Affective and Pleasurable Design, Vol. 123, 2024, 120–130
https://doi.org/10.54941/ahfe1004683

Development of an AI Literacy Scale Using Multiple-Choice Questions

Xinyun Li, Qin Gao, and Pei-Luen Patrick Rau

Department of Industrial Engineering, Tsinghua University, Beijing 100084, China

ABSTRACT

With continuously emerging and developing artificial intelligence (AI) technologies, we now have more opportunities to interact with AI agents, to use AI applications to assist our jobs, and to assess the solutions provided by AI. This ability to properly identify, use, evaluate, and collaborate with AI-related products is referred to as AI literacy (Long & Magerko, 2020; Wang et al., 2022). The objective of the current study is to develop an instrument to measure general users' AI literacy by replacing subjective self-report questions with objective, multiple-choice questions. 12 questions were derived from four dimensions of AI literacy (i.e., awareness, evaluation, ethics, and future AI), and a total of 230 validated responses were collected through the online survey. After deleting an unqualified item, the explorative factor analysis revealed a 3-factor structure of the remaining 11 items in the AI literacy scale: interacting with AI products, understanding AI's capabilities, and understanding AI's limitations. Each sub-scale is of acceptable reliability and validity. Furthermore, we examined the relationships between AI literacy and actual use of AI products, digital literacy, the attitude towards AI agents, and individual characteristics such as gender and education. The results suggested that a higher level of overall AI literacy was associated with better digital literacy and a more positive attitude towards AI agents. The ability to interact with AI literacy, however, was correlated with more negative feelings about AI agents. Gender and differences in education were shown to have a significant impact on AI literacy. This exploratory study contributes to developing a more objective measurement of general users' AI literacy and provides some insights on users' attitudes towards AI.

Keywords: Artificial intelligence literacy, Digital literacy, Attitude towards artificial intelligence

INTRODUCTION

The continuously emerging and developing artificial intelligence (AI) technologies have greatly changed our lives. Thanks to the growing number of smart devices integrated with AI, we now have more opportunities than ever to interact with AI agents, to use AI applications to assist our jobs, and to assess the solutions that AI has to offer. This ability to properly identify, use, evaluate, and collaborate with AI-related products is referred to as AI literacy (Long & Magerko, 2020; Wang et al., 2017). Similar to computer literacy and digital literacy (Ng, 2012; Wilson et al., 2015), for general AI users, being AI literate does not require comprehensive understandings of the underlying concepts or algorithms. Instead, it stresses effectively and efficiently using AI products and applications to assist one's daily life and job.

For AI designers and researchers, it is important to understand general users' human-AI interaction (HAII), which gives rise to the need to quantitatively measure AI literacy. Yet the AI literacy measurement instruments now in use may fall short of this goal for two reasons. Firstly, while the definition of AI literacy implies an objective ability, the majority of AI measurements rely on self-report questionnaires, which limit their results to subjective assessment. Secondly, as the conceptual framework of AI literacy is continuously evolving, some of the measurements may be incomplete and require updating.

The objective of the current study is to develop a more objective instrument to measure general users' AI literacy. Theoretical models of AI literacy from earlier studies were adopted and combined as the basic structure of our AI literacy scale. Data was gathered through an online questionnaire survey in China with 230 valid responses, and the instrument's reliability and validity were evaluated. Based on the scores of our AI literacy scale, we then briefly analyzed the individual characteristics related to AI literacy.

Conceptual Frameworks of AI Literacy

Various conceptual frameworks of AI literacy have been proposed, either for a particular population (Cetindamar et al., 2022; Chiu et al., 2022; Kandlhofer & Steinbauer, 2018; Kim & Youngjun, 2022; Zhang et al., 2022), for a particular type of AI product (Wienrich & Carolus, 2021), or for general users' common usage of AI (Long & Magerko, 2020; Wang et al., 2022). The former two types of AI literacy frameworks usually rely heavily on the features of the population or AI product. For instance, AI literacy for students may emphasize knowledge about the technical concepts and processes of AI (Kandlhofer & Steinbauer, 2018; Zhang et al., 2022), including intelligent agents, automata, machine learning, etc. The AI literacy framework suggested for digital workplaces should include additional aspects and dimensions related to work, human-machine interaction, and learning (Cetindamar et al., 2022). Similarly, in Wienrich & Carolus's conversation agent literacy model (2021), they identified five dimensions that are all closely connected to users' understandings of technologies used in smart speakers. These AI literacy frameworks have the strength of being domain-specific, but this strength will turn into a problem when applying them to broader users and usages.

