Routledge
Taylor & Francis Group

Ethics & Behavior

ISSN: 1050-8422 (Print) 1532-7019 (Online) Journal homepage: www.tandfonline.com/journals/hebh20

Development and psychometric evaluation of the artificial intelligence ethical awareness scale

Yalçın Kanbay, Burcu Arkan, Aysun Akçam & Belgin Varol

ETHICS & BEHAVIOR
https://doi.org/10.1080/10508422.2025.2595433

Development and psychometric evaluation of the artificial intelligence ethical awareness scale

Yalçın Kanbay [a], Burcu Arkan [b], Aysun Akçam [c], and Belgin Varol [d]

[a]Psychiatric Nursing Department, Faculty of Health Sciences, Artvin Çoruh University; [b]Psychiatric Nursing Department, Uludağ University; [c]Turkish Education Department, Necmettin Erbakan University; [d]Psychiatric Nursing Department, Gulhane Nursing Faculty, University of Health Sciences

ABSTRACT

The rapid integration of artificial intelligence (AI) into daily life has created ethical challenges, highlighting the need for responsible use. This methodological study developed the AI Ethical Awareness Scale to measure individuals' awareness levels. Data were collected from 616 AI users, and analyses confirmed a three-factor, nine-item structure: Transparency, Integrity, and Awareness. The scale demonstrated strong reliability and validity, explaining 74.8% of the variance. Developed within the Turkish cultural context, it is a valid and reliable tool to assess AI ethical awareness. The AIEAS can be used to assess the effectiveness of ethics education programs or inform policy recommendations regarding responsible AI use. It also holds potential for adaptation across different cultural settings to support cross-cultural research.

KEYWORDS

Artificial intelligence; AI ethics; ethical awareness; scale development

INTRODUCTION

Today, artificial intelligence (AI) technologies are profoundly transforming human life in various domains, particularly in education, healthcare, security, communication, and decision-making processes (Buttazzo, 2023; Ocen et al., 2025). With the growing interest in AI, the number of users is rapidly increasing. This transformation not only brings improvements in functionality and efficiency but also raises ethical issues and uncertainties (Askin et al., 2023; Bélisle-Pipon & Victor, 2024). Ethical debates surrounding AI include its autonomy in decision-making processes, data privacy, algorithmic biases, impacts on human rights, and the chain of responsibility, all of which underscore the importance of individuals' ethical awareness (González et al., 2024). Furthermore, ethical awareness diverges from related constructs such as moral sensitivity or ethical judgment, as it refers specifically to individuals' ability to notice and interpret the ethical dimensions of AI use. This perspective aligns with Floridi's et al. (2018) digital ethics framework, which highlights beneficence, non-maleficence, autonomy, justice, and explicability as core principles guiding ethical reflection in AI contexts. Ethical awareness is critical not only for software developers and experts but also for users and citizens at all levels. The ability of individuals to make informed, sensitive, and ethically grounded decisions regarding AI systems is of great significance for both individual security and societal well-being (UNESCO, 2021). In this context, measuring individuals' levels of ethical awareness in their interactions with AI technologies has become an urgent need for both scientific research and policy development.

International organizations have also provided various recommendations in this regard. UNESCO (2021) emphasizes the need to enhance ethical sensitivity, digital skills, and information literacy at the societal level to ensure that individuals can use AI systems consciously and ethically. Similarly, the

CONTACT Belgin Varol varolbelgin@yahoo.com Sağlık Bilimleri Üniversitesi, Gülhane Hemşirelik Fakültesi, Emrah mah, Keçiören, Ankara 06010, Türkiye

ethical guidelines published by the European Union highlight that AI should be human-centered, with particular emphasis on principles such as transparency, environmental sensitivity, accountability, privacy, and data security (European Parliament, 2019). A comprehensive study on AI ethics further identified transparency, fairness, responsibility, non-maleficence, and privacy as the universal ethical values that stand out across contexts (Huang et al., 2023). These principles demonstrate that ethical informed judgment should be addressed as a multidimensional construct.

Several measurement tools have been developed in the literature to assess ethical attitudes or knowledge levels related to AI. For example, the scale developed by Kim and Ko (2022) focuses on system-level ethical principles such as transparency, fairness, safety, and accountability. However, this scale is considered to lack sufficient content regarding individual ethical behaviors, user responsibility, and sensitivity to societal impacts. Similarly, the Attitudes Toward the Ethics of Artificial Intelligence scale was designed exclusively to measure university students' ethical attitudes, which limits its applicability to the general population (Jang et al., 2022). The AI Ethical Awareness and Academic Integrity scale, on the other hand, was developed specifically for health sciences students (Abuadas & Albikawi, 2025). Furthermore, in the Artificial Intelligence Literacy Scale for Chinese College Students, ethics is addressed only as one of the subdimensions and in a rather limited scope (Ma & Chen, 2024). In addition, a generative AI awareness scale developed for secondary school students in Türkiye focuses primarily on students' general awareness and does not address broader user-centered ethical sensitivity (Semerci Şahin et al., 2025). A common feature of these instruments is that they mostly evaluate general ethical tendencies and do not sufficiently address user-centered ethical issues specific to AI. This creates limitations for both the advancement of scientific research and the development of effective policies. However, existing instruments are considered to insufficiently reflect user-centered ethical reflexivity, especially individuals' ability to recognize ethical risks, reflect on societal impacts of AI, and make ethically informed decisions. Therefore, there is an increasing need for valid and reliable psychometric tools capable of measuring individuals' ethical sensitivity in their interactions with AI-based systems, their capacity to recognize ethical risks, and their level of ethical reflexivity toward these technologies.

The primary aim of this study is thus to develop a valid and reliable scale that can measure individual levels of ethical awareness regarding AI. The developed scale is expected to contribute to multiple areas of application, such as assessing the effectiveness of ethics education programs, conducting user-centered ethical evaluations of AI systems, and informing the development of ethical policy recommendations. Furthermore, it holds theoretical significance by providing an academic foundation for understanding individuals' relationship with technology and their **ethical sensitivity** in digital environments.

MATERIALS AND METHODS

Aim of the Study

The primary aim of this research was to develop a measurement tool to assess individuals' levels of ethical awareness regarding artificial intelligence (AI) technologies and to examine its psychometric properties, including validity and reliability.

Study Design

This study employed a cross-sectional methodological design (DeVellis, 2017).

Ethical Consideration

To develop the AI Ethical Awareness Scale (AIEAS) within the Turkish cultural context, ethical approval was obtained from the Ethics Committee of Health Sciences at Çankırı Karatekin

University (Meeting No: 20, Date: 19.04.2025). Informed consent was obtained from all participants. ChatGPT-5.0 was used only for language editing and formatting; it did not contribute to the conceptual development, analysis, or interpretation of the study. All AI-assisted edits were carefully reviewed and verified by the authors to ensure originality, accuracy, and consistency with the intended meaning.

Scale Development Process

The development of the AIEAS was carried out in two phases and seven steps (Figure 1). Phase I focused on defining the conceptual framework, while Phase II involved psychometric evaluations.

