European Journal of Engineering Education

ISSN: 0304-3797 (Print) 1469-5898 (Online) Journal homepage: www.tandfonline.com/journals/ceee20

Measuring attitudes toward responsible AI in engineering – development and validation of the RAISE scale

Marie Mirsch, Nils Knoth, Ben Schultz, Katharina Bata, Johannes Schleiss & Leicht-Scholten Carmen

To cite this article: Marie Mirsch, Nils Knoth, Ben Schultz, Katharina Bata, Johannes Schleiss & Leicht-Scholten Carmen (27 Jan 2026): Measuring attitudes toward responsible AI in engineering – development and validation of the RAISE scale, European Journal of Engineering Education, DOI: 10.1080/03043797.2026.2620482

To link to this article: https://doi.org/10.1080/03043797.2026.2620482

Published online: 27 Jan 2026.

EUROPEAN JOURNAL OF ENGINEERING EDUCATION
https://doi.org/10.1080/03043797.2026.2620482

Measuring attitudes toward responsible AI in engineering – development and validation of the RAISE scale

Marie Mirsch [a], Nils Knoth [b], Ben Schultz [a], Katharina Bata [c], Johannes Schleiss [d] and Leicht-Scholten Carmen [a]

[a]Gender and Diversity in Engineering, RWTH Aachen University, Aachen, Germany; [b]Department of Psychology, Universität Kassel, Kassel, Germany; [c]Scientific Computing & Mathematics, Karlsruhe Institute of Technology, Karlsruhe, Germany; [d]Artificial Intelligence Lab, Otto-von-Guericke-University Magdeburg, Magdeburg, Germany

ABSTRACT

As Artificial Intelligence (AI) technologies increasingly shape the field of engineering, ethical considerations are becoming essential for fostering responsible innovation. However, a validated instrument to assess attitudes toward Responsible AI, specifically in the engineering domain, is still missing to date. This study presents the development and validation of the Responsible AI Attitudes Specific to Engineering (RAISE) scale. After refining the item pool of a validated test instrument with expert input from the engineering domain, we conducted confirmatory factor analysis on data from 235 engineering students and professionals in Germany. The resulting 15-item scale measures engineers' self-reported attitudes along three core Responsible AI dimensions: do-no-harm, transparency, and privacy. It demonstrates acceptable model fit, internal consistency, and measurement invariance across demographic groups. The RAISE scale can serve as a diagnostic and evaluative tool in engineering education and training programmes, helping to inform and assess efforts to foster Responsible AI engagement.

ARTICLE HISTORY
Received 28 August 2025
Accepted 17 January 2026

KEYWORDS
Responsible AI; AI ethics; engineering ethics; attitudes; scale development

Nomenclature

Abbreviation	Full Term
ABET	Accreditation Board for Engineering and Technology
AI	Artificial Intelligence
ANOVA	Analysis of Variance
AT-EAI	Attitudes Toward the Ethics of Artificial Intelligence
CDIO	Conceive–Design–Implement–Operate
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
EU	European Union
EVR	Expert Validation Round
F	F-value (Levene's Test)
FVI	Face Validity Index
I-CVI	Item-Level Content Validity Index
M	Mean
MLR	Maximum Likelihood Estimator with Robust Standard Errors
n/N	Sample Size (N = Full Sample, n = Sub Sample)
OECD	Organisation for Economic Co-operation and Development

CONTACT Marie Mirsch marie.mirsch@gdi.rwth-aachen.de www.linkedin.com/in/marie-mirsch-fairnessai

p	*p*-value
RAISE	Responsible AI Attitudes Specific to Engineering
RMSEA	Root Mean Square Error of Approximation
RWTH	Rheinisch-Westfälische Technische Hochschule (Rhenish-Westphalian Technical University)
SD	Standard Deviation
SRMR	Standardised Root Mean Squared Residual
t	t-value (t-test)
W	Test Statistic (Wilcoxon Test)
ρ	Correlation Coefficient

1. Introduction

The increasing integration of Artificial Intelligence (AI)-based systems[1] is accelerating the digital transformation of the engineering domain. The capacity to model complex and nonlinear processes, perform data-driven predictions without prior assumptions, and analyze vast datasets has garnered significant attention from engineering researchers and practitioners (Baghbani et al. 2022). Not least, the widespread use of generative AI has led to the integration of AI into our everyday lives and, consequently, into the engineering workplace. For instance, in Civil Engineering, AI supports structural design, construction management, and infrastructure maintenance (Bao and Li 2019; Demertzis, Demertzis, and Iliadis 2023; Liu et al. 2024), eventually leading to more cost-effective project execution (Nithyakarpagam and Mathiraj 2025). In the energy sector, AI enables smart grids and sustainable real-time energy management (Sureshraj et al. 2023). Across domains, AI supports complex decision-making and system optimisation (Yüksel et al. 2023) and enhances collaborative workflows (Gyory et al. 2022). Accessible toolkits and generative AI technologies have lowered the threshold for creating customised software solutions, even for those with limited programming skills (Schmidt 2023). Consequently, engineers are likely to be or become users and, increasingly, also designers of AI systems.

Alongside these numerous benefits, the deployment of AI systems in engineering is not without risks for companies (i.e. in terms of *public reputation and trust*), society (i.e. in terms of *harms*), and practitioners (i.e. in terms of *deskilling*). The opacity of many predictive AI systems makes it difficult to understand, explain, or contest AI-driven decisions (Burrell 2016), and uncertainty in risk assessment challenges their responsible use (Metcalf et al. 2021). As a consequence, unreproducible errors raise questions of accountability (Santoni de Sio and Mecacci 2021), which are particularly pressing due to a certain degree of autonomy and scalability. Moreover, an increasing overreliance on AI systems can diminish human oversight in critical contexts (Holzinger, Zatloukal, and Müller 2025), and particularly overreliance on generative AI systems may impact cognitive abilities, with individuals favouring fast AI-generated solutions over more deliberate reasoning (Zhai, Wibowo, and Li 2024).

To tackle these challenges, the field of *Responsible AI* discusses legal and technical, but primarily ethical considerations[2] throughout the AI lifecycle (Dignum 2020). These ethical considerations have given rise to a new interdisciplinary research field centred on developing technical means to implement principles such as fairness, transparency, and privacy in AI systems. The critique of these techno-solutionist approaches (Barocas and Selbst 2016; Sloane and Moss 2019) for being too narrow and 'oversimplif[ying] the philosophical field of ethics' while 'glorif[ying it] as a designated source of truth about what is right' (McFadden and Alvarez 2024, 786) recently led to an emerging third wave of AI ethics (Bolte and van Wynsberghe 2025; Häußermann and Lütge 2022; Huang et al. 2022; Kind 2020; Wegner, Decker, and Leicht-Scholten 2024). The research in this emerging third wave highlights structural, organisational, and systemic factors, suggesting that ethical AI practice must encompass more than technical solutions. In this sense, the conversation shifts from *creating ethical AI* to *using AI ethically*. While this shift was initially driven primarily by experiences with predictive and decision-making systems, it has gained renewed urgency with the rapid diffusion of generative AI, which introduces distinct yet related ethical challenges concerning data privacy and consent, stereotyping, and deskilling (Johri, Lindsay, and Qadir 2023). This shift underscores the need for a workforce equipped not only with technical expertise but also with the capacity for moral reflection and responsible decision-making in the development and deployment of AI systems.

Regulatory frameworks like the EU AI Act (Article 4) explicitly mandate AI literacy in companies. In parallel, educational standards like the CDIO Syllabus 3.0 (Malmqvist et al. 2022) and ABET curricula (ABET 2024), as well as professional guidelines such as those issued by the Association of German Engineers (VDI Verein Deutscher Ingenieure e.V. 2021) emphasise the importance of ethical AI. While ethics education is already well-established within engineering curricula (Martin, Conlon, and Bowe 2021), dedicated training in AI-specific ethics remains in its infancy (Decker et al. 2024; Moreno, Decker, and Leicht-Scholten 2024). However, the demand for such competencies extends beyond formal university education. As AI technologies increasingly reshape engineering practice, continuous professional development (i.e. lifelong learning and vocational training) becomes essential to ensure that not only students, but also practicing professionals, remain equipped to design and use AI responsibly. Current efforts to update curricula must therefore address both academic education and professional training, ensuring continuous skill development across career stages.

One's *attitudes* correlate with a predisposition for moral behaviour and are integral to moral decision-making. Attitudes encompass traits such as an individual's mindset, motivation, and adaptability (Allport 1935). Positive attitudes guide the motivation to engage with challenges (White 1959) and influence one's self-efficacy (Bandura 1997). Therefore, they can significantly affect one's capacity to apply knowledge and skills responsibly through expectations and interpretative frameworks as well as their effectiveness (Katz 1960; Schwarz 2007; Weinert 2001). As Knoth et al. (2024, 5) note, 'competent behaviours may not be performed if the necessary skills are present, but personal attitudes counteract performance.' In the context of Responsible AI, attitudes emerge as a necessary prerequisite for moral deliberation and moral behaviour (Walker 2012). They can take on many facets, for example, judging the relative importance of ethical principles such as fairness, transparency, privacy (Ghotbi, Ho, and Mantello 2022; Kieslich, Keller, and Starke 2022). They particularly influence one's adaptability and problem-solving skills, which are prerequisites for engaging with ethical questions and moral dilemmas (Ajzen 1985, 1991). Therefore, fostering ethical attitudes alongside technical expertise is fundamental to ensuring that engineers are not only aware of ethical principles but also committed to applying them in practice.

To facilitate targeted training and education, a better understanding of learners' attitudes is required – both in higher education and in professional contexts. Without *reliable tools,* it is difficult to assess the impact of interventions or monitor progress over time. Therefore, robust assessment frameworks are not only necessary for quality assurance in education but also play a key role in promoting Responsible AI practices in engineering disciplines. Yet, while some tools measure general attitudes toward AI (Grassini 2023; Schepman and Rodway 2020; Sindermann et al. 2021), these are not tailored to ethical dimensions. Similarly, existing instruments for moral reasoning in engineering (Howland et al. 2024; Zhu et al. 2014) lack specificity for AI ethics. Also, measures for assessing attitudes toward AI ethics (Jang, Choi, and Kim 2022) remain too generic for engineering-specific applications.

Knoth et al. (2024) highlight the importance of developing *domain-specific* AI literacy assessments to complement existing generic instruments. While generic AI literacy captures basic knowledge and skills relevant across contexts, it does not sufficiently address the specific applications, data types, and ethical challenges encountered in professional domains such as engineering. The authors argue that domain-specific AI literacy is essential for understanding how AI can be effectively and responsibly applied within a given field, taking into account domain-related expertise, practices, and regulatory requirements. In engineering, for example, the use cases, data characteristics, and potential societal impacts of AI differ significantly from other sectors, therefore necessitating tailored assessment approaches. The framework proposed by Knoth et al. (2024) distinguishes between generic AI literacy, domain-specific AI literacy, and AI ethics literacy, each measured along cognitive, behavioural, and attitudinal dimensions. This structure enables the development of instruments that more accurately reflect the competencies engineers require when working with AI in their specific professional environments.

