Computers and Education: Artificial Intelligence 4 (2023) 100126

Contents lists available at ScienceDirect

Computers and Education: Artificial Intelligence

journal homepage: www.sciencedirect.com/journal/computers-and-education-artificial-intelligence

Delphi study for the development and preliminary validation of an item set for the assessment of non-experts' AI literacy

Matthias Carl Laupichler *, Alexandra Aster, Tobias Raupach

Institute of Medical Education, University Hospital Bonn, Bonn, Germany

ARTICLE INFO

Keywords:
AI literacy
Assessment
Scale
Questionnaire
Delphi study
Validity

ABSTRACT

Artificial intelligence literacy is a concept that has been the focus of exhaustive research recently. However, there are very few psychometrically sound and thoroughly evaluated instruments that attempt to assess AI literacy in a valid way. Therefore, this study presents an item set to assess the AI literacy of non-experts. In the context of a Delphi expert study, 53 subject matter experts participated in three iterative questionnaire rounds to generate potential AI literacy items and assess their content validity. In addition, the experts made suggestions on how the items' wording accuracy could be improved and evaluated the wording suggestions of the other experts. Of 47 potential items, 38 were judged relevant for inclusion in a final AI literacy questionnaire. The result is one of the first freely available AI literacy item sets and represents an important first step in assessing AI literacy and its subconstructs. Finally, the development of the items through the execution of an iterative Delphi study and the strong focus on content validity contribute to the advancement of AI literacy theory.

1. Introduction

1.1. AI literacy and its relevance

Although the importance of AI literacy research has increased in recent years, there is still no commonly accepted definition of AI literacy. However, one of the most commonly cited definitions is that of Long & Magerko, who describe AI literacy 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, p. 2).

The importance of an AI literate population is growing as more and more aspects of personal and professional life are permeated by AI. On the one hand, individuals engage (consciously or unconsciously) with AI-based applications in their spare time, such as smart speakers (Bentley et al., 2018), face recognition (Adjabi et al., 2020), or recommender systems for web-applications (Zhang et al., 2019). On the other hand, AI applications are also increasingly finding their way into the workplace, and employees have to learn how to deal with these novel systems (Chowdhury et al., 2022) to ensure they can continue to work in decent work environments. (Braganza, Chen, Canhoto, & Sap, 2021).

1.2. Assessing AI literacy and related concepts

Several efforts have been made to develop scales to capture constructs related to AI literacy. However, they mostly deal with the affective component of AI collaboration and cannot be used to capture AI literacy itself. Examples include the “Attitudes Towards Artificial Intelligence” Scale (Sindermann et al., 2021), the “General Attitudes Towards Artificial Intelligence” Scale (Schepman & Rodway, 2022), as well as the “Artificial Intelligence Anxiety Scale” (Wang & Wang, 2022).

In contrast to the assessment of attitudes towards AI, there are few research projects that seek to advance the psychometrically valid measurement of AI literacy. In order to capture the current state of research on methods of AI literacy assessment, we conducted a brief literature review with five search terms synonymous with “ai literacy scale” in five different databases. We ran a first search in April 2022, and a second search with the same terms in the same databases in October 2022. The initial search yielded ten results, whereof two were published in another language than English and two called for the creation of AI literacy scales (Ng et al., 2021a, 2021b). In the remaining six papers, an AI literacy scale was developed as a means to evaluate the learning outcomes of specific educational programs. However, these scales have not been psychometrically evaluated (de Souza, 2021; Kong et al., 2021) and were often developed specifically for particular courses (Dai et al.,

* Corresponding author. Institute of Medical Education, University Hospital Bonn, Venusberg-Campus 1, 53127, Bonn, Germany.
E-mail addresses: matthias.laupichler@ukbonn.de (M.C. Laupichler), alexandra.aster@ukbonn.de (A. Aster), tobias.raupach@ukbonn.de (T. Raupach).

https://doi.org/10.1016/j.caeai.2023.100126
Received 21 November 2022; Received in revised form 10 January 2023; Accepted 10 January 2023
Available online 12 January 2023

2020). In addition, some authors seemed to have a different understanding of AI literacy. For example, Lin et al. (2021) and Shih et al. (2021) reported an AI literacy scale that contains the two factors "teamwork" and "attitudes toward AI" and thus does not reflect the central aspects of the AI literacy definitions reported above. Finally, although some authors provide examples of the items in the scale (Kong et al., 2021), most papers do not include the entire scale, making it difficult for other researchers to replicate the results.

The second search yielded one additional result. Wang et al. (2022) introduced the first psychometrically evaluated “Artificial Intelligence Literacy Scale”, consisting of twelve items on four dimensions (i.e., “awareness”, “usage”, “evaluation”, and “ethics”). In many ways, this scale represents a significant improvement over the previously developed scales in that it approaches the development of an instrument to measure AI literacy in a structured and methodically sound manner. It must be mentioned, however, that the “Artificial Intelligence Literacy Scale” was developed as “a valid and reliable scale to measure people’s AI literacy for future [human-AI interaction] research” (Wang et al., 2022, p. 5). Human-AI interaction (HAII) research has traditionally viewed the use of AI from the perspective of program design rather than user capabilities (Amershi et al., 2019). While focusing on HAII is a legitimate approach, the extent to which the “Artificial Intelligence Literacy Scale” is valid outside the HAII research domain is debatable. A general AI literacy scale that can be used universally should be applicable in all research areas related to AI literacy. As an illustration, a general method for measuring AI could be used to assess AI literacy before and after attending an introductory AI-course.

The underlying definition, which was formulated by the authors themselves, differs in some respects from established definitions such as that of Long and Magerko (2020). Most importantly, the strong focus on AI awareness is certainly relevant, but somewhat neglects AI-knowledge and -understanding, which constitutes a main aspect of most AI literacy definitions (Ng et al., 2021b). Thus, we see the need for a universally applicable AI literacy scale.

1.3. Non-experts as target group

Similar to other technological literacies like data literacy (Wolff et al., 2016) or computational literacy (Jacob & Warschauer, 2018), AI literacy is commonly used to describe the competencies of non-experts rather than (AI-) professionals. Non-experts are defined as individuals who have not received formal training in AI and are using AI applications rather than developing them. Thus, in general, all adults interacting with sophisticated and modern digital applications can be considered non-experts, as it can be assumed that most of today’s digital applications are at least partially based on AI algorithms. In our interpretation, a non-expert is on one of the two lower levels of the framework proposed by Faruqe et al. (2021). Therefore, he or she is either a “consumer […] who uses the outputs of AI to improve their work or life” or a “co-worker [who] knows the basics of how the AI systems work and uses AI outputs in the work” (Faruqe et al., 2021, p. 1). Although (younger) children are no AI experts, they are not included in the target group of the items either, as they are not yet consumers or co-workers.

1.4. Aim of this study

The aim of this study was to generate a face and content valid AI literacy item set that can be used to develop a scale to assess the AI literacy of non-experts. To evaluate the relevance of the different items for the assessment of AI literacy and to determine its content validity, we conducted an expert Delphi study with subject matter experts (SMEs). Content validity can be defined as “the degree to which elements of an assessment instrument are relevant to and representative of the targeted construct for a particular assessment purpose.” (Haynes et al., 1995, p. 238).

