Computers in Human Behavior: Artificial Humans 4 (2025) 100132

Contents lists available at ScienceDirect

Computers in Human Behavior: Artificial Humans

journal homepage: www.journals.elsevier.com/computers-in-human-behavior-artificial-humans

A multinational assessment of AI literacy among university students in Germany, the UK, and the US

Marie Hornberger [a,*], Arne Bewersdorff [a], Daniel S. Schiff [b], Claudia Nerdel [a]

[a] Technical University of Munich, Germany
[b] Purdue University, USA

ARTICLE INFO

Keywords:
AI literacy
AI education
Higher education
Item response theory
Artificial intelligence

ABSTRACT

AI literacy is one of the key competencies that university students – future professionals and citizens – need for their lives and careers in an AI-dominated world. Cross-national research on AI literacy can generate critical insights into trends and gaps needed to improve AI education. In this study, we focus on Germany, the UK, and the US given their leadership in AI adoption, innovation, and proactive engagement in AI policy and education. We assessed the AI literacy of 1,465 students across these three countries using a knowledge test previously validated in Germany. We additionally measure AI self-efficacy, interest in AI, attitudes towards AI, AI use, and students' prior learning experiences. Our analysis based on item response theory demonstrates that the AI literacy test remains effective in measuring AI literacy across different languages and countries. Our findings indicate that the majority of students have a foundational level of AI literacy, as well as relatively high levels of interest and positive attitudes related to AI. Students in Germany tend to have a higher level of AI literacy compared to their peers in the UK and US, whereas students in the UK tend to have more negative attitudes towards AI, and US students have higher AI self-efficacy. Based on these results, we offer recommendations for educators on how to take into account differences in characteristics of students such as attitudes towards AI and prior experiences to create effective learning opportunities. By validating an existing AI literacy test instrument across different countries and languages, we provide an instrument and data which can orient future research and AI literacy assessment.

1. Introduction

Artificial intelligence (AI) is gaining importance in our workplaces and daily lives. To navigate in this AI-dominated world, students need to develop *AI literacy* (Long & Magerko, 2020; Ng et al., 2021). AI literacy is increasingly considered to be a key competency for university students to prepare for their future careers (Černý, 2024; Laupichler et al., 2022; Ng et al., 2021). This is not only relevant for students in technical disciplines but for students in all disciplines, as AI is gaining importance in all domains (Lane et al., 2023). For example, in marketing contexts, AI facilitates market analysis and provides predictive insights that support decision-making (Kopalle et al., 2022). In the arts, AI systems contribute to the creation of innovative visual artwork and musical compositions, sparking new forms of artistic expression and collaboration between humans and computers (Ardeliya et al., 2024). Therefore, universities are now faced with the challenge of preparing their students for an AI-dominated workforce and society (D. Lee, Oh, & Hong, 2024; Southworth et al., 2023). Besides AI literacy, related variables that reflect affective orientations towards AI, like AI self-efficacy, interest in AI, and attitudes towards AI, are relevant to consider in higher education as they can play a role in the process of learning about AI (Bewersdorff et al., 2025; Eccles & Wigfield, 2002; Knoth et al., 2024).

Due to the growing importance of AI literacy, many universities have created AI programs and courses to foster AI literacy among all students. To design effective AI courses and programs, understanding the current state of AI literacy among students is an essential precursor (Hornberger et al., 2023). Research on student competencies and potential gaps is needed to inform not only educators but also professional associations, commercial developers, and policymakers on how to improve AI education in universities and beyond. However, much of the existing research is restricted to single settings, such as individual courses, disciplines, or universities, limiting its generalizability and comparability (Kong et al., 2021; e.g., Kong et al., 2022; Y.-J. Lee, Oh, & Hong, 2024). Moreover, there has been minimal multinational research on AI literacy

* Corresponding author. Arcisstrasse 31, 80333, Munich, Germany.
E-mail address: marie.hornberger@tum.de (M. Hornberger).

https://doi.org/10.1016/j.chbah.2025.100132
Received 28 October 2024; Received in revised form 23 February 2025; Accepted 23 February 2025
Available online 25 February 2025

in higher education despite the growing need to diversify our understanding (LaFrance et al., 2024; Nemorin, 2024).

To address this gap, we conducted the first major multinational study to date to examine the state of AI literacy among university students. We focus on three countries – Germany, the United Kingdom (UK), and the United States (US) – that are leaders in AI adoption and innovation and that have been active in considering AI policy and educational responses. All three countries have issued numerous reports on the integration of AI in society (see Cath et al., 2018), though with some differences. For instance, while all three countries consider similar economic, ethical, and social implications of AI, the US and UK approaches tend to emphasize innovation and economic growth, while the EU has advanced more precautionary and ethics-forward regulatory frameworks.

Regarding education in particular, all three countries have begun implementing national AI strategies and initiatives aimed explicitly at enhancing AI education (UK: Department for ScienceInnovation and TechnologyOffice for Artificial Intelligence, 2021; Germany: The Federal Government, 2018; Federal Ministry of Education and Research, 2021; US: U.S. Department of EducationOffice of Educational Technology, 2023), including guidance for educators as well as developers and researchers. Importantly, the socio-cultural, political, and economic landscapes play a major role in their educational systems and approach to AI (Cath et al., 2018; Cave & Dihal, 2023). While a full account of these differences and their interaction with AI literacy is beyond the scope of this paper, this study does offer insights for all three countries. For instance, for educators, insights from this study can help inform curriculum development and instructional strategies in light of their unique educational ecosystems and student needs. Policymakers and international organizations can leverage this cross-national comparison to inform AI education policies and frameworks, aiming to more effectively support student learning and workforce preparation goals.

Finally, much of the existing research is limited to evaluations or considerations of single courses, disciplines, or institutions, leading to test instruments that are only validated within specific contexts. This hinders global comparisons of AI literacy. By providing a validated assessment tool and cross-national data, this study contributes to building a foundation for informed decision-making in AI education and policy.

2. Research questions

To advance these goals in the multinational context, we examine multiple key variables that can be systemized by the ABC (affective, behavioral, cognitive) approach, which has been used by various researchers to study AI education (Knoth et al., 2024; Ng et al., 2024). AI literacy is our main cognitive variable (basic knowledge). We consider AI self-efficacy, interest in AI, and attitudes towards AI as key affective variables. For the behavioral dimension, we examine the use of AI tools and prior AI learning experiences. Critically, while studies have shown that cognitive, affective, and behavioral aspects of interaction with AI are related to one another (Bewersdorff et al., 2025; Gado et al., 2022; Hornberger et al., 2023; Y.-J. Lee, Oh, & Hong, 2024; Stöhr et al., 2024), robust cross-cultural research on these topics is lacking. This study seeks to close that gap by assessing these variables among university students in Germany, the UK, and the US. Further, we aim to investigate if an AI literacy test previously validated on a German sample is useable in cross-national surveys.

Specifically, this study aims to answer the following research questions.

RQ1. Does the English translation of an updated AI literacy test previously validated on a German sample also demonstrate reliability (through measurement accuracy and internal consistency), validity (through construct validity), and fairness (regarding country and gender) in the UK and US?

RQ2. What is the current state of cognitive (AI literacy), affective (AI self-efficacy, interest in AI, attitudes towards AI), and behavioral (use of AI, prior experiences) variables related to AI among students in Germany, the UK, and the US?

RQ3. Are there meaningful cross-country differences in students' AI literacy, AI self-efficacy, interest in AI, attitudes towards AI, AI usage, and prior AI learning experiences?

Addressing these three research questions enables us to do the following: First, we are able to provide a multinational validated AI literacy test that allows various stakeholders to assess (or benchmark) the AI literacy of students across countries. Second, we provide an understanding of some of the affective and behavioral characteristics that may be associated with AI literacy in order to better understand student status and needs. Third, we provide an initial descriptive (though not causal) account of the state of AI literacy and the associated variables across three leading AI countries.

3. Theoretical background

3.1. AI literacy

AI literacy can be defined as "a set of competencies that enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace" (Long & Magerko, 2020). In the context of higher education, AI literacy has been investigated mostly in the context of individual courses. For example, Kong et al. (2021, 2023) have developed and evaluated AI courses for university students with diverse backgrounds. Prior studies on AI literacy in higher education indicate that university students possess little knowledge of AI (Černý, 2024; Hornberger et al., 2023; Teng et al., 2022). To date, there is little research on university students' AI literacy *across different countries*. Korte et al. (2024) evaluated the effectiveness of an AI workshop with students in different countries. However, there is only one study that compares AI literacy levels between countries that is known to the authors: O'Dea et al. (2024) investigated AI literacy among students in the UK and Hong Kong using a self-assessment scale. They found that the country of study and prior learning about AI influenced self-perceived AI literacy.

Furthermore, while multiple assessments of AI literacy have been developed (Lintner, 2024); most are self-assessment scales (Carolus et al., 2023; Laupichler et al., 2023; Lin et al., 2023; Ng et al., 2023; Pinski & Benlian, 2023; Wang et al., 2022). Recently, however, some performance-based tests have also been developed and validated (Hornberger et al., 2023; Soto-Sanfiel et al., 2024; Zhang et al., 2024). However, there are no tests that have been tested for cross-cultural validity (Lintner, 2024), limiting their effective and reliable adoption in multiple countries and hindering high-quality comparison and trend analysis, a common goal prominently prioritized by international organizations and assessments like PISA (OECD, 2024b).

3.2. Affective variables related to AI

Affective variables like AI self-efficacy, interest in AI, and attitudes towards AI also shape students' interactions with AI (Gado et al., 2022). Some conceptualizations of AI literacy take a holistic approach to AI literacy and view affective, behavioral, cognitive, and ethical aspects to be inherent parts of AI literacy (Ng et al., 2024). For example, Knoth et al. (2024) argue that attitudes and self-efficacy are important volitional factors that influence both cognition and behavior. Research in other fields has shown that self-efficacy, interest, and attitudes are related to performance (Bybee & McCrae, 2011; Chang et al., 2014; Eccles & Wigfield, 2002; Krapp & Prenzel, 2011). For instance, previous research has shown that both positive and negative attitudes towards AI are prevalent among students, with a tendency towards positive attitudes (Y.-J. Lee, Oh, & Hong, 2024; Stöhr et al., 2024). Some studies further suggest that interest in AI is relatively high among university students (Almaraz-López et al., 2023; Y.-J. Lee, Oh, & Hong, 2024). Y.-J. Lee, Arnold, et al. (2024) further found that participants have higher confidence in using AI (self-efficacy) than their level of knowledge would suggest. To gain a more complete picture of AI literacy, we include affective variables in this cross-national research on AI literacy.