General AI literacy models, in contrast, do not focus on particular populations or products. The conceptual framework built by Long and Magerko (2020) covers five basic questions about AI: What is AI? What can AI do? How does AI work? How should AI be used? How do people perceive AI? These 5 questions then led to 17 core competencies for AI users and 15 design considerations for AI designers. More recently, in the AI literacy framework proposed by Wang et al. (2022), AI literacy was decomposed into four dimensions: (1) awareness, (2) skills, (3) evaluation, and (4) ethics. Although these two models can be applied to more general AI users, they fail to include users' expectations and understandings of the evolution of AI in the future—an ability referred to as future literacy (Liveley, 2022)—in their framework. This ability is a significant competency for general AI users because understanding where future AI technology is heading will allow them to anticipate and prepare for the impacts that AI will have on society, careers, and daily activities (Kim & Youngjun, 2022; Zhang et al., 2022).

Measurement Tools for AI Literacy

The large body of AI literacy measurements has been developed for high school or college students to meet the practical needs of evaluating the educational outcomes of AI curricula. The primary focus of these instruments, in line with the learning goals, is students' AI-related knowledge and their ability to program, read, and evaluate AI algorithms (Chiu et al., 2022; Kim & Youngjun, 2022; Zhang et al., 2022). Some of these measurement tools additionally incorporate ethical issues and practices that may arise from AI as well as the societal impacts of AI (Kim & Youngjun, 2022; Zhang et al., 2022) for a more comprehensive assessment of students' AI literacy. The problem with these tools, though, is that they do not put enough emphasis on the ability to interact with AI products.

Another type of AI literacy measurement focuses on the AI literacy specific to particular AI products. For instance, Wienrich & Carolus (2021) developed a conversation agent literacy scale (CALS) with 29 multiple-choice items based on Long and Magerko's framework (2020). The criterion validity and internal consistency of CALS were examined and found to be acceptable. Yet the items in this scale are mostly limited to conversation agents, which can hardly be generalized to other AI products. Further, this scale does not address the ethical issues of using AI products or the impacts of future AI technology.

Different from the measurement tools designed for the target population or product, the tools for general AI literacy should measure not only the knowledge of AI technology but also the capability of properly using AI products to aid daily activities. Wang et al. (2022) have developed the AI literacy scale (AILS) for this purpose. The AILS contains 12 items from four dimensions (i.e., awareness, usage, evaluation, and ethics), and its reliability and validity have been examined. Despite being a well-developed tool for measuring general AI literacy, the AILS requires some modification to cover the future literacy dimension.

Furthermore, the aforementioned tools, except for CALS, all have the common limitation that they were constructed using self-report questions. Despite the convenience of the self-report question, its potential for bias needs to be noted (Wang et al., 2022; Wienrich & Carolus, 2021). According to Wang et al. (2022), respondents in their study tended to receive high scores in the AILS, probably as a result of overconfidence. This result suggests that self-report questions may not be the ideal approach to measuring AI literacy because they only assess subjective, perceived self-ability rather than objective AI literacy.

METHODS

Questionnaire Development

The current study developed an AI literacy scale with four subscales—awareness (AW), evaluation (EV), ethics (ET) (B. Wang et al., 2022), and future AI (FU) (Liveley, 2022)—each with three items. The items were adapted from existing AI literacy measurement tools (Kim & Youngjun, 2022; Wang et al., 2022), but the statements were changed to multiple-choice questions to assess respondents' AI literacy in a more objective manner. Each multiple-choice question contained five options, which were statements about AI technology, applications, and common use cases. These options were derived from popular science publications as well as trending topics such as AI painting, in an attempt to avoid making the questions too abstruse for average AI users.