We developed a primary research question, which was divided into two subquestions (1a and 1b). The first question was related to the content validity of potential items on AI competence. Therefore, research question 1a was.

RQ1a. Which items are relevant for and representative of AI literacy?

The second subquestion related to the wording of the items. Since the items on AI literacy must contain all necessary information without including superfluous or irrelevant aspects, research question was 1b.

RQ1b. How can the items be rephrased to most accurately represent the construct of AI literacy?

2. Method

An iterative, three-round Delphi expert study (Hsu & Sandford, 2007) was conducted to develop a face and content valid AI literacy item set. Using a Delphi study allowed for very elaborate validity testing. This is primarily due to the fact that the Delphi methodology is a multistage, iterative procedure. This has the advantage that the experts involved can take into account each other’s opinion and thus a consensus can be reached. In addition, the participant pool in Delphi surveys consists of several experts. This means that it is not the opinions of individual persons that count, but the assessments of a large group that is very well versed in the field.

Ethical approval to perform this study was granted by the Ethics Committee of the University of Bonn, Germany (Reference 194/22).

2.1. Expert panel

2.1.1. Expert panel selection

We contacted a total of 471 potential SMEs in the field of AI literacy and AI education by email (see Fig. 1). The SMEs’ contact details were retrieved from three sources. 400 contacts originated from the "Networking Event: AI in University Education 2022" and 50 from the "AI Networking Event North Rhine-Westphalia 2022", both of which were organized by the German Federal Ministry of Education and Research (BMBF). The other 21 people were members of smaller AI working groups in which the first author participated. Since the participants of these events were engaged in AI education on a professional basis (mostly as researchers or lecturers), they could be considered AI experts. Nevertheless, the actual expertise was assessed at a later stage (see section 3.1). In addition to AI expertise, it can be assumed that the experts had good pedagogical and didactic skills, since most of them were either lecturers from the university sector or worked at the intersection of AI and education. Although some participants may have been more knowledgeable or skilled in one of the two areas (i.e., AI or pedagogy), this population still provided the best opportunity to reach a reasonably large sample of participants. Of all those contacted, 85 prospective participants (18%) completed a brief registration survey. In the actual Delphi study rounds, 59 (Round 1), 55 (Round 2), and 53 (Round 3) SMEs participated. Thus, the dropout rate was 5% between rounds 1 and 2 and 4% between rounds 2 and 3.

2.1.2. Experts’ characteristics

Most participants (N = 53) answered all or part of the sociodemographic questions asked at the beginning of Round 3. About two in five (41%, N = 22) of the SMEs identified as being female, 53% (N = 28) as being male, 2% (N = 1) as “other”, and two did not wish to disclose their gender. Most SMEs (43%, N = 23) were between 30 and 39 years old, with the youngest participants being between 18 and 29 years old and the oldest being between 50 and 65 years old. The majority of study participants had a Master’s degree (55%, N = 29), while 26% (N = 14) had doctorate degrees and 15% (N = 8) were professors. Nearly two thirds of the SMEs worked at a university (60%, N = 32) and 23% (N = 12) worked at a university of applied sciences. In addition, some SMEs did not work at a research or educational institution (13%, N = 7) or worked at another form of research or educational institution (4%, N =

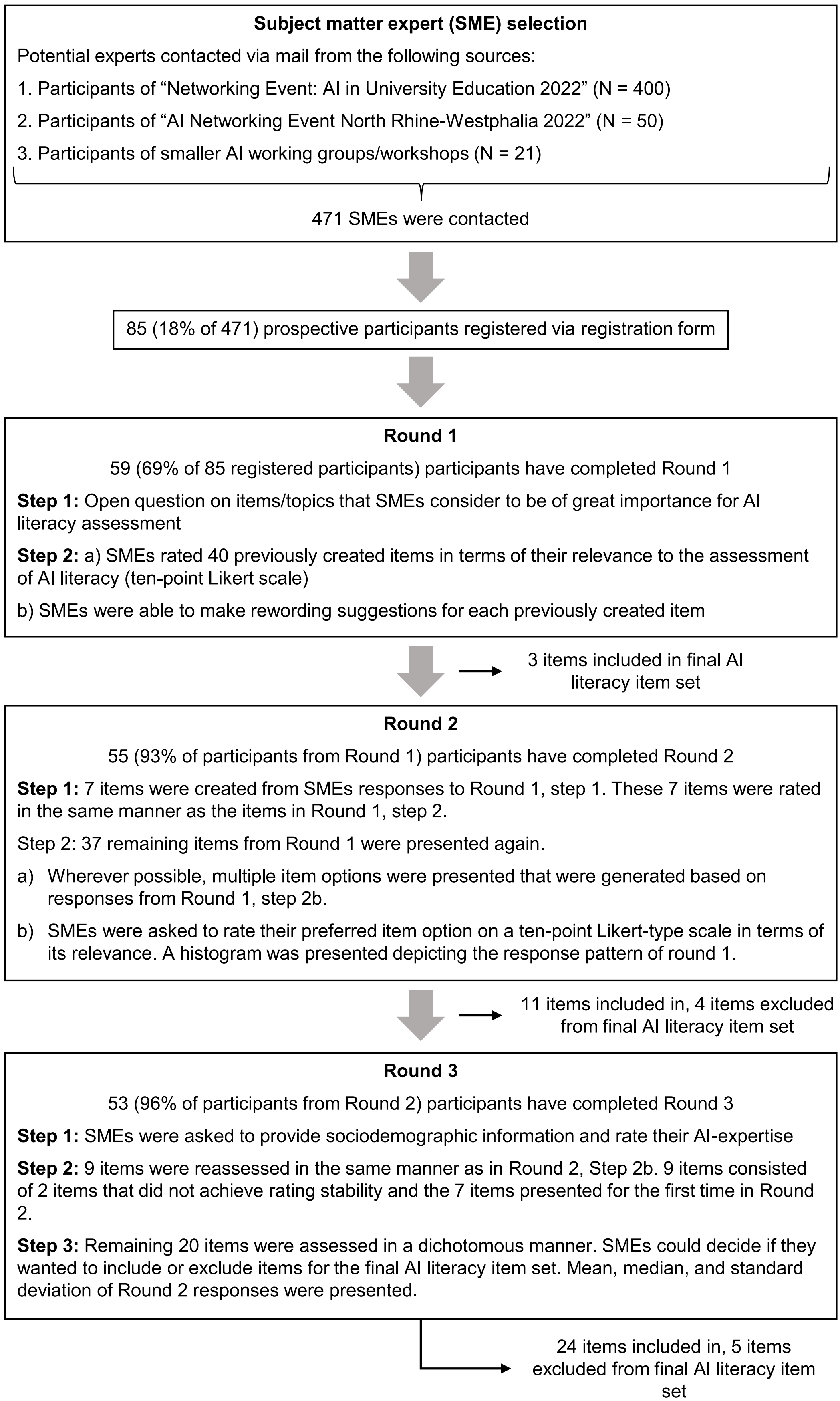

Fig. 1. Delphi procedure across three iterative rounds, including subject matter expert selection.