3.3. Prior experience with AI

Although AI tools have been integrated into everyday technology for several years, their use has often gone unnoticed, perhaps with the exception of autonomous vehicles and smart assistants like Alexa. However, the release of ChatGPT marked a significant shift as it sparked widespread attention on AI, encouraging a more deliberate engagement with AI technologies (Dreksler et al., 2025). Multiple studies conducted after the publication of ChatGPT indicate that use of AI (mostly large language model-based chatbots) is widespread among students (Anthology, 2023; Freeman, 2024; Garrel & Mayer, 2023; Nam, 2023; Stöhr et al., 2024). Studies find that students in technical disciplines report more AI use than students in other disciplines (Garrel & Mayer, 2023; Stöhr et al., 2024). However, it is still unknown what effects this increased interaction with AI has on AI literacy. We might imagine that more use of AI leads to more information-seeking and learning behavior. However, research about "digital natives" reveals that mere *exposure* to a technology does not necessarily lead to competent use of that technology (Margaryan et al., 2011). To gain a broader picture of students' interactions with AI, we are thus interested in how the *use of AI* might differ between countries.

Besides interaction with AI tools, formal and informal learning experiences about AI are also likely to influence AI literacy, AI use, and AI-related affective variables. As universities start to introduce more AI courses for students of all disciplines (Southworth et al., 2023), we expect that a growing number of students in non-technical disciplines will have attended AI courses along with partaking in various informal learning opportunities (O'Dea et al., 2024). Prior learning experiences can have an impact on various AI-related variables. For example, Y.-J. Lee, Arnold, et al. (2024) found that students with more prior learning experiences have higher knowledge and higher confidence (AI self-efficacy) and more awareness of the negative effects of AI (negative attitudes). Along these lines, low levels of skill in mathematics or computer science can be barriers to AI literacy (Kong et al., 2023; Long & Magerko, 2020). Therefore, we examine *prior learning experiences about AI* to compare the prerequisites of students across the three countries.

This paper emphasizes psychometric validation and a descriptive multinational comparison of AI literacy and related affective and behavioral variables. For a comprehensive theoretical framework and analysis connecting the introduced variables resulting in a path model, see our publication on the influences of AI literacy, attitudes towards AI, interest in AI, and AI use on AI self-efficacy (Bewersdorff et al., 2025).

4. Method

4.1. Sample

We conducted a cross-sectional and multinational study[1] in three countries, recruiting university students (undergraduate and graduate) from Germany, the UK, and the US through the sample provider Prolific. Participation in the study was voluntary, took approximately 23 min on average, and participants received 4.50£ (equal to $6, and 5.40€, respectively) as compensation. The study was conducted in December 2023. A total of 1558 students from more than 512 universities participated. Before analyzing the data, we excluded 93 invalid cases.[2] The final sample size was $N = 1{,}465$ with $N_{US} = 494$, $N_{UK} = 499$, and $N_{GER} = 472$.

The mean age of the participants was $M = 28.4$ ($SD = 10.3$). Among the participants, 747 students (51.0%) were male, 680 (46.4%) were female, 33 (2.3%) were non-binary, and 5 (0.3%) preferred not to disclose their gender. Regarding educational level, 1010 students were pursuing a bachelor's or similar degree (68.9%), 393 (26.8%) were enrolled in a master's or equivalent program, and 61 (4.2%) were participating in other programs, such as graduate or Ph.D. studies.

This study examines students from three countries with three distinct educational contexts. The current state of students' AI literacy can inform us on how education and policy can be improved to support national goals like competitiveness and innovation. As the status quo is relevant for these stakeholders, the samples do not necessarily need to be balanced between countries in this research. Further, as discipline has been shown to be a relevant predictor of AI literacy (Hornberger et al., 2023), we display the composition of disciplines by country in Fig. 1, which indicates a higher representation of Engineering and Technology students in Germany and a higher proportion of Medicine and Health and Social Science students in the US and UK.

4.2. Study design

The study is based on previous research on AI literacy in Germany (Hornberger et al., 2023). The survey was translated and augmented with additional questions on AI-related affective variables, use of AI, prior AI learning experiences, and socioeconomic/demographic characteristics. The complete list of questions can be seen in Appendix A.

4.2.1. AI literacy

To assess AI literacy, we adapted a test previously validated with 1,286 students in Germany (Hornberger et al., 2023). In the process of translation, four of the original German items were altered to improve readability, clarity, and equivalence with the English version (see Appendix B). Another key question we addressed is the validity given the dynamic nature of AI technology; the original instrument was developed and tested in 2022 before the release of ChatGPT. With expert feedback, we therefore updated 12 items to make question wordings and answer choices more 'immune' to technological change (see Appendix B). Importantly, the test, built on the competencies by Long and Magerko (2020), focuses on fundamental AI concepts, such as machine learning principles, rather than popular trends or specific AI applications. This approach facilitates the assessment's ongoing relevance even as AI technologies, popular use cases, or terms-of-art continue to evolve. The final test consists of 30 items, 29 of which are multiple-choice items and one of which is a sorting task. An overview can be found in Table 1, and the full-text version can be obtained from the supplementary material (Appendix A).

4.2.2. Affective variables

To measure affective variables – AI self-efficacy, interest in AI, and attitudes towards AI – we used scales previously validated by Hornberger et al., 2023. The response format of all scales consisted of a 5-point Likert scale from 1 = strongly disagree to 5 = strongly agree. To measure self-efficacy, we used a scale consisting of eight items (e.g., "I have a good understanding of the basic principles of AI."). To measure interest in AI, we used a scale adapted from Mang et al. (2019). The scale

[1] This study is part of a broader research project. Bewersdorff et al. (2025) analyze the same data to investigate how cognitive, affective, and behavioral variables related to AI build AI self-efficacy.

[2] Fourteen participants were excluded because they failed both attention checks, and an additional 75 cases were excluded because they did not finish the survey. Lastly, 3 participants were excluded because they completed the survey too quickly (less than 5 min, $Mdn = 19.1$), which made it implausible that they responded thoughtfully.

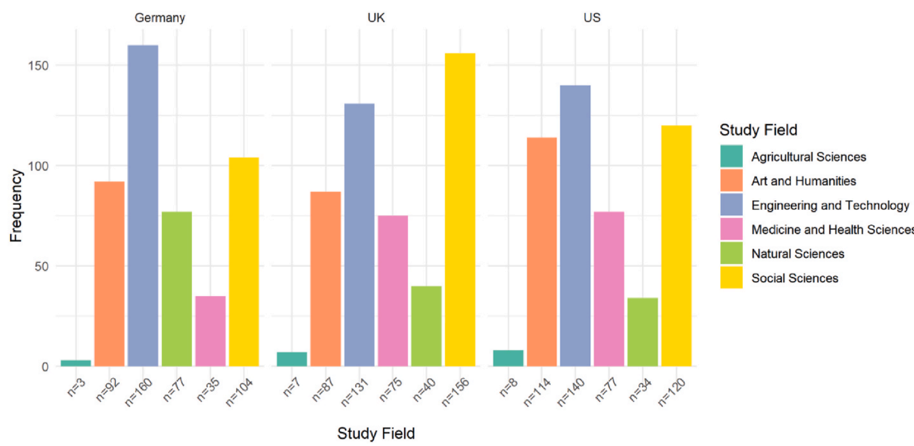

Fig. 1. Proportion of students from different disciplines by country.

Table 1
AI literacy item overview.

Competency (Long & Magerko, 2020)	Items
Recognizing AI	Typical applications (01), Recognizing a chatbot (02)
Interdisciplinarity	AI systems (03), Interdisciplinary research fields (04)
Understanding Intelligence	Intelligence of AI (05), Intelligence of AI 2 (06)
General vs. Narrow	Weak and strong AI (07), Capabilities of weak AI (08)
AI's Strengths & Weaknesses	Superiority of AI (09), Superiority of humans (10)
Representations	Knowledge representations 1 (11), Knowledge representations 2 (12)
Decision-Making [of AI]	Decision-making (13), Optimization (14), Supervised and unsupervised learning (15)
ML Steps	Iterative process (16), Steps in supervised learning (17)*, Training and test data (18)
Human Role in AI	Human influence (19), Human influence 2 (20)
Programmability	Programmability (21)
Data Literacy	Visualization of data (22)
Learning from Data	Learning from data (23), Learning form user data (24)
Critically Interpreting Data	Representativeness of data (25)
Ethics	Ethical principles (26), Black box (27), Societal challenges (28), Risks of AI (29), Legal challenges (30)

consists of five items (e.g., "I generally have fun when I am learning about Artificial Intelligence."). We assessed attitudes towards AI with eight items adapted from the General Attitudes towards AI scale by Schepman and Rodway (2020). We used four items measuring positive attitudes (e.g., "There are many beneficial applications of Artificial Intelligence.") and four items measuring negative attitudes (e.g., "I think Artificial Intelligence is dangerous.").

4.2.3. Prior experience with AI

To assess prior experience with AI, students were asked to indicate if they are currently studying computer science or have studied computer science in the past and if they had taken a course on AI. In addition, they were asked to indicate how often they learn about AI in their free time (never/once or twice a year/less than once a month/once or twice a month/once a week/two or three times a week/every day) to indicate their level of informal engagement or learning with respect to AI. Furthermore, participants rated how much their coursework addresses issues in computer science and technology and estimated their experience in each of computer science, programming, and ethics (none, some, a lot). To assess usage of AI, we asked for the frequency (never, rarely, from time to time, often, very often) of using AI for education and, separately, in their private lives (i.e., besides education).

4.2.4. Sociodemographic variables

We assessed the following sociodemographic background variables: age, gender, pursued degree, and parents' highest educational degree.

4.3. Analysis

4.3.1. Assessing the validity of the AI literacy test

As a first step, we modeled the response patterns of the 30 items in the AI literacy test using item response theory (IRT), a framework widely used in psychometrics for assessing the relationship between individuals' latent traits and their item responses (Embretson & Reise, 2009). One of the key strengths of IRT is that it allows for item-level analysis, enabling more precise measurement and better handling of diverse item properties. This advantage is particularly useful in the context of our study as it enables analysis of how different items function across varying populations. To assess whether the test exhibits consistent properties across the three countries, we applied IRT models to four distinct subsets of our data: the full dataset as well as Germany-only, UK-only, and US-only subsets.