A pilot study was conducted to determine whether the expressions and the difficulty of the questions were appropriate before the questionnaire was distributed. 5 volunteers participated in the pilot study. All of them are proficient users of digital devices and hold bachelor's degrees in science or engineering. Based on their feedback, we modified some of the questionnaire's expressions and adjusted the difficulty of the three multiple-choice questions within a subscale from easy to normal to challenging. The final version of the AI literacy scale is presented in Table 1.

Table 1. The AI literacy scale (without specific item options).

Item	Item Content
AW1	Select from the following options: the smart device(s).
AW2	Select from the following options: correct match(es) of AI technology and the help it can offer.
AW3	Select from the following options: the correct match(es) of AI technology and the applications or products it was used in.
EV1	Select from the following options: the correct combination(s) of AI application or product—its capability—its limitation.
EV2	Select from the following options: the correct statement(s) of the solutions provided by AI applications or products.
EV3	Select from the following options: the correct match(es) of AI application or product and the task it can solve.
ET1	Select from the following options: the ethical use(s) of AI applications and products.
ET2	Select from the following options: the use case(s) of AI applications and products in which you may have privacy or information security concerns.
ET3	Select from the following options: the use case(s) in which you think the AI technology is abused.
FU1	Select from the following options: the capabilities of AI technology that may be realized in the future but not realized yet.
FU2	Select from the following options: the potential impacts of AI technology on the society or the possible changes it may bring to your daily life.
FU3	Select from the following options: the correct statements of the possible influence of AI technology on future careers and the changes it may bring.

In addition to objective AI literacy, we examined AI usage, digital literacy and attitudes towards AI in the questionnaire. AI usage was assessed with 3 self-report questions, which capture whether respondents' use AI applications to assist their daily lives. Digital literacy was measured by the digital literacy scale (Ng, 2012), including 10 items in three dimensions: the technical dimension (6 items), the cognitive dimension (2 items), and the social-emotional dimension (2 items). This construct refers to the literacy associated with the use of digital technologies (Ng, 2012; Wilson et al., 2015), and it was found to be strongly associated with AI literacy in previous studies (Wang et al., 2022; Wienrich & Carolus, 2021). Attitudes towards AI were evaluated through the Negative Attitude towards Robots Scale (NARS), which was initially developed to measure people's negative attitude towards communication robots (Nomura et al., 2008). It measured attitude in 14 items from three constructs: interaction with robots (S1, 6 items), social influence of robots (S2, 5 items), and emotional interactions with robots (S3, a reversely coded sub-scale with 3 items). The word "robots" was replaced by "AI agents" in our questionnaire. We measured the respondents' attitude towards AI because it was also found to be related to AI literacy (Wang et al., 2022).

Item Scoring

For each multiple-choice question, 1 point was awarded/deducted for correctly/incorrectly judging each of the 5 options, yielding a total score that might range between −5, −3, −1, +1, +3, +5. This scoring technique was used to correct the bias generated from guessed answers. The items of AI usage, digital literacy, and NARS were rated on a 5-point Likert scale.

Questionnaire Distribution

The questionnaire was distributed online. An attentiveness question and a reversed question were included to filter out the random responses. The remaining responses were checked by the researcher to ensure that the answers were thoughtfully chosen. The attentive respondents were then given a small incentive payment. A total of 230 validated responses were collected, with a relative even distribution in gender (male = 127, female = 103). The age of the respondents ranged from 18 to 74, with a mean age of 27.83 (SD = 8.84). Most of the respondents (85.7%) had a bachelor's degree or above. Table 2 displays the details of the demographic information.

Table 2. Demographic information of respondents.

