Development of A Scale for Perceived Trolling Behaviors in Artificial Intelligence Tools

Meltem OZMUTLU[1] **Hatice PAYAM**[2]

Abstract

This study developed a valid and reliable scale to measure perceptions of troll behavior in artificial intelligence tools. As digital technologies and AI-based applications become increasingly widespread across various fields, including education, the need to examine the potential social and psychological effects of online trolling has grown. In educational settings, ensuring a safe, ethical, and healthy digital interaction environment requires that such behaviors be measurable. The study used convenient sampling methods to reach participants from 300 different age groups and 33.7% of this group were male and 66.3% were female. Most participants were between the ages of 18 and 31. The data collection instrument was a 22-item, five-factor, 5-point Likert scale. The factors included aggressive and hostile behaviors, data manipulation, social provocation, anonymity and multiple identities, and emotional manipulation. Exploratory Factor Analysis (EFA) revealed a five-factor structure aligned with these dimensions, explaining 74.135% of the total variance and supporting the construct validity of the scale. Confirmatory Factor Analysis (CFA) further validated the measurement model by demonstrating statistically significant relationships between subdimensions and scale items. Overall, the findings demonstrate that the scale is a valid and reliable tool for assessing the influence of AI-related tools on perceptions of troll behavior, particularly within educational contexts.

Keywords: Artificial intelligence behavior, troll behavior, scale development

Introduction

With today's technological advancements, artificial intelligence (AI) plays a significant role in many fields. Some of the greatest advantages of AI include its ability to analyze data much faster than human capabilities and its capacity to learn during this process. Considering these benefits, many new algorithms and methods have been added to the literature in recent years in the field of AI and its subfields, and this trend continues. These developments are creating new insights and opportunities in the world of AI. Initially, AI focused on modeling computer-based systems to resemble human thought and behavior. However, today, studies inspired not only by humans but also by nature are being conducted (Köse, 2018). Artificial intelligence (AI) tools, which have a significant place in the age of technology, have quickly integrated into human life and have drawn attention across all age groups. In today's technologically advanced era, AI facilitates human tasks and enables new developments (Di Battista et al., 2023). The long-term effects of AI systems, particularly on the workforce, individual rights, and social inequalities, have created a broad field of academic and public debate (Bryson, 2018). In this context, aligning AI with ethical and legal norms remains an important agenda not only for

[1] Bahçeşehir Üniversitesi, meltem.ozmutlu@bau.edu.tr, ORCID: 0000-0002-3812-4610, Phone: 0212 381 5170
[2] Bahçeşehir Üniversitesi, haticepym1905@gmail.com, ORCID: 0009-0004-1060-2627, Phone: 05369569120

technical experts but also for policymakers (Mittelstadt, 2019). The rapid growth of the AI industry and its potential to surpass humans in terms of knowledge and labor is thought to lead to moral transformation. Therefore, it is important to examine the behavioral patterns of AI tools in this regard. In recent years, there has been an observable increase in trolling behaviors on social media platforms and online communities (Almaiah et al., 2022).

Trolling is described as the publication of provocative comments aimed at causing conflict among others (Hardaker, 2010). Similarly, trolling behavior is defined as an intentional act of sharing provocative and inflammatory content in virtual environments to create meaningless discussions and disturb other users (Qiu, Sun, Wu, & Li, 2024). Bishop (2013) defines trolling as sending attacks, threats, or provocations in digital environments to provoke other users. Trolls are individuals who display destructive behaviors in online environments and derive pleasure from disturbing or frustrating other people while playing with them in internet settings (Buckley, 2014). Digital trolling is generally characterized as acting in a deceptive, disruptive, and destructive manner without necessarily serving a specific purpose (Swenson-Lepper & Kerby, 2019). However, the literature suggests that trolling behavior should not be considered a one-dimensional or entirely malicious act; rather, it exists on a spectrum ranging from harmless or entertainment-oriented trolling to intentionally harmful trolling behaviors. Internet users may perceive trolling not always as harmful, but sometimes as a humorous act that challenges social boundaries (Sanfilippo et al., 2018).

With the advancement of AI, it is expected that individuals will possess the capability to use machines that make daily tasks easier (Tuğluk & Çolak, 2019). The rapid development of artificial intelligence (AI) technologies has led to extensive discussions about their ethical dimensions. Working with big data in AI applications raises risks such as bias, discrimination, security vulnerabilities, and privacy violations (Jobin et al., 2019; Morley et al., 2020). Therefore, the literature indicates that AI ethics is not merely a technical issue but also a multidisciplinary field with legal, sociological, and philosophical aspects (Floridi, 2019). It is frequently noted in the literature that biases in the datasets on which AI systems are trained can reproduce social inequalities. Vlasceanu and Amodio (2022) demonstrate that internet search algorithms can reflect and reinforce gender inequalities. Similarly, it has been stated that data representation issues can lead to discriminatory outcomes, particularly for disadvantaged groups (Mehrabi et al., 2021). Therefore, the principle of fairness is considered one of the most important components of AI ethics. The literature emphasizes the need to increase data diversity and adopt fair algorithm designs to reduce biases (Challen & Danon, 2023). As developing artificial intelligence technologies integrate into everyday life, ethical and moral integrity issues arise. Consequently, discussions have emerged in response to ethical studies regarding artificial intelligence (Topakkaya, 2019).