Thus, while Responsible AI is a growing field in engineering, there is currently no domain-specific, validated instrument designed to assess engineers' attitudes toward Responsible AI. This study aims to address this gap by developing and validating the RAISE scale to assess Responsible AI Attitudes Specific to the Engineering domain, while ensuring applicability across demographic groups (e.g. gender, engineering field, employment status). Additionally, we explore whether knowledge serves as a predictor for these attitudes, offering insights into how educational and vocational training programmes might be effectively structured to promote Responsible AI attitudes and use across both academic and professional settings.

2. Methods and materials

Our item and scale development process followed a three-phase framework, including (1) item development, (2) content and face validation, and (3) construct validation through confirmatory factor analysis. An overview of the method steps is given in Figure 1. The number of items in each phase is depicted in Table 1 below. The following subsections describe this process, the demographics of our sample, and the descriptive and inferential analyses conducted.

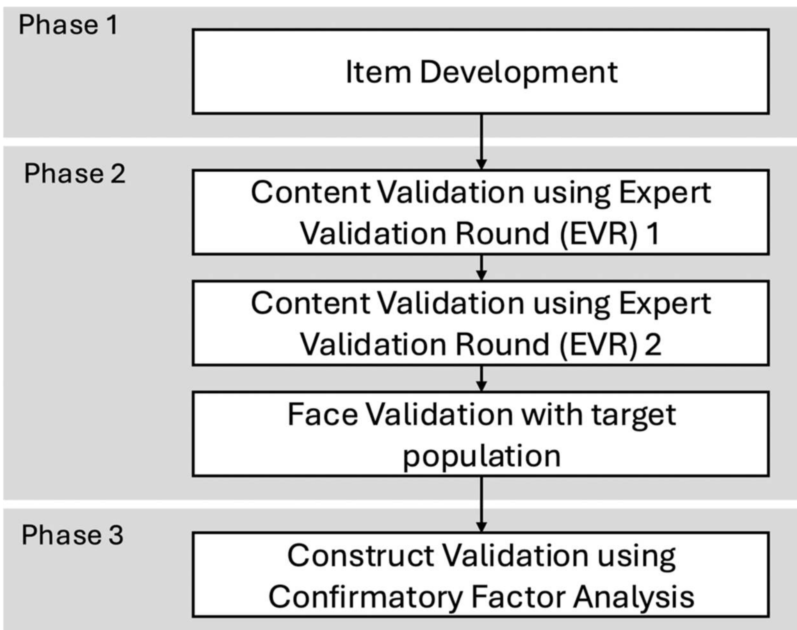

Figure 1. Development and validation steps in the RAISE scale development.

2.1. Item development

To develop the items, we used a mixed deductive-inductive approach, as proposed by Boateng et al. (2018). Deductively, we started with Jang, Choi, and Kim's (2022) validated scale for assessing students' attitudes toward ethical AI. In their study, they derive five core ethical AI dimensions, namely fairness, transparency, non-maleficence, privacy, and responsibility, based on AI ethics guidelines and literature reviews. Although recent advances in generative AI have intensified public and academic debates about ethical AI since late 2022, the underlying normative dimensions of common principles have remained remarkably stable. What has evolved are the manifestations, risk constellations, and application contexts of these ethical principles, rather than the principles themselves. For instance, issues of stereotyping are discussed in the context of fairness (Xiang 2024), and deepfakes and data ownership in the context of privacy (Lee et al. 2024; Lovato et al. 2024). Transparency will be critical in addressing issues of hallucinations (Adel and Alani 2025), and non-maleficence and responsibility pertain across all systems (e.g. Golda et al. 2024). Furthermore, recent post-genAI scale development studies with a similar aim, such as Wang et al. (2025), which assess AI ethical reflection, also rely on the initial five core ethical dimensions. To capture greater nuance than the initial scale and to adapt it to the engineering domain, we inductively refined and expanded Jang, Choi, and Kim's (2022) initial item pool by drawing on literature relevant to each dimension and incorporating domain-specific expert input. For *fairness*, we expanded the item pool by considering the fairness dimensions of procedural and distributive fairness by Colquitt and Rodell (2015). This step was taken to measure the distributive nature that engineered systems can have on their environment and the procedural effects that require consideration during the engineering of those systems. In practice, fairness becomes relevant throughout the whole AI lifecycle from design, data collection, model selection, and training to deployment and end-of-life procedures. This includes dimensions of accessibility, representativeness, correctability, accuracy, decision control, as well as equity and equality. Erasmus, Brunet, and Fisher (2021) define *transparency* through three interconnected concepts: explainability, understandability, and interpretability. They co-construct each other: interpretability depends on both understandability and explainability. Transparency in practice means making design and development decisions, as well as specific outputs of AI systems, visible and understandable for the intended users. This openness aims to foster trust, even if providing such transparency may sometimes come at the cost of reduced system accuracy. Items for the constructs of non-maleficence and privacy are aligned with Jobin, Ienca, and Vayena's (2019) analysis on AI ethics guidelines. *Non-maleficence* describes that 'AI should never cause foreseeable or unintentional harm' (Jobin, Ienca, and Vayena 2019, 394). Thus, they argue, non-maleficence includes fostering positive outcomes of AI usage; avoiding privacy infringements; prohibiting discrimination; preventing physical, psychological, or sexual harm; as well as maintaining trust in the system. Additionally, users should not become victims of skill degradation due to excessive use of AI (e.g. creative skills). The adherence to non-maleficence also includes the consideration of dual-use (e.g. for military operations) or misuse cases, as well as security measures against malicious hacking. *Privacy,* as considered by Jobin, Ienca, and Vayena (2019), is a value that many consider a right to be protected. Practically, it is discussed in relation to data protection and data security, but it is also bound to concepts of freedom and trust. Measures cover a range of categories, including modes of achievement, technical measures, further research, transparency, and regulatory approaches. For practitioners, this implies a responsibility to consider privacy and undergo training to implement it. Finally, *responsibility* addresses the allocation and assumption of accountability across the AI lifecycle and is particularly salient in complex socio-technical systems. Drawing on Santoni de Sio and Mecacci (2021), responsibility in the context of AI extends beyond backward-looking liability to encompass forward-looking obligations to ensure meaningful human control. This includes clearly defined roles and responsibilities among designers, engineers, organisations, and users; the capacity to anticipate and mitigate harmful outcomes; and the establishment of organisational structures that enable accountability, contestability, and corrective action. In engineering contexts, responsibility thus implies active engagement with the societal consequences of AI systems throughout design, deployment, and operation.

The process of expanding, refining, and adapting resulted in an initial item pool of 54 items (*v1*): 18 items for fairness, 8 items for transparency, 12 items for non-maleficence, 8 items for privacy, and 8 items for responsibility. Reverse-worded items were included to reduce acquiescence bias.

2.2. Content and face validation

Content validity was assessed in two expert validation rounds (EVR) with interdisciplinary expert panels, and face validity in a group of engineering students.

2.2.1. Content validation

In the first expert validation round (EVR 1), 12 experts assessed the initial 54 items (*v1*) for relevance and clarity using 4-point Likert scales (1 = not relevant/clear, 4 = highly relevant/clear), following Yusoff's (2019) guidelines. The panel was composed of experts in AI, engineering, ethics, and psychology/behavioural sciences to ensure comprehensive coverage of all relevant topic areas. To reflect this disciplinary diversity, we indicate each expert's domain in Table A1 (Appendix). Each item was accompanied by a domain definition adopted from Jang, Choi, and Kim (2022) and a brief explanation of content validity (Guion 1977). Based on qualitative feedback from EVR1, we retained positively evaluated items and revised items with comments for improvement. The revised version (*v2*) consisted of 52 items: 18 items for fairness, 8 items for transparency, 10 items for non-maleficence, 7 items for privacy, and 9 items for responsibility. We then conducted a second expert validation round (EVR 2) with nine additional experts (see Table A1 (Appendix) for domain details). Following EVR2, we retained items with an item-level content validity index (I-CVI) of ≥ 0.78 for both clarity and relevance. Items with an I-CVI ≥ 0.70 for relevance but < 0.78 for clarity were modified, while items with a relevance score below 0.70 were excluded. Based on expert input, one item was reclassified from non-maleficence to privacy. This iterative process resulted in a 37-item scale (*v3*), comprising 10 items for fairness, 5 items for transparency, 8 items for non-maleficence, 7 items for privacy, and 7 items for responsibility.

2.2.2. Face validation

Following the content validation, the items were analyzed for face validity (FVI) by the target population on an item- and scale-level, as proposed by Yusoff (2019).

The target population consisted of engineering students enrolled in the Innovation & Diversity seminar at RWTH Aachen University. Prior to the seminar, 22 students completed a pre-test of the scale and provided open-ended feedback. Subsequently, 21 students participated in a structured group discussion on November 7, 2024. The session proceeded in several steps: first, students reviewed the item content and scale structure. Next, they examined data distributions to identify any problematic response patterns, such as floor or ceiling effects, that might indicate issues with item formulation. They then discussed their interpretation of each item, focusing on aspects of interpretability, relevance, and clarity to identify any potentially suggestive or biased wording. Finally, they made recommendations on whether each item should be retained, revised, or removed. Based on this comprehensive feedback, we revised the wording of several items for improved clarity and excluded others, resulting in a refined 31-item scale (*v4*). This version included a balanced number of reverse-coded items to reduce response bias and was subsequently used for construct validation, as described in the following sections.

2.3. Participants

We collected data on *v4* via an online survey administered through SoSci Survey. Participants were recruited from 21 German comprehensive universities and universities of applied sciences, as well as through professional engineering networks. All participants were informed about the study's objectives, data confidentiality, and the voluntary nature of participation. Data cleaning involved excluding incomplete responses, response time outliers, and cases of patterned responding. Furthermore, data cleaning involved identifying and limiting socially desirable responses (Durmaz, Dursun, and Kabadayi 2020; Randall and Fernandes 1991). To do so, our survey included the *brief social desirability* *scale* containing four yes-no items (Haghighat 2007). As recommended by the authors, we excluded responses that were answered positively three or four times ($n = 38$).

This process yielded 235 complete responses, exceeding the recommended minimum sample size of 5–10 participants per item (Hair et al. 2013) and aligning with Jang, Choi, and Kim (2022) as an adequate sample size for a confirmatory factor analysis (CFA).

Out of the $N = 235$ participants, 56 participants identified as female (23.8%), 162 as male (68.9%), one as diverse (0.4%), and 16 opted not to disclose their gender (6.8%). The participants' mean age was 26.4 years, ranging from 18 to 64 years, with the largest cohorts being 20–24 years ($n = 82$, 34.9%) and 25–29 years ($n = 73$, 31.1%). Smaller proportions were aged 30–34 ($n = 28$, 11.9%) and 35–39 ($n = 18$, 7.7%). Each of the 5-year age groups between 40 and 64 contained three or four participants, while six individuals (2.6%) did not report their age. Participants' disciplinary backgrounds reflected diverse engineering fields, with multiple selections permitted. The most common disciplines were mechanical engineering ($n = 69$), civil engineering ($n = 65$), and information technology-related fields such as computer science and software engineering ($n = 61$). Further disciplines included electrical engineering ($n = 35$), science and mathematics ($n = 24$), architecture ($n = 10$), and georesources and materials engineering ($n = 9$). Additionally, 39 participants selected 'Other' and provided free-text responses, including environmental engineering, aerospace/aeronautical engineering, and data science. Concerning professional status, the majority of participants ($n =$ 140, 59.6%) were university students, interns, or apprentices. Another 88 participants (37.4%) were employed or self-employed, while 7 individuals (3.0%) selected 'Other.'