Phase I. Conceptual framework

The first phase consisted of three steps: generating the item pool, establishing content validity, and conducting a pilot study.

Step 1. Item Pool Generation. The initial step in developing the AIEAS was to generate a comprehensive item pool that represented the construct to be measured. To this end, a systematic review of national and international literature on AI ethics and ethical awareness was conducted. The review focused on themes such as ethical awareness, algorithmic bias, data privacy, transparency, human rights, human – machine relations, AI responsibility, and accountability (Bayram, 2025; Floridi et al., 2018; Hagendorff, 2020; Huang et al., 2023; Jobin et al., 2019; Mittelstadt et al., 2016). Based on these themes, cognitive, affective, and behavioral items were developed to reflect individuals' capacities for ethical evaluation and their ethical sensitivity regarding AI technologies.

In addition, during the item pool generation process, the five ethical domains proposed by Floridi et al. (2018)—beneficence, non-maleficence, autonomy, justice, and explicability/transparency – were adopted as the conceptual framework. Aligning the item pool with these domains ensured that the scale reflected contemporary digital ethics principles and addressed fundamental ethical considerations relevant to AI use., Expert opinions from different disciplines were also sought during the item generation process. A total of seven academics specializing in ethics, artificial intelligence, communication, and educational sciences were consulted to contribute to item development and provide topic-specific recommendations. Following their feedback, some items were added, others were revised, and items considered inadequate in terms of content were removed. As a result, a draft form consisting of 28 original items was developed to capture the subdimensions of the construct of ethical awareness.

Step 2. Content Validity. Following the generation of the item pool, a content validity analysis was planned through expert evaluation. The aim of this process was to determine the extent to which the items represented the intended construct and to assess their adequacy in terms of content. Accordingly, feedback was sought from a total of ten experts with diverse backgrounds, representing disciplines such as artificial intelligence, ethics, measurement and evaluation, educational technologies, and psychology. The evaluation was conducted using the Content Validity Ratio (CVR) technique developed by Lawshe (1975). For each item, a CVR value was calculated and compared against Lawshe's minimum acceptance thresholds. Items that were deemed inadequate were either removed from the pool or revised. In addition to item-level CVR values, the Content Validity Index (CVI) was also calculated to provide an overall index of content validity for the scale.

Step 3. Pilot Study. The purpose of the pilot study was to test the suitability of the developed scale for the target population. Therefore, the selected sample needed to represent the target population as closely as possible by considering factors such as age, gender, education level, and occupation (Erkuş, 2012; Trochim & Donnelly, 2006). It has been suggested that the sample size for a pilot study should typically range between 30 and 50 participants (Boateng et al., 2018). This number is considered sufficient to test the draft version of the scale at the initial stage and to make necessary revisions. Accordingly, the trial form, which had already undergone content validity analysis, was administered to a sample of 55 participants with characteristics similar to the target population, in order to determine whether the items were properly understood.

Phase II. Psychometric evaluation

This phase consisted of the following steps: defining the population and sample, data collection, data analysis, and examination of validity and reliability.

Step 1. Population and sample. As this was a scale development study, it was necessary to employ factor analysis techniques. Initially, exploratory factor analysis (EFA) was conducted to identify the underlying structure, followed by confirmatory factor analysis (CFA) to validate the emerging structure. Therefore, two separate but comparable samples were required. Achieving an adequate sample size is essential in scale development studies. In general, the sample size for factor analysis should be at least five times the number of items, although larger samples are often recommended. For instance, Comrey and Lee (1992) suggested that a minimum of 300 participants is appropriate for a robust factor analysis. The data collection process in this study was carried out in two stages. In the first stage, data were collected to identify the underlying factor structure of the scale, and an Exploratory Factor Analysis (EFA) was conducted on this dataset. In the second stage, additional data were collected from an independent sample to confirm the factor structure, and a Confirmatory Factor Analysis (CFA) was performed. Through these two phases, a total of 616 participants were reached, of whom 310 were used for EFA and 306 for CFA.

In the EFA sample, 50.3% were male, and 64.2% were undergraduate students or graduates. A total of 49.3% reported daily digital device use of 4–6 hours, while 33.3% reported frequent use of AI tools during that time. Nearly half of the participants (48%) stated that they had a moderate level of knowledge about ethical issues related to AI use, yet only 6% had received formal training on the subject. The mean age of the EFA sample was 29.9 ± 10.4 years (range: 18–56). In the CFA sample, 50.3% were female, and 68.3% were undergraduate students or graduates. A total of 51.6% reported daily digital device use of 4–6 hours, with 33.7% frequently using AI tools. More than half (50.7%) reported having a moderate level of knowledge about ethical issues related to AI use, while only 6.2% had received training on the subject. The mean age of the CFA sample was 28.8 ± 10.2 years (range: 18–62).

Step 2. Data collection. The data collection process was carried out using Google Forms, an online tool that enables access to large audiences and allows participants to complete the survey at their own convenience. The survey link was distributed through e-mail groups and social media platforms to reach potential participants. Data were collected between [29 April- 07 July 2025], and participation was entirely voluntary. Prior to completing the survey, participants were informed about the purpose of the study, and anonymity and confidentiality of the responses were ensured.

Step 3. Data analysis. In the data analysis phase, descriptive statistics and EFA were performed using SPSS version 26, while CFA was conducted with AMOS version 23. Descriptive statistics included frequencies, means, and percentages. Correlation analysis was employed to examine relationships, whereas construct validity was assessed through EFA and CFA. Before conducting the factor analysis, the normality of the dataset was examined, and the skewness and kurtosis values for all items were evaluated. The analysis showed that the skewness and kurtosis values of the items were within the range of −2 to +2. This range is widely accepted in the literature as indicative of approximate normality and demonstrates that the data are suitable for factor analysis. For convergent validity, formulas developed in Microsoft Excel were used to calculate the Average Variance Extracted (AVE) and Composite Reliability (CR) values. The reliability of the scale was examined through Cronbach's alpha coefficients and split-half reliability analysis.

Step 4. Validity and reliability. To determine the construct validity of the scale, EFA was performed using the principal components method. Since the factors were assumed to be correlated, an oblique rotation technique, specifically direct oblimin, was applied. CFA was subsequently conducted to test the fit of the factor structure obtained from the EFA, and goodness-of-fit indices were examined. The following thresholds were considered acceptable: .05 <RMSEA <.10; .90 ≤ CFI ≤.95; .90 ≤ GFI ≤.95; AGFI > .90 (Meydan & Şeşen, 2011; Schermelleh-Engel & Moosbrugger, 2003; Wang & Wang, 2012).

For convergent and discriminant validity, AVE and CR values were evaluated. Convergent validity was considered to be supported when AVE values were ≥.50 (Bagozzi & Yi, 1988), while CR values ≥.70 indicated adequate construct reliability (Hair et al., 2020). Regarding reliability, Cronbach's α coefficients were examined, with values above .70 regarded as evidence of sufficient reliability for research instruments (DeVellis, 2017).