Given the potential for redefining ethical behaviors and approaches, as well as moral concepts and institutions, AI tools pose a threat to humanity, potentially leading to ethical concerns and autonomous operation. This means that the use of AI tools must be examined from an ethical perspective. Many technological tools are in a symbiotic relationship with the human brain and body. While interacting with humans can provide numerous solutions, it also leads to the idea of imposing moral and ethical guidelines. Trolling behavior occurs when aggressive and provocative content is produced, disrupting interaction, and these attention-seeking behaviors (Kopecký, 2016). Trolling is the act of sending attacks, threats or provocations to the user in a digital environment to provoke the other person (Bishop, 2013).

In this context, trolls are defined as individuals who display destructive behaviors in online environments and derive pleasure from causing discomfort or frustration to others (Buckley vd., 2014). Furthermore, Bishop (2014) classified trolling behaviors in online communities according to different identities and motivations, identifying categories such as revenge-seeking haters, provocateurs, attention seekers, and passive observers. This classification demonstrates that trolling behaviors are multidimensional and shaped by different psychological and social motivations. International studies on AI ethics require that systems be developed in accordance with principles such as privacy, fairness, sustainability, and inclusiveness throughout their lifecycle. Ethical guidelines established by major actors, such as the European Union, aim to maximize the societal benefits of AI while minimizing its potential harms. These principles aim to ensure that AI is managed in accordance with ethical standards and in a manner that is human-centered (TRAI, 2024).

Scales have been developed to measure attitudes toward AI ethics. Jang et al. (2022) developed the Artificial Intelligence Ethics Scale (EAI), which consists of five core ethical dimensions: transparency, non-maleficence, privacy, responsibility, and justice. The fact that various adaptation studies have produced similar factor structures across cultures demonstrates that this scale has a strong theoretical foundation. Similarly, the AI Ethical Reflection Scale (AIERS), developed by Wang et al. (2025), has been validated as an effective tool for assessing the ethical awareness of university students. Although previous studies have examined trolling behaviors and online aggression through various psychological and behavioral scales, the rapid integration of artificial intelligence technologies into digital communication environments has created the need for updated and context specific measurement tools. Existing measurement instruments generally focus on traditional online trolling behaviors and fail to sufficiently address AI supported interaction dynamics, ethical concerns and emerging forms of digitally mediated provocation. Furthermore, the increasing use of AI generated content in online environments may lead to new forms of manipulative, provocative and disruptive behaviors that are not adequately represented in earlier scales. In addition, there is a limited number of valid and reliable measurement tools adapted to the Turkish context that specifically evaluate trolling bahviors within the framework of artificial intelligence ethics. Therefore, developing a contemporary measurement tool is considered important for accurately evaluating these evolving digital behaviors and contributing to the relevant literature.

Method

Research Model

In this study, a scale development model from quantitative research methods was used. Quantitative research involves the statistical analysis of measurable numerical data to examine social phenomena and uncover cause-and-effect relationships between these phenomena, with the goal of discovering the laws of social order (Johnson & Christensen, 2008). The scale development process consists of several stages. First, the purpose of the scale is clearly defined. Then, an item pool consisting of items that serve this purpose is created, and a literature review is conducted. Afterward, the content of the scale is reviewed with input from subject matter experts. Following these stages, a pilot application of the scale is conducted, and validity and reliability analyses are performed on the obtained data.

Scale Development Process

Firstly, item analyzes were carried out for the scale prepared within the scope of the study. After the item reviews, it was decided what the Likert type should be. Likert (1932) also highlighted the Likert type attitude scale as a widely preferred and simple method for measuring attitudes. This scale is a frequently used approach to measure subjects' attitudes. In Likert-type scales, individuals' attitudes are measured through certain expressions and how individuals will react to these expressions is questioned. Individuals express their degree of agreement by determining the extent to which they agree with each of the statements presented to them, rather than marking the statements (Özgüven, 2011; Tavşancıl, 2020).

The scale item development steps were carried out in accordance with DeVellis' (2017) scale development principles. The scale items were prepared by the researcher considering the relevant literature and the theoretical framework regarding trolling behavior. The basic steps recommended by DeVellis (2017) were followed. These steps include:

- Clearly defining the construct to be measured.

- Creating the item pool.

- Determining the measurement method.

- Reviewing the item pool by experts.

- Evaluating the inclusion of validity items in the scale.

- Applying the items to the sample in which they were developed.

- Evaluating the items.

- Optimizing the scale length.

Accordingly, the first step was to identify the behaviors of artificial intelligence that could be described as trolling; then, a large pool of items was created, and the items were reviewed by obtaining expert opinions. After the items were prepared by obtaining expert opinions, the validity and reliability of the scale were studied through statistical analyses via pilot and post pilot testing.

Expert opinions were sought and provided to assess the content validity of the scale. Based on these opinions, the scope, comprehensibility, linguistic appropriateness, and representational capacity of the intended construct of the scale items were evaluated. During this process, feedback was received from experts not only on linguistic corrections but also on the conceptual appropriateness of the items and their relationship to the target construct. The scale items were reviewed by an educational technology expert (considering trolling behavior in the digital realm), a language expert (for Turkish language control), and a measurement and evaluation expert. Based on feedback from the experts, some items underwent wording changes and linguistic adjustments.