2.4. Construct validation through confirmatory factor analysis

The items generated through the iterative process (*v4*) were psychometrically validated with the data from the online survey. To do so, a CFA was initiated using the lavaan R package (Rosseel 2012). This analysis assessed the extent to which our collected data aligned with our proposed five-factor model, comprising sub-factors of fairness, transparency, non-maleficence, privacy, and responsibility, as outlined by Jang, Choi, and Kim (2022). We used multiple indices to evaluate the model: the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardised root mean residual (SRMR) (Kline 2005). These indices were used to judge how well the model matched the data. The CFI weighs the proposed model against an independent (null) model, adjusting for sample size and complexity. Values above .90 usually indicate an acceptable fit (Hu and Bentler 1999). The RMSEA shows how closely the model's covariance structure approximates that of the population. Scores below .08 are generally viewed as an acceptable fit,

Table 1. Item counts throughout the three development phases.

Phase	Development Step	Fairness	Trans-parency	Non-Maleficence	Privacy	Responsi-bility	$\sum$
Phase I	Item Development	18 items	8 items	12 items	8 items	8 items	54 items *v1*
Phase II	Content Validation with EVR 1	18 items	8 items	10 items	7 items	9 items	52 items *v2*
	Content Validation with EVR 2	10 items	5 items	8 items	7 items	7 items	37 items *v3*
	Face Validation	8 items	5 items	7 items	6 items	5 items	31 items *v4*
Phase III	Construct Validation using CFA	Summarized in the do-no-harm scale with 7 items	3 items	Summarized in the do-no-harm scale with 7 items	5 items	Summarized in the do-no-harm scale with 7 items	15 items *Final RAISE Scale*

Note. EVR stands for Expert Validation Round, CFA for Confirmatory Factor Analysis.

and scores below .05 are viewed as an excellent fit (Marsh, Hau, and Wen 2004). The SRMR captures the average gap between observed and predicted correlations, with figures under .08 indicating a good fit (Marsh, Hau, and Wen 2004). To examine internal consistency, we calculated McDonald's omega, which is a more robust alternative to Cronbach's alpha that is less prone to bias because it accounts for cases in which the assumption of tau-equivalence (equal factor loadings for all items) is violated (Graham 2006). McDonald's omega is therefore better suited to congeneric scales, where item loadings contribute differently to the total score (e.g. within a CFA) (Lucke 2005). As with Cronbach's alpha, McDonald's omega values of .70 or higher are typically considered adequate, while values down to .60 may still be deemed acceptable (Azmi et al. 2024; Shrout 1998).

The item development and validation process is depicted in Table 1.

2.5. Descriptive and inferential analyses

In addition to assessing the instrument's psychometric properties through CFA, we conducted group comparisons and correlational analyses to evaluate the scale's validity and practical applicability further.

First, group comparisons across gender, engineering domain, and professional status allowed us to examine whether the scale performs consistently across diverse subpopulations. Such analyses contribute to assessing structural stability and provide indications of measurement invariance, which is critical for the generalizability of psychometric instruments (Putnick and Bornstein 2016). Notable group differences may also reveal whether specific subscales or items are interpreted differently depending on demographic or contextual factors. To select appropriate statistical procedures for group comparisons, we first assessed the normality of the data using the Shapiro–Wilk test (Yazici and Yolacan 2007). A statistically significant result ($p < .05$) indicates a deviation from normality. In cases where the Shapiro–Wilk test suggested deviations from normality, we additionally considered skewness and kurtosis indices as recommended by Kline (2016). These indices were calculated by dividing the respective skewness and kurtosis statistics by their standard errors. According to them, absolute values exceeding $|3|$ for the skewness index or $|10|$ for the kurtosis index indicate *substantial* deviations from normality and may call for non-parametric analyses. To examine the assumption of homogeneity of variances, we employed Levene's test (Gastwirth, Gel, and Miao 2009). A non-significant result ($p > .05$) suggests that the variances between groups are sufficiently similar to satisfy the assumption of homogeneity of variances. For two-group comparisons (e.g. gender and professional status), we then applied independent samples t-tests under the assumption of equal variances, provided that both normality and homogeneity of variances could be assumed (Field 2024). The independent samples t-test evaluates whether the means of two independent groups differ significantly. In cases where normality was met but the assumption of homogeneity of variances was violated (as indicated by a significant Levene's test), we employed Welch's t-test as a robust alternative. Unlike the standard t-test, Welch's t-test adjusts the degrees of freedom to account for heterogeneity of variances and yields more accurate Type I error rates under such conditions (Ruxton 2006). When normality was violated – regardless of whether homogeneity of variances could be assumed – we opted for the Mann–Whitney U test (Nachar 2008). Instead of comparing means, the Mann–Whitney U test assesses whether the distributions of ranks differ systematically between two groups. It is particularly appropriate when the scale of measurement is ordinal or when the distribution of scores is skewed or contains outliers. A significant result in this test indicates that the central tendency of one group is systematically higher or lower than the other. Still, it does not allow for direct interpretation in terms of mean differences. For comparisons involving more than two groups (e.g. across multiple engineering domains), a one-way analysis of variance (ANOVA) was conducted, provided that both normality and homogeneity of variances were satisfied (Field 2024). A one-way ANOVA tests whether there is a statistically significant difference in means among three or more independent groups by comparing the ratio of between-group to within-group variance. A significant F-statistic indicates that at least one group mean differs from the others; however, it does not specify which groups differ. To mitigate the risk of inflated Type I error rates due to multiple testing, we applied the Holm correction to all *p*-values resulting from multiple pairwise comparisons (Holm 1979). The Holm procedure is a sequentially rejective method that controls the family-wise error rate more efficiently than the traditional Bonferroni correction by adjusting *p*-values in a stepwise manner. This enables greater statistical power while maintaining strict control over false positives.

Second, we explored correlations between participants' prior AI knowledge, specific Responsible AI knowledge, as well as engineering ethics knowledge and their attitudes toward Responsible AI. Specifically, we aimed to examine whether familiarity with AI is associated with more reflective or ethically aware attitudes. Previous research has shown mixed results in this regard. For example, Jang, Choi, and Kim (2022) found that prior AI education had a significant effect on fairness-related attitudes, but not on other ethical dimensions. Other studies have suggested that knowledge alone does not necessarily influence ethical attitudes (Fabrigar et al. 2006). Our analyses seek to contribute further empirical evidence to this debate and to assess the scale's sensitivity to differences in prior AI knowledge. We assessed AI knowledge using three overarching items derived from Pinski and Benlian (2023) AI literacy scale, capturing participants' general self-assessed competence in understanding and interacting with AI technologies. Each item was rated on a five-point Likert scale ranging from 1 ('I strongly disagree') to 5 ('I strongly agree'), and scores were averaged to form a composite AI knowledge score. In addition to that, to measure participants' self-perceived knowledge of Responsible AI principles, five single-item indicators were employed, each corresponding to one of the Responsible AI principles of the original scale by Jang, Choi, and Kim (2022) (fairness, transparency, privacy, responsibility, and non-maleficence). Participants rated their familiarity with each principle on a five-point Likert scale (1 = 'not knowledgeable at all', 5 = 'very knowledgeable'). To account for the merged do-no-harm scale in our final scale, the respective principles (fairness, non-maleficence, and responsibility) were averaged to form a composite score (see 3.1.2 for an explanation of why these principles were combined). A further single item asked participants to rate their overall knowledge of engineering ethics on the same five-point Likert scale. To decide on the mode of evaluation correlation, we first examined the distributional characteristics of the involved variables. Normality was assessed using the Shapiro–Wilk test (Yazici and Yolacan 2007) as well as standardised skewness and kurtosis indices, following the guidelines by Kline (2016). In all of our cases, one or both variables showed significant deviations from normality; thus, the Spearman rank correlation coefficient was used as a robust non-parametric measure (Field 2024). Visual inspection of scatterplots further supported the decision.

All statistical analyses were conducted using R. Subscale scores were calculated by averaging the items within each ethical dimension to enable comparisons and correlation analyses.

3. Results

This section presents the results from the three-stage validation of the instrument, beginning with expert-based content validation, followed by confirmatory factor analysis. The second part of this chapter contains our descriptive and inferential statistics results related to the final scale.

3.1. Validity of the instrument

The psychometric quality of the developed instrument was assessed by content, face, and construct validity.

3.1.1. Content and face validity

Content validity was established through two rounds of expert validation (EVR 1 and EVR 2) involving interdisciplinary experts in AI, engineering, ethics, and behavioural sciences. The first round (EVR 1, $n = 12$) qualitatively improved the items (leading to *v2*), and the second round (EVR 2; $n = 9$)

determined the item pool for face validity (leading to *v3*). To complement the expert-based assessment, face validity was examined with 21 participants from the target population. Through qualitative feedback, group discussions, and suggestions for improvement, items were evaluated regarding their comprehensibility, interpretability, and potential for response bias (leading to *v4*). As a result of this iterative validation process, scale *v4* comprises a total of 31 items, distributed across the five dimensions: fairness (8 items), transparency (5 items), non-maleficence (7 items), privacy (6 items), and responsibility (5 items). This version served as the basis for the subsequent CFA.

3.1.2. Construct validity through confirmatory factor analysis

We conducted an initial CFA with the proposed model of fairness, transparency, non-maleficence, privacy, and responsibility using the maximum likelihood estimator with robust corrections (MLR). However, the model demonstrated poor to mediocre fit to the data: $\chi^2(424) = 857.25$, $p < .001$, CFI = .661, RMSEA = .068 (90% CI [.061, .074]), and SRMR = .082. Although the RMSEA is within an acceptable range, the CFI is below the conventional threshold of .90 for good fit.

Therefore, we extracted the standardised factor loadings of all items on their respective factors to determine the source of this misfit. Indeed, many items had loadings below .40. CFA factor loadings of .40 or higher are typically considered acceptable, particularly in the social sciences. These loadings indicate a strong enough association between an observed variable and its underlying latent factor to meaningfully contribute to measuring the construct (Stevens 2001). Consequently, we eliminated all items with loadings below .40 and reran the CFA. The factors now consisted of the following number of items: fairness (2 items), transparency (4 items), non-maleficence (5 items), privacy (5 items), and responsibility (4 items). Rerunning the CFA with this model specification produced a

Table 2. Item wordings, means (M), standard deviations (SD), and factor loadings (λ), organised by their respective final subscales along with reliabilities (ω) of the subscales.