RESULTS

This section presents the findings related to the validity and reliability of the scale, the confirmation of the obtained factor structure, and the relevant statistical results.

Validity

The validity of the scale was examined through content validity, construct validity, and internal validity.

Content Validity

In this study, content validity was assessed through expert evaluations from 10 specialists. The critical value for 10 experts was calculated as .62. Three items (i6, i15, i25) had CVR values below the threshold of .62 and were therefore removed from the item pool. The remaining 25 items demonstrated CVR values ranging from .80 to 1.00. The CVI value for 20 items was calculated as .912, and it was found to be higher than the CVR value (CVI > CVR). These findings indicate that, following item elimination, the draft form consisting of 25 items demonstrated sufficient content validity as a whole (Table 1).

Table 1. Content validity findings.

Item	Ne	N	CVR ((Ne−N/2)/(N/2)	Critical Value	Decision	Item	Ne	N	CVR ((Ne−N/2)/(N/2)	Critical Value	Decision
i1	9	10	0.80	.62	Retain	***i15**	**8**	**10**	**0.60**	**.62**	**Remove**
i2	9	10	0.80	.62	Retain	**i16**	9	10	0.80	.62	Retain
i3	9	10	0.80	.62	Retain	**i17**	10	10	1.00	.62	Retain
i4	9	10	0.80	.62	Retain	**i18**	10	10	1.00	.62	Retain
i5	9	10	0.80	.62	Retain	**i19**	10	10	1.00	.62	Retain
***i6**	**8**	**10**	**0.60**	**.62**	**Remove**	**i20**	10	10	1.00	.62	Retain
i7	9	10	0.80	.62	Retain	**i21**	10	10	1.00	.62	Retain
i8	9	10	0.80	.62	Retain	**i22**	10	10	1.00	.62	Retain
i9	9	10	0.80	.62	Retain	**i23**	10	10	1.00	.62	Retain
i10	9	10	0.80	.62	Retain	**i24**	10	10	1.00	.62	Retain
i11	10	10	1.00	.62	Retain	***i25**	**7**	**10**	**0.40**	**.62**	**Remove**
i12	9	10	0.80	.62	Retain	**i26**	10	10	1.00	.62	Retain
i13	10	10	1.00	.62	Retain	**i27**	10	10	1.00	.62	Retain
i14	10	10	1.00	.62	Retain	**i28**	10	10	1.00	.62	Retain

Panel size = 10 experts

Critical CVR = .62

S-CVI = .912

Ne: number of experts rating the item as "Essential"; N: number of experts; CVR: Content Validity Ratio; S-CVI: Scale-level Content Validity Index. * Items below the critical CVR threshold (*N* = 10 → CVR_crit = .62) and therefore removed.

Construct Validity

When the item – total correlations were examined, four items (i2, i5, i9, i24) were found to have values below the .30 threshold and were therefore removed from the scale (Çokluk et al., 2014; Nunnally & Bernstein, 1994; Şencan, 2005). This left 22 items in the pool. The item – total correlations of the remaining items ranged between .469 and .663 (Table 2).

According to the results, the Kaiser – Meyer – Olkin (KMO) measure was .821. Bartlett's Test of Sphericity yielded a significant result ($\chi^2 = 1272.542$; df = 36; $p < .001$). These values, which fall within the recommended limits in the literature (Büyüköztürk, 2010), indicate that the sample size was adequate and the data matrix was suitable for factor analysis (Table 3).

For the EFA, the principal components method was used, and since the factors were assumed to be correlated, an oblique rotation technique (direct oblimin) was applied with six iterations. Items with factor loadings below .50 and those that cross-loaded on multiple factors were excluded from the analysis. The results yielded a three-factor structure consisting of nine items with eigenvalues greater than 1 (Table 3). The total variance explained by the scale was calculated as 74.8%.

First factor, Transparency, refers to the evaluation of principles such as proper citation of AI-generated content, respect for copyright, and openness and honesty in knowledge production. It measures individuals' ethical sensitivity regarding the disclosure of the origins of AI-assisted content, the legitimate use of information, and respect for intellectual property rights. Transparency is based on the idea that individuals should act openly and honestly both toward themselves and their audiences when producing digital content. In contexts where AI technologies

Table 2. Item means, corrected item – total correlations (CITC), and Cronbach's alpha if item deleted.

Item	Mean	SD	α	CITC	Item	Mean	SD	α	CITC
i1	3.71	1.04	0.916	0.542	i16	4.19	0.70	0.916	0.527
***i2**	**4.00**	**0.85**	**0.915**	**0.269**	i17	3.72	1.01	0.915	0.583
i3	3.67	1.00	0.917	0.579	i18	3.88	0.95	0.915	0.596
i4	2.92	1.15	0.918	0.504	i19	4.03	0.88	0.914	0.602
***i5**	**3.77**	**1.03**	**0.917**	**0.233**	i20	4.45	0.70	0.915	0.589
i7	3.61	0.99	0.916	0.493	i21	4.52	0.67	0.917	0.469
i8	4.22	0.79	0.916	0.499	i22	4.59	0.58	0.917	0.497
***i9**	**4.31**	**0.76**	**0.916**	**0.226**	i23	4.08	0.84	0.916	0.530
i10	4.08	0.86	0.913	0.662	***i24**	**4.43**	**0.81**	**0.917**	**0.256**
i11	4.35	0.69	0.915	0.576	i26	4.12	0.88	0.914	0.605
i12	4.04	0.87	0.915	0.566	i27	4.25	0.72	0.915	0.605
i13	3.94	0.92	0.916	0.531	i28	4.10	0.82	0.915	0.663
i14	3.73	1.00	0.914	0.632	–	–	–	–	–

SD: standard deviation; CITC: corrected item – total correlation. *Items with CITC <0.30 and therefore removed. "α if item deleted" represents the scale's Cronbach's alpha when the item is omitted.

Table 3. Efa findings for the factor structure of the scale.

No	Item	F1	F2	F3
i17	I respect copyright of the content I create with artificial intelligence.	.923		
i18	I cite sources in the content I create with artificial intelligence.	.883		
i14	I investigate the sources of AI-assisted content.	.649		
i22	I do not use artificial intelligence in ways that disturb others.		.937	
i21	I am against the use of fake identities/profiles created by artificial intelligence.		.905	
i20	I do not use AI-based chatbots to deceive people.		.771	
i23	I am aware of the negative effects of artificial intelligence on human relationships.			.871
i26	I know that artificial intelligence may pose risks in terms of social justice.			.859
i28	I take into account the negative impacts of artificial intelligence on the environment.			.844
	Explained Variance (%):	45.8	16.7	12.4
	Total Variance Explained: (%):		74.8	
	KMO: .821; Barlett's Test: 1272.542; df: 36; $p < .001$			
	Method: Principal Component Analyses; Rotation: Direct Oblimin (6 iterations)			

Factor 1: Transparency, Factor 2: Integrity, Factor 3: Awareness.