2).

Most participants stated that they had either a "good understanding of how AI works and where it is used" (43%, N = 23) or a "deep understanding of AI" (36%, N = 19) (see Table 1 for a detailed breakdown of the response frequencies). Participants reported dealing with AI once a week (28%, N = 15), almost every workday (36%, N = 19), or on a daily basis (23%, N = 12). The majority of the expert panel members had been working in the field of AI for "1–3 years" (45%, N = 24), followed by "3–10 years" (26%, N = 14). Nearly all of the participants were German native speakers, with one participant rating his or her German language skills at C2-level.[1] Moreover, 68% (N = 36) rated their English language skills to be on the C-level (36%, N = 19 for C1, and 32%, N = 17 for C2, respectively). Some participants self-assessed their English language skills to be on the B2-level (25%, N = 13). All participants lived and worked in Germany.

Table 1
Subject matter experts' characteristics (N = 53).

Response options	N	%
Own AI-Expertise		
No AI knowledge/experience at all	0	0
Basic idea of what AI is	7	13.2
Good understanding of how AI works, where it is used, etc.	23	43.4
Deep understanding of AI (conducted initial AI research/ development/knowledge accumulation)	19	35.8
Very deep understanding of AI (several years of intensive AI research/development/knowledge accumulation)	4	7.5
Frequency of engagement with AI		
Almost never	1	1.9
Less than once a week	6	11.3
Approximately once a week	15	28.3
Almost every (working) day	19	35.8
Every day	12	22.6
Duration of engagement with AI		
0 to 1 year	10	18.9
1 to 3 years	24	45.3
3 to 10 years	14	26.4
More than 10 years	5	9.4
Age		
18 to 29	12	22.6
30 to 39	23	43.4
40 to 49	9	17.0
50 to 65	9	17.0
Older than 65	0	0.0
Gender		
Female	22	41.5
Male	28	52.8
Other	1	1.9
Not specified	2	3.8
Highest level of education		
Secondary school leaving certificate	0	0.0
High school diploma	0	0.0
Bachelor's degree	1	1.9
Master's degree	29	54.7
Doctorate/PhD	14	26.4
Professorship	8	15.1
Other	1	1.9
Type of employment		
No employment at research or educational institution	7	13.2
University	32	60.4
University of Applied Sciences	12	22.6
Other type of educational/research institution	2	3.8

Note: N = Number of SMEs that chose this response option. % = Percentage of this response option in the total sample.

2.2. Procedure

In a first step preceding the Delphi study, an initial set of 40 AI literacy items was created. For this purpose, well-known and relevant AI literacy courses such as "Elements of AI" (University of Helsinki & MinnaLearn, 2018; www.elementsofai.com), "AI for Everyone" (Ng & DeepLearning.AI, 2022; www.coursera.org/learn/ai-for-everyone), "Introduction to AI" (Waldmann et al., 2020, www.ki-campus.org/courses/einfuehrungki2020) and books such as "Human + Machine" (Daugherty & Wilson, 2018) and "Artificial Intelligence: The Insights You Need from Harvard Business Review" (Davenport et al., 2019) were reviewed in an unsystematic manner to identify recurring content. Key terms from the various sources were collected and compared. Terms that appeared in at least two independent sources were transformed into items. In addition, Long & Magerko's (2020) AI literacy framework with its 16 AI competency domains was used as a further basis for item generation. To avoid a rigid classification of each item into one of the 16 competencies, the framework was only used as an implicit decision support tool. Although "it is both a common and an acceptable modification of the Delphi process format to use a structured questionnaire in Round 1 that is based upon an extensive review of the literature" (Hsu, 2007, p. 2), we wanted to ensure that the preliminary themes identified reflected the most important AI constructs. Therefore, the topics were discussed with a small convenience sample of AI experts (N = 5) to generate the items presented in round 1.

The actual Delphi study was conducted online via the questionnaire tool "evasys" (evasys Giannarou & Zervas, 2014).

In the first round, participants were given a common definition of AI literacy (definition by Long & Magerko, 2020) in order for all participants to be able to share one definition of the underlying construct (please find the questionnaires for all three rounds in the original German version as well as in the English translation in Supplementary Material 1). In addition, it was explained to the SMEs for which target group the questionnaire will be designed and what exactly the term "non-experts" means (see section 1.3). Subsequently, participants were asked to enter their own ideas regarding topics and items that would be highly relevant for an AI literacy scale in a text box (see Fig. 1). This question was presented before the evaluation of the initially generated items in order to avoid possible influencing effects. Accordingly, participants were then asked to rate the items in terms of their relevance to an AI literacy scale. Relevance was rated on a ten-point Likert-type scale from 1 ("not relevant at all") to 10 ("very relevant"). There was an option to abstain ("no answer"). After each item, participants could indicate whether they wanted to suggest a rewording ("Would you reword, change, add, or shorten the preceding item?"). If they clicked "Yes," a text box appeared in which they could enter their suggestions.

At the beginning of Round 2, items that were generated from the free-text responses at the beginning of Round 1 were presented. The rating procedure was the same as for the items in Round 1. Afterwards, all items from round 1 were presented again for which no final decision regarding inclusion/exclusion could be made (see Fig. 2 for an overview of the inclusion/exclusion decision process). The procedure was structured as follows. Wherever possible, multiple item options (i.e., slightly different versions of the same item) were created based on the rewording suggestions from Round 1. Participants could first select their preferred item wording and then rate the preferred version on a ten-point Likert scale. As additional information, the rating results from Round 1 were presented as a histogram.

In Round 3, the first step was to gather some information about the SMEs themselves. This information included age, gender, highest level of education, country of main affiliation, and if they worked in a research or education facility. In addition, the experts were asked to rate

[1] A1 stands for the lowest language proficiency level, C2 for the highest language proficiency level. Individuals at the "A" level are considered basic users, individuals at the "B" level are considered independent users, and individuals at the "C" level are considered proficient users. The meaning of each level was explained to the SMEs in the questionnaire.

