventh ACM Conference on Learning Scale, Atlanta, GA, USA, 18–20 July 2024; pp. 556–558.
22. Darko, C.O.; Karki, R.; Teye, E.; Liu, L. Assessment of the quality of design of human created and AI-based learning materials. In *Proceedings of the Society for Information Technology & Teacher Education International Conference*; Association for the Advancement of Computing in Education: Asheville, NC, USA, 2024; pp. 841–846.
23. Lund, B.D.; Wang, T.; Mannuru, N.R.; Nie, B.; Shimray, S.; Wang, Z. Chatgpt and a new academic reality: Artificial intelligence-written research papers and the ethics of the large language models in scholarly publishing. *J. Assoc. Inf. Sci. Technol.* **2023**, *74*, 570–581. [CrossRef]
24. Estrada-Molina, O.; Fuentes-Cancell, D.R.; Morales, A.A. The assessment of the usability of digital educational resources: An interdisciplinary analysis from two systematic reviews. *Educ. Inf. Technol.* **2022**, *27*, 4037–4063. [CrossRef]
25. Xie, K.; Tosto, G.D.; Chen, S.-B.; Vongkulluksn, V.W. A systematic review of design and technology components of educational digital resources. *Comput. Educ.* **2018**, *127*, 90–106. [CrossRef]
26. Rozo, H.; Real, M. Pedagogical guidelines for the creation of adaptive digital educational resources: A review of the literature. *J. Technol. Sci. Educ.* **2019**, *9*, 308–325. [CrossRef]
27. Lee, H.-J.; Han, C. Bringing digital resource decision-making up to date: Preparing the inaugural class of digitalised preservice teachers for informed evaluation of digital resources. *J. Educ. Teach.* **2023**, *49*, 445–460. [CrossRef]
28. Almendro, D.; Silveira, I.F. Quality assurance for open educational resources: The OERTrust framework. *Int. J. Learn. Educ. Res.* **2018**, *17*, 1–14. [CrossRef]
29. Jing, Y.; Wang, C.; Chen, Y.; Wang, H.; Yu, T.; Shadiev, R. Bibliometric mapping techniques in educational technology research: A systematic literature review. *Educ. Inf. Technol.* **2024**, *29*, 9283–9311. [CrossRef]
30. Watson, R.T. Beyond being systematic in literature reviews in IS. *J. Inf. Technol.* **2015**, *30*, 185–187. [CrossRef]
31. Labadze, L.; Grigolia, M.; Machaidze, L. Role of AI chatbots in education: Systematic literature review. *Int. J. Educ. Technol. High. Educ.* **2023**, *20*, 56. [CrossRef]
32. Waqar, A.; Othman, I.; Shafiq, N.; Mansoor, M.S. Applications of AI in oil and gas projects towards sustainable development: A systematic literature review. *Artif. Intell. Rev.* **2023**, *56*, 12771–12798. [CrossRef]
33. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Ann. Intern. Med.* **2009**, *151*, 264–269. [CrossRef]
34. Chen, X.; Hu, Z.; Wang, C. Empowering education development through AIGC: A systematic literature review. *Educ. Inf. Technol.* **2024**, *29*, 17485–17537. [CrossRef]
35. Ali, O.; Abdelbaki, W.; Shrestha, A.; Elbasi, E.; Alryalat, M.A.A.; Dwivedi, Y.K. A systematic literature review of artificial intelligence in the healthcare sector: Benefits, challenges, methodologies, and functionalities. *J. Innov. Knowl.* **2023**, *8*, 100333. [CrossRef]
36. Dalkey, N.; Helmer, O. An experimental application of the delphi method to the use of experts. *Manag. Sci.* **1963**, *9*, 458–467. [CrossRef]
37. Olsen, A.A.; Wolcott, M.D.; Haines, S.T.; Janke, K.K.; McLaughlin, J.E. How to use the delphi method to aid in decision making and build consensus in pharmacy education. *Curr. Pharm. Teach. Learn.* **2021**, *13*, 1376–1385. [CrossRef] [PubMed]

38. Zhao, P.; Md Ali, Z.; Ahmad, Y. Developing indicators for sustainable urban regeneration in historic urban areas: Delphi method and Analytic Hierarchy Process (AHP). *Sustain. Cities Soc.* **2023**, *99*, 104990. [CrossRef]
39. Phichetkunbodee, N.; Chantrawutikorn, M.; Sundaranaga, C.; Chaichana, C.; Pholporton, K.; Kridakorn Na Ayutthaya, T.; Rinchumphu, D. Using the Delphi method for selecting energy efficiency indicators in the Thailand context. *Energy Rep.* **2023**, *9*, 408–414. [CrossRef]
40. Khan, R.A.; Spruijt, A.; Mahboob, U.; Eraky, M.A.; van Merrienboer, J.J. Curriculum viability indicators: A Delphi study to determine standards and inhibitors of a curriculum. *Eval. Health Prof.* **2021**, *44*, 210–219. [CrossRef]
41. Seo, E.; Ryu, J.; Hwang, S. Building key competencies into an environmental education curriculum using a modified Delphi approach in South Korea. *Environ. Educ. Res.* **2020**, *26*, 890–914. [CrossRef]
42. Rowe, G.; Wright, G.; Bolger, F. Delphi: A reevaluation of research and theory. *Technol. Forecast. Soc. Change* **1991**, *39*, 235–251. [CrossRef]
43. Turoff, M. The design of a policy Delphi. *Technol. Forecast. Soc. Change* **1970**, *2*, 149–171. [CrossRef]
44. Rowe, G. *Expert Opinions in Forecasting: The Role of the Delphi Technique*; Kluswer Academic: Alphen aan den Rijn, The Netherlands, 2001.