Table 4. 27% lower–upper group comparison.

Factor	Group	n	M	SD	t	p
F1. Transparency	Lower Group	84	8.13	1.64	−30.344	<.001
	Upper Group		14.20	.82		
F2. Integrity	Lower Group	84	11.33	1.25	−26.784	<.001
	Upper Group		14.40	1.01		
F3. Awareness	Lower Group	84	9.50	1.52	−31.161	<.001
	Upper Group		14.82	.38		
AIEAS (total)	Lower Group	84	30.69	3.13	−32.199	<.001
	Upper Group		43.01	1.58		

M = Mean, SD = Standard Deviation, AIEAS = Artificial Intelligence Ethical Awareness Scale. Lower and upper groups were determined based on the 27% extreme groups method.

are used as producers, explicitly indicating the source of the content, respecting copyright, and ensuring the traceability of information are considered fundamental indicators of ethical awareness. Transparency reflects the extent to which individuals adhere to the principles of digital ethics, information ethics, and academic integrity. The variance explained by this factor was 45.8%.

Second factor, Integrity, measures the extent to which individuals adhere to ethical principles when using AI technologies, such as avoiding harm to others, refraining from creating deceptive content, and rejecting the production of fake identities. This dimension evaluates ethical decision-making and individuals' stance against the potential misuse of AI applications. Integrity represents individuals' sensitivity to the impact of their digital actions on others and their awareness of violations of ethical boundaries. It is directly related to cyber ethics, digital citizenship, resistance to deception, and digital moral autonomy. The variance explained by this factor was 16.7%.

Third factor, Awareness, encompasses ethical sensitivity toward the social, environmental, and relational impacts of AI technologies. The items reflect individuals' level of consciousness regarding the negative effects of AI on human relationships, social justice, and environmental sustainability. This dimension represents not only individuals' ethical practices but also their capacity to reflect on the systemic implications of AI. Awareness is closely linked to critical technology literacy, the relationship between technology and social structures, ethical foresight, and social responsibility. The variance explained by this factor was 12.4%.

Internal Validity

The internal validity of the items constituting the scale was tested using the 27% upper – lower group comparison method. This method evaluates the extent to which an item distinguishes between individuals with high and low scores. Participants were ranked according to their total scale scores, and the top 27% (upper group) and the bottom 27% (lower group) were compared. The results indicated statistically significant differences (p <.001) between the mean scores of the upper group and the lower group, both at the factor level and for the overall scale. These findings demonstrate that the structure successfully differentiates between individuals with high and low levels of ethical awareness, thus supporting the internal validity of the scale (Table 4).

2Reliability

The Cronbach's α reliability coefficients of the identified structure were calculated as .79 for the Transparency factor, .85 for the Integrity factor, .82 for the Awareness factor, and .84 for the overall scale. The Gutman values were .73, .81, .76, and .88, respectively. The Spearman – Brown split-half coefficients for the factors and the total scale ranged between .634 and .778. These values fall within the recommended thresholds in the literature (DeVellis, 2017). Based on these findings, it can be concluded that the scale demonstrated an adequate level of reliability (Table 5).

Table 5. Reliability findings.

Factor	r	Gutman	Cronbach's α
F1. Transparency	.675	.73	.79
F2. Integrity	.778	.81	.85
F3. Awareness	.648	.76	.82
AIEAS (total)	.634	.88	.84

r = Spearman-Brown Coefficient, AIEAS = Artificial Intelligence Ethical Awareness Scale

Confirmation of the Factor Structure

The construct validity of the scale was further examined through CFA. In evaluating the CFA results, goodness-of-fit indices such as RMSEA (Root Mean Square Error of Approximation), CFI (Comparative Fit Index), GFI (Goodness-of-Fit Index), AGFI (Adjusted Goodness-of-Fit Index), and NFI (Normed Fit Index) were taken into consideration. The three-factor structure of the AIEAS was tested using a second-order multifactor CFA model. Examination of the structural paths indicated that all three factors significantly contributed to the overall construct (Figure 2). Furthermore, all items significantly loaded onto their respective factors (p <.001).

The goodness-of-fit values obtained for the model were as follows: $\chi^2/df = 2.479$, RMSEA = .070, CFI = .97, GFI = .96, AGFI = .92, and NFI = .95. These indices were found to be within the acceptable thresholds recommended in the literature (Meydan & Şeşen, 2011; Schermelleh-Engel & Moosbrugger, 2003; Wang & Wang, 2012).

According to the CFA results, the standardized factor loadings for the Transparency factor ranged from .63 to .86, the loadings for the Integrity factor ranged from .72 to .85, and the loadings for the Awareness factor ranged from .72 to .82. The item reliability coefficients (R^2) corresponding to these loadings were found to be at acceptable levels, and the error variances were observed to be low. These findings indicate that the items adequately represent their respective factors and that the model has

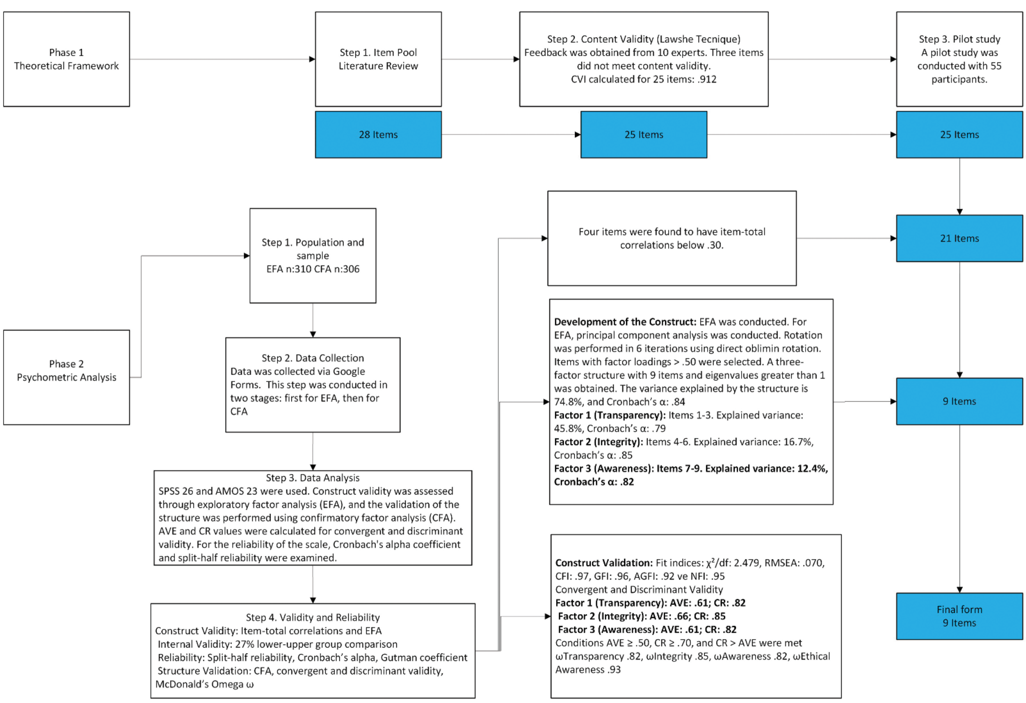

Figure 1. Research methodology.