Items	Wordings	M	SD	λ	ω
'Do-No-Harm'					.685
	I think that AI systems that I use should consider all stakeholders' perspectives (e.g. urban planning considering residents' and officials' perspectives), even if this increases costs and time.	3.91	0.99	.528	
	When using AI systems, I will strive to put it only to good use.	4.29	0.89	.411	
	I find it important to evaluate whether the AI systems that I use in my engineering work are simultaneously used in harmful applications by others.	3.46	1.18	.373	
	I would only use AI systems in my engineering work if responsibility for the outcomes (including potentially harmful ones) is clearly defined.	3.97	0.96	.445	
	I feel that it is my responsibility to thoroughly assess the possible consequences of using AI systems in any of my engineering work, even if this increases required time and costs.	4.19	0.79	.466	
	When using AI systems in engineering tasks that affect the public, I care about considering public concerns even if this increases required time and costs.	4.19	0.86	.632	
	When using AI systems in engineering tasks that affect the public, I would order external audits.	3.64	0.98	.596	
'Transparency'					.771
	AI systems that I use in engineering do not have to explain the reasoning behind their outcomes. (reversed)	4.24	0.91	.739	
	In my engineering work, I trust AI systems that cannot explain why they made certain decisions. (reversed)	4.05	1.03	.695	
	AI systems do not have to allow me as an engineer to trace the reasoning behind every outcome. (reversed)	4.01	1.04	.749	
'Privacy'					.710
	I would use AI systems that lack comprehensive data protection measures that safeguard sensitive information. (reversed)	3.82	1.10	.579	
	It is okay for me to use someone else's personal data to run AI systems. (reversed)	3.94	1.10	.412	
	I would only use AI systems in my engineering work if they consider privacy.	3.78	1.06	.648	
	Privacy training before working with AI systems is essential to me.	3.81	1.00	.527	
	I am committed to making the extra effort to follow AI security guidelines (such as access control and data security) to protect sensitive information and maintain system integrity.	4.14	0.94	.533	

Note. All Likert-scaled items (with values 1–5) follow the question '*What is your opinion on the following statements?*'; λ: factor loadings of the items on their respective subscales; ω = McDonald's Omega.

covariance matrix of latent variables that was not positive definite, indicating that there could be very high correlations between variables. Upon closer inspection, we observed that the fairness factor strongly correlated with non-maleficence ($r = .954$) and responsibility ($r = .985$). Considering these findings, we changed our a priori chosen factorial structure, merging the fairness, non-maleficence, and responsibility factors into a factor called 'do-no-harm.' This new factor consists of the items from its previous factors (11 items).

This new model demonstrated an improved, though still suboptimal, fit: $\chi^2(167) = 394.44$, $p < .001$, CFI = .760, RMSEA = .08, 90% CI [.069,.090], and SRMR = .082. While the RMSEA and SRMR fall within or near acceptable limits, the CFI remains below the commonly recommended threshold of .90, indicating room for improvement to the model. Therefore, the next step was to inspect the modification indices (Kaplan 1990; Sörbom 1989). Inspection of the modification indices led to the elimination of items that were redundant with other items in their scale, had high shared error variance with other items, or exhibited cross-loadings on factors to which they were not intended to load. Conversely, items that showed higher loadings on other factors to which they were not originally assigned, but that were theoretically meaningful, were retained. The final assignment of items to factors, their loadings, means, standard deviations, and wording can be found in Table 2.

The final model yielded an acceptable fit: $\chi^2(87) = 134.88$, $p = .001$, CFI = .917, RMSEA = .052 (90% CI [.034,.068]), and SRMR = .061. The CFI value meets the commonly recommended threshold of $\geq .90$, and both the RMSEA and the SRMR fall within the acceptable range. These results suggest that the model adequately represents the data. While not optimal, the reliability of each resulting subscale was also within an acceptable range: do-no-harm ($\omega = .69$), transparency ($\omega = .77$), and privacy ($\omega = .67$). Following the described procedure, we obtained a sound, practical, and theoretically meaningful scale that values domain-specificity in the area of AI when measuring engineers' attitudes toward Responsible AI.

3.2. Descriptive and inferential statistics

This section presents the results of conducting descriptive and inferential analyses on 235 responses from engineering students and professionals in Germany.

3.2.1. Descriptive statistics

Participant demographics are disclosed in section 2.3. Means, standard deviations, and scale reliabilities (McDonald's ω) for each subscale are reported in Table 2 (see Section 3.1.2). In summary, participants generally expressed high agreement with attitudes toward Responsible AI across all dimensions. The highest mean was observed for the transparency subscale (M = 4.10), indicating that participants strongly valued comprehensibility and explainability in AI systems. The do-no-harm subscale followed closely (M = 3.95), indicating a strong concern for mitigating potential harm and a straightforward assignment of accountability when using AI. The privacy subscale had the lowest mean among the three (M = 3.90), though still reflecting high agreement with data protection and ethical information handling. With all subscale means well above the scale midpoint of 3, these results suggest that the participants essentially value Responsible AI, with some variation in emphasis across the principles.

3.2.2. Group disparities in attitudes toward responsible AI

To investigate whether attitudes toward Responsible AI differed by gender, professional status, or disciplinary domain, we compared both descriptive and inferential statistics.

3.2.2.1. Gender differences. First, we compared responses along the gender dimension. We compared responses between women ($n = 56$) and men ($n = 162$), excluding participants who selected 'diverse' ($n = 1$) or preferred not to disclose ($n = 16$). We acknowledge the limitation that the female sample only represents one-third of the male sample – a limitation that is common in

Table 3. Descriptive statistics by gender.

Subscale	Female ($n = 56$)	Male ($n = 162$)
Do-No-Harm	M = 4.05, SD = 0.565	M = 3.92, SD = 0.553
Transparency	M = 4.11, SD = 0.904	M = 4.06, SD = 0.805
Privacy	M = 4.04, SD = 0.551	M = 3.83, SD = 0.712

engineering (Kahn and Ginther 2017). Descriptively, we find that females scored higher on average than males in all dimensions (see Table 3). Standard deviations indicate relatively stable estimates across groups. While means suggest gender-based trends in attitudes, this finding could not be confirmed statistically.

We first checked for normal distribution of the subscales per group using the Shapiro–Wilk test. For do-no-harm, scores indicated deviation from normality both for males (W = .978, $p = .010$) and females (W = .958, $p = .047$); for privacy, scores indicated deviation from normality for males (W = .965, $p = .000$), but not for females (W = .965, $p = .107$); and for transparency, both male (W = .916, $p < .001$) and female (W = .858, $p < .001$) scores indicated deviations from normality. We then computed standardised skewness and kurtosis indices. For the do-no-harm subscale, skewness indices ranged from −2.38 (male) to −0.46 (female), and kurtosis indices from 8.49 (male) to 4.57 (female). For transparency, skewness indices were −3.53 (male) and −2.61 (female), with kurtosis indices of 7.05 (male) and 3.86 (female). The privacy subscale yielded skewness indices of −2.52 (male) and −1.36 (female), and kurtosis indices of 7.43 (male) and 4.07 (female). Although some indices exceeded the recommended thresholds, the deviations were mild to moderate. Given the sufficiently large sample size ($N = 235$), which supports the robustness of parametric tests against moderate violations of normality (Ghasemi and Zahediasl 2012), we conducted tests that assumed normal distributions in the subsequent tests. Homogeneity of variances was assessed using Levene's test. The assumption was met for do-no-harm ($F(1, 216) = 0.019$, $p = .892$) and transparency ($F(1, 216) = 0.510$, $p = .476$), but was violated for privacy ($F(1, 216) = 4.286$, $p = .040$). Consequently, standard independent samples t-tests assuming equal variances were applied for do-no-harm and transparency, while Welch's t-test, which does not assume equal variances (Ruxton 2006), was used for privacy.

Independent samples t-tests revealed no statistically significant gender differences in do-no-harm, $t(216) = 1.50$, $p = .134$, nor transparency, $t(216) = 0.45$, $p = .656$. For privacy, there was a statistically significant difference in the uncorrected test, $t(123) = 2.23$, $p = .028$, with female participants rating privacy considerations more highly than males. However, after the Holm correction (Holm 1979), none of the comparisons reached statistical significance (adjusted *p*-values: .269, .656, and .083, respectively). Taken together, the results suggest that there were no statistically significant differences between female and male participants in their responses across the three subscales after controlling for multiple testing. The data indicate a slight trend toward higher ratings among women, particularly on the privacy subscale; however, this difference did not withstand correction for multiple comparisons.

3.2.2.2. Engineering domain differences. Second, we compared potential differences across the three most highly represented engineering domains among our respondents, namely civil engineering ($n = 62$), mechanical engineering ($n = 51$), and IT-related fields (hereafter referred to as technology; $n = 43$). Mean scores for each subscale across the three engineering domains were compared (see Table 4). Mechanical engineering scores are higher throughout all subscales, but this could not be confirmed statistically.

Table 4. Descriptive statistics by engineering domain.

Subscale	Civil Engineering ($n = 62$)	Mechanical Engineering ($n = 51$)	Technology ($n = 43$)
Do-No-Harm	M = 3.96, SD = 0.567	M = 4.06, SD = 0.551	M = 3.87, SD = 0.647
Transparency	M = 4.12, SD = 0.771	M = 4.30, SD = 0.746	M = 4.02, SD = 0.740
Privacy	M = 3.80, SD = 0.779	M = 4.07, SD = 0.644	M = 3.87, SD = 0.634

We first applied the Shapiro-Wilk test within each domain and subscale to examine the assumption of normality. The results indicated deviations from normality, particularly for the transparency subscale across all domains ($p < .01$), as well as for privacy in the civil ($p = .013$) and mechanical ($p = .043$) domains. For the do-no-harm subscale, the Shapiro-Wilk test was not significant in any domain (all $p > .05$), suggesting no statistical evidence against normality. Across all subscales and domains, skewness and kurtosis indices remained well within the thresholds, indicating that although some deviations from normality were statistically significant, the distributions did not exhibit severe skewness or kurtosis. Thus, despite some formal violations detected by the Shapiro-Wilk test, the overall distributional characteristics can be considered approximately normal, and parametric analyses were deemed appropriate. Homogeneity of variances was tested using Levene's test for each dependent variable. The assumption was met for all subscales: do-no-harm, $F(2, 153) = 0.56$, $p = .574$; transparency, $F(2, 153) = 0.17$, $p = .842$; privacy, $F(2, 153) = 1.29$, $p = .278$.

One-way ANOVAs revealed no statistically significant differences between domains on any of the three subscales, with results as follows: do-no-harm: $F(2, 153) = 1.30$, $p = .275$; transparency: $F(2, 153) = 1.68$, $p = .190$; privacy: $F(2, 153) = 2.13$, $p = .122$. In summary, there were no significant differences in participants' attitudes toward Responsible AI across the three largest engineering domains. Although descriptive statistics indicated minor differences in mean scores (with mechanical engineering having the highest scores in all three subscales), these did not reach statistical significance in any of the subscales.

3.2.2.3. Employment status differences.

Lastly, we compared attitudes between engineering students and practitioners. To do so, we grouped students and interns, henceforth referred to as students ($n = 140$), and employees and self-employed individuals, henceforth referred to as practitioners ($n = 88$). Mean scores for each subscale across the two groups were comparable (see Table 5).

The Shapiro-Wilk-test indicated deviations from normality for transparency in both groups (students: $W = 0.896$, $p < .001$; practitioners: $W = 0.922$, $p = .006$), as well as for privacy (students: $W = 0.950$, $p = .013$; practitioners: $W = 0.953$, $p = .043$). For do-no-harm, the tests showed no significant deviation from normality (students: $W = 0.964$, $p = .063$; practitioners: $W = 0.967$, $p = .165$). To further evaluate the extent of non-normality, we examined skewness and kurtosis indices. All subscales fell within recommended thresholds, except transparency among students, which showed a notable skewness index of −5.09, suggesting substantial non-normality.