Demographics		Sample Size	Percentage (%)
Gender	Female	103	44.78
	Male	127	55.22
Age	<20	7	3.04
	[20, 30)	167	72.61
	[30, 40)	29	12.61
	[40, 50)	15	6.52
	>=50	12	5.22

(Continued)

Table 2. Continued

Demographics		Sample Size	Percentage (%)
Education	High school or lower	6	2.61
	Junior college	27	11.74
	Bachelor	139	60.43
	Master	40	17.39
	Ph. D	18	7.83

RESULTS

Exploratory Factor Analysis

Exploratory Factor Analysis (EFA) was carried out to investigate the factorial structure of the 12 multiple-choice questions in the AI literacy scale. First, we checked the factorability of the data. The Kaiser-Meyer-Olkin (KMO) measure of the whole data was 0.85, above the recommended value of 0.6, indicating good sampling adequacy. Yet the item FU1 had a KMO value of 0.37, which was lower than the acceptable limit of 0.5 (Kaiser & Rice, 1974), and its correlations with other items were all below 0.3. Therefore, we removed this item from the scale. The KMO measure of the remaining items reached 0.87, and the Bartlett's test was significant ($\chi^2\,(105)\;=\;630.87, p <$ .001), which suggested that the remaining 11 items were suitable for factor analysis.

To identify the factorial structure with the best interpretability, a parallel analysis was conducted with 2-, 3-, and 4-factor structures, respectively. The 3-factor structure derived from principal component analysis with oblimin rotation was best suited, explaining 56% of the total variance. Oblique rotation was used since we assumed that the components should be correlated rather than independent of each other. The results of the EFA were displayed in Table 3.

The first factor (INT) we identified is related to the ability to interact with AI ethically and critically, which includes items such as identifying the ethical use of AI (ET1, 2, 3) and critically evaluating AI's applications and influences (AW1, EV3, FU3). The other two factors are both concerning the understanding of AI technologies. The capability factor (CAP) captures the knowledge of AI's capabilities, such as what technologies are used (AW2, AW3) and what they can do (EV2, FU2), whereas the limitation factor (LIM) is about the knowledge of AI's limitations (EV1).

Table 3. Results of the EFA.

			Loadings		
Constructs	Description	Item	INT	CAP	LIM
Interaction (INT)	The ability to interact with AI ethically and critically.	ET2	**0.77**	−0.08	0.01
		ET1	**0.76**	−0.15	−0.13
		FU3	**0.56**	0.22	0.21

(Continued)

Table 3. Continued

Constructs	Description	Item	Loadings INT	CAP	LIM
		EV3	**0.52**	0.12	0.36
		AW1	**0.52**	0.25	−0.16
		ET3	**0.50**	0.39	0.21
Capability (CAP)	The understanding of AI's capabilities.	AW2	0.06	**0.78**	−0.39
		AW3	−0.02	**0.66**	0.26
		EV2	−0.12	**0.63**	0.36
		FU2	0.31	**0.55**	0.11
Limitation (LIM)	The understanding of AI's limitations.	EV1	0.06	−0.01	**0.79**
Variance Explained			23%	21%	12%

Reliability and Validity Analysis

For each factor, we evaluated the internal consistency by Cronbach's alpha values, the convergent validity by composite reliability (CR) and average variance extracted (AVE) (Fornell & Larcker, 1981), and the criterion validity by its correlation with other factors (i.e., intercorrelations) and with AI usage, digital literacy, and attitudes towards AI. The results were presented in Table 4.

The values of Cronbach's alpha, CR, and AVE for each factor and the full scale were close to or higher than the recommended values (i.e., 0.7 for Cronbach's alpha and CR and 0.5 for AVE), suggesting acceptable internal consistency and convergent validity. The three factors of AI literacy were all significantly correlated with the overall AI literacy score, and their intercorrelations were acceptable, according to Clark & Watson (2016). With respect to the relationship between AI literacy and other constructs, in general, a higher level of AI literacy was found to be associated with a higher level of digital literacy and a more positive attitude towards AI, as we expected. It should be noted, however, that although the overall attitude went more positive with higher AI literacy, the emotional feelings towards AI were more negative with a higher INT factor score or higher AI literacy score.

Table 4. Analysis of reliability and validity.