45. De Loë, R.C.; Melnychuk, N.; Murray, D.; Plummer, R. Advancing the state of policy delphi practice: A systematic review evaluating methodological evolution, innovation, and opportunities. *Technol. Forecast. Soc. Change* **2016**, *104*, 78–88. [CrossRef]
46. Agrawal, T.K.; Kalaiarasan, R.; Olhager, J.; Wiktorsson, M. Supply chain visibility: A delphi study on managerial perspectives and priorities. *Int. J. Prod. Res.* **2024**, *62*, 2927–2942. [CrossRef]
47. Seely, R.L.; Iglarsh, H.; Edgell, D. Utilising the Delphi technique at international conferences: A method for forecasting international tourism conditions. *J. Travel Res.* **1980**, *1*, 30–35.
48. Gao, M.; Che, W.; Qi, S. The construction of a physical literacy assessment index system for junior high school students: A Chinese study. *Sci. Rep.* **2024**, *14*, 23938. [CrossRef] [PubMed]
49. Saaty, T.L. The analytic hierarchy process (AHP). *J. Oper. Res. Soc.* **1980**, *41*, 1073–1076.
50. Vafadarnikjoo, A.; Ahmadi, H.B.; Liou, J.J.; Botelho, T.; Chalvatzis, K. Analyzing blockchain adoption barriers in manufacturing supply chains by the neutrosophic analytic hierarchy process. *Ann. Oper. Res.* **2023**, *327*, 129–156. [CrossRef]
51. Vaidya, O.S.; Kumar, S. Analytic hierarchy process: An overview of applications. *Eur. J. Oper. Res.* **2006**, *169*, 1–29. [CrossRef]
52. Yildiz, S. Spatial multi-criteria decision making approach for wind farm site selection: A case study in Balıkesir, Turkey. *Renew. Sustain. Energy Rev.* **2024**, *192*, 114158. [CrossRef]
53. Chen, Y.; Wang, D.; Zhang, L.; Guo, H.; Ma, J.; Gao, W. Flood risk assessment of Wuhan, China, using a multi-criteria analysis model with the improved AHP-Entropy method. *Environ. Sci. Pollut. Res.* **2023**, *30*, 96001–96018. [CrossRef]
54. Aguarón, J.; Moreno-Jiménez, J.M. The geometric consistency index: Approximated thresholds. *Eur. J. Oper. Res.* **2003**, *147*, 137–145. [CrossRef]
55. Haktanır, E.; Kahraman, C. Integrated ahp & topsis methodology using intuitionistic z-numbers: An application on hydrogen storage technology selection. *Expert Syst. Appl.* **2024**, *239*, 122382.
56. Tanaka, O.M.; Gasparello, G.G.; Hartmann, G.C.; Casagrande, F.A.; Pithon, M.M. Assessing the reliability of chatgpt: A content analysis of self-generated and self-answered questions on clear aligners, tads and digital imaging. *Dent. Press J. Orthod.* **2023**, *28*, e2323183. [CrossRef]
57. Birks, S.; Gray, J.; Darling-Pomranz, C. Using artificial intelligence to provide a 'flipped assessment' approach to medical education learning opportunities. *Med. Teach.* **2024**, 1–8. [CrossRef] [PubMed]
58. Liu, X.; Xiao, Y. Chinese university teachers' engagement with generative ai in different stages of foreign language teaching: A qualitative enquiry through the prism of addie. *Educ. Inf. Technol.* **2024**, *30*, 485–508. [CrossRef]
59. Wei, Q. The effects of the AIGC software on English majors: Professional competence, individual perceptions, and career planning. *J. Educ. Humanit. Soc. Sci.* **2024**, *32*, 67–73. [CrossRef]
60. Maheshwari, G. Factors influencing students' intention to adopt and use ChatGPT in higher education: A study in the Vietnamese context. *Educ. Inf. Technol.* **2024**, *29*, 12167–12195. [CrossRef]
61. Lu, G.; Hussin, N.B.; Sarkar, A. Navigating the future: Harnessing artificial intelligence generated content (aigc) for enhanced learning experiences in higher education. In Proceedings of the 2024 International Conference on Advances in Modern Age Technologies for Health and Engineering Science, Shivamogga, India, 16–17 May 2024; pp. 1–12.