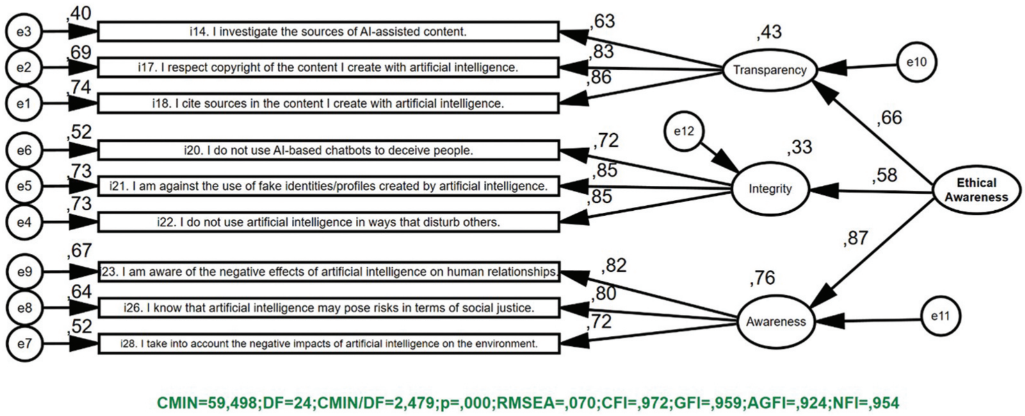

Figure 2. CFA results of the scale's factor structure.

a statistically robust structure. Overall, the CFA results confirm the three-factor structure of the scale, demonstrating that the items contribute meaningfully to the factor structure and that the model possesses strong validity.

Convergent and Discriminant Validity

Convergent and discriminant validity of the scale were examined using Average Variance Extracted (AVE) and Composite Reliability (CR) values. Convergent validity is considered to be supported when AVE values are ≥.50 and CR values are ≥.70 (Bagozzi & Yi, 1988; Hair et al., 2020).

For the three-factor structure of the scale, the AVE and CR values were as follows: Transparency (AVE = .61, CR = .82), Integrity (AVE = .66, CR = .85), and Awareness (AVE = .61, CR = .82). Since the AVE values for all three factors exceeded .50, the average variance extracted by the items was deemed significant. In addition, the CR values being greater than .70 indicated sufficient composite reliability. When AVE and CR were considered together, the criteria for convergent validity (AVE ≥ .50, CR ≥.70, and CR > AVE) were met, confirming that the scale demonstrated adequate convergent validity.

For discriminant validity, the square roots of the AVE values are presented in parentheses and in bold along the diagonal of the correlation matrix. As these values were higher than the corresponding inter-factor correlations, discriminant validity was established. Furthermore, based on the factor loadings obtained from the CFA, McDonald's Omega (ω) coefficients were calculated. High reliability values were observed for both the factors and the overall scale (ωŞeffaflık .82, ωSorumluluk .85, ωDuyarlılık .82 ve ωEtik Farkındalık .93) (Table 6).

Table 6. Convergent and discriminant validity.

Factor	n	AVE	CR	Omega (ω)
F1. Transparency	3	.61	.82	.82
F2. Integrity	3	.66	.85	.85
F3. Awareness	3	.61	.82	.82
AIEAS (total)	9	–	–	.93
Discriminant Validity				
	F1	F2	F3	
F1. Transparency	(.781)			
F2. Integrity	.379**	(.812)		
F3. Awareness	.498**	.446**	(.781)	

AVE: Average Variance Extracted; CR: Composite Reliability; ω: McDonald's Omega, *p* <.01.

Test–Retest Reliability

To assess the temporal stability of the scale, the test – retest method was employed. Test – retest reliability is used to determine whether a measurement instrument provides consistent results over time. For medium-term psychological constructs such as motivation and attitudes, an interval of 2 to 4 weeks between measurements is typically recommended (Hinkin, 1998). In line with this, the scale was re-administered to a group of 63 participants after a two-week interval. The correlation coefficient between the two measurements was found to be .83 (p <.001). This high correlation indicates that the scale provides stable and consistent measurements over time.

DISCUSSION

The present study aimed to develop and validate the AI Ethical Awareness Scale, and the findings were discussed in light of the existing literature. In this study, factor-analytic techniques were employed to identify the structure of the AIEAS. Prior to conducting factor analysis, it is essential to examine whether the items contribute adequately to the scale. For this purpose, item – total correlations were analyzed. Item – total correlation is a widely accepted method in the literature to enhance the validity and reliability of a scale. A correlation value of .30 or higher is generally considered acceptable, as it indicates that the item represents the overall construct well (Nunnally & Bernstein, 1994). Low item – total correlations suggest that an item may not sufficiently contribute to the construct being measured and may reduce the reliability of the scale (DeVellis, 2017). Therefore, items with item – total correlations below .30 are generally regarded as problematic and should either be revised or removed (Çokluk et al., 2014; Şencan, 2005). In this study, four items had item – total correlations below .30 and were removed from the scale. The remaining items, which had correlations above the .30 threshold, were deemed to contribute adequately to the construct and were retained for factor analysis.

According to the literature, appropriate conditions for conducting factor analysis require that the KMO test result be ≥.70 and that Bartlett's Test of Sphericity yield a significant result at p <.05 (Alpar, 2016; Erdoğan et al., 2014). In this study, the KMO value was above the recommended threshold (KMO = .821), and Bartlett's Test produced a significant result (χ^2 = 1272.542; df = 36; p <.001). These findings indicate that the sample size was adequate and that the dataset was suitable for factor analysis.

The construct validity of the scale was determined through exploratory factor analysis (EFA). The principal components method was employed, and given the assumption that the factors were correlated, an oblique rotation technique (direct oblimin) was applied. Factors with eigenvalues greater than 1 were retained, and the scree plot was also examined to determine factor structure. It was preferred that each factor explain at least 5% of the variance and that the total variance explained by the scale be at least 50%. For item elimination, a minimum factor loading of .50 was adopted, and items were considered cross-loading if the difference between loadings on different factors was less than .10; such items were removed from the scale (Büyüköztürk, 2010). As a result of the analysis, a three-factor structure with eigenvalues greater than 1 was obtained, and rotation was completed in six iterations. After the removal of items with factor loadings below .50 and items that cross-loaded on multiple factors, a three-factor structure explaining 74.8% of the variance was identified. Considering that in multifactor instruments the total variance explained should be at least 40%, and in unidimensional instruments at least 30% (Field, 2018), the explained variance obtained in this study indicates that the scale meets the adequacy criteria for multifactor measurement tools.