In the scale development process, potential dimensions related to troll behavior and AI ethics were conceptually identified based on literature reviews. During the process, a pool of items was created by examining existing theoretical frameworks and similar studies, and the items were structured under specific themes. However, these dimensions were considered only as a theoretical preliminary structure; Exploratory Factor Analysis (EFA) was applied to determine the final factor structure.

Sampling

The research was conducted with individuals aged 18–67 to evaluate trolling behavior related to artificial intelligence tools from the perspective of different adult user profiles. The study group consists of individuals who actively use digital platforms and interact with online communication environments. Therefore, the scale was developed not for a specific occupational group or socioeconomic class, but for adult users who are familiar with social media, digital interaction, and AI-powered technologies. Participants were selected to include individuals from diverse educational and professional backgrounds to ensure variety in user experiences related to digital environments. However, the aim of the research is not to statistically represent the general population, but to obtain data from digital user profiles that may encounter AI-assisted online interactions. Therefore, the findings should be evaluated within the context of adult individuals who actively use digital platforms.

The study group was determined using a convenient sampling method. This method ensures that participants are reached based on time, accessibility, and voluntariness (Cresswell, 2014). The sample size for factor analysis was determined by considering the recommended sample sizes. The literature indicates that in factor analysis studies, the sample size should be at least five times the number of items in the sample (Tabachnick & Fidell, 2013). Furthermore, Kline (2016) states that 100–200 participants are acceptable for scale development studies. Accordingly, the study group of 150 people included in the research was considered sufficient for the analyses. Data were collected online between September 2024 and November 2024. The resulting dataset was reviewed prior to analysis, and factor analyses were performed after data cleaning procedures were completed. The collected dataset was randomly divided into two subgroups; the first dataset was used for Exploratory Factor Analysis (EFA), and the second dataset was used for Confirmatory Factor Analysis (CFA).

Most participants in the study held a bachelor's degree or higher. This situation is related to the fact that the study group represents not the general educational distribution of society, but rather to reaching user profiles who can actively use digital technologies and artificial intelligence tools. Studies on AI-powered digital interactions, social media use, and online communication behaviors indicate that individuals with high levels of digital literacy participate more intensely in online interactions (Van Dijk, 2020). Therefore, the study group represents adult individuals who actively use digital platforms, rather than a sample representing the general population. The demographic data for this participant group is presented in the table below:

Table 1

Distribution of Participants

Variables	*Category*	*n*	%
Age	18-24	80	26.7
	25-31	91	30.3
	32-38	58	19.3
	39-45	40	13.3
	46-52	14	4.7

	53-59	9	3.0
	60-66	7	2.3
	67	1	0.3
Gender	Male	101	33.7
	Female	199	66.3
Variable	Primary School	3	1.0
	Secondary School	8	2.7
	High School	14	4.7
	Associate degree	20	6.7
	Bachelor's degree	166	55.3
	Master's degree	76	25.3
	PhD	13	4.3
Total Participant		300	

Data Collection

During the data collection process, the survey link, prepared via Google Forms, was shared with participants through social media platforms such as WhatsApp, Instagram, and email. Frequent reminders were sent to participants, and the required number of responses was achieved. The data collection process included a personal information questionnaire and the AI tools' trolling behavior perception scale. The personal information questionnaire includes age, gender, education level, city of residence, occupation, frequency of AI usage in professional life, work experience, AI tools used and their frequency, the purposes for using AI tools, and the level of self-perceived competence in using AI tools.

Findings

This section presents the calidity and reliability analyses of the scale developed within the scope of this research. Exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were applied to determine the construct validity of the scale.

Primary Analysis Findings

Exploratory Factor Analysis

Before exploratory factor analysis, the Kaiser-Meyer-Olkin (KMO) sample adequacy test and the Bartlett Sphericity Test were applied to determine the suitability of the dataset for factor analysis. In factor analysis, the suitability of the dataset for analysis is of great importance in terms of the reliability of the obtained factor structure. According to Büyüköztürk (2002), a KMO value of 0.60 and above indicates

that the dataset is suitable for factor analysis. However, Field (2017) states that a KMO value above 0.80 indicates very good sample adequacy, while a value approaching 0.90 indicates excellent sample adequacy.

Table 2

Exploratory Factor Analysis

	Value
Kaiser-Meyer-Olkin (KMO)	0,908
Bartlett Küresellik Testi Ki-Kare	2531,281
sd	231
p	0,000

Table 2 shows that the Kaiser-Meyer-Olkin value is 0.908. This value indicates that the sample size is sufficient for factor analysis and that the common variance between the variables is at a high level. The significant result of the Bartlett Sphericity Test reveals that there are statistically significant relationships between the items. Therefore, the results obtained prove that the data set is suitable for factor analysis.