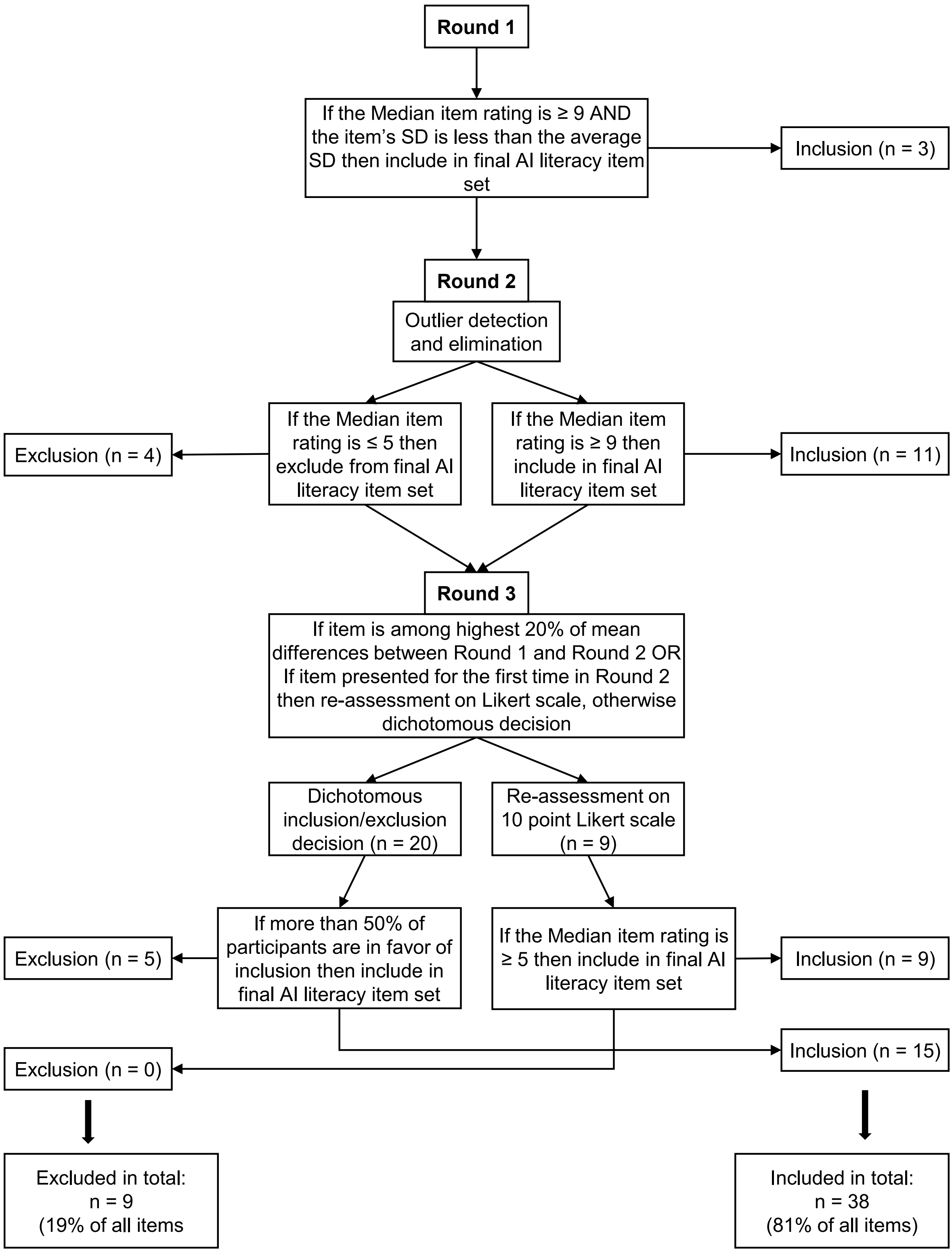

Fig. 2. Rules for including items in/excluding items from the final AI literacy item set and number of items included/excluded due to these rules.
Note. "Inclusion" means inclusion in the final AI literacy item set, "exclusion" means exclusion from the final AI literacy item set.

their English and German proficiency on a single question (from A1 to C2, according to the levels of the "Common European Framework of Reference", Council of Europe, 2022) since the instructions were presented in German but the items themselves in English. Finally, the participants were asked to rate their own AI expertise and to indicate since when and how regularly they have been involved with AI. Subsequently, items were reassessed whose response patterns had not achieved stability over the first two rounds. In addition, the items that were presented for the first time in Round 2 were reassessed. The remaining items, which could not yet be included or excluded from the final item set, but already showed a stable response pattern, were assessed in a dichotomous manner. This means that the SMEs were able to make a final decision on whether the items should be included or excluded. To support the decision, the mean, median, and standard deviation of the Round 2 ratings were provided (as text) next to the item text.

2.3. Consensus criteria

Initially, we had planned to use a fixed consensus criterion comparable to that proposed by Giannarou and Zervas (2014). In our case, this would have been a standard deviation of less than 1 (please refer to OSF-preregistration https://doi.org/10.17605/OSF.IO/B7R4H). However, during the evaluation of the first round, it turned out that the experts' opinions differed too much to apply a fixed consensus criterion. Therefore, we generated a set of hierarchically organized rules that determined whether an item should be presented again in the next round or not. In summary, an item was presented again if it could not yet be definitively decided whether the respective item should be included in or excluded from the final item set (see Fig. 2 for a detailed description of the consensus rules).

2.4. Data analysis

The mean, median, and standard deviation for each item were calculated after each round. In addition, histograms were created to summarize the response pattern of the previous round for the participants in a structured way. To calculate the stability of the response pattern, the mean difference between round 1 and round 2 was calculated (as an absolute value). This calculation was performed to determine if there was a stable dissent of expert responses (i.e., SMEs stick to their assessments despite differing opinions). If this was the case, further presentation of the items was considered unnecessary and they were included or excluded based on the criteria described above. After Round 2, a boxplot was created for each item to identify and subsequently eliminate potential outliers. All values that fell outside of the whiskers (max. 1.5 x interquartile range) were treated as outliers. Data analysis was performed using SPSS Statistics (IBM, 2022).

3. Results

3.1. Pool of potentially useful items

Forty items were generated by analyzing AI introductory courses and books. The items included were chosen to assess core competencies of AI that appeared repeatedly in the various popular science courses and books. Examples that occurred in at least two of the sources mentioned in section 2.2 and were therefore included as items in the preliminary item pool included "I can tell if the things I use frequently are supported by artificial intelligence." or “I can describe what a Turing test is supposed to find out." (both initial item wording prior to rewording suggestions). In addition to these 40 items, another sample of items was generated by analyzing the SMEs’ responses to an open question posed at the very beginning of Round 1. Somewhat unsurprisingly, most topics or items that were suggested by the SMEs were already covered by items generated in advance. However, seven items that were not included in the initial item set were added to the pool of items which could be potentially relevant for assessing AI literacy. Thus, a total of 47 items were evaluated by the SMEs regarding their relevance throughout the three rounds.

3.2. Relevance of potential items and decision on inclusion in the final scale

44 of the 47 items were rated at least twice on a ten-point Likert scale. The remaining three items were rated only once, as the evaluation resulted in a median of 10 in the first round, while showing low variability. Thus, in the case of the three items, the experts agreed very early on that they were important for the assessment of AI literacy. The three items were: “I can describe risks that may arise when using artificial intelligence systems.”, “I can explain why data plays an important role in the development and application of artificial intelligence.”, and “I can identify ethical issues surrounding artificial intelligence.”

After Round 2 and the elimination of outlier values, 11 items whose median was $\geq$9 were included in the final item set, whereas four items whose median was $\leq$5 were excluded from the final item set.

In Round 3, two items had to be presented again due to a lack of response stability (i.e., high difference between rounds). Those items were “I can describe the potential impact of artificial intelligence on the future.” and “I can explain how sensors are used by computers to collect data that can be used for AI purposes.” Both were subsequently included in the final item set. Of the 20 items assessed in a dichotomous manner (i.e., include or exclude), 15 items were included in the final item set. Finally, all of the seven items presented for the first time in Round 2 were included in the item set, since all of them had a median of $\geq$6.