Consequently, we used a Mann–Whitney U test to compare groups on the transparency subscale. The test revealed a statistically significant difference between students and practitioners ($W = 5113.5$, $p = .029$), with students reporting higher transparency scores. However, this effect did not remain significant after Holm correction ($p = .112$). For the other two subscales (do-no-harm and privacy), Levene's tests confirmed homogeneity of variances across groups (all $p > .21$), justifying the use of independent samples t-tests assuming equal variances. Results showed no significant group differences for do-no-harm, $t(226) = -0.41$, $p = .680$, nor for privacy, $t(226) = -0.21$, $p = .834$.

Table 5. Descriptive statistics by students and practitioners.

Subscale	Students ($n = 140$)	Practitioners ($n = 88$)
Do-No-Harm	M = 3.97, SD = 0.547	M = 3.94, SD = 0.583
Transparency	M = 4.22, SD = 0.787	M = 3.99, SD = 0.820
Privacy	M = 3.91, SD = 0.685	M = 3.89, SD = 0.695

3.2.3. Disparities in attitudes toward responsible AI based on prior knowledge

To explore whether AI education is associated with more reflective attitudes, we examined correlations between participants' attitudes toward Responsible AI and their AI knowledge, Responsible AI knowledge, and engineering ethics knowledge.

3.2.3.1. Correlations with AI knowledge. To assess the relationship between participants' AI knowledge (assessed in terms of AI literacy, see section 2.4) and their attitudes toward Responsible AI, we first examined the distribution of the respective variables. Shapiro-Wilk tests indicated significant deviations from normality for all variables (do-no-harm: W = 0.979, $p = .002$; transparency: W = 0.898, $p < .001$; privacy: W = 0.962, $p < .001$). While the AI Literacy scale fell within acceptable thresholds (Skewness Index = –0.82; Kurtosis Index = 7.99), the subscales do-no-harm, privacy, and particularly transparency (Skewness Index = –4.80) indicated significant non-normal distributions. Consequently, non-parametric Spearman rank-order correlations were used to examine associations between AI literacy and attitudes toward Responsible AI.

The correlations between AI literacy and the three attitude subscales were weak and not statistically significant: do-no-harm ($\rho = -0.040$, $p = .547$), transparency ($\rho = -0.059$, $p = .366$), and privacy ($\rho = -0.034$, $p = .604$). These results suggest that AI literacy is not meaningfully associated with attitudes toward Responsible AI in this sample and may not serve as a reliable predictor of alignment with Responsible AI principles.

3.2.3.2. Correlations with responsible AI knowledge. To assess the relationship between participants' Responsible AI knowledge and their attitudes toward Responsible AI, we first examined the distribution of the respective variables. The Shapiro-Wilk test indicated deviations from normality for all Responsible AI knowledge indicators (do-no-harm: W = 0.969, $p < .001$; transparency: W = 0.902, $p < .001$; privacy: W = 0.915, $p < .001$), as well as for all attitude subscales (see previous section). However, skewness and kurtosis indices for the Responsible AI knowledge indicators fell within acceptable thresholds (|Skewness Index| < 3; |Kurtosis Index| < 10), indicating no substantial univariate non-normality. Given the ordinal nature of the knowledge indicators and the significant results from the Shapiro-Wilk tests, we still opted for non-parametric Spearman rank-order correlations.

The analysis revealed a small but statistically significant positive correlation between Responsible AI knowledge related to privacy and attitudes toward privacy ($\rho = .169$, $p = .009$), indicating that greater self-reported knowledge of privacy-related principles was associated with more favourable attitudes toward privacy in AI contexts. In contrast, correlations between knowledge and attitudes in the domains of do-no-harm ($\rho = .071$, $p = .278$) and transparency ($\rho = .045$, $p = .497$) were weak and not statistically significant. These results suggest that Responsible AI knowledge may be selectively associated with positive attitudes, particularly in domains where public discourse and individual awareness (e.g. privacy) are more pronounced.

3.2.3.3. Correlations with engineering ethics knowledge

The Shapiro–Wilk tests indicated significant deviations from normality for all variables ($p < .001$ for most), which was further supported by the calculated skewness and kurtosis indices. Notably, the transparency and do-no-harm subscales showed pronounced departures from normality, while the engineering ethics knowledge scale exhibited a slight negative skew (Skewness Index = −0.53). Due to these violations of normality assumptions, Spearman rank-order correlations were conducted to examine the relationships between engineering ethics knowledge and the three ethical attitude subscales.

The results revealed no statistically significant correlations between engineering ethics knowledge and the attitude subscales: do-no-harm ($\rho = 0.10$, $p = .11$), transparency ($\rho = 0.04$, $p = .59$), and privacy ($\rho = 0.09$, $p = .16$).

4. Discussion

We developed and validated the RAISE scale to assess attitudes toward Responsible AI within the engineering domain. Building on the five-factor model proposed by Jang, Choi, and Kim (2022), we expanded the item pool based on expert input and conducted a confirmatory factor analysis using data from 235 complete responses from engineering students and practitioners. The confirmatory factor analysis resulted in a three-factor structure, comprising the subscales do-no-harm, transparency, and privacy, resulting in a final scale with 15 items. The model demonstrated acceptable fit indices and satisfactory internal consistency for each subscale, providing initial evidence for the reliability and construct validity of the instrument. Moreover, the scale performed consistently across demographic groups (i.e. gender, engineering domain, and professional status).

4.1. Discussion of the RAISE scale validity

Interestingly, our confirmatory factor analysis suggested a consolidation of the dimensions fairness, non-maleficence, and responsibility into a single construct. Observing that all items pointed toward the prevention of harm caused by the system or the attribution of responsibility when such harm occurs, we consequently labelled this factor the do-no-harm subscale. It captures attitudes toward preventing harm by considering diverse stakeholder perspectives, designing AI systems for beneficial use, evaluating them for potentially harmful applications, and acknowledging responsibility for their outcomes. The transparency subscale addresses the importance of understanding how AI systems operate and produce their outputs – an aspect that is not necessarily distinct from harm, but rather a legal and moral prerequisite (Decker, Wegner, and Leicht-Scholten 2025). Finally, the privacy subscale addresses the protection of sensitive data, both one's own and that of others.

It is important to be clear about what the scale can and cannot do. Specifically, it aims to measure self-reported attitudes toward acting responsibly in contexts where engineers engage with AI systems in their academic or professional lives. That is, it reflects the degree to which individuals *believe* that the ethical principles of do-no-harm, transparency, and privacy *should be considered*. Lind (2019) refers to this affective dimension as moral orientation. However, moral orientation is not a predictor of moral *competence*, defined as 'the ability to solve problems and conflicts through deliberation and discussion based on moral principles' (Lind 2019, 7), and thus reflects a more deliberative and skill-based approach. This distinction is crucial: while moral orientation reflects what individuals *believe* should be done, moral competence concerns their ability to *apply* those principles in concrete (real or hypothetical) situations. In this sense, moral orientation may be more vulnerable to social desirability bias, capturing what individuals think they should believe (e.g. one *should* not harm) rather than how they would reason through morally complex situations. Moral competence, by contrast, involves the practical application of one's moral principles and requires the cognitive and dialogical skills to engage in ethical reasoning. Moreover, possessing moral competence does not necessarily *guarantee* moral behaviour. Even individuals who know what is morally appropriate in general (moral orientation) and in specific cases (moral competence) may fail to act accordingly due to other mediating factors, for example, an inability to recognise morally salient features of a situation, or fear of negative personal consequences. The scale does not capture these behavioural dimensions. This limitation reflects the well-documented intention–behaviour gap (Ajzen 1991; Blake 1999), which is particularly salient in moral psychology, where context and situational constraints strongly influence action, as illustrated by classic studies such as the Good Samaritan experiment (Darley and Batson 1973).

Apart from that, while our initial validation results are promising, several limitations must be acknowledged. While we had a satisfactory number of respondents from 21 comprehensive universities and universities of applied sciences in Germany, the distribution among these universities is not uniform. Furthermore, the focus on Germany is a limitation in itself. Although the model showed configural stability across gender, engineering domain, and employment status, further testing could incorporate different cultural contexts, career stages, and other engineering disciplines to further establish the scale's robustness. Second, like all self-report instruments, the scale is potentially subject to social desirability bias, which may obscure respondents' true attitudes (Durmaz, Dursun, and Kabadayi 2020; Randall and Fernandes 1991). In particular, attitudes toward Responsible AI may be influenced by normative expectations, especially in educational or professional settings where 'correct' responses are implicitly known or assumed. We tried to minimise this influence by erasing respondents who were shown to be prone to socially desirable answers using the brief social desirability scale (Haghighat 2007). Finally, while our factor structure was derived based on a theoretically informed item pool that extended the scale by Jang, Choi, and Kim (2022), the CFA made major changes, adding an exploratory aspect to our scale development process. Further validation on an independent sample would be worthwhile to confirm the scale's construct validity and practical utility more firmly.

The key distinctions between our scale and other instruments reviewed earlier are as follows: contrary to established tools for assessing general attitudes toward AI – such as those presented by Grassini (2023), Schepman and Rodway (2023), Sindermann et al. (2021), or Stein et al. (2024) – our scale explicitly focuses on attitudes toward *Responsible* AI. While other tools focus on general ethical reasoning within engineering – such as those by Zhu et al. (2014) and Howland et al. (2024) – our scale explicitly focuses on ethical challenges of AI systems. Lastly, our scale is developed for and validated within the engineering domain, specifying the focus of scales like the Attitudes Toward the Ethics of Artificial Intelligence (AT-EAI) scale by Jang, Choi, and Kim (2022). Thus, our scale addresses the attitudes toward Responsible AI *within* engineering, offering a theoretically grounded, domain-specific instrument for use in both research and educational practice and evaluation.

4.2. Discussion of the descriptive and inferential statistics

While this study aimed to develop and validate the RAISE scale, the descriptive findings also offer an initial insight into how engineering students and professionals in Germany perceive the importance of Responsible AI. Overall, the attitudes toward the do-no-harm, transparency, and privacy principles were rated positively, suggesting a generally positive attitude toward ethical considerations in AI development and use. However, variability in responses, especially in the do-no-harm subscale, indicates differences in how respondents prioritise specific concerns. Weak or non-significant correlations between attitudes and knowledge suggest that simply knowing about Responsible AI principles does not strongly predict the degree to which individuals endorse them as necessary. This dissociation supports theoretical models distinguishing between *knowing* what is ethically appropriate and *valuing* it (Ajzen 1991; Blake 1999). Interestingly, only the relationship between privacy knowledge and privacy attitudes showed a statistically significant (albeit small) positive correlation, possibly reflecting heightened public awareness of data protection concerns (Hallinan, Friedewald, and McCarthy 2012; Trepte et al. 2015). This highlights the importance of fostering targeted knowledge in AI ethics, particularly in domains where ethical concerns are more salient to the general public, such as data privacy. Furthermore, our findings suggest that general knowledge about engineering ethics is not strongly related to ethical attitudes toward the highlighted principles, emphasising the potential disconnect between ethical awareness in engineering and specific ethical stance. This suggests that engineering ethics training may not necessarily lead to more favourable ethical stances toward AI technologies.