Similarly, Bayram (2025), in a scale development study, obtained a four-factor structure consisting of 12 items (bias, transparency, accountability, and data privacy), with each factor including three items and a total variance explained of 63%. In the present study, however, a three-factor structure with nine items (three per factor) was obtained, explaining 74.8% of the variance. This higher level of explained variance and the parsimonious factor structure indicate that the developed scale is both practical to use and conceptually robust in capturing the construct of AI ethical awareness.

To assess the internal validity of the scale, a 27% upper – lower group comparison was conducted. Item discrimination is evaluated by examining the significance of the difference between the lower and upper groups (p <.05) (DeVellis, 2017). In this study, statistically significant differences were observed (p <.001) between the scores of the lower group (with low scale scores) and the upper group (with high scale scores), both for the overall scale and for each factor. This finding was considered an important indicator of internal validity (Table 4).

For the reliability of the scale, Cronbach's α, Gutman reliability coefficients, and split-half reliability were examined. A Cronbach's α coefficient of .70 or higher is recommended for research instruments (DeVellis, 2017). Split-half reliability, one of the methods used to assess internal consistency, ranges between 0 and 1, with values ≥.70 indicating acceptable reliability. In Bayram's (2025) study, Cronbach's α values for the factors ranged from .62 to .72 (bias: .67, transparency: .74, accountability: .62, and data privacy: .72), while the total scale reliability was reported as .84. Since the α values for the bias and accountability factors were below .70, the reliability levels for these factors were considered weak. In contrast, in the present study, Cronbach's α values ranged between .79 and .85 for the factors, and .84 for the overall scale, all of which are sufficient. The split-half reliability results further indicated positive and strong correlations between the two halves of the scale. Taken together, these findings suggest that the scale demonstrated a high level of internal validity.

The literature emphasizes that factor structures identified through EFA should be confirmed with CFA. To determine whether a model is validated, various model fit indices are examined. However, as noted in previous studies, there are numerous fit indices, and it is difficult to establish a single standard set of indices to be universally accepted (Çokluk et al., 2014; Koyuncu & Kılıç, 2019; Orcan, 2018). In the present study, indices such as χ^2/df, RMSEA, CFI, GFI, AGFI, and NFI were used, and the acceptable ranges suggested in the literature were taken as reference (Meydan & Şeşen, 2011; Schermelleh-Engel & Moosbrugger, 2003; Wang & Wang, 2012). The fit indices for the model were χ^2/df = 2.479, RMSEA = .070, CFI = .97, GFI = .96, AGFI = .92, and NFI = .95 (Figure 2). These values fall within the recommended thresholds in the literature. Based on these results, the three-factor structure of the scale was confirmed.

AVE and CR are two important indicators used in structural equation modeling and scale development studies to evaluate the validity and reliability of a measurement tool. AVE refers to the proportion of the total variance explained by a construct through its items. In other words, AVE indicates the extent to which the items of a scale represent the overall construct. An AVE value of .50 or above is considered acceptable, as it shows that the items account for at least 50% of the variance in the construct. An AVE value below .50 suggests weak explanatory power of the construct and that the items may need to be revised (Fornell & Larcker, 1981).

CR, on the other hand, is a reliability coefficient that assesses the internal consistency of a construct. It is regarded as an alternative to Cronbach's alpha, as it calculates reliability by taking into account the factor loadings of the items. A CR value of .70 or higher indicates adequate internal consistency. While AVE evaluates convergent and discriminant validity, CR assesses the internal consistency of the scale items (Hair et al., 2020). For a measurement tool to be considered valid and reliable, AVE values should be ≥.50 and CR values ≥.70.

In this study, the AVE values were greater than .50 and CR values greater than .70 for all factors, and the square roots of the AVE values were higher than the corresponding inter-factor correlations, confirming both convergent and discriminant validity of the scale. Additionally, based on the factor loadings obtained from the CFA, McDonald's Omega (ω) values were calculated. McDonald's Omega is an internal consistency coefficient that indicates the proportion of variance in the items explained by the latent construct. Unlike Cronbach's alpha, which assumes tau-equivalence (equal contribution of items), Omega relaxes this assumption by incorporating factor loadings, thus providing a more realistic estimate of reliability. In this study, Omega values were calculated as ωTransparency .82, ωIntegrity .85, ωAwareness .82 and ωAIEAS.93. These values indicate high reliability for both the individual factors and the overall scale (Dunn et al., 2014).

The results of the independent samples t-test by gender are presented in Table 7. According to the analyses, female participants scored significantly higher than males on the Transparency factor (t = 2.956, p = .003). Similarly, females obtained significantly higher scores on the Awareness factor

Table 7. Differences in Mean scores by gender.

Factor	Gender	n	X	Sd.	t	p
Transparency	Female	154	11.8	2.3	2.956	.003
	Male	152	10.9	2.6		
Integrity	Female	154	13.5	1.7	.297	.767
	Male	152	13.4	1.8		
Awareness	Female	154	12.6	2.1	2.672	.008
	Male	152	11.9	2.2		
AIEAS (total)	Female	154	37.9	4.9	2.680	.008
	Male	152	36.3	5.1		

compared to males ($t = 2.672$, $p = .008$). When examining the total score of the scale, female participants also had significantly higher AIEAS total scores than male participants ($t = 2.680$, $p = .008$). These findings suggest that females may have a higher level of ethical awareness regarding artificial intelligence. In contrast, no significant difference was observed between females and males in the Integrity factor ($t = .297$, $p = .767$), indicating that awareness related to the principle of integrity does not vary by gender.

Although the AIEAS does not specifically aim to measure broader areas such as digital ethics, digital citizenship, or media ethics, the scale's emphasis on transparency, responsibility, and awareness resonates with these wider conceptual frameworks. Previous research shows that responsible technology use is closely linked to individuals' digital citizenship skills, ethical online behavior, and awareness of digital risks (Özbay et al., 2021). In this sense, the AIEAS may provide a useful starting point for future educational or organizational efforts that aim to embed AI ethics within broader digital ethics initiatives. Moreover, the structure and content of the AIEAS suggest that it could be used as a practical tool in educational and organizational ethics programs, particularly in AI-based or digital learning environments, where raising awareness of transparency, responsibility, and ethical use is increasingly essential. Conclusion

In this study, a measurement tool was developed and validated to assess the ethical awareness levels of AI users within the Turkish cultural context. Content validity, construct validity, convergent and discriminant validity, reliability analyses, and confirmation of the factor structure demonstrated that the scale possesses adequate validity and reliability and provides consistent measurement in the Turkish context. Future studies may adapt and validate this scale in different cultural contexts to broaden its applicability.

Strengths of the study

The scale development process was grounded in the AI ethics literature, ensuring conceptual integrity by considering key issues such as algorithmic bias, data privacy, transparency, human rights, and ethical responsibility. This foundation strengthened the theoretical validity of the scale. The instrument was comprehensively evaluated through content validity (CVR, CVI), construct validity (EFA, CFA), convergent and discriminant validity (AVE, CR), internal consistency (Cronbach's α, Gutman, McDonald's Omega), and test – retest analyses. The three-factor structure (Transparency, Integrity, Awareness) explained 74.8% of the total variance, which is considered high. Cronbach's α coefficients ranged from .79 to .85, while McDonald's Omega values ranged from .82 to .93, indicating strong internal consistency.