In summary, a total of 47 items were evaluated regarding their relevance for inclusion in an AI literacy scale. Of these, 38 items were rated as relevant and representative for AI literacy, while nine items were not included in the final item set due to lack of relevance (see Table 2).

3.3. Validity of item wording

In addition to generating potentially relevant AI literacy items and evaluating them, the Delphi study had a third purpose. To verify that the items proposed by the research team or SMEs were worded as clearly, concisely, and validly as possible, the SMEs were given the opportunity to make rewording suggestions for each item. The rewording suggestions were evaluated by members of the research team. The first step was to check whether the entry was an actual rewording or improvement proposal. An example of a comment that was interesting but did not contain a rewording suggestion was "I think many experts don’t know this term". In a second step, it was examined whether the suggestion met the purpose of the proposed assessment. For example, some SMEs suggested changing certain items to an item that would test respondents’ knowledge of AI. While this would also be an interesting research project, it goes beyond the scope of the work presented here. Finally, the number of times a particular rewording suggestion was made was counted. Frequent mentions were presented as alternative item options. As an example for the different item options presented to the SMEs, one can look at item #6, which originally said: “I can distinguish media representations of AI (e.g., in movies or video games) from realistic AI.”. The alternative item options generated based on the SMEs’ responses were “I can distinguish science fiction representations of AI (e.g., in movies or video games) from real AI.” and “I can evaluate whether media representations of AI (e.g., in movies or video games) go beyond the current capabilities of AI technologies.” Alternative item options could not be generated for each item. Nevertheless, at least one to a maximum of three alternative options were created for 66% of all items ($N_{item} = 31$). Of these alternatives, the preferred version was selected by each participant in the following rounds.

This approach further increased the scale’s content validity by ensuring that no important item content was omitted or that the inclusion of unnecessary item content negatively affected the relevance of the item.

4. Discussion

We conducted a Delphi expert study to generate an item set that supports the development of a scale for the assessment of non-experts’ AI literacy.

Table 2
Mean (M), median (Mdn), and standard deviation (SD) for all items across all 3 rounds.

	Item	Round 1			Round 2			Round 3			Inclusion/ Exclusion,
	I can …	*M*	*Mdn*	*SD*	*M*	*Mdn*	*SD*	*M*	*Mdn*	*SD*	Round
1	tell if the technologies I use are supported by artificial intelligence.	8.9	9	1.49	9.0	9	1.26	n.a., f.	n.a., f.	n.a., f.	Included, Round 2
2	name examples of technical applications that are supported by artificial intelligence.	8.2	9	2.31	9.1	9	0.99	n.a., f.	n.a., f.	n.a., f.	Included, Round 2
3	explain the differences between human and artificial intelligence.	7.6	8	2.52	8.6	9	1.22	n.a., f.	n.a., f.	n.a., f.	Included, Round 2
4	describe how artificial intelligence consists of an interplay of complex algorithms and mathematical formulas.	5.4	6	2.65	4.6	4	2.07	n.a., f.	n.a., f.	n.a., f.	Excluded, Round 2
5	explain the difference between general (or strong) and narrow (or weak) artificial intelligence	7.2	8	2.23	7.1	7	1.69	n.a., d.	n.a., d.	n.a., d.	Included, Round 3
6	evaluate whether media representations of AI (e.g., in movies or video games) go beyond the current capabilities of AI technologies.	7.2	8	2.32	7.8	8	1.32	n.a., d.	n.a., d.	n.a., d.	Included, Round 3
7	explain what is meant by the term singularity in the context of artificial intelligence.	4.8	4	2.60	3.4	3	1.54	n.a., f.	n.a., f.	n.a., f.	Excluded, Round 2
8	name weaknesses of artificial intelligence.	8.6	9	1.94	8.9	9	1.04	n.a., f.	n.a., f.	n.a., f.	Included, Round 2
9	name strengths of artificial intelligence.	8.2	9	1.88	8.6	9	1.23	n.a., f.	n.a., f.	n.a., f.	Included, Round 2
10	describe risks that may arise when using artificial intelligence systems.	9.2	10	1.25	n. a., f.	n.a., f.	n.a., f.	n.a., f.	n.a., f.	n.a., f.	Included, Round 1
11	describe advantages that can come from using artificial intelligence systems.	8.0	8	1.88	8.1	8	1.22	n.a., d.	n.a., d.	n.a., d.	Included, Round 3
12	describe the potential impact of artificial intelligence on the future.	6.2	6	2.77	7.1	7	1.82	7.8	8	1.46	Included, Round 3
13	distinguish AI applications that already exist from AI applications that are still in the future.	7.3	8	2.23	7.7	8	0.94	n.a., d.	n.a., d.	n.a., d.	Included, Round 3
14	describe what knowledge representation means.	6.0	7	2.71	6.3	7	2.15	n.a., d.	n.a., d.	n.a., d.	Excluded, Round 3
15	explain how AI applications make decisions.	7.3	8	2.15	7.8	8	1.33	n.a., d.	n.a., d.	n.a., d.	Included, Round 3
16	explain how AI-expert systems work.	6.3	6	2.48	5.7	6	2.24	n.a., d.	n.a., d.	n.a., d.	Excluded, Round 3
17	explain how machine learning works at a general level.	7.2	8	2.26	7.8	8	1.44	n.a., d.	n.a., d.	n.a., d.	Included, Round 3
18	describe how machine learning models are trained, validated, and tested.	6.8	7	2.46	7.3	7	1.73	n.a., d.	n.a., d.	n.a., d.	Included, Round 3
19	explain the difference between ’supervised learning’ and ’unsupervised learning’ (in the context of machine learning).	7.0	7	2.25	7.3	8	1.96	n.a., d.	n.a., d.	n.a., d.	Included, Round 3
20	explain how ’reinforcement learning’ works on a basic level (in the context of machine learning).	6.2	7	2.51	6.5	7	1.82	n.a., d.	n.a., d.	n.a., d.	Included, Round 3
21	explain how deep learning relates to machine learning.	6.3	7	2.67	7.2	7	1.59	n.a., d.	n.a., d.	n.a., d.	Included, Round 3
22	explain what the term ’artificial neural network’ means.	6.9	7	2.24	7.6	8	1.01	n.a., d.	n.a., d.	n.a., d.	Included, Round 3
23	critically evaluate the implications of artificial intelligence applications in at least one subject area.	8.1	8	1.93	8.6	9	1.02	n.a., f.	n.a., f.	n.a., f.	Included, Round 2
24	explain why data plays an important role in the development and application of artificial intelligence.	9.2	10	1.32	n. a., f.	n.a., f.	n.a., f.	n.a., f.	n.a., f.	n.a., f.	Included, Round 1
25	describe why humans play an important role in the development of artificial intelligence systems.	8.4	9	2.12	9.0	9	1.13	n.a., f.	n.a., f.	n.a., f.	Included, Round 2
26	describe how some artificial intelligence systems can act in their environment and react to their environment.	7.1	7	2.08	6.8	7	1.71	n.a., d.	n.a., d.	n.a., d.	Included, Round 3
27	explain how sensors are used by computers to collect data that can be used for AI purposes.	6.2	7	2.72	7.3	8	1.8	6.4	6	2.06	Included, Round 3
28	name applications in which AI-assisted computer vision is used.	6.6	7	2.61	6.5	7	1.81	n.a., d.	n.a., d.	n.a., d.	Excluded, Round 3
29	name applications in which AI-assisted natural language processing/ understanding is used.	6.9	7.5	2.66	7.3	7	1.72	n.a., d.	n.a., d.	n.a., d.	Included, Round 3
30	identify ethical issues surrounding artificial intelligence.	9.1	10	1.36	n. a., f.	n.a., f.	n.a., f.	n.a., f.	n.a., f.	n.a., f.	Included, Round 1
31	explain what the term ’black box’ means in relation to artificial intelligence systems.	7.9	9	2.29	8.7	9	1.23	n.a., f.	n.a., f.	n.a., f.	Included, Round 2
32	describe how biases arise in AI systems.	8.4	9	1.86	9.3	10	1.00	n.a., f.	n.a., f.	n.a., f.	Included, Round 2
33	critically reflect on the potential impact of artificial intelligence on individuals and society.	7.6	8	2.37	8.7	9	1.23	n.a., f.	n.a., f.	n.a., f.	Included, Round 2
34	give a short overview about the history of artificial intelligence.	4.2	4	2.44	2.4	2	1.22	n.a., f.	n.a., f.	n.a., f.	Excluded, Round 2
35	explain what the term ‘artificial intelligence winter’ means.	3.8	3	2.46	1.7	2	0.65	n.a., f.	n.a., f.	n.a., f.	Excluded, Round 2
36	explain why AI has recently become increasingly important.	7.1	7	1.94	7.3	8	1.57	n.a., d.	n.a., d.	n.a., d.	Included, Round 3