	INT	CAP	LIM	AI literacy[a]
	Reliability and validity metrics			
Cronbach's alpha	.77	.68	—	.82
CR	.82	.81	.64	.91
AVE	.45	.51	.64	.49
	Correlations			
Interaction	—	**.56**[b]	**.26**	**.93**
Capability	—	—	**.29**	**.82**
Limitation	—	—	—	**.42**
AI usage	.01	**.18**	.09	.09

(Continued)

Table 4. Continued

	INT	CAP	LIM	AI literacy[a]
Digital literacy (DL)	−.04	**.17**	.13	.06
DL-technical	−.09	**.13**	.12	.01
DL-cognitive	−.07	.13	.10	.01
DL-social emotional	**.14**	**.26**	**.13**	**.22**
NARS	−.01	**−.17**	**−.21**	−.10
NARS-interaction	**−.32**	**−.35**	**−.25**	**−.39**
NARS-influence	.11	−.00	−.10	.06
NARS-emotional (R)[c]	**−.36**	−.08	.03	**−.27**

a. The AI literacy score was calculated as the average score of the items in the INT, CAP, and LIM factors.
b. The significant correlations were marked by bold texts.
c. The emotional dimension of NARS was a reversed dimension. Higher scores indicate a more positive attitude.

AI Literacy Scores

The descriptive statistics of AI literacy scores are displayed in Table 5. According to the results, the mean scores of each factor and the overall AI literacy scale were beyond the midpoint, suggesting the respondents had a relatively high level of AI literacy. The item difficulty indices of the subscales and the overall scale were around 0.8, indicating that for each item, approximately 20% of the responses were correct (i.e., received +5 points) on average.

Figure 1 shows the relationship between AI literacy scores and demographic variables. Female respondents had significantly higher AI literacy scores compared to male respondents ($M_1 = 2.11, M_2 = 1.57, t = 3.10, p = .002, d = .41$). Respondents older than 25 years (i.e., the sample median) had slightly higher AI literacy scores than the younger adults ($M_1 = 1.96, M_2 = 1.62, t = 1.89, p = .060, d = .25$). The respondents' education level was also found to positively correlate with their AI literacy score ($\rho = .15, p = .027$). Their learning experience did not affect the AI literacy score, however.

Table 5. Descriptive statistics of AI literacy scores.

	INT	CAP	LIM	AI literacy
Min	−2	−2.50	−3	−1.18
Max	5	5	5	5
Mean[a]	1.81	1.91	1.45	1.81
SD	1.66	1.50	1.84	1.35
Item Difficulty[b]	.77	.82	.91	.80

a. The mean factor score is calculated as the mean score of the factor items.
b. Item difficulty is calculated as (1 − *count of correct responses/all responses*).

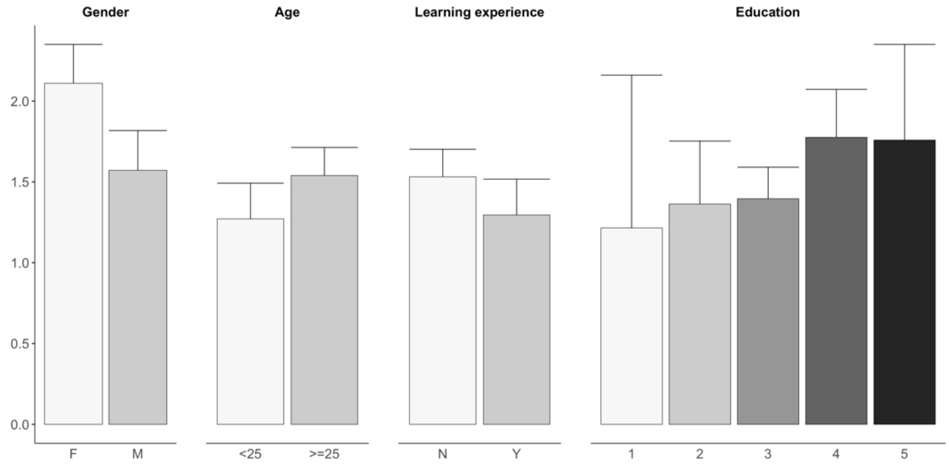

Figure 1: Relationship between AI literacy and gender, age, learning experience, and education. Learning experience: N—no, Y—yes. Education: 1—high school or lower, 2—junior college, 3—bachelor, 4—master, 5—Ph.D.