The CFA fit indices (e.g., CFI = .97, GFI = .96, RMSEA = .070) demonstrated that the model showed a high degree of fit with the data. The study reached a sample of 616 individuals with diverse age, educational, and user experience backgrounds, supporting the generalizability of the findings to a broader population. Full compliance with ethical principles was ensured, with both ethics committee approval and participant consent obtained in accordance with scientific ethical standards. Additionally, the transformation of scale scores into a 0–100 range enhances the comparability of results across different studies, thereby increasing the applicability of the scale in various research and practice contexts.

Limitations

The sample was recruited on a voluntary basis through Google Forms. Consequently, the over-representation of individuals with digital access and an interest in technology may limit the generalizability of the findings to the broader population. This also introduces a potential self-selection bias, as individuals who voluntarily choose to participate online may systematically differ from those who do not. Since the survey instrument relied on self-reporting, the risk of social desirability bias was present, as participants may have responded in a way that portrayed themselves as more ethically sensitive than they actually are.

The scale was developed only in Turkish and within the cultural context of Türkiye, which may restrict its applicability in other cultural settings. In line with this, no universal cross-cultural applicability is claimed, and the scale should be interpreted within its cultural context. Although an English version of the scale (Appendix 1) was prepared for dissemination purposes, no formal translation or back-translation procedures were conducted. Accordingly, the current psychometric evidence applies solely to the Turkish version, and the English form should not be used for empirical research until a full linguistic validation is completed. Cross-cultural validation and adaptation studies are therefore needed. Moreover, given the rapid advances in AI technologies, perceptions and attitudes toward ethical awareness may shift over time. To ensure the continued relevance of the scale, periodic reassessment is recommended. Future research should also consider conducting longitudinal validation studies and multi-group measurement invariance testing to further strengthen the psychometric evidence of the scale.

The scale was not examined in relation to other psychological variables (e.g., ethical decision-making tendency, digital literacy, or duration of technology use), which limits its level of external validity. In addition, criterion validity was not assessed, and future studies should examine correlations between the AIEAS and related constructs such as AI literacy, ethical judgment, or moral sensitivity to further strengthen its external validity. Additionally, as the data were collected online, it was not possible to control the environment in which participants completed the survey, the level of attention they devoted, or potential external influences, which may have affected data quality.

ACKNOWLEDGMENTS

We sincerely thank all participants for their valuable time and contributions to this study.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author(s).

ORCID

Yalçın Kanbay http://orcid.org/0000-0002-8025-9877
Burcu Arkan http://orcid.org/0000-0002-7285-6196
Aysun Akçam http://orcid.org/0000-0001-9428-3942
Belgin Varol http://orcid.org/0000-0001-7138-9007

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

USE OF AI TOOLS

The authors used ChatGPT, a large language model developed by OpenAI, to support language editing and enhance clarity during the manuscript preparation process. All content was critically reviewed and approved by the authors to ensure accuracy and originality.

REFERENCES

Abuadas, M., & Albikawi, Z. (2025). AI ethical awareness and academic integrity in higher education: Development and validation of a new scale. *Ethics & Behavior*. Advance online publication. 1–18. https://doi.org/10.1080/10508422.2025.2511336

Alpar, R. (2016). *Uygulamalı istatistik ve geçerlik güvenirlik: Spor, sağlık ve eğitim bilimlerinden örneklerle*. Detay Yayıncılık.

Askin, S., Burkhalter, D., Calado, G., & El Dakrouni, S. (2023). Artificial intelligence applied to clinical trials: Opportunities and challenges. *Health and Technology*, *13*(2), 203–213. https://doi.org/10.1007/s12553-023-00738-2

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, *16*(1), 74–94. https://doi.org/10.1007/BF02723327

Bayram, V. (2025). Yapay zekâ Etiği: Bir Ölçek Geliştirme Çalışması. *Uluslararası İktisadi Ve İdari Akademik Araştırmalar Dergisi*, *5*(1), 123–148.

Bélisle-Pipon, J.-C., & Victor, G. (2024). Ethics dumping in artificial intelligence. *Frontiers in Artificial Intelligence*, *7*, 1426761. https://doi.org/10.3389/frai.2024.1426761

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, *6*, 149. https://doi.org/10.3389/fpubh.2018.00149

Buttazzo, G. (2023). Rise of artificial general intelligence: Risks and opportunities. *Frontiers in Artificial Intelligence*, *6*, 1226990. https://doi.org/10.3389/frai.2023.1226990

Büyüköztürk, Ş. (2010). *Sosyal Bilimler İçin Veri Analizi El Kitabı* (2nd ed.). Pegem Akademi.

Çokluk, Ö., Şekercioğlu, G., & Büyüköztürk, Ş. (2014). *Sosyal Bilimler için çok değişkenli istatistik: SPSS ve LISREL*. Pegem Akademi.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Lawrence Erlbaum Associates.

DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). Sage Publications.

Dunn, T. J., Baguley, T., & Brunsden, V. (2014, August). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399–412. https://doi.org/10.1111/bjop.12046

Erdoğan, S., Nahcivan, N., & Esin, N. (Eds.). (2014). Hemşirelikte araştırma: Süreç, uygulama ve kritik (1. baskı). In *N. Esin, Veri toplama yöntem ve araçları & veri toplama araçlarının güvenirlik ve geçerliği* (pp. 193–231). Nobel Tıp Kitapevleri.

Erkuş, A. (2012). Psikolojide ölçme ve ölçek geliştirme-i: Temel kavramlar ve işlemler. Pegem Akademi Yayınevi. s. 56, 60, 144.

European Parliament. (2019). Eu guidelines on ethics in artificial intelligence: Context and implementation [Briefing No. 640163]. European Parliamentary Research Service. https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI(2019)640163_EN.pdf

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). Sage Publications.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V.Vayena, E. & Vayena, E. (2018). Ai4people-an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39–50. https://doi.org/10.1177/002224378101800104

González, A. L., Moreno, M., Román, A. C. M., Fernández, Y. H., & Pérez, N. C. (2024). Ethics in artificial intelligence: An approach to cybersecurity. *Inteligencia Artificial*, *27*(73), 38–54. https://doi.org/10.4114/intartif.vol27iss73pp38-54

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, *30*(1), 99–120. https://doi.org/10.1007/s11023-020-09517-8

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2020). *Multivariate data analysis* (8th ed.). Cengage Learning.