(*continued on next page*)

Table 2 (*continued*)

	Item	Round 1			Round 2			Round 3			Inclusion/ Exclusion,
	I can …	*M*	*Mdn*	*SD*	*M*	*Mdn*	*SD*	*M*	*Mdn*	*SD*	Round
37	describe what a Turing test is supposed to find out.	5.9	6	2.73	5.2	6	2.36	n.a., d.	n.a., d.	n.a., d.	Excluded, Round 3
38	explain how rule-based systems differ from machine learning systems.	7.1	7	2.47	7.6	7.5	1.61	n.a., d.	n.a., d.	n.a., d.	Included, Round 3
39	explain how decision tree systems work.	6.4	7	2.35	7.1	7	1.57	n.a., d.	n.a., d.	n.a., d.	Excluded, Round 3
40	assess if a problem in my field can and should be solved with artificial intelligence methods.	7.5	8	2.05	8.4	9	1.55	n.a., f.	n.a., f.	n.a., f.	Included, Round 2
41	describe what artificial intelligence is.	n.a., 2nd	n.a., 2nd	n.a., 2nd	8.3	9	1.94	9.5	10	0.64	Included, Round 3
42	describe the concept of explainable AI.	n.a., 2nd	n.a., 2nd	n.a., 2nd	7.8	8	1.69	8	8	1.44	Included, Round 3
43	* explain why data security must be considered when developing and using artificial intelligence applications. & explain why data privacy must be considered when developing and using artificial intelligence applications.	n.a., 2nd	n.a., 2nd	n.a., 2nd	8.4	9	1.94	8.75	9	1.59	Included, Round 3
44	describe the concept of big data.	n.a., 2nd	n.a., 2nd	n.a., 2nd	7.8	8	2.04	8.2	8	1.26	Included, Round 3
45	give examples from my daily life (personal or professional) where I might be in contact with artificial intelligence.	n.a., 2nd	n.a., 2nd	n.a., 2nd	8.8	9	1.58	9	9	0.98	Included, Round 3
46	explain what an algorithm is.	n.a., 2nd	n.a., 2nd	n.a., 2nd	7.5	8	2.15	8	8	1.6	Included, Round 3
47	describe potential legal problems that may arise when using artificial intelligence.	n.a., 2nd	n.a., 2nd	n.a., 2nd	7.1	7.5	2.40	7.4	8	1.65	Included, Round 3

n.a., f. = Not applicable, final decision was made. n.a., d. = Not applicable, dichotomous decision. n.a., 2nd = Not applicable, item created after first round.
Note: The items presented here represent the final item options which were selected by the participants. The final seven items were generated from the initial responses from Round 1, which is why the statistical characteristics for these items are reported starting with Round 2. After Round 1, three items were included in the final questionnaire version because the median score was ten. After Round 2, 11 items were included in the final questionnaire version because the median score was at least nine. In addition, after Round 2, four items were excluded from the final questionnaire version because the median score was five or lower. *At the beginning, "data security" and "data privacy" were combined in one item. However, the SMEs decided that this item should be divided into two items.

4.1. *Significance of the findings*

As described earlier, a strong increase in scientific AI literacy publications and popular scientific AI literacy courses, books, etc. has been observed in recent years (Laupichler et al., 2022; Long & Magerko, 2020; Ng et al., 2021a). While there are various efforts to improve AI literacy of non-experts, there is no way to assess individuals' AI literacy, which has a detrimental effect on AI literacy research. Therefore, this research project was conducted to support the development of one of the first measurement tools to assess AI literacy in non-experts.

The primary concern in this study was to assess the content validity of the items in a reliable manner. While content validity is the basic prerequisite for the existence of a meaningful questionnaire and should accordingly be given the highest priority (Zamanzadeh et al., 2014), it is often only evaluated through methodologically problematic procedures, or disregarded completely. Especially for a topic as new and complex as AI literacy, simply assessing content validity by a small sample of SMEs would be problematic. This is especially true when the experts are not selected from a large pool of potential participants, but are personally chosen by the researchers, which can, for example, lead to selection bias (Blackwell & Hodges, 1957). To circumvent this problem, we contacted over 450 potential experts, of whom 53 contributed their heterogeneous opinions.

It must be mentioned in this regard that the experts rated 81% of all items (38 out of 47) as relevant for capturing AI literacy. On the one hand, this could mean that the expert evaluation or the exclusion criteria were too insensitive. On the other hand, it could also be that this large number of items is necessary to validly capture the rather complex model of AI literacy.

Another interesting finding is that attitudes or affective components toward AI do not appear in the item set generated in this study. This is true for both the initial 40 items and the items suggested by the SMEs. Thus, the item set differs from the AI scales presented in the theory section. While some of these were developed specifically to assess AI attitudes (Sindermann et al., 2021; Schepman & Rodway, 2022), even the scales primarily developed to assess AI literacy often contain some items covering affective components. For example, Lin et al. (2021) and Shih et al. (2021) reported an AI literacy scale with two factors, "teamwork" and "attitudes toward AI." The item set presented here contains some items that could be loosely connected to the "teamwork" factor, for example "I can describe why humans play an important role in the development of artificial intelligence systems." or "I can assess if a problem in my field can and should be solved with artificial intelligence methods." However, no items from this item set seem fit to the proposed "attitudes toward AI"-factor, although this has to be examined further by conducting factor analyses. This is consistent with the content of most AI literacy definitions, which are more concerned with knowledge and understanding of AI, its application, evaluation, and creation, and AI ethics (Ng et al., 2021b).