4.3. Implications for teaching and practice

The scale can be applied in various contexts. First, it can serve as one of the first diagnostic tools in *educational settings* to assess the baseline attitudes of engineering students toward Responsible AI. Educators and curriculum designers can use the results to identify blind spots and tailor teaching interventions accordingly. This can help identify overly negative (or, in the less prominent case, overly positive) attitudes toward Responsible and ethical AI (Martin, Conlon, and Bowe 2021) and to intervene to foster recognition of the topic's complexity. Empirical pre-, inter-, and post-training assessments can then inform the design and evaluation of responsibility initiatives in AI training programmes for engineers (Hess and Fore 2017). Second, the scale can be employed in *applied research* to examine how various factors such as disciplinary engineering background, professional experience, or exposure to moral dilemmas influence attitudes toward Responsible AI. Third, the scale may be helpful in *industry settings* for workforce development and training. It provides a structured way for organisations to assess how practitioners perceive key ethical responsibilities related to AI, thereby informing the design of corporate training programmes, internal guidelines, work procedures, or industry standards (Prem 2023). However, as sketched above, it is essential to note that the scale is not intended to assess actual ethical decision-making competence or behaviour in complex AI-related dilemmas. Nor should it be used in isolation to evaluate ethical maturity or professional integrity. Instead, it should be understood as a complementary tool, providing insights into what individuals believe *should* matter when working with AI.

5. Conclusion

This study introduces and validates the RAISE scale as a new instrument to assess attitudes toward Responsible AI among engineering students and professionals. Drawing from theoretical foundations in Responsible AI (focused on ethics) and adapting an existing scale to measure attitudes toward Responsible AI, we developed a concise, three-factor scale encompassing do-no-harm, transparency, and privacy subscales. The RAISE scale exhibits acceptable psychometric properties, demonstrating reliability across subgroups and providing initial evidence of construct validity. Beyond its methodological contribution, the scale offers practical utility, allowing educators, researchers, and organisations to gauge how engineers value ethical principles in AI contexts. This is particularly relevant in an era when AI technologies are rapidly transforming engineering practices and raising new ethical challenges. Understanding and supporting engineers' moral orientation is a foundational step toward fostering a culture of Responsible AI development and use. Future research should investigate how the RAISE scale performs across cultural contexts and over time, as well as its predictive value in relation to ethical decision-making or behaviour in real-world scenarios. Longitudinal studies could investigate how ethical attitudes evolve throughout academic training or professional experience. Additionally, integrating the scale into intervention studies may help identify which pedagogical strategies most effectively foster Responsible AI engagement among future engineers.

Notes

1. In this paper, we refer to the AI in the broadest scope following the OECD updated working definition: 'An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.' OECD (2024).
2. While the term Responsible AI originated from addressing ethical implications, it has since been watered down. In the context of this paper, we will use the term primarily to address ethical aspects.

Acknowledgments

This work was funded by the Federal Ministry of Education and Research (BMBF) and the Ministry of Culture and Science of the German State of North Rhine-Westphalia (MKW) under the Excellence Strategy of the Federal Government and the Länder. We thank the KI-Campus research fellows for their support and the SEFI community for providing valuable comments during the presentation of earlier work that led to this scale.

Disclosure statement

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

Funding

This work was supported by Ministry of Culture and Science of the German State of North Rhine-Westphalia (MKW); Federal Ministry of Education and Research (BMBF).

Notes on contributor

Marie Mirsch (née Decker), M.Sc., is a research assistant and doctoral candidate at RWTH Aachen University. She conducts research at the intersection of mathematics, ethics, and social sciences, intending to anchor diversity perspectives in technology. Her interdisciplinary research focuses on intersectionality in the context of algorithmic fairness – a central topic of current AI research. She also considers aspects of procedural fairness and participatory approaches, such as including citizens' perspectives. As part of her work at the Research Group 'Gender and Diversity in Engineering', she also deals with ethical issues relating to using artificial intelligence in engineering. A Research Fellowship from the BMBF-funded AI Campus supports her research. As project manager of the Responsible Research and Innovation (RRI) Hub at RWTH Aachen University, she coordinates and implements national and international projects to strengthen social responsibility in technology development as part of the 'ENHANCE – European Universities of Technology Alliance'. Her teaching activities include the seminar 'Responsible AI for Engineers' at RWTH, the online course 'Responsible Innovators for Tomorrow' as part of the ENHANCE alliance, and the workshop 'Responsible AI' at the University of Koblenz.

Nils Knoth, M.Sc., is a doctoral candidate at the University of Kassel researching at the intersection of psychology and artificial intelligence. His work focuses on modelling and assessing AI and digital literacy, and on experimental studies of the competencies needed to interact effectively with AI-based systems in hybrid intelligence contexts, including prompt engineering.

Ben Schultz, M.Sc., holds a bachelor's degree in Business Informatics and a master's degree in Computational Social Systems from RWTH Aachen University. With his interdisciplinary background, his work focuses on ethical, human-centered design and the practical steps towards responsible AI for academia, industry, and the public sector. He is dedicated to developing innovative yet responsible solutions for AI systems.

Dr. Katharina Bata, is team leader of a junior research group in the field of mathematics education at the Karlsruhe Institute of Technology. Her research is on the preparation and teaching of machine learning methods and mathematical modeling in school and university. She holds a PhD in Mathematics Education (2024) and Master of Science in Applied Mathematics from University of Cologne (2017) as well as a Bachelor of Science in Mathematics from TU Dortmund (2015).

Dr. Johannes Schleiss, is a researcher at the Artificial Intelligence Lab of the Otto-von-Guericke University Magdeburg, Germany. His work focuses on developing innovative and applied educational concepts for Artificial Intelligence (AI) and integrating AI technologies into education. He holds a PhD in AI and Education (2025) and Master of Science in Digital Engineering from Otto-von-Guericke University Magdeburg (2020) as well as a Bachelor of Engineering from the Nuremberg Institute of Technology, Germany (2017).

Univ.-Prof. Dr. Carmen Leicht-Scholten, a political scientist by training, is the Director of the RRI Hub at RWTH Aachen University and holds the Chair of Gender and Diversity in Engineering. She is based at the Faculty of Civil Engineering and has a second seat at the Faculty of Arts and Humanities. The Chair for Gender and Diversity investigates, places, and publishes intersectional perspectives on gender in engineering and technology in research and innovation. As a bridging professorship, the chair aims to strengthen interdisciplinary communication and exchange between technology and societal issues. The aim is to integrate gender and diversity perspectives into the broad field of science and technology in research to realise socially responsible innovations and technologies. By integrating these perspectives, the team sensitises engineering students to the social embeddedness of technical artefacts in society and their interdependencies. This enables students to reflect on these perspectives in their future work. She acts as an expert in many national and international research projects and associations and is engaged in various scientific boards as well as in business associations to integrate gendered perspectives in science and technology.

Data availability statement

The dataset supporting the findings of this study is available in the Figshare repository with DOI 10.6084/m9.figshare.30001513. Access is currently private and will be automatically made public upon acceptance of the manuscript for publication. During peer review, the dataset can be accessed via the Figshare link provided to reviewers.

Ethical compliance statement

This study is based on quantitative data collected through an online survey administered to participants in an academic context. Participation was entirely voluntary, and all respondents provided informed consent prior to beginning the survey. The data were collected and analyzed in anonymized form. According to the guidelines of the Interfaculty Ethics Committee of RWTH Aachen University, the nature of this research does not require formal ethical approval. All procedures complied with relevant data protection regulations and ethical standards, ensuring confidentiality and informed participation.

Statement on the use of artificial intelligence tools

The authors used Grammarly® for proofreading the manuscript, limited to correcting grammar and improving spelling. The authors take full responsibility for the content of the manuscript.

ORCID

Marie Mirsch http://orcid.org/0000-0002-9138-231X
Nils Knoth http://orcid.org/0009-0000-9403-2991
Ben Schultz http://orcid.org/0000-0003-4171-0204
Katharina Bata http://orcid.org/0009-0009-5968-4260
Johannes Schleiss http://orcid.org/0009-0006-3967-0492
Leicht-Scholten Carmen http://orcid.org/0000-0003-2451-6629

References

ABET. 2024. Criteria for Accrediting Computing Programs: Effective for Reviews during the 2025–2026 Accreditation Cycle. Effective for Reviews During the 2025–2026 Accreditation Cycle. https://www.abet.org/wp-content/uploads/2024/11/2025-2026_CAC_Criteria.pdf?.

Adel, A., and N. Alani. 2025. "Can Generative AI Reliably Synthesise Literature? Exploring Hallucination Issues in ChatGPT." *AI & SOCIETY* 40 (8): 6799–6812. https://doi.org/10.1007/s00146-025-02406-7.

Ajzen, I. 1985. "From Intentions to Actions: A Theory of Planned Behavior." In *Action Control: From Cognition to Behavior*, edited by J. Kuhl, 11–39. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-69746-3_2.

Ajzen, I. 1991. "The Theory of Planned Behavior." *Organizational Behavior and Human Decision Processes* 50 (2): 179–211. https://doi.org/10.1016/0749-5978(91)90020-T.

Allport, G. W. 1935. "Attitudes." In *Handbook of Social Psychology*, edited by C. Murchison, 798–844. Worcester, MA: Clark University Press.

Azmi, A. D., S. Wahab, N. B. Abdul Kadir, N. A. Abdul Wahab, and R. A. Razak. 2024. "The Development and Validation of a Knowledge, Attitude, and Practice Questionnaire of Methamphetamine Use." *Scientific Reports* 14 (1): 21835. https://doi.org/10.1038/s41598-024-73126-z.

Baghbani, A., T. Choudhury, S. Costa, and J. Reiner. 2022. "Application of Artificial Intelligence in Geotechnical Engineering: A State-of-the-Art Review." *Earth-Science Reviews* 228:103991. https://doi.org/10.1016/j.earscirev.2022.103991.

Bandura, A. 1997. *Self-efficacy: The Exercise of Control. Self-efficacy: The Exercise of Control*. New York: W H Freeman/Times Books/ Henry Holt & Co.

Bao, Y., and H. Li. 2019. "Artificial Intelligence for Civil Engineering; [人工智能时代的土木工程]." *Tumu Gongcheng Xuebao/China Civil Engineering Journal* 52 (5): 1–11. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85072529341&partnerID=40&md5=1c363c3b021678f4e039418f099fa724.

Barocas, S., and A. D. Selbst. 2016. Big Data's Disparate Impact. 104 California Law Review 671. Advance online publication. https://doi.org/10.2139/ssrn.2477899.

Blake, J. 1999. "Overcoming the 'Value-Action Gap' in Environmental Policy: Tensions between National Policy and Local Experience." *Local Environment* 4 (3): 257–278. https://doi.org/10.1080/13549839908725599.

Boateng, G. O., T. B. Neilands, E. A. Frongillo, H. R. Melgar-Quiñonez, and S. L. Young. 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.

Bolte, L., and A. van Wynsberghe. 2025. "Sustainable AI and the Third Wave of AI Ethics: A Structural Turn." *AI and Ethics* 5 (2): 1733–1742. Advance online publication. https://doi.org/10.1007/s43681-024-00522-6.

Burrell, J. 2016. "How the Machine 'thinks': Understanding Opacity in Machine Learning Algorithms." *Big Data & Society* 3 (1): 1–12. https://doi.org/10.1177/2053951715622512.