We chose to fit the 1-Parameter Logistic (PL), 2-PL, and 3-PL models using the *mirt* package (Chalmers, 2012) in *R* (R Core Team, 2022). The 1-PL model (also known as Rasch model) was included to evaluate item difficulty across the different subsets, providing a baseline for comparison. The 2-PL model allowed us to further examine item discrimination, giving insight into how well each item differentiates between individuals with varying levels of the underlying trait. Finally, the 3-PL model was used to account for guessing behavior, which is particularly relevant for multiple-choice formats. We used a fixed guessing parameter ($g = .25$), which corresponds to a guessing rate of 25% due to the one-out-of-four multiple-choice response format. We verified the assumptions of unidimensionality and local independence that need to be fulfilled to use the chosen IRT models.[3]

To determine which of the three IRT models best describes the data, we compared multiple model fit and item fit indices. Model fit assesses the degree to which a statistical model accurately captures the underlying data structure. Item fit assesses how well individual item responses align with model predictions, ensuring that each item functions as

[3] To test unidimensionality, we fitted a unidimensional model using confirmatory factor analysis with the package *lavaan* (Rosseel, 2012) for *R* (R Core Team, 2022). We evaluated the fit of this unidimensional model using the following commonly used indices and cutoff values that would indicate a unidimensional structure of the response patterns: $X^2/df < 2$–3; RMSEA $<.05$; SRMR $<.1$ (Backhaus et al., 2015; Brown, 2015). We evaluated the assumption of local independence using the Q3 statistic (Yen, 1984) and examined if the correlation of the residuals of a pair of items is higher than the cutoff value .2 (Chen & Thissen, 1997).

expected in measuring the latent trait. We evaluated model fit using the M_2 statistic (Maydeu-Olivares & Joe, 2006), a limited information statistic suitable for sparse data. We used the cutoff $\leq .05$ for RMSEA and SRMR suggested by Maydeu-Olivares (2013). Additionally, we examined the TLI and CFI indices using $\geq .95$ as the cutoff. To evaluate item fit, we used the signed chi-squared (S - X_2) index proposed by Orlando and Thissen (2003). This item fit statistic is based on response patterns rather than on sum statistics and is therefore also suitable for IRT models other than the 1-PL model.

4.3.2. Differential item functioning

To ensure that the test measures fairly across the three countries and is therefore valid for cross-national comparison, we analyzed the items with regard to differential item functioning (DIF; Fischer & Karl, 2019). Items with DIF "function differently" between populations, meaning they are easier for one group to solve compared to another, even when individuals from both groups possess the same underlying ability. Uniform DIF occurs when the difference in item performance between groups is consistent across all levels of the underlying trait (in this study: AI literacy), while non-uniform DIF occurs when this difference varies depending on the level of the trait. DIF-affected items thus undermine the accuracy of measurement and can result in misleading conclusions based on biased test results.

In the present study, we investigated DIF with regard to country, given our purpose of comparing results between the three investigated countries, drawing valid conclusions, and creating a valid multinational, multi-lingual instrument. Additionally, we examined gender-related DIF, a common concern in test development, including for STEM contexts (Le, 2009; Liu & Wilson, 2009). To evaluate DIF, we used a logistic regression method (Fischer & Karl, 2019; Swaminathan & Rogers, 1990), and tested for uniform and non-uniform DIF based on IRT estimations using the package *lordif* (Choi et al., 2011). As significance rates can be inflated with increasing sample size (Zumbo, 1999), we used McFadden's pseudo R^2 as a magnitude measure with .035 as the cutoff (Jodoin & Gierl, 2001). Significant results would indicate that an item is easier or harder to solve for one group independent of a person's ability. Items flagged for DIF are removed from further analyses and from the resulting instrument.

4.3.3. Validation of IRT models for affective variables

For AI self-efficacy, interest in AI, and attitudes towards AI, we also conducted an IRT analysis using rating scale models (Andrich, 1978) due to the Likert-type response format. To determine if the models fit well, we looked at reliability estimations and checked whether item fit (infit/outfit) was in an acceptable range of 0.5–1.5 (Linacre, 2002).

4.3.4. Cross-country comparisons

To answer RQ3, we compared the levels of AI literacy, AI self-efficacy, interest in AI, and attitudes towards AI between countries using the IRT scores. We conducted Anova and Tukey multiple comparisons of means with a 95% family-wise confidence level. To account for differences in the composition of the country subsets, we added the following additional predictors to the model to function as control variables: age, gender, pursued degree, parents' degree, and discipline.

5. Results

5.1. Cross-national validation of the AI literacy test

5.1.1. Descriptive item statistics

On average, the participants correctly solved $M = 15.03$ ($SD = 4.82$) of the 30 items. Based on their responses, we calculated indices for item difficulty and discrimination. Item difficulty refers to the proportion of participants that solved an item (with high values translating to low difficulty). A test should ideally capture a wide range of difficulties between .05 and .95 to reliably assess a broad spectrum of ability levels without ceiling or floor effects. Item discrimination describes the correlation of an item with the whole test and indicates how well the test distinguishes between persons with different levels of AI literacy. Item discrimination should be positive and is optimally between .4 and .7 (Moosbrugger & Kelava, 2020). Note that because of the multiple-choice format, it can be expected that, on average, participants should solve approximately 7.5 items by guessing alone. Therefore, in addition to calculating a simple difficulty index for each item, we also construct an adjusted item-level index accounting for the 25% guessing probability. Table 2 shows descriptive statistics for all used AI literacy items. Item difficulty (corrected for guessing) ranged from .03 to .78, representing a wide range of difficulty. Item discrimination varied from .11 to .50, indicating adequate discrimination properties. However, some items fall below the optimal range, which may slightly reduce the test's ability to differentiate between participants.

Table 2
Descriptive item statistics for the AI literacy items.

Item	Item label	Difficulty index	Difficulty index corrected for guessing	Discrimination index
01	Typical applications	.37	.17	.38
02	Recognizing a chatbot	.58	.44	.30
03	AI systems	.39	.18	.21
04	Interdisciplinary research fields	.61	.48	.43
05	Intelligence of AI	.68	.58	.34
06	Intelligence of AI 2	.83	.77	.27
07	Weak and strong AI	.31	.08	.41
08	Capabilities of weak AI	.46	.28	.43
09	Superiority of AI	.41	.21	.42
10	Superiority of humans	.28	.04	.20
11	Knowledge representations	.29	.06	.24
12	Knowledge representations 2	.28	.05	.32
13	Decision-making	.39	.19	.30
14	Optimization	.46	.27	.25
15	Supervised and unsupervised learning	.32	.09	.34
16	Iterative process	.49	.32	.40
17	Steps in supervised learning	.43	.24	.45
18	Training and test data	.44	.25	.34
19	Human influence	.27	.03	.31
20	Human influence 2	.52	.36	.27
21	Programmability	.84	.78	.37
22	Visualization of data	.55	.39	.40
23	Learning from data	.69	.59	.40
24	Learning form user data	.51	.35	.35
25	Representativeness of data	.81	.74	.42
26	Ethical principles	.45	.27	.11
27	Black box	.51	.35	.42
28	Societal challenges	.79	.72	.43
29	Risks of AI	.62	.50	.50
30	Legal challenges	.44	.25	.26

5.1.2. IRT model selection

As outlined in section 4.3, we fitted three IRT (item response theory) models – 1-PL, 2-PL, and 3-PL – to four different subsets of the data: Germany, the UK, the US, and the entire dataset, after evaluating the assumptions of unidimensionality and local independence.[4] We evaluated multiple model fit and item fit indices to evaluate how well each model fits the data and determine which model provides the best fit overall.

Table 3 shows that for the German subset, the 3-PL model fits the data well and best overall, with only the TLI index slightly missing the cutoff and with the BIC slightly favoring the 1-PL model. In the UK dataset, similarly, while no model fulfills all fit indices, and the AIC and BIC suggest different models, the 3-PL model is closest to meeting the cutoffs. For the US subset, all indices except SRMSR indicate a good fit and best fit with the 3-PL model. For the entire dataset, the 3-PL model narrowly misses the CFI and TLI cutoffs, yet both AIC and BIC suggest it as the best model. Considering all model fit indices, the 3-PL model appears to be the most suitable choice across all countries and the entire dataset.

We also examined item fit using the signed chi-squared (S - X^2) index (Orlando & Thissen, 2000). Due to multiple testing, the *p*-values were corrected for false discovery rate using the method proposed by Benjamini and Hochberg (1995). Detailed tables for all four subsets evaluated with respect to all three IRT models can be viewed in Appendix C. Again, the 3-PL model was the best-fitting model, with only four significant items in the entire dataset (indicating lack of fit) and no significant items in each country subset. Therefore, we used the 3-PL model to calculate and compare parameter estimations.

5.1.3. Testing fairness (DIF analysis)

To investigate if the test functions similarly (or 'fairly') across countries and genders, we conducted a DIF analysis as outlined in 4.3. Comparing the three countries, two items were flagged for uniform DIF, indicating that the difficulty of these items varies depending on the participants' country of residence. As shown in Appendix C, the German subset has a higher probability of answering item 12 (R^2 = .047) correctly, and the UK and US subsets have a higher chance of solving item 26 (R^2 = .055). Both R^2 values indicate a moderate magnitude of DIF (<.070). We therefore excluded both items from the following analyses and the revised AI literacy test published with this paper. To examine potential DIF regarding gender, we performed the same analysis for a subset with only participants who identified as female or male (N = 1,427). This analysis yielded no DIF items, indicating the robustness of the AI literacy test regardless of gender.

5.1.4. Final model fit and estimation of person ability scores

In the final model, we removed items 12 and 26 due to identified DIF (see section 5.1.3) and then refitted the 3-PL model (parameter estimations can be found in Appendix C). We calculated person ability estimates, representing each individual's AI literacy score based on the IRT model. We examined multiple reliability measures, all of which fell within an acceptable range: marginal reliability was .76, EAP reliability was .77, and Cronbach's α was .74. Fig. 2 shows a Wright Map, which presents item difficulties against the distribution of person abilities. The Wright Map illustrates that the items are fairly evenly distributed across the range of person ability (M = 0.00, SD = 0.88), with item 6 (intelligence of AI 2) being the easiest and item 3 (AI systems) the most difficult. However, there are fewer items at the extremes of person ability, which may reduce measurement accuracy at these levels (i.e., for extremely high or extremely low performers). As can be seen in Fig. 3, AI literacy is normally distributed in the overall sample as well as the three different country subsamples. However, the peak of the AI literacy distribution seems to be higher in Germany than in the other countries.

5.2. Affective variables

We calculate IRT scores for the affective variables using fitted rating scale models (Andrich, 1978). All scales demonstrated good reliability (Cronbach's α: .81 - .91; EAP reliability: .81 - .90). Also, the infit and outfit measures for all items of all variables fell within the required range (>0.5 and <1.5). Table 4 shows descriptive statistics for the updated AI literacy test and the affective scales. Compared to the midpoint of the Likert scale (3), participants show relatively high AI self-efficacy, interest in AI, and positive attitudes towards AI. Negative attitudes towards AI are slightly below the midpoint of the scale.

5.3. Cross-country comparison

As outlined in 4.3, we compared the levels of AI literacy, AI self-efficacy, interest in AI, and attitudes towards AI across countries. We used IRT scores instead of raw mean scores as they provide a more exact measurement. A depiction of the distribution of each variable by country with an indication of significant differences can be found in Fig. 4. The means and standard deviations of all variables by country can be viewed in Appendix C.