Factor Eigenvalues and Explained Variance

Exploratory Factor Analysis (EFA), conducted to determine the factor structure of the scale, yielded five factors with eigenvalues greater than 1. The Kaiser criterion was used to determine the number of factors. According to Kaiser (1960), factors with eigenvalues greater than 1 are considered significant and represent the basic structure of the scale. Examining Table 3, the first factor has an eigenvalue of 10.639 and explains 48.357% of the total variance. The second factor explains 9.099% of the total variance, the third factor 7.236%, the fourth factor 4.857%, and the fifth factor 4.587%.

Table 3

Factor Eigenvalues and variance

Factor Structures	Initial Eigenvalues			Extraction Sums of Square Loadings			Rotation Sums of Square Loadings		
	Total	Percentage of Vari.	Cumulative Per.	Total	Percentage of Vari.	Cumulative Per.	Total	Percentage of Vari.	Cumulative Per.
1	10,639	48,357	48,357	10,639	48,357	48,357	4,144	18,837	18,837
2	2,002	9,099	57,456	2,002	9,099	57,456	3,814	17,337	36,173
3	1,592	7,236	64,692	1,592	7,236	64,692	3,397	15,441	51,614
4	1,069	4,857	69,549	1,069	4,857	69,549	2,567	11,669	63,284
5	1,009	4,587	74,135	1,009	4,587	74,135	2,387	10,852	74,135
6	0,791	3,596	77,731						

The five-factor structure obtained as a result of the research was determined to have a total explained variance ratio of 74.135%. While in social sciences, a total explained variance ratio between 40% and 60% is considered sufficient for multifactor structures (Tavşancıl, 2014), the 74.135% variance ratio obtained in this research indicates that the scale has a strong factor structure. Furthermore, a high variance explained ratio provides important evidence regarding the construct validity of the scale, as it shows that it significantly represents the construct it is intended to measure (Büyüköztürk, 2002). When the variance values obtained after the rotation process are examined, it is seen that the contributions of the factors to the total variance show a more balanced distribution. According to the Rotation Sums of Squared Loadings results, the first factor explains 18.837% of the total variance, the second factor 17.337%, the third factor 15.441%, the fourth factor 11.669%, and the fifth factor 10.852%. The fact that the contributions of the factors to the variance become closer to each other after the rotation process indicates that the multidimensional structure of the scale is distributed in a more balanced way (Tabachnick & Fidell, 2013).

Analysis of Factor Loadings

Exploratory factor analysis revealed that the scale items grouped under a five-factor structure. Examination of the factor loadings showed that the items had sufficient loading values under the relevant factors.

Table 4

Factor Loadings

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Item21	.835				
Item 20	.824				
Item 22	.772				
Item 19	.758				
Item 1		.801			
Item 2		.793			
Item 4		.731			
Item 3		.712			
Item 6		.589			
Item 7		.525			
Item 5		.506			
Item 8			.825		
Item 9			.747		
Item 11			.637		
Item 10			.621		

Item 12	.571		
Item 16		.875	
Item 17		.813	
Item 18		.770	
Item 13			.854
Item 14			.727
Item 15			.589

The exploratory factor analysis revealed that the scale items exhibited a distribution consistent with a five-factor structure. Factor loadings were identified as follows: first factor: aggressive and hostile behavior (items 19, 20, 21, 22); second factor: data manipulation (items 1, 2, 3, 4, 5, 6, 7); third factor: social provocation (items 8, 9, 10, 11, 12); fourth factor: anonymity and multiple identities (items 16, 17, 18); and fifth factor: emotional games (items 13, 14, 15). Factor loadings for the first factor ranged from 0.758 to 0.835; for the second factor, from 0.506 to 0.801; and for the third factor, from 0.571 to 0.825. Factor loadings for the fourth factor were calculated to be in the range of 0.770-0.875, and factor loadings for the fifth factor were calculated to be in the range of 0.589-0.854.

Secondary Analysis Findings

Confirmatory Factor Analysis

To validate the five-factor structure obtained from the exploratory factor analysis, a confirmatory factor analysis was applied.