Colquitt, J. A., and J. B. Rodell. 2015. "Measuring Justice and Fairness." In *The Oxford Handbook of Justice in the Workplace*, edited by R. S. Cropanzano and M. L. Ambrose, 187–202. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199981410.013.8.

Darley, J. M., and C. D. Batson. 1973. ""from Jerusalem to Jericho": A Study of Situational and Dispositional Variables in Helping Behavior." *Journal of Personality and Social Psychology* 27 (1): 100–108. https://doi.org/10.1037/h0034449.

Decker, M., J. Schleiss, B. Schultz, S. G. Moreno, S. Stober, and C. Leicht-Scholten. 2024. "Towards Responsible AI – Competencies for Engineers: An Explorative Literature Review on Existing Frameworks." In Proceedings of the 52nd Annual Conference of SEFI. https://doi.org/10.5281/zenodo.14256815.

Decker, M., L. Wegner, and C. Leicht-Scholten. 2025. "Procedural Fairness in Algorithmic Decision-Making: The Role of Public Engagement." *Ethics and Information Technology* 27 (1): 1–16. https://doi.org/10.1007/s10676-024-09811-4.

Demertzis, K., S. Demertzis, and L. Iliadis. 2023. "A Selective Survey Review of Computational Intelligence Applications in the Primary Subdomains of Civil Engineering Specializations." *Applied Sciences* 13 (6): 3380. https://doi.org/10.3390/app13063380.

Dignum, V. 2020. *Responsible ARTIFICIAL INTELLIGENCE: How to Develop and Use AI in a Responsible Way*. Cham: Springer. https://link.springer.com/book/10.1007978-3-030-30371-6.

Durmaz, A., İ Dursun, and E. T. Kabadayi. 2020. "Mitigating the Effects of Social Desirability Bias in Self-Report Surveys." In *Advances in Library and Information Science (ALIS) Book Series: 2020: 4. Challenges of Globalization and Inclusivity in Academic Research*, edited by A. Ippolito, C. Inglese, and S. Chakraborty, 146–185. Hershey: IGI Global. https://doi.org/10.4018/978-1-7998-1025-4.ch007.

Erasmus, A., T. D. P. Brunet, and E. Fisher. 2021. "What Is Interpretability?" *Philosophy & Technology* 34 (4): 833–862. https://doi.org/10.1007/s13347-020-00435-2.

Fabrigar, L. R., R. E. Petty, S. M. Smith, and S. L. Crites. 2006. "Understanding Knowledge Effects on Attitude-Behavior Consistency: The Role of Relevance, Complexity, and Amount of Knowledge." *Journal of Personality and Social Psychology* 90 (4): 556–577. https://doi.org/10.1037/0022-3514.90.4.556.

Field, A. 2024. *Discovering Statistics Using IBM SPSS Statistics*. 6th edn. London, Southern Oaks, California, New Delhi, and Singapore: Sage Publisher.

Gastwirth, J. L., Y. R. Gel, and W. Miao. 2009. "The Impact of Levene's Test of Equality of Variances on Statistical Theory and Practice." *Statistical Science* 24 (3): 343–360. https://doi.org/10.1214/09-STS301.

Ghasemi, A., and S. Zahediasl. 2012. "Normality Tests for Statistical Analysis: A Guide for Non-Statisticians." *International Journal of Endocrinology and Metabolism* 10 (2): 486–489. https://doi.org/10.5812/ijem.3505.

Ghotbi, N., M. T. Ho, and P. Mantello. 2022. "Attitude of College Students towards Ethical Issues of Artificial Intelligence in an International University in Japan." *AI & SOCIETY* 37 (1): 283–290. https://doi.org/10.1007/s00146-021-01168-2.

Golda, A., K. Mekonen, A. Pandey, A. Singh, V. Hassija, V. Chamola, and B. Sikdar. 2024. "Privacy and Security Concerns in Generative AI: A Comprehensive Survey." *IEEE Access* 12:48126–48144. https://doi.org/10.1109/ACCESS.2024.3381611.

Graham, J. M. 2006. "Congeneric and (Essentially) Tau-Equivalent Estimates of Score Reliability." *Educational and Psychological Measurement* 66 (6): 930–944. https://doi.org/10.1177/0013164406288165.

Grassini, S. 2023. "Development and Validation of the AI Attitude Scale (AIAS-4): A Brief Measure of General Attitude Toward Artificial Intelligence." *Frontiers in Psychology* 14:1191628. https://doi.org/10.3389/fpsyg.2023.1191628.

Guion, R. M. 1977. "Content Validity—The Source of My Discontent." *Applied Psychological Measurement* 1 (1): 1–10. https://doi.org/10.1177/014662167700100103.

Gyory, J. T., N. F. Soria Zurita, J. Martin, C. Balon, C. McComb, K. Kotovsky, and J. Cagan. 2022. "Human versus Artificial Intelligence: A Data-Driven Approach to Real-Time Process Management during Complex Engineering Design." *Journal of Mechanical Design* 144 (2): Article 021405. https://doi.org/10.1115/1.4052488.

Haghighat, R. 2007. "The Development of the Brief Social Desirability Scale (BSDS)." *Europe's Journal of Psychology* 3 (4). https://doi.org/10.5964/ejop.v3i4.417.

Hair, J., R. Anderson, B. J. Babin, and W. C. Black. 2013. *Multivariate Data Analysis: Pearson New International Edition (7. Auflage)*. London: Pearson Education Limited. https://elibrary.pearson.de/book/99.150005/9781292035116.

Hallinan, D., M. Friedewald, and P. McCarthy. 2012. "Citizens' Perceptions of Data Protection and Privacy in Europe." *Computer Law & Security Review* 28 (3): 263–272. https://doi.org/10.1016/j.clsr.2012.03.005.

Häußermann, J. J., and C. Lütge. 2022. "Community-in-the-loop: Towards Pluralistic Value Creation in AI, or—Why AI Needs Business Ethics." *AI and Ethics* 2 (2): 341–362. https://doi.org/10.1007/s43681-021-00047-2.

Hess, J. L., and G. Fore. 2017. "A Systematic Literature Review of US Engineering Ethics Interventions." *Science and Engineering Ethics* 24 (2): 551–583. https://doi.org/10.1007/s11948-017-9910-6.

Holm, S. 1979. "A Simple Sequentially Rejective Multiple Test Procedure." *Scandinavian Journal of Statistics* 6 (2): 65–70. http://www.jstor.org/stable/4615733.

Holzinger, A., K. Zatloukal, and H. Müller. 2025. "Is Human Oversight to AI Systems Still Possible?" *New Biotechnology* 85:59–62. https://doi.org/10.1016/j.nbt.2024.12.003.

Howland, S. J., B. K. Jesiek, S. Claussen, and C. B. Zoltowski. 2024. "Measures of Ethics and Social Responsibility Among Undergraduate Engineering Students: Findings from a Longitudinal Study." *Science and Engineering Ethics* 30 (1): 5. https://doi.org/10.1007/s11948-024-00462-8.

Hu, L., and P. M. Bentler. 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. https://doi.org/10.1080/10705519909540118.

Huang, LT-L, H-Y Chen, Y-T Lin, T-R Huang, T-W Hun,, LT-L Huang,, and T-W Hung,. 2022. "Ameliorating Algorithmic Bias, or Why Explainable AI Needs Feminist Philosophy." *Feminist Philosophy Quarterly* 8 (3/4): 1–31. https://doi.org/10.5206/fpq/2022.3/4.14347.

Jang, Y., S. Choi, and H. Kim. 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. https://doi.org/10.1007/s10639-022-11086-5.

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

Johri, A., E. Lindsay, and J. Qadir. 2023. Ethical Concerns and Responsible Use Of Generative Artificial Intelligence In Engineering Education. https://doi.org/10.21427/0T6R-FZ62.

Kahn, S., and D. Ginther. 2017. Women and STEM. https://doi.org/10.3386/w23525.

Kaplan, D. 1990. "Evaluating and Modifying Covariance Structure Models: A Review and Recommendation." *Multivariate Behavioral Research* 25 (2): 137–155. https://doi.org/10.1207/s15327906mbr2502_1.

Katz, D. 1960. "The Functional Approach to the Study of Attitudes." *Public Opinion Quarterly* 24 (2, Special Issue: Attitude Change): 163. https://doi.org/10.1086/266945.

Kieslich, K., B. Keller, and C. Starke. 2022. "Artificial Intelligence Ethics by Design. Evaluating Public Perception on the Importance of Ethical Design Principles of Artificial Intelligence." *Big Data & Society* 9 (1): 205395172210929. https://doi.org/10.1177/20539517221092956.

Kind, C. 2020. The Term 'ethical AI' Is Finally Starting to Mean Something. VentureBeat, August 23, 2020. https://venturebeat.com/ai/the-term-ethical-ai-Is-Finally-starting-to-mean-Something/.

Kline, R. B. 2016. *Principles and Practice of Structural Equation Modeling*. 4th edn. New York and London: The Guilford Press. https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=4000663.

Kline, T. J. 2005. *Psychological Testing: A Practical Approach to Design and Evaluation*. Thousand Oaks: SAGE Publisher. https://doi.org/10.4135/9781483385693.

Knoth, N., M. Decker, M. C. Laupichler, M. Pinski, N. Buchholtz, K. Bata, and B. Schultz. 2024. "Developing a Holistic AI Literacy Assessment Matrix – Bridging Generic, Domain-Specific, and Ethical Competencies." *Computers and Education Open* 6:100177. https://doi.org/10.1016/j.caeo.2024.100177.

Lee, H-P, Y-J Yang, T. S. von Davier, J. Forlizzi, and S. Das. 2024. "Deepfakes, Phrenology, Surveillance, and More! A Taxonomy of AI Privacy Risks." In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, edited by F. F. Mueller, P. Kyburz, J. R. Williamson, C. Sas, M. L. Wilson, P. T. Dugas, and I. Shklovski, 1–19. New York: ACM. https://doi.org/10.1145/3613904.3642116.

Lind, G. 2019. *How to Teach Moral Competence*. 2nd edn. Berlin: Logos Verlag. http://www.content-select.com/index.php?id=bib_view&ean=9783832588304.

Liu, H., F. Zhang, Z. Chen, and L. Wang. 2024. "Applied Research Status and Prospects of Artificial Intelligence in Civil Engineering Field; [人工智能在土木工程领域的应用研究现状及展望]." *Tumu Yu Huanjing Gongcheng Xuebao/Journal of Civil and Environmental Engineering* 46 (1): 14–32. https://doi.org/10.11835/j.issn.2096-6717.2022.016.

Lovato, J., J. W. Zimmerman, I. Smith, P. Dodds, and J. L. Karson. 2024. "Foregrounding Artist Opinions: A Survey Study on Transparency, Ownership, and Fairness in AI Generative Art." *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* 7:905–916. https://doi.org/10.1609/aies.v7i1.31691.

Lucke, J. F. 2005. "The α and the ω of Congeneric Test Theory: An Extension of Reliability and Internal Consistency to Heterogeneous Tests." *Applied Psychological Measurement* 29 (1): 65–81. https://doi.org/10.1177/0146621604270882.

Malmqvist, J., U. Lundqvist, A. Rosén, K. Edström, R. Gupta, H. Leong, S. M. Cheah, et al. 2022. "The CDIO Syllabus 3.0 – an Updated Statement of Goals." In 18th CDIO international conference proceedings – full papers.