The effect of country on AI literacy was significant (F(2,1431) = 75.74, p < .001). The differences between Germany and the UK (diff = -0.50) and Germany and the US (diff = -0.61) were statistically significant (ps < .001), while the difference between the UK and the US is not statistically significant (diff = -0.11, p = .081). Next, we found that the effect of country on AI self-efficacy was significant (F(2,1431) = 11.02, p < .001). More specifically, German students had a significantly lower self-efficacy score than the UK (diff = 0.24, p = .002) or US students (diff = 0.32, p < .001). The difference between the UK and the US was not significant (diff = 0.08, p = .475). Furthermore, the effect of country on interest in AI was significant (F(2,1431) = 4.66, p = .010). More specifically, US students had a significantly higher interest in AI score than the UK (diff = 0.37, p = .015) or German students (diff = 0.33, p = .036). The difference between Germany and the UK was not significant (diff = -0.04, p = .960).

The effect of country on positive attitudes towards AI was significant (F(2,1431) = 16.45, p < .001). All inter-country comparisons were statistically significant. German students had a higher score of positive attitudes than UK students (diff = -0.55, p < .001) and US students (diff = -0.31, p = .003). Students in the US had a higher score of positive attitudes than students in the UK (diff = 0.23, p = .038). The effect of country on negative attitudes towards AI was significant (F(2,1431) = 7.31, p < .001). The UK students had significantly higher negative attitudes than the German (diff = 0.23, p = .010) or US students (diff =-0−.28, p < .001). The difference between Germany and the US was not significant (diff = -0.05, p = .799).

In summary, students in Germany demonstrate higher AI literacy and more positive attitudes towards AI compared to students in the UK and US. In contrast, students in the US exhibit higher AI self-efficacy and interest in AI than their counterparts in the UK and Germany. Furthermore, students in the UK show the highest level of negative attitudes towards AI. Fig. 5 illustrates the mean IRT scores for these variables across each country.

Next, we compare the prior experience of students between the three countries. Fig. 6 shows how often students use AI in their education and in their personal lives. Students in Germany state that they use AI as part of their education more frequently (43.8% use AI often or very often) than students in the UK (33%) and the US (29.5%). Meanwhile, there seem to be only small differences in self-reported use of AI in students' personal lives across these three countries.

[4] We verified the assumption of unidimensionality by performing a confirmatory factor analysis with a one-factor model for the whole dataset as well as for the country-specific subsets. For all datasets, the model fitted the data well (see Appendix C). Therefore, we can say that the assumption of unidimensionality is fulfilled, implying that the data can be modeled by unidimensional IRT models. We verified the assumption of local independence for all subsets using the Q3 statistic (Yen, 1984) based on the 3 PL model. Therefore, we examined the correlation between the residuals of all items. Since all correlations were smaller than .2, local independence was not violated.

Table 3
Model fit indices.

Subset	Model	M_2	RMSEA	SRMSR	TLI	CFI	AIC	BIC
Ger	1-PL	692	.036	.065	.860	.860	16834	16963
	2-PL	535	.026	.047	.924	.930	16746	16995
	3-PL	496	.022	.048	.947	.951	16730	16979
UK	1-PL	778	.040	.065	.790	.791	18194	18324
	2-PL	602	.031	.048	.872	.880	18081	18333
	3-PL	532	.025	.051	.917	.923	18093	18346
US	1-PL	838	.043	.069	.797	.798	18198	18328
	2-PL	610	.032	.049	.890	.897	18052	18304
	3-PL	498	.022	.051	.950	.953	18030	18282
All Data	1-PL	1635	.044	.061	.818	.818	53898	54062
	2-PL	1004	.032	.036	.903	.909	53403	53721
	3-PL	803	.026	.038	.935	.940	53347	53664

Note. RMSEA = root mean square error of approximation; SRMSR = standardized root mean squared residuals, TLI = Tucker–Lewis index; CFI = comparative fit index; AIC = Akaike information criterion; BIC = Bayesian information criterion.

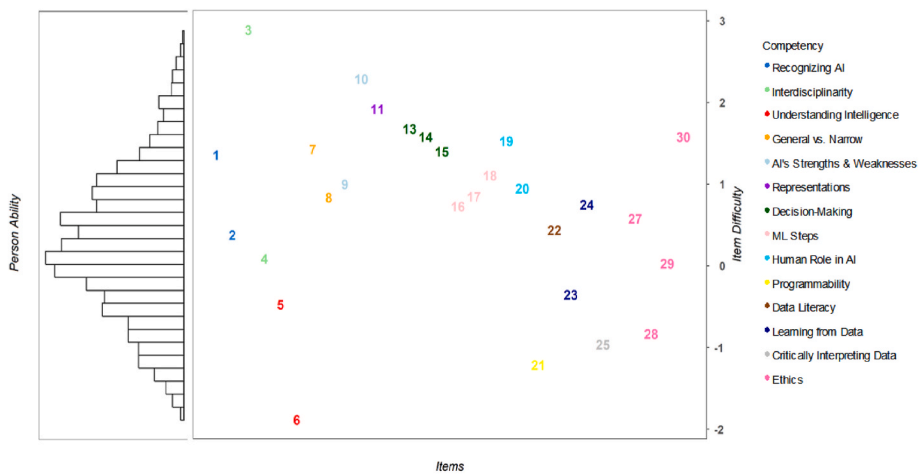

Fig. 2. Wright map.

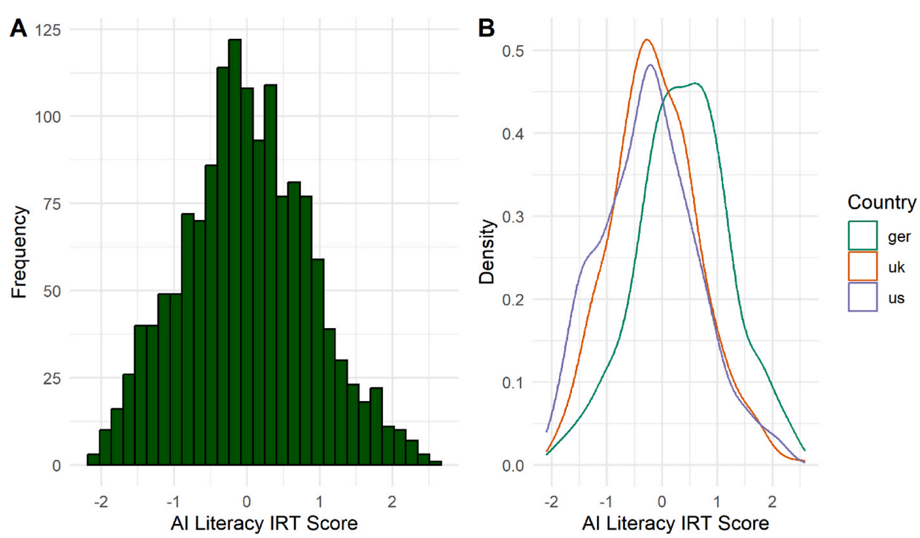

Fig. 3. AI literacy IRT score distribution overall (A) and by country (B).

Fig. 7 shows that more students in the US (39.3%) report that they have studied computer science (CS) than have students in Germany (30.1%) or the UK (31.9%). 26.5% of the German students have attended a course about AI in comparison to only 15.8% of UK and 14.0% of US students. Meanwhile, two-thirds of students in the US (65.8%) learn about AI in their leisure time at least once a month, followed by Germany (61.5%) and UK students (56.1 %) (see Fig. 8).

Fig. 9 shows the participants' prior coursework-based experience with technology, computer science, programming, and ethics. German students indicate more experience in computer science and technology,

Table 4
Descriptive statistics for cognitive and affective scales.

	M	SD
AI literacy	15.0	4.8
AI self-efficacy	3.8	0.6
Interest in AI	3.7	0.8
Positive attitudes towards AI	3.9	0.7
Negative attitudes towards AI	2.7	0.9

Note. Range AI literacy: 0–28; other scales: 1-5.

and programming compared to students in the UK or the US. However, the UK and US students report a higher relation of their coursework to ethics.

6. Discussion

The aim of our study was twofold: to assess the current state of AI literacy and surrounding variables among university students in Germany, the UK, and the US; and to extend an AI literacy test for use in cross-national studies. Specifically, we aimed to evaluate if an AI literacy test previously validated in Germany is robust enough to extend to multinational and multilingual settings. To that end, we modified and revalidated the instrument, including assessing cross-national validity.

Our findings indicate that the revised AI literacy test is suitable for cross-national comparisons, and holds up well when comparing between the 2022 version and 2023 version used in this study. Next, when comparing AI literacy, affective variables, and prior experience, we found significant differences between the three countries. Overall, German students have a higher level of AI literacy, which might be due to more learning experiences with AI. This suggests that AI programs at universities do have a positive effect on students' AI literacy.

6.1. Discussion of the cross-national validation

In this study, we used an updated version of an AI literacy test created in 2022, modifying some questions and answer choices, and creating an English version of the test. Our analysis demonstrates that the IRT model used in the original validation of the test, the 3-PL model, is applicable to both the international sample and the country-specific subsamples in the present study. This finding confirms that both language versions of the test have similar properties in the three investigated countries, supporting their viability for international use and comparison as well as extension to still other countries and languages in the future. Notably, we identified two items affected by country-specific DIF, which were subsequently removed from further analyses. We therefore recommend using this updated test for international German and English-speaking samples. Reliability estimates are acceptable (Tavakol & Dennick, 2011) but slightly lower than in the original validation. As reliability is not a fixed property of a test, but dependent on the sample, this is likely due to the more heterogeneous sample in the present study (Streiner, 2003).

To the best of our knowledge, the present study is the first cross-cultural validation of an AI literacy test (Lintner, 2024). Although this validation is limited to three countries, the publication of a validated English version of the test can enhance its international use and facilitate further research. As our study was conducted across multiple universities in three countries, the findings can be used to inform future multinational research and benchmarking efforts.

6.2. Understanding the level of AI literacy and related variables in Germany, the UK, and the US

On average, the students correctly solved around half of the questions of the AI literacy test. This suggests a generally low level of AI literacy and highlights a gap between current knowledge and the goal of achieving a more comprehensive understanding. This finding aligns with previous research that found minimal knowledge of AI among students (Černý, 2024; Hornberger et al., 2023; Teng et al., 2022). Interestingly, the level of AI literacy is similar to research done *prior* to the release of ChatGPT (Hornberger et al., 2023). This might imply that more interaction or awareness does not automatically lead to higher literacy, which is consistent with the phenomenon of "digital natives" (Margaryan et al., 2011).