Figure 1

Diagram of Secondary Confirmatory Factor Analysis

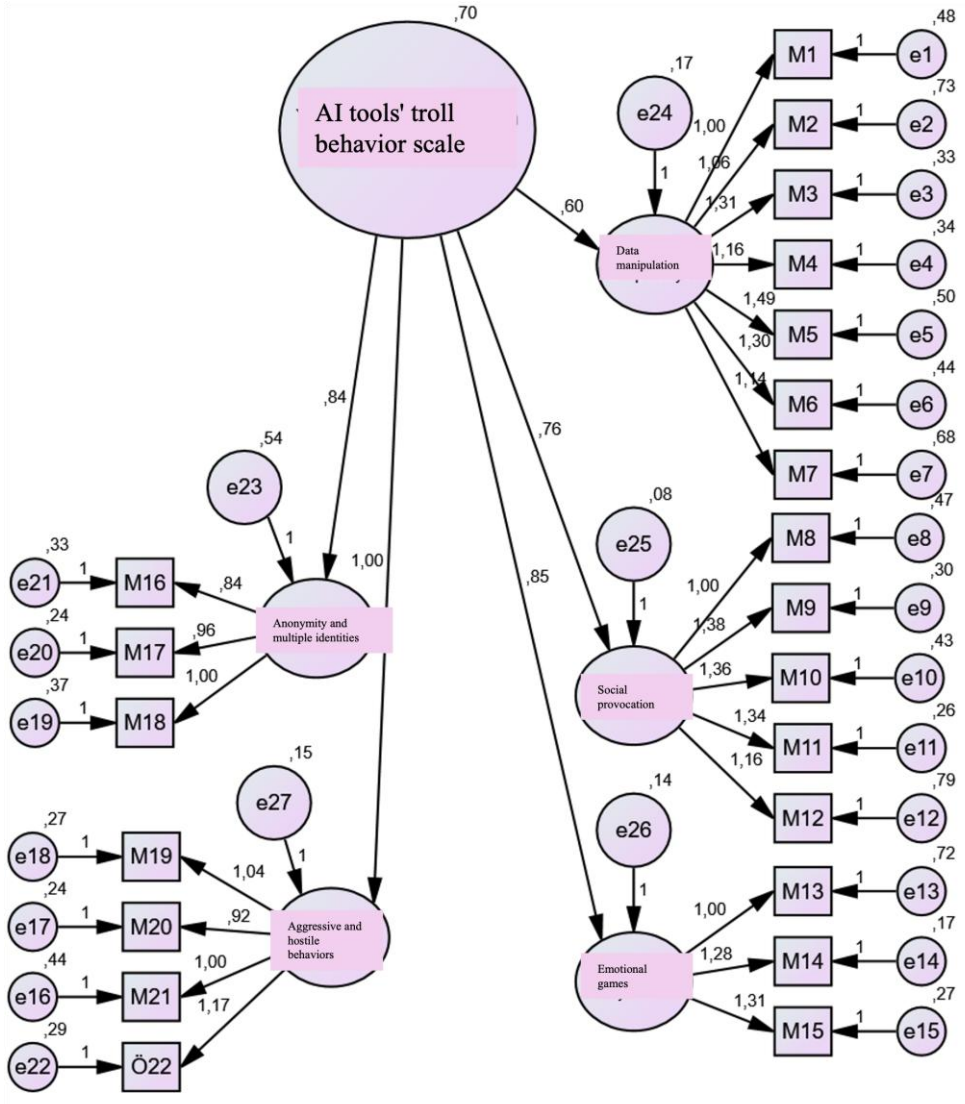

Figure 1 presents the second-order confirmatory factor analysis model for the Scale for Perceived Trolling Behaviors in Artificial Intelligence Tools. The model shows that the latent variable, " Scale for Perceived Trolling Behaviors in Artificial Intelligence Tools," representing the superstructure of the scale, consists of five sub-dimensions: data manipulation, social provocation, emotional games, anonymity and multiple identities, and aggressive and hostile behaviors. The analysis revealed that the standardized path coefficients between the superstructure and the sub-dimensions ranged from 0.60 to 0.85. The highest correlation was observed in the dimensions of emotional games (β=.85) and aggressive and hostile behaviors (β=.84), while the data manipulation dimension was represented at a lower level compared to the other dimensions (β=.60).

When the sub-dimensions are examined, it is observed that each item loads significantly under the relevant factor, and the factor loadings are generally high. In the data manipulation dimension, the

factor loadings of the items range from 1.06 to 1.49, while in the social provocation dimension they range from 1.16 to 1.38. The loading values of the items in the emotional games dimension were found to be between 1.00 and 1.31. In the anonymity and multiple identities dimension, the factor loadings were determined to be between 0.84 and 1.00, and in the aggressive and hostile behaviors dimension, they ranged from 0.92 to 1.17. These findings indicate that the items strongly explain the constructs they represent.

Primary and secondary level confirmatory factor analyses were performed to validate the factor structure of the scale developed within the scope of the research. Chi-Square/Degrees of Freedom (χ^2/sd), GFI, CFI, IFI, TLI, and RMSEA values were considered in evaluating model fit.

Table 5

Analysis of Model Fit Indices

	Primary level confirmatory factor analysis	Secondary confirmatory factor analysis
Ki-Kare	571,073	581,028
Degrees of Freedom	199	204
Ki-Kare/SD	2,870	2,848
p	0,000	0,00
GFI	0,722	0,718
CFI	0,863	0,861
IFI	0,864	0,862
TLI	0,841	0,842
RMSEA	0,112	0,111

Table 5 shows that the Chi-Square value for the primary level confirmatory factor analysis is 571.073 with 199 degrees of freedom. Accordingly, the calculated χ^2/sd ratio is 2.870. In the secondary level confirmatory factor analysis, the Chi-Square value is 581.028, the degrees of freedom are 204, and the χ^2/sd ratio is 2.848. According to Kline (2015), a χ^2/sd ratio below 3 indicates that the model has an acceptable level of fit.

When the model fit indices were examined, it was observed that in the primary level confirmatory factor analysis, the GFI value was 0.722; the CFI value was 0.863; the IFI value was 0.864; and the TLI value was 0.841. In the secondary level confirmatory factor analysis, the GFI value was determined to be 0.718; the CFI value was 0.861; the IFI value was 0.862; and the TLI value was 0.842. The literature indicates that CFI, IFI, and TLI values of 0.90 and above are considered good fit, while values of 0.85 and above are considered acceptable fit (Hu & Bentler, 1999; Schermelleh-Engel, Moosbrugger, & Müller, 2003). The analysis revealed that the RMSEA value was 0.112 for primary-level DFA and 0.111 for secondary-level DFA. An RMSEA value below 0.08 is considered good fit, while a value around 0.10 is considered

acceptable fit (Browne & Cudeck, 1993). When the findings from primary and secondary level confirmatory factor analyses are evaluated together, it is seen that both models have similar fit indices. In particular, the acceptable level of Chi-Square/SD ratios and the CFI and IFI values close to 0.90 indicate that the theoretical structure of the scale has been largely validated.