The most recent scale, which is also the only one that has been psychometrically studied ("Artificial Intelligence Literacy Scale", Wang et al., 2022), does not include attitude items. However, as already described, it does not contain many items that test understanding or knowledge about AI. The item set presented here has several items that can be interpreted as enabling individuals to assess their knowledge about AI and its most important subfields (e.g., machine learning). Exemplary items that are concerned with AI understanding would be "I can describe the concept of explainable AI." or "I can explain how deep learning relates to machine learning.". Since most researchers include a knowledge and understanding component in their definitions of AI literacy, it should be included in an AI literacy scale.

4.2. *Strengths*

This research project developed the first freely available item set for assessing the AI literacy of non-experts. This work thus forms the basis for the development of a psychometrically evaluated, generally applicable AI literacy assessment scale.

The primary strength of the research presented here is the elaborate face and content validation of the item set. While measures of external validity (i.e., construct validity, criterion validity) are usually evaluated in relative detail, too little attention is paid to content validity in the development of tests and questionnaires. By involving more than 50 experts and repeatedly evaluating the relevance, we achieved a high content validity of the item set, ensuring the representativeness of the items for AI literacy.

Another advantage of the item set presented here is that all of the items are listed in this article and can therefore be considered “open access”. This is not the case with other AI literacy scales, as they describe the use of the scales but do not report the content (i.e., the items). Thus, other researchers cannot use the items for their own research or replicate the corresponding studies. Moreover, in our case, both the included and excluded items were reported, so that readers can evaluate whether they agree or disagree about the correctness of the SMEs' decisions.

4.3. Limitations

Even though the main objective of this study was to develop and validate an AI literacy item set, it can be considered a limitation that no factor analysis was performed using a test sample. Conducting an exploratory factor analysis would have the advantage of identifying the common factors underlying AI literacy (Mulaik, 2010). In addition to different benefits for questionnaire development and presentation, this could even support the development of AI literacy theories, as most proposed AI literacy subfields are currently based on purely theoretical considerations. Furthermore, with the help of factor analysis it would be possible to reduce the total number of variables (Wirtz & Nachtigall, 2004), which in turn would have a positive effect on participants' commitment and reduce respondent fatigue (Schatz et al., 2012). Therefore, it must be reiterated that the item set presented here is not a definitive AI literacy scale, but an item set whose applicability as an AI literacy scale in real-world settings can only be evaluated through future research.

Another issue that all AI literacy questionnaires encounter is the selection of the most appropriate AI literacy definition. Since a valid AI literacy item set is, by its nature, intended to "measure a representative sample of the subject matter" (APA, 2022), the definition of the item must be as precise and unchallengeable as possible. However, due to the plethora of different AI literacy definitions (e.g., Kandlhofer et al., 2016; Long & Magerko, 2020; Ng et al., 2021b), it is impossible to use a single universally valid definition as a basis. Theoretically, instead of using Long & Magerko's (2020) definition, we could have presented the SMEs with any other definition as a starting point. However, the choice of this particular AI literacy representation was not arbitrary. Rather, we used it because it is the most widely accepted and most frequently cited definition. This does not necessarily mean that it could not be improved, but at least it provides a commonly accepted foundation.

Finally, two methodological limitations have to be considered. First, the SME selection method resulted in a sample that was predominantly from academia and higher education. However, the opinions of representatives from other subpopulations, such as industry or secondary education, might reveal slightly different AI literacy items. Future research projects should therefore investigate the extent to which the item set can be usefully applied in areas outside of higher education. Second, the choice of the consensus criteria is rather uncommon when compared to other Delphi studies (see Table 1 in Giannarou & Zervas, 2014). Although the rules described in Fig. 2 reflect empirically based decisions, they nevertheless have the disadvantage of being based, at least in part, on decisions made by the research team. This, however, is due to the fact that the initially planned measure of consensus turned out to be infeasible in the context of this study (as described in section 2.3), which is why the alternative had to be deployed.

4.4. Future research directions

The next major step should be to distribute the item set to a larger normative sample. The data obtained from this can be used to further test the psychometric properties of item set and to develop a final (i.e. non-preliminary) AI literacy scale. This would entail conducting a factor analysis and reliability testing. In addition to psychometric evaluations, other questions arise. For example, it could be examined whether the finalized scale could also be used as a pre/post or then/post assessment for the evaluation of AI literacy courses. In addition, it would be useful to examine the extent to which AI literacy and attitudes toward AI or trust in AI are related. It can be hypothesized that an increase in AI literacy correlates with higher trust in AI, a relationship that has been found for scientific literacy as well (Einsiedel, 1994). Last but not least, the participating SMEs in this study themselves pointed out an interesting research direction, namely the development of a knowledge or skills test (as opposed to a psychological questionnaire). The item set presented in this work has the goal to enable the development of a scale for the assessment of the AI literacy of non-experts. In the future, however, it may become equally important to *test* the AI literacy of individuals, for example in the sense of a classic multiple choice knowledge test. Companies, among others, could use this knowledge or skills test to assess the AI literacy of applicants without bias, avoiding social desirable responses.

5. Conclusion

With the generation of the AI literacy item set, we responded to the call for ways to assess AI literacy, which was expressed by several researchers. The purpose of this study was to generate a set of potential items for assessing AI literacy and to test its representativeness for the AI literacy construct. Future research will examine the further psychometric properties of the item set. This concerns both an additional evaluation of validity by distributing the questionnaire to a sample population, as well as the testing of reliability and objectivity. We therefore want to encourage other research teams to use the item set as an preliminary assessment tool to further evaluate the questionnaire in an iterative manner.

Funding

This work was supported by the Open Access Publication Fund of the University of Bonn. The fund is paying the costs of the article processing charge and had no influence, substantive or otherwise, on the scientific process or content of this work.

Ethical approval

The study was conducted between July 2022 and October 2022 by the University Hospital Bonn and the University of Bonn, Bonn, Germany. Participation in the study was voluntary and participants gave their written informed consent to participate in the study. The study was approved by the Research Ethics Committee of the University of Bonn (Reference 194/22).

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

First and foremost, we would like to thank the subject matter experts for sharing their valuable and thoughtful insights with us. Furthermore, we would like to thank Mike Bernd, Elena Trunz, Marko Jovanovic and

Stephan Jonas for their support in the development of the initial item set. Last but not least, we would like to thank all colleagues who helped with the organization and evaluation of the study as well as with feedback regarding the manuscript: Manuel Müller, Jana Schirch, Saskia Zimmer, and all others.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.caeai.2023.100126.