Our findings indicate that students in Germany demonstrate higher AI literacy than those in the UK and the US. Based on our data, this may be due to more prior learning experience with AI. For instance, a greater number of German students report having attended courses on AI than to their counterparts in the UK or US. Additionally, they indicate having more experience in AI, computer science, and programming, as well as a more frequent active use of AI in their education. This aligns with prior research that highlights the importance of course participation for AI literacy (Y.-J. Lee, Oh, & Hong, 2024; O'Dea et al., 2024). Moreover, German students show greater interest in AI and more positive attitudes towards AI compared to students in the UK and US. Students with a higher interest in AI and more positive attitudes may seek more opportunities to learn about AI, thereby enhancing their AI literacy (Bybee & McCrae, 2011; Krapp & Prenzel, 2011). However, it is also possible that students with higher literacy have higher interest and positive attitudes (Kim, 2023).

Importantly, the results in the study are best understood as providing a descriptive account on cognitive (AI literacy), affective (AI self-efficacy, interest in AI, attitudes towards AI), and behavioral (AI use and prior learning experiences) variables, and how this aggregates at the level of countries, rather than a full causal account. This study did not measure nor can it fully account for the wide array of socio-cultural, political, demographic, infrastructural, and other factors that may play critically important roles in explaining AI literacy within and across countries. Indeed, differences between countries may reflect numerous social, political, economic, demographic, and cultural differences, most of which cannot be fully addressed within the context of any study alone. Countries may, for instance, devote a greater allotment of support to higher education or AI-related courses, in absolute or relative terms. They may culturally hold greater or lesser value with respect to the role of education, respect accorded to educators, or the importance of topics like AI. Countries also have different common educational and labor market pathways, such as how many years of education are needed before students specialize. The demographic composition of students differs across countries as well, such that comparisons of whole student populations are not straightforward. These topics and their relationship with AI literacy are all worthy of future research, and can be approached across different levels of scale (e.g., individual schools, municipalities, or countries) using methods like process tracing, econometric analysis, and interviews. Overall, however, the observed differences in this study are not too large, which can be explained by cultural (Cave & Dihal, 2020), economic (OECD, 2024a), and political (Cath et al., 2018) similarities in the three investigated countries.

Another finding is that, despite lower levels of literacy, the level of AI self-efficacy is relatively high, which fits with Y.-J. Lee, Arnold, et al. (2024)'s finding of high confidence. US students report the highest level of AI self-efficacy, although they do not have the highest level of AI knowledge. This disconnect confirms that, although AI literacy and AI self-efficacy are modestly correlated, high knowledge does not necessarily translate to high self-efficacy, and vice versa (for a detailed analysis of the relationship between AI literacy, AI self-efficacy, and other variables, see Bewersdorff et al., 2025). This result may be due to cultural differences, such as a tendency among US students to make more optimistic self-assessments (Klassen, 2004) or due to the phenomenon that learners tend to be overconfident in newly learned skills (Dunning et al., 2004), which has also been found in the field of AI (Y.-J. Lee, Oh, & Hong, 2024). More research is needed to understand this phenomenon and how to address it in the field of AI to enable people to

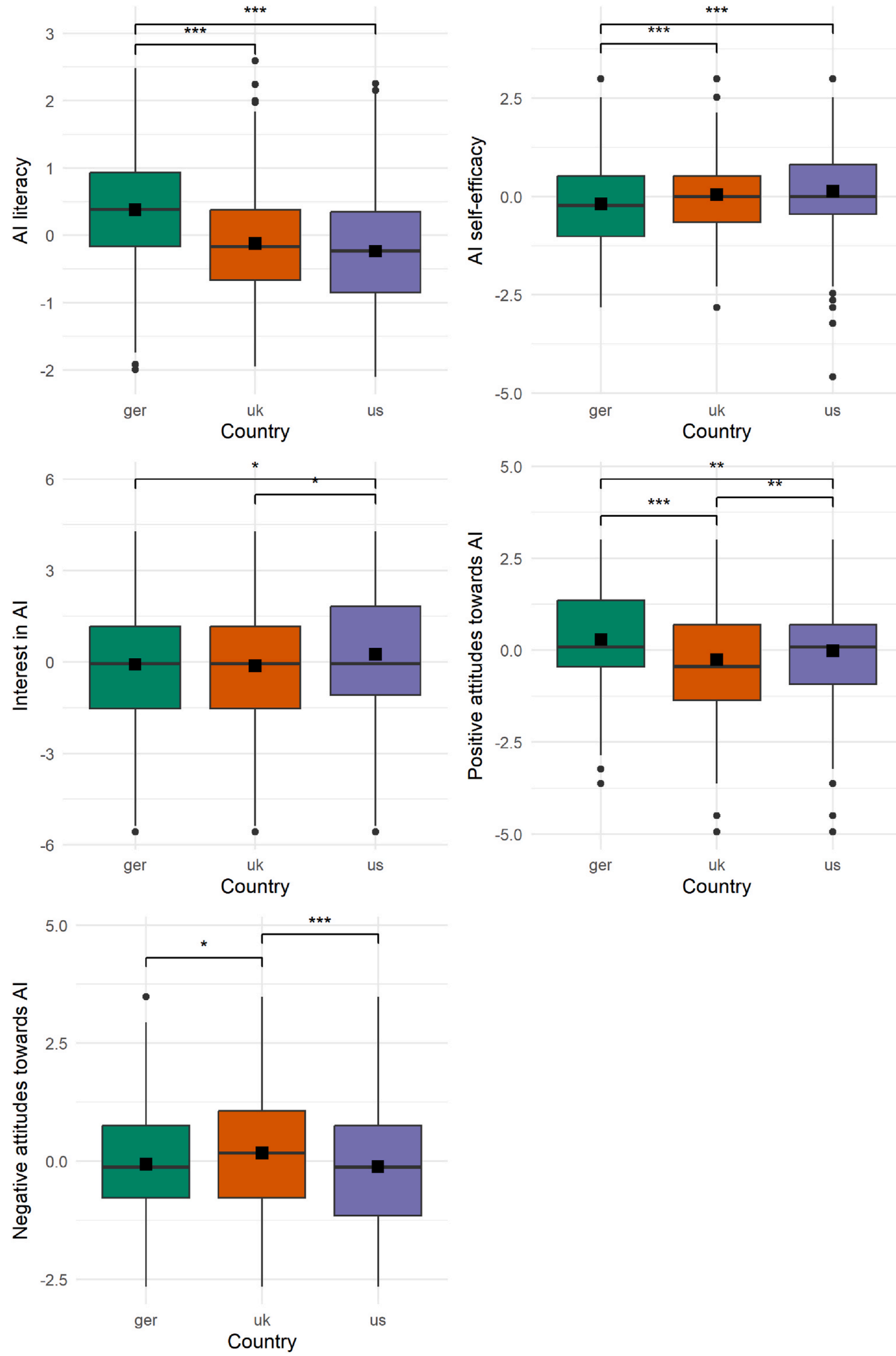

Fig. 4. Score distribution of key variables as a function of country
Note. Black squares represent group means. *p < .05, **p < .01, ***p < .001.

make well-informed decisions about the integration of AI into our society rather than being influenced by misinformation and unnecessary fear (Bewersdorff et al., 2023).

Furthermore, students report high interest in AI, reflecting prior research (Almaraz-López et al., 2023; Y.-J. Lee, Oh, & Hong, 2024). This could be explained by the finding that students are feeling unprepared and thinking of the need to use and understand AI in their careers (Pucchio et al., 2022; Teng et al., 2022). Further, our findings indicate a higher level of positive than negative attitudes towards AI, which confirms previous results (Y.-J. Lee, Oh, & Hong, 2024; Stöhr et al., 2024). German students report the highest level of positive attitudes towards AI, followed by US students, while UK students report the lowest level. Also, UK students report the highest level of negative attitudes towards AI, whereas US and German students demonstrate lower levels of negative attitudes towards AI. These differences may be explained by more experience and higher AI literacy of German students, which might

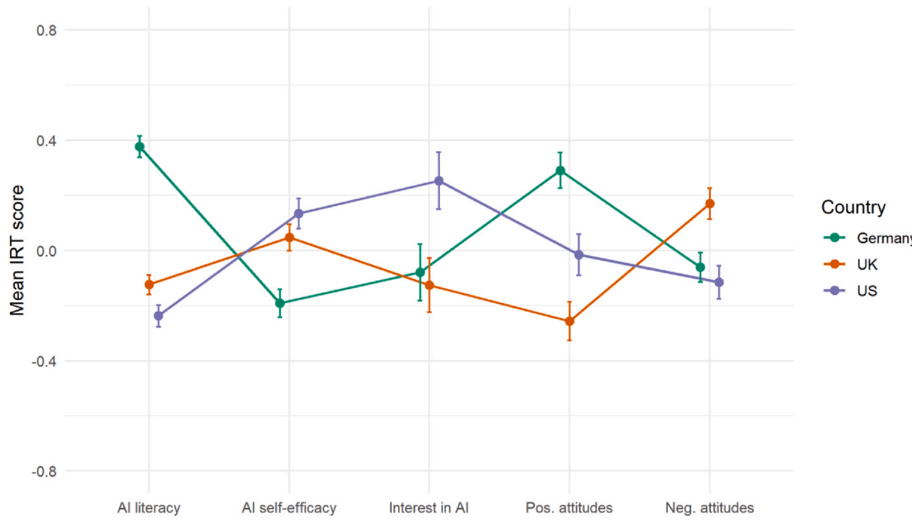

Fig. 5. Mean IRT Scores for key variables by country with standard errors.

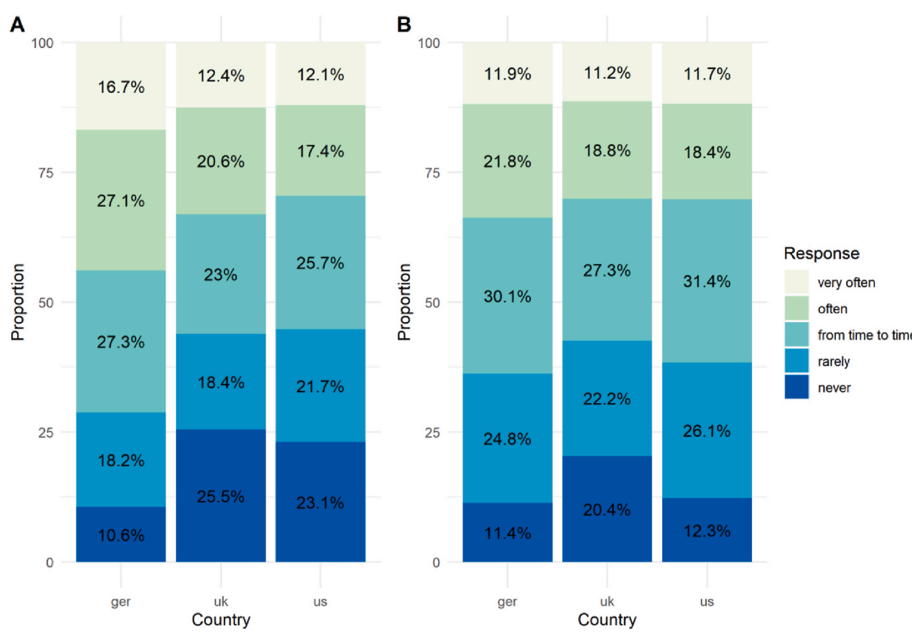

Fig. 6. How often do you actively use AI systems like ChatGPT as part of your education (A) or as part of your personal life (besides education) (B).?.

increase positive attitudes towards AI and reduce negative attitudes towards AI (Kim, 2023). Additionally, UK and US students report having more experience in ethics, which could contribute to them being more aware of the negative aspects of AI. The observed differences may also stem from cultural differences in attitudes towards AI (Neudert et al., 2020). Cave and Dihal (2023) describe how each country has a unique cultural imprint that shapes perceptions of AI. While these perceptions exist *within* Western views (Cave & Dihal, 2020), they remain distinct within their own cultural contexts. Future research could contribute to understanding these cultural variations and their implications for teaching.