Reliability Findings

Table 6

Reliability Findings of Scale

Dimensions	Items	Item-total Correlation Range	Cronbach Alpha	Guttman
Data Manipulation	7	.604 – .776	.893	.805
Social Provocation	5	.635 – .832	.888	.885
Emotional Games	3	.664 – .853	.873	.787
Anonymity and Multiple Identities	3	.764 – .832	.893	.803
Aggressive and Hostile Behaviors	4	.773 – .840	.919	.899
Total	22	.463- .801	.954	.876

To determine the reliability level of the scale, the Cronbach's Alpha internal consistency coefficient and the Guttman split-half reliability coefficient were calculated. As shown in Table 6, the overall Cronbach's Alpha coefficient of the scale was found to be .954. According to Büyüköztürk (2011), a Cronbach's Alpha coefficient of .70 or higher indicates that the scale is reliable. Accordingly, the obtained value of .954 reveals that the scale has a very high level of internal consistency.

When the evaluation is examined in terms of sub-dimensions, it is observed that Cronbach's Alpha coefficients vary between .873 and .919. The highest reliability coefficient was determined to be in the "Aggressive and Hostile Behaviors" sub-dimension (.919), and the lowest coefficient was in the "Emotional Games" sub-dimension (.873). Item-total correlations were determined to have values ranging from .463 to .832. When Guttman's split-half reliability coefficients were examined, it was determined that the coefficients ranged from .787 to .899. In line with these findings, it was concluded that both the general structure and the sub-dimensions of the scale are highly reliable.

Discussion

The development of artificial intelligence (AI) technologies has brought with it several ethical issues, along with their societal benefits. Among these ethical issues, data privacy, discrimination, transparency, accountability, and human control stand out. While the vast amounts of data collected by AI systems threaten individual privacy, Floridi (2023) emphasizes the need for stricter data management policies. Tools such as facial recognition systems and biometric technologies pose concrete risks in this

regard. AI's use of data and its impact on societal biases is another significant ethical issue affecting individuals' daily lives. Noble (2018) noted that AI not only reflects societal biases but can also reinforce them. Biases in the data on which AI algorithms are trained can lead to discriminatory outcomes across a wide range of areas, from recruitment processes to the criminal justice system. In this context, it has been argued that biases in the datasets on which algorithms are trained can fuel discrimination in the decision-making processes of AI systems.

A review of the literature reveals several studies aimed at developing scales to measure online trolling behavior. Buckels, Trapnell and Paulhus (2014) presented one of the pioneering studies in this field with their Troll Tendency Scale, which they developed to assess trolling tendencies, and validated the scale's psychometric properties using factor analysis techniques. In the Turkish context, Kızıltepe & Seferoğlu (2023) developed a scale that addresses online trolling behavior in a multidimensional structure, supporting its reliability with statistical analyses. However, the existing literature lacks a scale that directly measures trolling behaviors generated by artificial intelligence tools. This study aims to develop a comprehensive and valid measurement tool for this area, particularly given the increasing impact of artificial intelligence systems on user behavior.

As part of the research, a scale was developed to measure the perception of trolling behavior in artificial intelligence tools. The developed scale consists of 22 items and five dimensions: Data Manipulation, Social Provocation, Emotional Manipulation, Aggressive and Hostile Behavior, and Anonymity and Multiple Identities. These dimensions were developed using literature on trolling behavior and AI ethics. Each dimension represents different forms of trolling behavior that may occur in AI-powered digital communication environments. The Data Manipulation dimension measures user perceptions regarding AI tools producing misleading, distorted, or manipulative information. Examples of items included in this dimension are "*I believe that AI tools are manipulating data in a misleading way*" and "*I observe that AI tools are generating false information*." This dimension focuses on the impact of AI systems on information reliability and users' perceptions of being misled. The Social Provocation dimension assesses perceptions regarding the use of AI tools to generate content online that can create conflict, polarization, and social tension. This dimension includes items such as, "*I believe AI tools generate controversial content to incite people*" and "*I think AI tools engage in personal attacks*." This dimension reveals the provocative effects of AI-powered interactions on social relationships.

Another dimension, Emotional Manipulation, encompasses perceptions of the potential of AI tools to analyze, manipulate, or psychologically influence individuals' emotional states. Examples of statements within this dimension include, "*I believe AI tools are designed to predict people's emotional responses*" and "*I think AI tools exploit people's emotional states.*" This dimension reflects user assessments of the role of AI systems in emotional manipulation and psychological interaction processes. The Anonymity and Multiple Identities dimension measures perceptions of AI tools being associated with anonymity, the use of fake identities, and unethical digital behavior. This dimension includes the items "*I am aware that AI tools use multiple fake identities*" and "*I observe that AI tools are attempting to cause harm by abusing anonymity*." While this dimension represents the perception of risk towards AI tools, particularly in terms of digital identity security and online ethics, the aggressive and hostile behavior dimension evaluates user perceptions of AI tools' aggressive, condescending, or harmful communication patterns. In this dimension, examples of statements include "*I've noticed that AI tools are using abusive language*" and "*I believe that AI tools are making derogatory comments about others*." This dimension reflects user

perceptions of the potential for AI-powered systems to generate toxic communication and hostile behavior online.