Marsh, H. W., K-T Hau, and Z. Wen. 2004. "In Search of Golden Rules: Comment on Hypothesis-Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu and Bentler's (1999) Findings." *Structural Equation Modeling: A Multidisciplinary Journal* 11 (3): 320–341. https://doi.org/10.1207/s15328007sem1103_2.

Martin, D. A., E. Conlon, and B. Bowe. 2021. "A Multi-level Review of Engineering Ethics Education: Towards a Socio-Technical Orientation of Engineering Education for Ethics." *Science and Engineering Ethics* 27 (5): 60. https://doi.org/10.1007/s11948-021-00333-6.

McFadden, Z., and L. Alvarez. 2024. "Performative Ethics from within the Ivory Tower: How CS Practitioners Uphold Systems of Oppression." *Journal of Artificial Intelligence Research* 79:777–799. https://doi.org/10.1613/jair.1.15423.

Metcalf, J., E. Moss, E. A. Watkins, R. Singh, and M. C. Elish. 2021. "Algorithmic Impact Assessments and Accountability." In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 735–746. New York: ACM. https://doi.org/10.1145/3442188.3445935.

Moreno, S. G., M. Decker, and C. Leicht-Scholten. 2024. "Implementing AI Ethics Education: Engineering Competencies for Sustainable Societal Impact." In *20th International CDIO Conference. Symposium Conducted at the Meeting of ESPRIT*, edited by L. Bettaieb, Ajailia Nadia, J. Bennedsen, K. Edström, Singh, Nivan Mar, R. Lyng, and F. Byström, 643–660. Tunisia: Chalmers University of Technology.

Nachar, N. 2008. "The Mann-Whitney U: A Test for Assessing Whether Two Independent Samples Come from the Same Distribution." *Tutorials in Quantitative Methods for Psychology* 4 (1): 13–20. https://doi.org/10.20982/tqmp.04.1.p013.

Nithyakarpagam, A., and S. P. Mathiraj. 2025. "Exploring the Impact of Artificial Intelligence on Innovation and Efficiency in the Construction Industry." *Journal of Information Systems Engineering and Management* 10 (7s): 643–655. https://doi.org/10.52783/jisem.v10i7s.961.

OECD. 2024. *Explanatory Memorandum on the Updated OECD Definition of an AI System* (OECD Artificial Intelligence Papers No. 8). Paris: OECD Publishing. https://doi.org/10.1787/623da898-en.

Pinski, M., and A. Benlian. 2023. "Ai Literacy – towards Measuring Human Competency in Artificial Intelligence." In *Proceedings of the 56th Annual Hawaii International Conference on System Sciences: January 3-6, 2023*, edited by T. X. Bui, 165–174. Manoa: Department of IT Management Shidler College of Business University of Hawaii. https://hdl.handle.net/10125/102649.

Prem, E. 2023. "From Ethical AI Frameworks to Tools: A Review of Approaches." *AI and Ethics* 3 (3): 699–716. https://doi.org/10.1007/s43681-023-00258-9.

Putnick, D. L., and M. H. Bornstein. 2016. "Measurement Invariance Conventions and Reporting: The State of the Art and Future Directions for Psychological Research." *Developmental Review* 41:71–90. https://doi.org/10.1016/j.dr.2016.06.004.

Randall, D. M., and M. F. Fernandes. 1991. "The Social Desirability Response Bias in Ethics Research." *Journal of Business Ethics* 10 (11): 805–817. https://doi.org/10.1007/BF00383696.

Rosseel, Y. 2012. "lavaan: AnRPackage for Structural Equation Modeling." *Journal of Statistical Software* 48 (2): 1–36. https://doi.org/10.18637/jss.v048.i02.

Ruxton, G. D. 2006. "The Unequal Variance t-Test is an Underused Alternative to Student's t-Test and the Mann–Whitney U Test." *Behavioral Ecology* 17 (4): 688–690. https://doi.org/10.1093/beheco/ark016.

Santoni de Sio, F., and G. Mecacci. 2021. "Four Responsibility Gaps with Artificial Intelligence: Why They Matter and How to Address Them." *Philosophy & Technology* 34 (4): 1057–1084. https://doi.org/10.1007/s13347-021-00450-x.

Schepman, A., and P. Rodway. 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., and P. Rodway. 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/10.1080/10447318.2022.2085400.

Schmidt, A. 2023. "Speeding up the Engineering of Interactive Systems with Generative AI." In *ACM Digital Library, Companion Proceedings of the 2023 ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, edited by A. Dix, M. Winckler, and M. Jones, 7–8. Swansea: Association for Computing Machinery. https://doi.org/10.1145/3596454.3597176.

Schwarz, N. 2007. "Attitude Construction: Evaluation in Context." *Social Cognition* 25 (5): 638–656. https://doi.org/10.1521/soco.2007.25.5.638.

Shrout, P. E. 1998. "Measurement Reliability and Agreement in Psychiatry." *Statistical Methods in Medical Research* 7 (3): 301–317. https://doi.org/10.1177/096228029800700306.

Sindermann, C., P. Sha, M. Zhou, J. Wernicke, H. S. Schmitt, M. Li, R. Sariyska, M. Stavrou, B. Becker, and C. Montag. 2021. "Assessing the Attitude towards Artificial Intelligence: Introduction of a Short Measure in German, Chinese, and English Language." *KI – Künstliche Intelligenz* 35 (1): 109–118. https://doi.org/10.1007/s13218-020-00689-0.

Sloane, M., and E. Moss. 2019. "AI's Social Sciences Deficit." *Nature Machine Intelligence* 1 (8): 330–331. https://doi.org/10.1038/s42256-019-0084-6.

Sörbom, D. 1989. "Model Modification." *Psychometrika* 54 (3): 371–384. https://doi.org/10.1007/BF02294623.

Stein, J-P, T. Messingschlager, T. Gnambs, F. Hutmacher, and M. Appel. 2024. "Attitudes towards AI: Measurement and Associations with Personality." *Scientific Reports* 14 (1): 2909. https://doi.org/10.1038/s41598-024-53335-2.

Stevens, J. P. 2001. *Applied Multivariate Statistics for the Social Sciences*. New York: Psychology Press. https://doi.org/10.4324/9781410604491.

Sureshraj, S. P., N. Duraisamy, R. D. Palaniappan, S. Sureshkumar, M. Priya, J. Britto Pitchai, M. Badcha Yakoob, S. Karthikeyan, G. Sundarajan, and S. Muthuveerappan. 2023. "Recent Innovations in Artificial Intelligence (AI) Algorithms in Electrical and Electronic Engineering for Future Transformations." In *Marvels of Artificial and Computational Intelligence in Life Sciences*, edited by T. Sivaraman, V. Subramanian Thangarasu, and G. Balakrishnan, 9–21. Sharjah: Bentham Science Publishers. https://doi.org/10.2174/9789815136807123010006.

Trepte, S., D. Teutsch, P. K. Masur, C. Eicher, M. Fischer, A. Hennhöfer, and F. Lind. 2015. "Do People Know about Privacy and Data Protection Strategies? Towards the "Online Privacy Literacy Scale" (OPLIS)." In *Law, Governance and Technology Series. Reforming European Data Protection Law*. Vol. 20, edited by S. Gutwirth, R. Leenes, and P. de Hert, 333–365. Dordrecht, NL: Springer Netherlands. https://doi.org/10.1007/978-94-017-9385-8_14.

VDI Verein Deutscher Ingenieure e.V. 2021. Ethische Grundsätze des Ingenieurberufs. www.vdi.de/ethischegrundsaetze.

Walker, M. 2012. "Evaluating the Intervention of an Ethics' Class in Students' Ethical Decision-Making: A Summative Review." *International Education Studies* 6 (1): 10–25. https://doi.org/10.5539/ies.v6n1p10.

Wang, Z., C-S Chai, J. Li, and V. W. Y. Lee. 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): 1–16. https://doi.org/10.1186/s41239-025-00519-z.

Wegner, L., M. Decker, and C. Leicht-Scholten. 2024. "Beyond Distributions: A Systematic Review on Relational Algorithmic Justice." In *Proceedings of the 3rd European Workshop on Algorithmic Fairness (EWAF24)*, edited by Mattia Cerrato, Alesia Vallenas Coronel, Petra Ahrweiler, Michele Loi, Mykola Pechenizkiy, and Aurelia Tamò-Larrieux, 1–2. Aachen: CEUR Workshop Proceedings. http://ceur-ws.org/Vol-3908/#paper_32.

Weinert, F. E. 2001. “Concept of Competence: A Conceptual Clarification.” In *Defining and Selecting key Competencies*, edited by D. S. Rychen and L. H. Salganik, 45–65. Göttingen: Hogrefe & Huber Publishers.
White, R. 1959. “Motivation Reconsidered: The Concept of Competence.” *Psychological Review* 66 (5): 297–333. https://doi.org/10.1037/h0040934.
Xiang, A. 2024. “Fairness & Privacy in an Age of Generative AI.” *Science and Technology Law Review* 25 (2): 288–312. https://doi.org/10.52214/stlr.v25i2.12765.
Yazici, B., and S. Yolacan. 2007. “A Comparison of Various Tests of Normality.” *Journal of Statistical Computation and Simulation* 77 (2): 175–183. https://doi.org/10.1080/10629360600678310.
Yüksel, N., H. R. Börklü, H. K. Sezer, and O. E. Canyurt. 2023. “Review of Artificial Intelligence Applications in Engineering Design Perspective.” *Engineering Applications of Artificial Intelligence* 118:105697. https://doi.org/10.1016/j.engappai.2022.105697.
Yusoff, M. S. B. 2019. “ABC of Content Validation and Content Validity Index Calculation.” *Education in Medicine Journal* 11 (2): 49–54. https://doi.org/10.21315/eimj2019.11.2.6.
Zhai, C., S. Wibowo, and L. D. Li. 2024. “The Effects of Over-Reliance on AI Dialogue Systems on Students’ Cognitive Abilities: A Systematic Review.” *Smart Learning Environments* 11 (1): 1–37. https://doi.org/10.1186/s40561-024-00316-7.
Zhu, Q., C. Zoltowski, M. Feister, P. Buzzanell, W. Oakes, and A. Mead. 2014. The Development of an Instrument for Assessing Individual Ethical Decisionmaking in Project-Based Design Teams: Integrating Quantitative and Qualitative Methods. In 2014 ASEE Annual Conference & Exposition Proceedings (24.1197.1-24.1197.12). ASEE Conferences. https://doi.org/10.18260/1-2–23130.

Appendix

Table A1. Panel of experts in EVR 1 and EVR 2, with respective area of expertise.

Round	Profession	Area of Expertise			
		AI	Engineering	Ethics	Psychology / Behavioural Sciences
EVR 1	Researcher	x			x
	Researcher			x	
	Graduate Student	x			x
	Researcher				x
	Researcher	x			
	Practitioner		x		
	Researcher		x		
	not given	x		x	x
	Researcher	x			x
	Researcher	x			x
	Practitioner		x		
	not given	x		x	
EVR 2	Researcher				x
	Consultant	x			x
	Researcher	x		x	
	Researcher	x			
	Practitioner		x		
	Practitioner		x		
	Practitioner		x		
	not given	x		x	x
	not given	x	x	x	x