6.3. Implications for educators and policymakers

Our findings suggest several implications for educators. First, although the use and awareness of AI tools have increased since the release of ChatGPT, AI literacy has not changed much according to our AI literacy test administered both before and after the release of ChatGPT. One possible explanation is that, despite increased informal attention, only a minority of students in our study have taken a course on AI. This calls for greater efforts in developing and offering learning opportunities for university students.

Furthermore, our findings imply that affective variables like AI self-efficacy, interest in AI, and attitudes towards AI are likely to play a role in developing AI literacy. This aligns with prior research on the interrelationships between these variables (Gado et al., 2022; Y.-J. Lee, Oh, & Hong, 2024). Compared to other technologies, AI is surrounded by a lot of fears and hopes (Bewersdorff et al., 2023). This calls for a more comprehensive approach to fostering AI literacy that integrates affective dimensions alongside cognitive ones. For example, psychological intervention strategies (Kim, 2023) could be used to decrease fears and enhance learning. Additionally, direct engagement with AI systems may be essential to cultivating a sense of wonder and sustained interest, and when combined with interventions attuned to students' positive or negative affect, can pave the way for deeper and more effective learning (Bewersdorff et al., 2025). Also, use of AI and prior

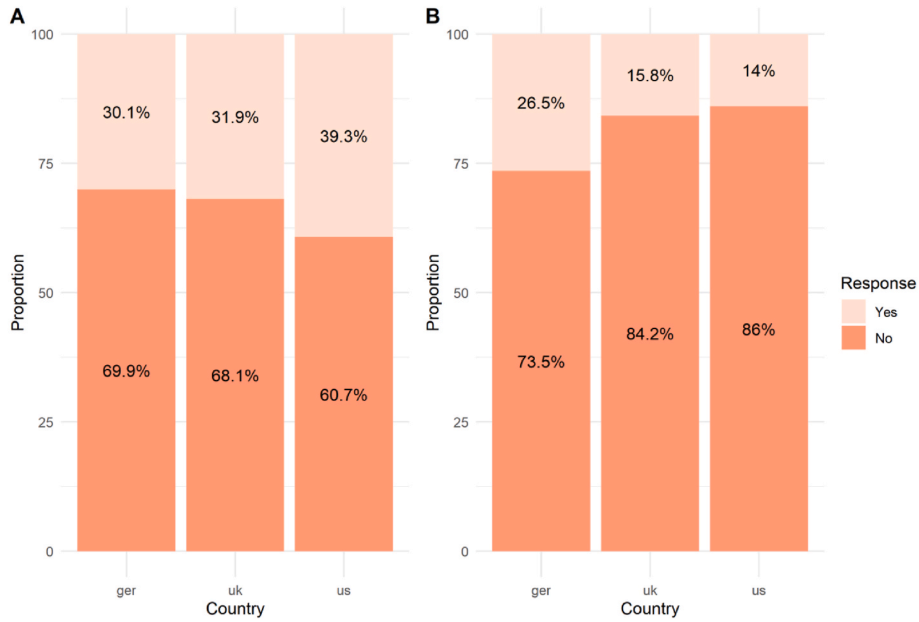

Fig. 7. Have you (A) formally studied computer science or a subject related to computer science (e.g., Business of IT, Data Science)/been enrolled in such a program in the past, or (B) ever taken a course about AI at a university/college.?.

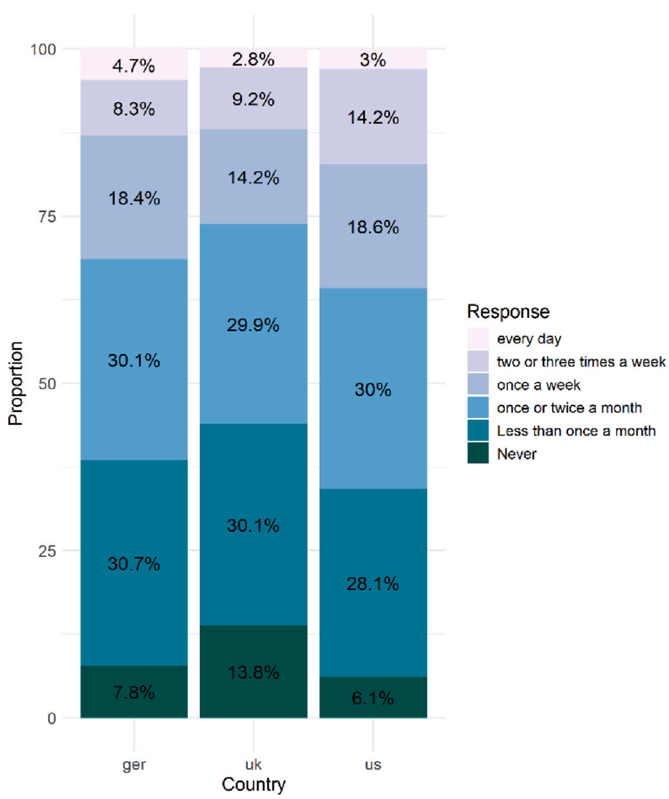

Fig. 8. How often do you learn about AI in your free time (e.g., book, podcast).?.

learning experiences are relevant variables to consider when fostering AI literacy. We advocate to not only consider the discipline of the target group, but to also be aware that individual learning experiences, including informal learning, can influence AI literacy and AI-related affective variables.

Overall, then, we recommend providing targeted learning opportunities that account for students' characteristics according to the ABC approach (Knoth et al., 2024; Ng et al., 2024). Using tests like those implemented in this study can help to gain insights on groups of learners and enable educators to better adapt to their needs (Bewersdorff et al., 2025). Furthermore, interdisciplinary courses or programs where students from different disciplines can learn from each other are a promising option to diminish differences between disciplines and enable all students to develop a critical understanding of AI. Moreover, providing hands-on experiences (e.g., programming assignments) might help create in-depth practical experience with AI and therefore, develop interest in AI and ultimately leading to AI literacy (Bewersdorff et al., 2025; Y.-J. Lee, Oh, & Hong, 2024). For example, AI courses typically offered for students of humanities tend to cover more ethical topics and affective aspects, while courses traditionally offered for STEM students often include practical assignments. This might be reflected in the differences in AI literacy and attitudes towards AI found between Germany (where more students have technical prior knowledge) and the UK and US (with more students having knowledge in ethics) in this study. Opening both types of courses for all students could help to enhance ethical AI literacy for STEM students while providing experiences of actual AI use for non-technical students (Y.-J. Lee, Oh, & Hong, 2024).

Our findings suggest several implications for policymakers. Enabling educators and their institutions to develop effective educational experiences requires meaningful investment by national and international actors. Overall, while policymakers have rightly identified the importance of boosting AI literacy generally (e.g., The Federal Government, 2018; U.S. Department of EducationOffice of Educational Technology, 2023), this study points to the need to emphasize a broader understanding of the various AI-related variables (e.g., behavioral, affective, cognitive, sociotechnical), and facilitate conditions to enable these aims. Policymakers should embrace a pluralistic definition of AI literacy beyond computational skills alone, in line with most AI literacy frameworks that recognize the interdisciplinary and sociotechnical nature of AI (Long & Magerko, 2020; Ng et al., 2021). To achieve these goals then, they must invest in relevant infrastructure, professional development, long-term planning, and support of AI literacy across the entire educational pipeline. AI literacy education should start in early childhood education and continue through primary and secondary education to higher and adult education (Touretzky et al., 2019; Laupichler et al., 2022). Policymakers are in a good position to encourage and fund local to international benchmarking studies, and to encourage the

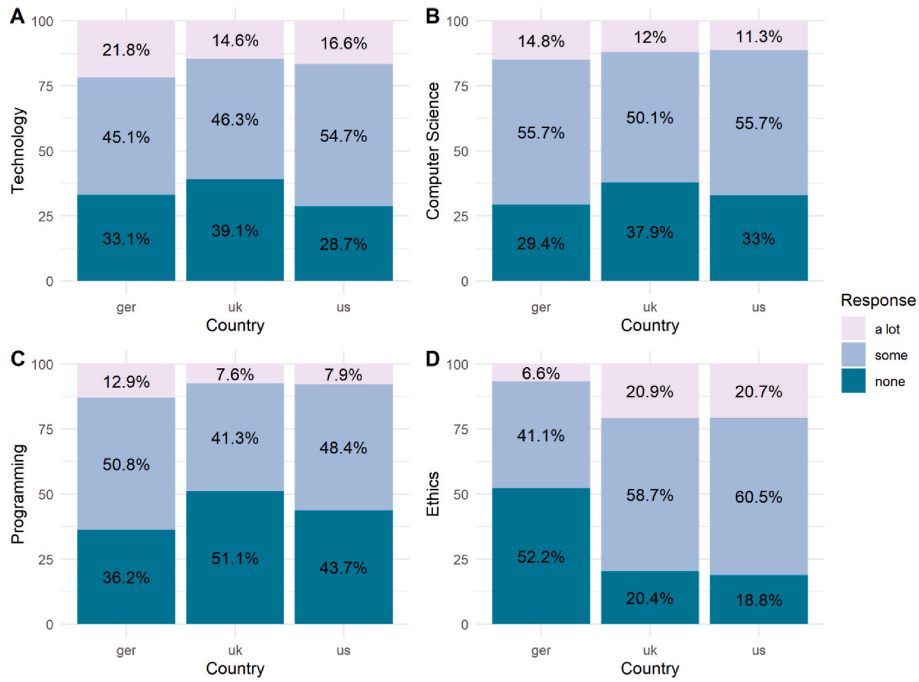

Fig. 9. Coursework addresses issues in Computer Science and Technology (A), Experience in Computer Science (B), Experience in Programming (C), Experience in Ethics (D).

identification of key gaps, best practices, and needs through research, task forces, and targeted regulation.