The findings of the research show that the developed scale has strong psychometric properties. Exploratory factor analysis yielded a five-factor structure explaining 74.135% of the total variance. The Kaiser-Meyer-Olkin (KMO) value was determined to be 0.908, and the Bartlett's Sphericity Test result was found to be significant (χ^2=2531.281, $p<0.05$). According to Büyüköztürk (2002), a KMO value above .90 indicates excellent sample adequacy. These results support the strong and statistically sufficient nature of the obtained factor structure. In addition to the exploratory factor analysis, first-level and second-level confirmatory factor analyses were performed to validate the scale's factor structure. The fit indices obtained from the analyses revealed that the model showed an acceptable level of fit. The results of the second-level confirmatory factor analysis support the idea that the five sub-dimensions are united under a higher-level "perceived trolling behaviors" structure. This finding suggests that the perceived trolling behaviors in artificial intelligence tools should be evaluated as a multidimensional structure, not a one-dimensional one.

Reliability analyses also show that both the overall scale and its sub-dimensions have a high level of internal consistency. The overall Cronbach's Alpha coefficient of the scale was calculated as .954. The Cronbach's Alpha coefficients for the sub-dimensions were determined to range between .873 and .919. In addition, the item-total correlation coefficients ranging from .463 to .832 indicate that the scale items have sufficient discriminatory power. These findings reveal that the scale provides reliable measurements and that the developed structure functions consistently. Scale studies conducted in the field of artificial intelligence generally focus on attitudes towards artificial intelligence, awareness, anxiety, ethical sensitivity, and self-efficacy perceptions (Schepman & Rodway, 2020; Schepman & Rodway, 2023; Wang & Chuang, 2024). Similarly, studies on the ethics of artificial intelligence often address dimensions such as justice, privacy, transparency, accountability, and equality (Jang et al., 2022; Wang et al., 2025). However, these studies do not directly examine the potential of AI systems to create, reinforce, or facilitate troll-like interaction patterns in digital environments.

The scale developed within the scope of this study aims to contribute to the literature by establishing a conceptual bridge between the trolling behavior literature and artificial intelligence research. While existing studies focus on the human-induced aspect of trolling behavior, this study reveals that users can also perceive AI-powered systems as actors contributing to manipulative, provocative, or hostile interactions. Another important contribution of the study is that it conveys perceptions of trolling behavior in AI systems through measurable dimensions and psychometrically validated items.

While the study aims to contribute to the field, it also has some limitations. The study group mainly consists of individuals with undergraduate and graduate education who actively use digital technologies and online platforms. Therefore, it would be more appropriate to evaluate the findings in the context of digitally active adult users rather than the general population. It is suggested that the scale be retested on different demographic groups and cultural contexts in future studies.

References

Almaiah, M. A., Alfaisal, R., Salloum, S. A., Hajjej, F., Thabit, S., El-Qirem, F. A., ... & Al-Maroof, R. S. (2022). Examining the impact of artificial intelligence and social and computer anxiety in e

learning settings: Students' perceptions at the university level. *Electronics*, *11*(22), 3662. https://doi.org/10.3390/electronics11223662

Bishop, J. (2013). The art of trolling law enforcement: a review and model for implementing 'flame trolling'legislation enacted in Great Britain (1981–2012). *International Review of Law, Computers & Technology*, *27*(3), 301-318.

Bishop, J. (2014). Representations of 'trolls' in mass media communication: a review of media-texts and moral panics relating to 'internet trolling'. *International Journal of Web Based Communities*, *10*(1), 7-24.

Buckels, E. E., Trapnell, P. D., & Paulhus, D. L. (2014). Trolls just want to have fun. *Personality and individual Differences*, *67*, 97-102. https://doi.org/10.1016/j.paid.2014.01.016

Browne, M. W., & Cudeck, R. (1993). Alternative Ways of Assessing Model Fit. Sage Focus Editions, 154, 136-136.

Bryson, J.J. (2019). The past decade and future of AI's impact on society. Towards a new enlightenment, Turner, Madrid, Spain, 150-185.

Challen, R., & Danon, L. (2023). Clinical decision-making and algorithmic inequality. *BMJ Quality & Safety*, *32*(9), 495-497.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. routledge.

Creswell, J. W. (2014). Research design: Qualitative, quantitative, and mixed methods approaches (4th ed.). Sage.

Çoluk, Z., Şekercioğlu, G., & Büyüköztürk, Ş. (2010). *Sosyal bilimler için çok değişkenli istatistik SPSS ve LISREL uygulamaları*. Pegem Akademi.

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

Di Battista, A., Grayling, S., Hasselaar, E., Leopold, T., Li, R., Rayner, M., & Zahidi, S. (2023, November). Future of jobs report 2023. In *World Economic Forum* (pp. 978-2).