62. Zhou, X.; Zafarani, R. A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Comput. Surv.* **2020**, *53*, 109. [CrossRef]
63. Sun, J. AIGC fusion exploration: The intersecting path of digital humanities and artificial intelligence. *J. Electr. Syst.* **2024**, *20*, 327–335.
64. Juschten, M.; Omann, I. Evaluating the relevance, credibility and legitimacy of a novel participatory online tool. *Environ. Sci. Policy* **2023**, *146*, 90–100. [CrossRef]

65. Cohen, S.A.; Fisher, A.C.; Pershing, S. Analysis of the readability and accountability of online patient education materials related to glaucoma diagnosis and treatment. *Clin. Ophthalmol.* **2023**, *17*, 779–788. [CrossRef]
66. Park, J.; Choo, S. Generative AI prompt engineering for educators: Practical strategies. *J. Spec. Educ. Technol.* **2024**. [CrossRef]
67. Liu, R.; Pang, W.; Chen, J.; Balakrishnan, V.A.; Chin, H.L. The application of scaffolding instruction and AI-driven diffusion models in children's aesthetic education: A case study on teaching traditional Chinese painting of the twenty-four solar terms in Chinese culture. *Educ. Inf. Technol.* **2024**, 1–32. [CrossRef]
68. Ayanwale, M.A.; Molefi, R.R. Exploring intention of undergraduate students to embrace chatbots: From the vantage point of lesotho. *Int. J. Educ. Technol. High. Educ.* **2024**, *21*, 20. [CrossRef]
69. Ouyang, F.; Zheng, L.; Jiao, P. Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. *Educ. Inf. Technol.* **2022**, *27*, 7893–7925. [CrossRef]
70. Sartori, L.; Bocca, G. Minding the gap (s): Public perceptions of AI and socio-technical imaginaries. *AI Soc.* **2023**, *38*, 443–458. [CrossRef]
71. Wang, Y.-C.; Xue, J.; Wei, C.; Kuo, C.-C.J. An overview on generative ai at scale with edge-cloud computing. *IEEE Open J. Commun. Soc.* **2023**, *4*, 2952–2971. [CrossRef]
72. Yang, X.; Ding, J.; Chen, H.; Ji, H. Factors affecting the use of artificial intelligence generated content by subject librarians: A qualitative study. *Heliyon* **2024**, *10*, e29584. [CrossRef]
73. Jiang, Q.; Zhang, Y.; Wei, W.; Gu, C. Evaluating technological and instructional factors influencing the acceptance of aigc-assisted design courses. *Comput. Educ. Artif. Intell.* **2024**, *7*, 100287. [CrossRef]
74. Taleb, I.; Serhani, M.A.; Bouhaddioui, C.; Dssouli, R. Big data quality framework: A holistic approach to continuous quality management. *J. Big Data* **2021**, *8*, 76. [CrossRef]
75. Sastry, G.; Heim, L.; Belfield, H.; Anderljung, M.; Brundage, M.; Hazell, J.; O'Keefe, C.; Hadfield, G.K.; Ngo, R.; Pilz, K.; et al. Computing power and the governance of artificial intelligence. *arXiv* **2024**, arXiv:2402.08797.
76. Wu, J.; Gan, W.; Chen, Z.; Wan, S.; Lin, H. Ai-generated content (AIGC): A survey. *arXiv* **2023**, arXiv:2304.06632.
77. Atenas, J.; Havemann, L. Questions of quality in repositories of open educational resources: A literature review. *Res. Learn. Technol.* **2014**, *22*, 20889. [CrossRef]
78. Mhouti, A.E.; Nasseh, A.; Erradi, M. How to evaluate the quality of digital learning resources. *Int. J. Comput. Sci. Res. Appl.* **2013**, *3*, 27–36.
79. Cooper, G. Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. *J. Sci. Educ. Technol.* **2023**, *32*, 444–452. [CrossRef]
80. Javaid, M.; Haleem, A.; Singh, R.P.; Khan, S.; Khan, I.H. Unlocking the opportunities through ChatGPT tool towards ameliorating the education system, BenchCouncil Transactions on Benchmarks. *Stand. Eval.* **2023**, *3*, 100115.
81. Jiang, J.; Liu, X.; Wang, Z.; Ding, W.; Zhang, S. Large group emergency decision-making with bi-directional trust in social networks: A probabilistic hesitant fuzzy integrated cloud approach. *Inf. Fusion* **2024**, *102*, 102062. [CrossRef]


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.