6.4. Conclusion

In this study, we validated an English version of a previously validated AI literacy test from Germany, demonstrating its reliability across three countries. This tool can be effectively utilized by researchers and practitioners in higher education contexts to assess AI literacy levels. Additionally, our research provides initial insights into the international comparison of AI literacy and related variables among university students. Our main findings are that across countries, students have a foundational knowledge of AI and score relatively high on positive affective AI-related variables. The findings of this study are limited to three Western countries with similar economic and cultural characteristics. More research among students in different parts of the world is needed to broaden our understanding of students' AI literacy and surrounding variables.

Furthermore, our results are limited by the sample and recruitment strategy. Even though we collected data from a large number of universities and students of all disciplines, the samples are not representative of the populations of students in the different countries, e.g., with regard to socio-economic status. Furthermore, as AI continues to develop rapidly, certain applications, terms, and technologies referenced in this study may become outdated. However, since the test assesses core AI literacy *concepts* rather than temporary trends, its overall validity should be resilient to many changes. Nevertheless, future research should continuously evaluate and, if necessary, refine this AI literacy assessment to reflect significant conceptual advancements or terminological updates in the field. Future research should also consider the interplay of AI literacy, AI-related affective variables, and prior experiences within a comprehensive framework (e.g., Bewersdorff et al., 2025). It should also assess the effectiveness of different small and large-scale interventions on AI literacy and examine strategies tailored to different student groups. Additionally, researchers should consider the impact of socio-cultural, political, and other factors on AI literacy efforts.

CRediT authorship contribution statement

Marie Hornberger: Writing – original draft, Visualization, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Arne Bewersdorff:** Writing – review & editing, Project administration, Investigation, Conceptualization. **Daniel S. Schiff:** Writing – review & editing, Project administration, Investigation, Funding acquisition, Conceptualization. **Claudia Nerdel:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT (www.chat.openai.com) as well as Grammarly (www.grammarly.com) in order to improve the readability and language of single sentences. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Funding statement

This work was supported in part by funding from Google Research. The project on which this paper is based was partly funded by the German Federal Ministry of Education and Research under the funding code 16DHBKI051. The responsibility for the content of this publication lies with the authors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to express our gratitude to Paulina Weyh and Sarah-

Alina Günzer for their support with the data analysis for this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chbah.2025.100132.

References

Almaraz-López, C., Almaraz-Menéndez, F., & López-Esteban, C. (2023). Comparative study of the attitudes and perceptions of university students in business administration and management and in education toward artificial intelligence. *Education Sciences, 13*(6), 609. https://doi.org/10.3390/educsci13060609

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*, 561–573. https://doi.org/10.1007/BF02293814

Anthology. (2023). AI in Higher Ed: Hype, Harm, or Help: The perceptions and realities of AI among university leaders and students. *Anthology Report.* https://www.anthology.com/paper/ai-in-higher-ed-hype-harm-or-help.

Ardeliya, V. E., Taylor, J., & Wolfson, J. (2024). Exploration of artificial intelligence in creative fields: Generative art, music, and design. *International Journal of Cyber and IT Service Management, 4*(1), 40–46. https://doi.org/10.34306/ijcitsm.v4i1.149

Backhaus, K., Erichson, B., & Weiber, R. (2015). *Fortgeschrittene Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*. Springer-Verlag.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B, 57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bewersdorff, A., Hornberger, M., Nerdel, C., & Schiff, D. S. (2025).). AI advocates and cautious critics: How AI attitudes, AI interest, use of AI, and AI literacy build university students' AI self-efficacy. *Computers and Education: Artificial Intelligence, 8*, Article 100340. https://doi.org/10.1016/j.caeai.2024.100340.

Bewersdorff, A., Zhai, X., Roberts, J., & Nerdel, C. (2023). Myths, pre- and misconceptions of artificial intelligence: A review of the literature. *Computers and Education: Artificial Intelligence, 4*, Article 100143. https://doi.org/10.1016/j.caeai.2023.100143.

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research* (2nd ed.). Guilford publications.

Bybee, R., & McCrae, B. (2011). Scientific literacy and student attitudes: Perspectives from PISA 2006 science. *International Journal of Science Education, 33*(1), 7–26. https://doi.org/10.1080/09500693.2010.518644

Černý, M. (2024). University students' conceptualisation of AI literacy: Theory and empirical evidence. *Social Sciences, 13*(3), 129. https://doi.org/10.3390/socsci13030129

Carolus, A., Koch, M. J., Straka, S., Latoschik, M. E., & Wienrich, C. (2023). Mails - meta AI literacy scale: Development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change- and meta-competencies. *Computers in Human Behavior: Artificial Humans, 1*(2), Article 100014. https://doi.org/10.1016/j.chbah.2023.100014

Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). Artificial intelligence and the 'good society': The US, EU, and UK approach. *Science and Engineering Ethics, 24*(2), 505–528. https://doi.org/10.1007/s11948-017-9901-7

Cave, S., & Dihal, K. (2020). The whiteness of AI. *Philosophy & Technology, 33*(4), 685–703. https://doi.org/10.1007/s13347-020-00415-6

Cave, S., & Dihal, K. (2023). *Imagining AI: How the world sees intelligent machines*. Oxford University Press.

Chalmers, R. P. (2012). Mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*(6), 1–29. https://doi.org/10.18637/jss.v048.i06.

Chang, C.-S., Liu, E. Z.-F., Sung, H.-Y., Lin, C.-H., Chen, N.-S., & Cheng, S.-S. (2014). Effects of online college student's Internet self-efficacy on learning motivation and performance. *Innovations in Education & Teaching International, 51*(4), 366–377. https://doi.org/10.1080/14703297.2013.771429

Chen, W.-H., & Thissen, D. (1997). Local dependence indexes for item Pairs using item response theory. *Journal of Educational and Behavioral Statistics, 22*(3), 265–289. https://doi.org/10.3102/10769986022003265.

Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software, 39*(8), 1–30. https://doi.org/10.18637/jss.v039.i08

Department for Science, Innovation and Technology, Office for Artificial Intelligence. (2021). National AI strategy. https://assets.publishing.service.gov.uk/media/614db4d1e90e077a2cbdf3c4/National_AI_Strategy_-_PDF_version.pdf.

Dreksler, N., Law, H., Ahn, C., Schiff, D., Schiff, K. J., & Peskowitz, Z. (2025). What does the public think about AI? An overview of the public's attitudes towards AI and a resource for future research. *Social Science Research Network* (SSRN Scholarly Paper 5108572) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5108572.

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for Health, education, and the workplace. *Psychological Science in the Public Interest: A Journal of the American Psychological Society, 5*(3), 69–106. https://doi.org/10.1111/j.1529-1006.2004.00018.x

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*, 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Embretson, S. E., & Reise, S. P. (2009). Item response theory for psychologists. *Multivariate applications series*. Psychology Press.

Federal Ministry of Education and Research. (2021). Richtlinie zur Bund-Länder-Initiative zur Förderung der Künstlichen Intelligenz in der Hochschulbildung. https://www.bmbf.de/bmbf/shareddocs/bekanntmachungen/de/2021/02/3409_bekanntmachung.html.

Fischer, R., & Karl, J. A. (2019). A primer to (Cross-Cultural) multi-group invariance testing possibilities in R. *Frontiers in Psychology, 10*, 1507. https://doi.org/10.3389/fpsyg.2019.01507

Freeman, J. (2024). Provide or punish? Students' views on generative AI in higher education. In *Hepi policy note* (Vol. 51)Higher Education Policy Institute. https://www.hepi.ac.uk/wp-content/uploads/2024/01/HEPI-Policy-Note-51.pdf.

Gado, S., Kempen, R., Lingelbach, K., & Bipp, T. (2022). Artificial intelligence in psychology: How can we enable psychology students to accept and use artificial intelligence? *Psychology Learning and Teaching, 21*(1), 37–56. https://doi.org/10.1177/14757257211037149

Garrel, J. von, & Mayer, J. (2023). Artificial Intelligence in studies—use of ChatGPT and AI-based tools among students in Germany. *Humanities and Social Sciences Communications, 10*(1), 1–9. https://doi.org/10.1057/s41599-023-02304-7

Hornberger, M., Bewersdorff, A., & Nerdel, C. (2023). What do university students know about Artificial Intelligence? Development and validation of an AI literacy test. *Computers and Education: Artificial Intelligence, 5*, Article 100165. https://doi.org/10.1016/j.caeai.2023.100165.

Jodoin, M. G., & Gierl, M. J. (2001). Evaluating type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection. *Applied Measurement in Education, 14*(4), 329–349. https://doi.org/10.1207/S15324818AME1404_2

Kim, S.-W. (2023). Change in attitude toward artificial intelligence through experiential learning in artificial intelligence education. *International Journal of Advanced Science, Engineering and Information Technology, 13*(5), 1953–1959. https://doi.org/10.18517/ijaseit.13.5.19039

Klassen, R. M. (2004). Optimism and realism: A review of self-efficacy from a cross-cultural perspective. *International Journal of Psychology, 39*(3), 205–230. https://doi.org/10.1080/00207590344000330

Knoth, N., Decker, M., Laupichler, M. C., Pinski, M., Buchholtz, N., Bata, K., & Schultz, B. (2024). Developing a holistic AI literacy assessment matrix – bridging generic, domain-specific, and ethical competencies. *Computers and Education Open, 6*, Article 100177. https://doi.org/10.1016/j.caeo.2024.100177

Kong, S.-C., Cheung, W. M.-Y., & Zhang, G. (2021). Evaluation of an artificial intelligence literacy course for university students with diverse study backgrounds. *Computers & Education: Artificial Intelligence, 2*, Article 100026. https://doi.org/10.1016/j.caeai.2021.100026

Kong, S.-C., Cheung, W. M.-Y., & Zhang, G. (2022). Evaluating artificial intelligence literacy courses for fostering conceptual learning, literacy and empowerment in university students: Refocusing to conceptual building. *Computers in Human Behavior Reports, 7*, Article 100223. https://doi.org/10.1016/j.chbr.2022.100223

Kong, S.-C., Cheung, W. M.-Y., & Zhang, G. (2023). Evaluating an artificial intelligence literacy programme for developing university students' conceptual understanding, literacy, empowerment and ethical awareness. *Educational Technology & Society, 26*, 16–30. https://www.jstor.org/stable/48707964.