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, *1*(1), 104–121. https://doi.org/10.1177/109442819800100106

Huang, C., Zhang, Z., Mao, B., & Yao, X. (2023). An overview of artificial intelligence ethics. *IEEE Transactions on Artificial Intelligence*, *4*(4), 799–819. https://doi.org/10.1109/TAI.2022.3194503

Jang, Y., Choi, S., & Kim, H. (2022). Development and validation of an instrument to measure undergraduate students' attitudes toward the ethics of artificial intelligence (AT-EAI) and analysis of its difference by gender and experience of ai education. *Education and Information Technologies*, *27*(6), 7901–7921. https://doi.org/10.1007/s10639-022-11086-5

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kim, D., & Ko, Y. (2022). Development and validation of ethical awareness scale for AI technology. *Journal of Digital Convergence*, *20*(1), 71–86. https://doi.org/10.14400/JDC.2022.20.1.071

Koyuncu, İ., & Kılıç, A. F. (2019). The use of exploratory and confirmatory factor analyses: A document analysis. *Education and Science*, *44*(198), 361–388. https://doi.org/10.15390/EB.2019.7665

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, *28*(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Ma, S., & Chen, Z. (2024). The development and validation of the artificial intelligence literacy scale for Chinese college students (AILS-CCS). *IEEE Access*, *12*, 146419–146429. https://doi.org/10.1109/ACCESS.2024.3468378

Meydan, H., & Şeşen, H. (2011). *Yapısal Eşitlik Modellemesi amos Uygulamaları*. Detay Yayıncılık.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2), 1–21. https://doi.org/10.1177/2053951716679679

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

Ocen, S., Elasu, J., Aarakit, S. M., & Olupot, C. (2025). Artificial intelligence in higher education institutions: Review of innovations, opportunities and challenges. *Frontiers in Education*, *10*, 1530247. https://doi.org/10.3389/feduc.2025.1530247

Orcan, F. (2018). Açımlayıcı ve doğrulayıcı faktör analizi: İlk hangisi kullanılmalı? Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, *9*(4), 414–421. https://doi.org/10.21031/epod.394323

Özbay, Ö., Doğan, U., Yıldız, E., & Seferoğlu, S. S. (2021). Examination of pre-service teachers' digital ethics levels in terms of various variables. *TRT Akademi*, *6*(12), 494–513. https://doi.org/10.37679/trta.913015

Schermelleh-Engel, K., & Moosbrugger, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, *8*(2), 23–74.

Semerci Şahin, R., Özbay, Ö., Çınar Özbay, S., & Durmuş Sarıkahya, S. (2025). Development of the generative artificial intelligence awareness scale for secondary school students in Türkiye. *European Journal of Pediatrics*, *184*(9), 585. https://doi.org/10.1007/s00431-025-06435-8

Şencan, H. (2005). *Sosyal ve Davranışsal Ölçümlerde Güvenilirlik ve Geçerlilik* (1. b. ed.). Seçkin Yayıncılık.

Trochim, W. M., & Donnelly, J. P. (2006). *The research methods knowledge base* (3rd ed.). Atomic Dog.

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. United Nations Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000381137

Wang, J., & Wang, X. (2012). *Structural equation modeling: Applications using Mplus (methods and applications)*. Wiley.

APPENDIX

Appendix 1. AI Ethical Awareness Scale Instruction

The AI Ethical Awareness Scale was developed to assess the ethical awareness levels of artificial intelligence (AI) users aged 18 and above. As a result of exploratory and confirmatory factor analyses, a three-factor and nine-item structure was obtained. The scale explains 74.8% of the variance, with Cronbach's $\alpha = .84$, Gutman = .88, and McDonald's Omega $\omega = .93$.

FACTOR 1 (TRANSPARENCY)

This factor refers to the evaluation of principles such as citation of AI-generated content, respect for copyright, and openness and honesty in knowledge production. It measures individuals' ethical sensitivity regarding disclosure of the origin of AI-assisted content, legitimate use of information, and respect for intellectual property rights. Transparency is grounded in the idea that individuals should act openly and honestly toward both themselves and their audience in the process of digital content production. In contexts where AI technologies are used as producers, explicitly stating the source of the content, respecting copyright, and ensuring traceability of information are considered fundamental indicators of ethical awareness. This factor reflects the extent to which individuals adhere to the principles of digital ethics, information ethics, and academic integrity. The variance explained by this factor was 45.8%, with Cronbach's $\alpha = .79$, Gutman = .73, and McDonald's Omega $\omega = .82$.

FACTOR 2 (INTEGRITY)

This factor measures the extent to which individuals adhere to ethical responsibility principles when using AI technologies, such as avoiding harm to others, refraining from creating deceptive content, and rejecting the generation of fake identities. It evaluates ethical decision-making and individuals' stance against the potential misuse of AI applications. Integrity represents sensitivity to the impact of digital actions on others and awareness of the violation of ethical boundaries. This factor is directly related to cyber ethics, digital citizenship, opposition to deception, and digital moral autonomy. The variance explained by this factor was 16.7%, with Cronbach's $\alpha = .85$, Gutman = .81, and McDonald's Omega $\omega = .85$.

FACTOR 3 (AWARENESS)

This factor encompasses ethical sensitivity toward the social, environmental, and relational impacts of AI technologies. The items reflect individuals' level of consciousness regarding the negative effects of AI on human relationships, social justice, and environmental sustainability. The Awareness dimension represents not only individuals' ethical practices but also their capacity to reflect on the systemic implications of AI. It is closely linked to critical technology literacy, the relationship between technology and social structures, ethical foresight, and social responsibility. The variance explained by this factor was 12.4%, with Cronbach's $\alpha = .82$, Gutman = .76, and McDonald's Omega $\omega = .82$.

SCALE INSTRUCTION

The AI Ethical Awareness Scale is a five-point Likert-type instrument (1 = Strongly Disagree, 5 = Strongly Agree). There are no reverse-scored items. The total scale score is obtained by summing all item scores, yielding a raw score between 9 and 45. In accordance with the instructions, the raw score should then be standardized to a 0–100 range. Higher scores indicate higher levels of AI ethical awareness.

STANDARDISATION

In order to facilitate consistent and comparable results across samples in studies using this scale and in future adaptation studies, standardization has been applied to the scoring system of the AI Ethical Awareness Scale.

To convert the raw scores into a 0–100 standardized score, the following formula should be used:

Standardized Score = ((Raw Score − 9)/36) × 100

For example, if an individual obtains a raw score of 27, the standardized score is calculated as follows:

Standardized Score = ((27 − 9)/36) × 100 = 50

Thus, a raw score of 27 corresponds to a standardized score of 50. In all studies, it is mandatory to use the standardized scores of the scale.

Artificial Intelligence Ethical Awareness Scale

Please indicate to what extent the following statements are appropriate for you.

Item	Strongly disagree (1)	Disagree (2)	Neutrol (3)	Agree (4)	Strongly Agree (5)
I respect copyright of the content I create with artificial intelligence.					
I cite sources in the content I create with artificial intelligence.					
I investigate the sources of AI-assisted content.					
I do not use artificial intelligence in ways that disturb others.					
I am against the use of fake identities/profiles created by artificial intelligence.					
I do not use AI-based chatbots to deceive people.					
I am aware of the negative effects of artificial intelligence on human relationships.					
I know that artificial intelligence may pose risks in terms of social justice.					
I take into account the negative impacts of artificial intelligence on the environment.					