References

Adjabi, I., Ouahabi, A., Benzaoui, A., & Taleb-Ahmed, A. (2020). Past, present, and future of face recognition: A review. *Electronics, 9*(8), 1–53. https://doi.org/10.3390/electronics9081188

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction. In *Conference on human factors in computing systems - proceedings*. https://doi.org/10.1145/3290605.3300233. May 2.

APA. (2022). Content validity. *APA Dictionary of Psychology*. https://dictionary.apa.org/content-validity.

Bentley, F., Luvogt, C., Silverman, M., Wirasinghe, R., White, B., & Lottridge, D. (2018). Understanding the Long-term Use of smart speaker Assistants. In *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* (pp. 1–24). https://doi.org/10.1145/3264901, 2(3).

Blackwell, D., & Hodges, J. L. (1957). Design for the Control of selection bias. *The Annals of Mathematical Statistics, 28*(2), 449–460.

Braganza, A., Chen, W., Canhoto, A., & Sap, S. (2021). Productive employment and decent work: The impact of AI adoption on psychological contracts, job engagement and employee trust. *Journal of Business Research, 131*, 485–494. https://doi.org/10.1016/j.jbusres.2020.08.018

Chowdhury, S., Budhwar, P., Dey, P. K., Joel-Edgar, S., & Abadie, A. (2022). AI-employee collaboration and business performance: Integrating knowledge-based view, socio-technical systems and organisational socialisation framework. *Journal of Business Research, 144*, 31–49. https://doi.org/10.1016/j.jbusres.2022.01.069

Council of Europe. (2022). Global scale - table 1 (CEFR 3.3): Common reference levels. *Common European Framework of Reference for Languages (CEFR)*. https://www.coe.int/en/web/common-european-framework-reference-languages/table-1-cefr-3.3-common-reference-levels-global-scale.

Dai, Y., Chai, C. S., Lin, P. Y., Jong, M. S. Y., Guo, Y., & Qin, J. (2020). Promoting students'well-being by developing their readiness for the artificial intelligence age. *Sustainability, 12*(16). https://doi.org/10.3390/su12166597

Daugherty, P. R., & Wilson, H. J. (2018). *Human+ machine: Reimagining work in the age of AI*. Harvard Business Press.

Davenport, T. H., Brynjolfsson, E., McAfee, A., & Wilson, H. J. (2019). *Artificial intelligence: The insights you need from Harvard business review*. Harvard Business Press.

Einsiedel, E. F. (1994). Mental Maps of science: Knowledge and attitudes among Canadian adults. *International Journal of Public Opinion Research, 6*(1), 35–44. https://doi.org/10.1093/ijpor/6.1.35

Faruqe, F., Watkins, R., & Medsker, L. (2021). *Competency model approach to AI literacy: Research-based path from initial framework to model*. arXiv preprint arXiv:2108.05809.

Giannarou, L., & Zervas, E. (2014). Using Delphi technique to build consensus in practice. *Journal of Business Science and Applied Management, 9*(2).

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A Functional approach to concepts and methods. *Psychological Assessment, 7*(3).

Hsu, C.-C., & Sandford, B. A. (2007). The Delphi technique: Making sense of consensus. *Practical Assessment, Research and Evaluation, 12*, 10. https://doi.org/10.7275/pdz9-th90

Jacob, S. R., & Warschauer, M. (2018). Computational thinking and literacy. *Journal of Computer Science Integration, 1*(1). https://doi.org/10.26716/jcsi.2018.01.1.1

Kandlhofer, M., Hirschmugl-Gaisch, S., & Huber, P. (2016). Artificial intelligence and computer science in education: From Kindergarten to university. In *2016 IEEE frontiers in education conference (FIE)* (pp. 1–9).

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

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

Lin, C.-H., Yu, C.-C., Shih, P.-K., & Wu, L. Y. (2021). International Forum of educational Technology & Society STEM based artificial intelligence learning in general education for non-engineering Undergraduate students. *Technology in Society, 24*(3), 224–237. https://doi.org/10.2307/27032867

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Conference on human factors in computing systems - proceedings*. https://doi.org/10.1145/3313831.3376727. April 21.

Mulaik, S. A. (2010). *Foundations of factor analysis* (2nd ed.). Taylor & Francis.

Ng, A., & DeepLearning.AI. (2022). In *AI for Everyone*. www.coursera.org/learn/ai-for-everyone.

Ng, D. T. K., Leung, J. K. L., Chu, K. W. S., & Qiao, M. S. (2021a). AI literacy: Definition, teaching, evaluation and ethical issues. In *Proceedings of the association for information science and technology* (pp. 504–509). https://doi.org/10.1002/pra2.487, 58(1).

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021b). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence, 2*. https://doi.org/10.1016/j.caeai.2021.100041

Schatz, R., Egger, S., & Masuch, K. (2012). The impact of test Duration on user fatigue and reliability of subjective Quality ratings. *Journal of the Audio Engineering Society, 60*(1), 63–73.

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

Shih, P. K., Lin, C. H., Wu, L. Y., & Yu, C. C. (2021). Learning ethics in AI-teaching non-engineering undergraduates through situated learning. *Sustainability, 13*(7). https://doi.org/10.3390/su13073718

Sindermann, C., Sha, P., Zhou, M., Wernicke, J., Schmitt, H. S., Li, M., & Montag, C. (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.

de Souza, C. E. C. (2021). *What if AI is not that fair? - understanding the impact of fear of algorithmic bias and AI literacy on information disclosure*. BI Norwegian Business School. Master thesis.

University of Helsinki, & MinnaLearn. (2018). In *Elements of AI*. https://www.elementsofai.com/.

Waldmann, A., Liebl, A., & Gerbert, P. (2020). In *Einführung in die KI*. www.ki-campus.org/courses/einfuehrungki2020.

Wang, B., Rau, P. L. P., & Yuan, T. (2022). In *Measuring user competence in using artificial intelligence: Validity and reliability of artificial intelligence literacy scale*. Behaviour and Information Technology. https://doi.org/10.1080/0144929X.2022.2072768.

Wang, Y. Y., & Wang, Y. S. (2022). Development and validation of an artificial intelligence anxiety scale: An initial application in predicting motivated learning behavior. *Interactive Learning Environments, 30*(4), 619–634. https://doi.org/10.1080/10494820.2019.1674887

Wirtz, M., & Nachtigall, C. (2004). *Deskriptive statistik* (3rd ed.). Juventa Verlag.

Wolff, A., Gooch, D., Cavero Montaner, J. J., Rashid, U., & Kortuem, G. (2016). Creating an understanding of data literacy for a data-driven Society. *Journal of Community Informatics, 12*(3), 9–26. www.ci-journal.net/index.php/ciej/article/view/1286.

Zamanzadeh, V., Rassouli, M., Abbaszadeh, A., Majd, H. A., Nikanfar, A., & Ghahramanian, A. (2014). Details of content validity and objectifying it in instrument development. *Nursing Practice Today, 1*(3), 163–171. http://npt.tums.ac.ir.

Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. In *ACM computing surveys*. Association for Computing Machinery. https://doi.org/10.1145/3285029. Vol. 52, Issue 1.