Kopalle, P. K., Gangwar, M., Kaplan, A., Ramachandran, D., Reinartz, W., & Rindfleisch, A. (2022). Examining artificial intelligence (AI) technologies in marketing via a global lens: Current trends and future research opportunities. *International Journal of Research in Marketing, 39*(2), 522–540. https://doi.org/10.1016/j.ijresmar.2021.11.002

Korte, S.-M., Cheung, W. M.-Y., Maasilta, M., Kong, S.-C., Keskitalo, P., Wang, L., Lau, C. M., Lee, J. C. K., & Gu, M. M. (2024). Enhancing artificial intelligence literacy through cross-cultural online workshops. *Computers and Education Open, 6*, Article 100164. https://doi.org/10.1016/j.caeo.2024.100164

Krapp, A., & Prenzel, M. (2011). Research on interest in science: Theories, methods, and findings. *International Journal of Science Education, 33*(1), 27–50. https://doi.org/10.1080/09500693.2010.518645

LaFrance, J., Johnson, J., & Harmon, H. (2024). Harnessing artificial intelligence's potential in rural education research: An application of ChatGPT-4. *Futurity Education, 4*(2), 86–105. https://doi.org/10.57125/FED.2024.06.25.05

Lane, M., Williams, M., & Broecke, S. (2023). The impact of AI on the workplace: Main findings from the OECD AI surveys of employers and workers. *OECD social, employment and migration working papers*. Advance online publication. https://doi.org/10.1787/ea0a0fe1-en

Laupichler, M. C., Aster, A., Haverkamp, N., & Raupach, T. (2023). Development of the "Scale for the assessment of non-experts' AI literacy" – an exploratory factor analysis. *Computers in Human Behavior Reports, 12*, Article 100338. https://doi.org/10.1016/j.chbr.2023.100338

Laupichler, M. C., Aster, A., Schirch, J., & Raupach, T. (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. *Computers & Education: Artificial Intelligence, 3*, Article 100101. https://doi.org/10.1016/j.caeai.2022.100101

Le, L. T. (2009). Investigating gender differential item functioning across countries and test languages for PISA science items. *International Journal of Testing, 9*(2), 122–133. https://doi.org/10.1080/15305050902880769

Lee, D., Arnold, M., Srivastava, A., Plastow, K., Strelan, P., Ploeckl, F., Lekkas, D., & Palmer, E. (2024). The impact of generative AI on higher education learning and teaching: A study of educators' perspectives. *Computers & Education: Artificial Intelligence, 6*, Article 100221. https://doi.org/10.1016/j.caeai.2024.100221

Lee, Y.-J., Oh, J., & Hong, C. (2024). Exploratory research on understanding university students' artificial intelligence literacy in a Korean university. *The Online Journal of*

Communication and Media Technologies, 14(3), Article e202440. https://doi.org/10.30935/ojcmt/14711
Lin, X.-F., Zhou, Y., Shen, W., Luo, G., Xian, X., & Pang, B. (2023). Modeling the structural relationships among Chinese secondary school students' computational thinking efficacy in learning AI, AI literacy, and approaches to learning AI. *Education and Information Technologies*, 1–27. https://doi.org/10.1007/s10639-023-12029-4
Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16*(2). https://www.rasch.org/rmt/rmt162.pdf.
Lintner, T. (2024). A systematic review of AI literacy scales. *Npj Science of Learning, 9*(1), 50. https://doi.org/10.1038/s41539-024-00264-4
Liu, O. L., & Wilson, M. (2009). Gender differences in large-scale math assessments: PISA trend 2000 and 2003. *Applied Measurement in Education, 22*(2), 164–184. https://doi.org/10.1080/08957340902754635
Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems*.
Mang, J., Ustjanzew, N., Leßke, I., Schiepe-Tiska, A., & Reiss, K. (2019). *PISA 2015 Skalenhandbuch. Dokumentation der Erhebungsinstrumente*. Waxmann. https://doi.org/10.25656/01:17449
Margaryan, A., Littlejohn, A., & Vojt, G. (2011). Are digital natives a myth or reality? University students' use of digital technologies. *Computers & Education, 56*(2), 429–440. https://doi.org/10.1016/j.compedu.2010.09.004
Maydeu-Olivares, A. (2013). Goodness-of-Fit assessment of item response theory models. *Measurement: Interdisciplinary Research & Perspective, 11*(3), 71–101. https://doi.org/10.1080/15366367.2013.831680
Maydeu-Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika, 71*(4), 713–732. https://doi.org/10.1007/s11336-005-1295-9
Moosbrugger, H., & Kelava, A. (2020). *Testtheorie und Fragebogenkonstruktion* (3rd ed.). Springer. https://doi.org/10.1007/978-3-662-61532-4
Nam, J. (2023). 56% of college students have used AI on assignments or exams. *BestColleges*. https://www.bestcolleges.com/research/most-college-students-have-used-ai-survey/.
Nemorin, S. (2024). Towards decolonising the ethics of AI in education. *Globalisation, Societies and Education*, 1–13. https://doi.org/10.1080/14767724.2024.2333821
Neudert, L.-M., Knuutila, A., & Howard, P. N. (2020). *Global attitudes towards AI, machine learning & automated decision making.* Oxford Commission on AI & Good Governance. https://oxcaigg.oii.ox.ac.uk/wp-content/uploads/sites/11/2020/10/GlobalAttitudesTowardsAIMachineLearning2020.pdf.
Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers & Education: Artificial Intelligence, 2*, Article 100041. https://doi.org/10.1016/j.caeai.2021.100041
Ng, D. T. K., Wu, W., Leung, J. K. L., Chiu, T. K. F., & Chu, S. K. W. (2024). Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach. *British Journal of Educational Technology, 55*(3), 1082–1104. https://doi.org/10.1111/bjet.13411
Ng, D. T. K., Wu, W., Leung, J. K. L., & Chu, S. K. W. (2023). Artificial intelligence (AI) literacy questionnaire with confirmatory factor analysis. In *2023 IEEE international conference on advanced learning technologies (ICALT)*. https://www.researchgate.net/profile/tsz-kit-ng/publication/372284548_artificial_intelligence_ai_literacy_questionnaire_with_confirmatory_factor_analysis.
O'Dea, X., Tsz Kit Ng, D., O'Dea, M., & Shkuratskyy, V. (2024). Factors affecting university students' generative AI literacy: Evidence and evaluation in the UK and Hong Kong contexts. In *Policy futures in education.* Advance online publication, Article 14782103241287401. https://doi.org/10.1177/14782103241287401.
OECD. (2024a). *OECD economic outlook* (Vol. 2024). Paris: OECD Publishing. https://doi.org/10.1787/d8814e8b-en. Issue 2.
OECD. (2024b). *PISA 2022 technical report*. OECD Publishing. https://doi.org/10.1787/01820d6d-en
Orlando, M., & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement, 24*(1), 50–64. https://doi.org/10.1177/01466216000241003
Orlando, M., & Thissen, D. (2003). Further investigation of the performance of S - X2: An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement, 27*(4), 289–298. https://doi.org/10.1177/0146621603027004004
Pinski, M., & Benlian, A. (2023). Ai literacy - towards measuring human competency in artificial intelligence. In *56th annual Hawaii international conference on system sciences*. https://scholarspace.manoa.hawaii.edu/handle/10125/102649.
Pucchio, A., Rathagirishnan, R., Caton, N., Gariscsak, P. J., Del Papa, J., Nabhen, J. J., Vo, V., Lee, W., & Moraes, F. Y. (2022). Exploration of exposure to artificial intelligence in undergraduate medical education: A Canadian cross-sectional mixed-methods study. *BMC Medical Education, 22*(1), 815. https://doi.org/10.1186/s12909-022-03896-5
R Core Team. (2022). *R: A Language and Environment for Statistical Computing [Computer software]*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.
Rosseel, Y. (2012). Lavaan : An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02.
Schepman, A., & Rodway, P. (2020). Initial validation of the general attitudes towards artificial intelligence scale. *Computers in Human Behavior Reports, 1*, Article 100014. https://doi.org/10.1016/j.chbr.2020.100014
Soto-Sanfiel, M. T., Angulo-Brunet, A., & Lutz, C. (2024). The scale of artificial intelligence literacy for all (SAIL4ALL): A tool for assessing knowledge on artificial intelligence in all adult populations and settings. https://doi.org/10.31235/osf.io/bvyku.
Southworth, J., Migliaccio, K., Glover, J. J., Glover, J Ja Net, Reed, D., McCarty, C., Brendemuhl, J., & Thomas, A. (2023). Developing a model for AI across the curriculum: Transforming the higher education landscape via innovation in AI literacy. *Computers & Education: Artificial Intelligence, 4*, Article 100127. https://doi.org/10.1016/j.caeai.2023.100127
Stöhr, C., Ou, A. W., & Malmström, H. (2024). Perceptions and usage of AI chatbots among students in higher education across genders, academic levels and fields of study. *Computers & Education: Artificial Intelligence, 7*, Article 100259. https://doi.org/10.1016/j.caeai.2024.100259
Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment, 80*(1), 99–103. https://doi.org/10.1207/S15327752JPA8001_18
Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement, 27*(4), 361–370. https://doi.org/10.1111/j.1745-3984.1990.tb00754.x
Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd
Teng, M., Singla, R., Yau, O., Lamoureux, D., Gupta, A., Hu, Z., Hu, R., Aissiou, A., Eaton, S., Hamm, C., Hu, S., Kelly, D., MacMillan, K. M., Malik, S., Mazzoli, V., Teng, Y., Laricheva, M., Jarus, T., & Field, T. S. (2022). *Health Care Students' Perspectives on Artificial Intelligence: Countrywide Survey in Canada, 8*(1), Article e33390. https://doi.org/10.2196/33390
The Federal Government. (2018). Artificial intelligence strategy. https://www.ki-strategie-deutschland.de/.
Touretzky, D., Gardner-McCune, C., Martin, F., & Seehorn, D. (2019). Envisioning AI for K-12: What should every child know about AI? *Proceedings of the AAAI Conference on Artificial Intelligence, 33*(1), 9795–9799. https://doi.org/10.1609/aaai.v33i01.33019795
U.S. Department of Education, Office of Educational Technology. (2023). Artificial intelligence and future of teaching and learning: Insights and recommendations. https://www2.ed.gov/documents/ai-report/ai-report.pdf.
Wang, B., Rau, P.-L. P., & Yuan, T. (2022). Measuring user competence in using artificial intelligence: Validity and reliability of artificial intelligence literacy scale. *Behaviour & Information Technology*, 1–14. https://doi.org/10.1080/0144929X.2022.2072768
Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement, 8*(2), 125–145. https://doi.org/10.1177/014662168400800201.
Zhang, H., Perry, A., & Lee, I. (2024). Developing and validating the artificial intelligence literacy concept inventory: An instrument to assess artificial intelligence literacy among middle school students. In *International journal of artificial intelligence in education*. Advance online publication. https://doi.org/10.1007/s40593-024-00398-x.
Zumbo, B. D. (1999). A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores. https://faculty.educ.ubc.ca/zumbo/dif/handbook.pdf.