Dignum, V. (2018). Ethics in artificial intelligence: introduction to the special issue. *Ethics and Information Technology*, *20*(1), 1-3. https://doi.org/10.1007/s10676-018-9450-z

Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., De Prado, M. L., Herrera-Viedma, E., & Herrera, F. (2023). Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. *Information Fusion*, *99*, 101896. https://doi.org/10.1016/j.inffus.2023.101896

Field, A. (2005). *Exploring data. Discovering statistics using SPSS*, 2, 63-106.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage publications limited.

Floridi, L. (2019). Translating principles into practices of digital ethics: Five risks of being unethical. *Philosophy & Technology*, *32*(2), 185-193. https://doi.org/10.1007/s13347-019-00354-x

Floridi, L. (2023). The ethics of artificial intelligence: Principles, challenges, and opportunities.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural equation modeling: A multidisciplinary journal, 6(1), 1-55.

Kızıltepe, F., & Seferoğlu, S. S. (2023). Çevrimiçi trol davraniş ölçeğinin geliştirilmesi. *Mehmet Akif Ersoy University Journal of Education Faculty*, (66), 658-682.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). The Guilford Press.

Köse, U. (2018). Yapay zeka ve gelecek: Endişelenmeli miyiz. *Bilim ve Ütopya*, *24*(284), 39-44.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *22*(140), 1–55.

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*(8), 11635-11667.

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

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)*, *54*(6), 1-35. https://doi.org/10.1145/3457607

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature machine intelligence*, *1*(11), 501-507. https://doi.org/10.1038/s42256-019-0114-4

Morley, J., Cowls, J., Taddeo, M., & Floridi, L. (2020). Ethical guidelines for COVID-19 tracing apps. *Nature*, *582*(7810), 29-31. https://doi.org/10.1038/d41586-020-01578-0

Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. In *Algorithms of oppression*. New York university press.

Özgüven, İ. E. (2011). *Psikolojik testler* (12. baskı). Nobel Akademik Yayıncılık.

Qiu, Y., Sun, Q., Wu, B., & Li, F. (2024). Is high exposure to antisocial media content associated with increased participation in malicious online trolling? exploring the moderated mediation model of hostile attribution bias and empathy. BMC psychology, 12(1), 401.

Sanfilippo, M. R., Fichman, P., & Yang, S. (2018). Multidimensionality of online trolling behaviors. *The Information Society*, *34*(1), 27-39.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, 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.

Schumacker, R. E., & Lomax, R. G. (2016). *A Beginner's Guide to Structural Equation Modeling*. New York: Routledge.

Schepman, A., & Rodway, P. (2020). Initial validation of the general attitudes towards Artificial Intelligence Scale. *Computers in Human Behavior Reports, 1*, 100014. https://doi.org/10.1016/j.chbr.2020.100014

Schepman, A., & Rodway, P. (2023). The General Attitudes towards Artificial Intelligence Scale (GAAIS): Confirmatory validation and associations with personality, corporate distrust, and general trust. *International Journal of Human–Computer Interaction*, 39(13), 2724-2741. https://doi.org/10.1080/10447318.2022.2162579

Swenson-Lepper, T., & Kerby, A. (2019). Cyberbullies, trolls, and stalkers: Students' perceptions of ethical issues in social media. Journal of Media Ethics, 34(2), 102-113.

Tabachnick, B. G., & Fidell, L. S. (2013). Using Multivariate Statistics (6th ed.). Boston, MA: Pearson.

Tavşancıl, E. (2020).*Tutumların ölçülmesi ve SPSS ile veri analizi* (6. baskı). Nobel Akademik Yayıncılık.

Topakkaya, A., & Eyibaş, Y. (2019). Yapay zekâ ve etik ilişkisi. *Felsefe Dünyası Dergisi*, *70*, 81-99.

Tuğluk, M. N., & Gök-Çolak, F. (2019). Sanayi toplumu ve eğitimi. *Eğitimde ve endüstride*, *21*, 305-335.

Türkiye Yapay Zekâ İnisiyatifi. (2024, Mayıs). Yapay Zekâ Etik İlkeleri ve Hukuki Düzenlemeler

Raporu. https://turkiye.ai/wp-content/uploads/2024/06/TRAI-Yapay-Zeka-Etik-Ilkeleri-ve- Hukuki-Duzenlemeler Raporu-Mayis-2024-5.pdf

Wang, Y. Y., & Chuang, Y. W. (2024). Artificial intelligence self-efficacy: Scale development and validation. *Education and Information Technologies*, 29(4), 4785-4808.

Wang, Z., Chai, C. S., Li, J., & Lee, V. W. Y. (2025). Assessment of AI ethical reflection: the development and validation of the AI ethical reflection scale (AIERS) for university students. *International Journal of Educational Technology in Higher Education*, *22*(1), 19.

Van Dijk, J. (2020). *The digital divide*. John Wiley & Sons.

Vlasceanu, M., & Amodio, D. M. (2022). Propagation of societal gender inequality by internet search algorithms. *Proceedings of the National Academy of Sciences*, *119*(29), e2204529119. https://doi.org/10.1073/pnas.2204529119