DISCUSSION

The current study developed a scale composed of 11 multiple-choice questions and 3 self-report questions in an attempt to measure AI literacy from a more objective perspective. The EFA suggested a 3-factor structure of the 11 items, which we labeled as interaction (INT), capability (CAP), and limitation (LIM). Subsequent analysis suggested acceptable reliability and validity for this explorative scale.

Our respondents' AI literacy score was, in general, positively correlated with their digital literacy and their attitude towards AI, consistent with previous studies (Wang et al., 2022; Wienrich & Carolus, 2021). A closer examination of the correlation patterns allowed us to reveal some new insights. First, we did not find any significant correlations between the respondents' self-reported AI usage and their scores on the INT factor, the LIM factor, or the overall AI literacy scale. This decoupling suggests that, compared to the subjective self-report measures, our objective AI literacy scores may reflect distinct underlying aspects of AI literacy. It also demonstrated the necessity to assess users' AI literacy based on their real knowledge and proficiencies with AI instead of how well they perceived themselves to know and behave. Furthermore, we found higher scores on AI literacy and the INT factor were connected to a more negative attitude towards emotional interaction with AI agents, contrary to the results found with subjective measures of AI literacy (Wang et al., 2022). This could be explained by the fact that a higher level of AI literacy would allow a more objective evaluation of AI agents and hinder the development of emotional connections.

CONCLUSION

This current study developed a new instrument to measure general users' AI literacy from a more objective prospective than the self-report measures.

This AI literacy scale allows AI researchers and designers to evaluate users' knowledge of AI technologies as well as their proficiency with AI agents in a quantitative manner. Although we managed to incorporate some recent cases of AI products and applications into the scale, when applying this scale, these cases are subject to modifications according to respondent characteristics and new developments in the AI field. It is also recommended to combine the cases with more specialized AI domains to better serve the research goal.

The limitations of the current study should be noted. First, based on our sample, we found gender, age, and education to be the potential influencing factors of AI literacy, but this result needs further validation with larger and more diverse samples. Second, we used self-report questions to measure AI usage in this study, but future research can consider other approaches such as direct observation and experimentation to further eliminate the influence of subjective answers. Finally, our AI literacy scale does not fit in the proposed 4-factor conceptual framework, possibly due to the limited number of items. Therefore, future studies can add more items to the scale to examine the conceptual framework of AI literacy.

REFERENCES

Cetindamar, D., Kitto, K., Wu, M., Zhang, Y., Abedin, B., & Knight, S. (2022). Explicating AI Literacy of Employees at Digital Workplaces. *IEEE Transactions on Engineering Management*, 1–14. https://doi.org/10.1109/TEM.2021.3138503

Chiu, T. K. F., Meng, H., Chai, C.-S., King, I., Wong, S., & Yam, Y. (2022). Creation and Evaluation of a Pretertiary Artificial Intelligence (AI) Curriculum. *IEEE Transactions on Education*, *65*(1), 30–39. https://doi.org/10.1109/TE.2021.3085878

Clark, L. A., & Watson, D. (2016). *Constructing validity: Basic issues in objective scale development* (p. 203). American Psychological Association. https://doi.org/10.1037/14805-012

Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, *18*(1), 39–50. https://doi.org/10.1177/002224378101800104

Kaiser, H. F., & Rice, J. (1974). Little Jiffy, Mark Iv. *Educational and Psychological Measurement*, *34*(1), 111–117. https://doi.org/10.1177/001316447403400115

Kandlhofer, M., & Steinbauer, G. (2018). A Driving License for Intelligent Systems. *Proceedings of the AAAI Conference on Artificial Intelligence*, *32*(1). https://doi.org/10.1609/aaai.v32i1.11399

Kim, S.-W., & Youngjun, L. (2022). The Artificial Intelligence Literacy Scale for Middle School Students. *Journal of The Korea Society of Computer and Information*, *27*(3), 225–238. https://doi.org/10.9708/jksci.2022.27.03.225

Liveley, G. (2022). AI Futures Literacy. *IEEE Technology and Society Magazine*, *41*(2), 90–93. https://doi.org/10.1109/MTS.2022.3173357

Long, D., & Magerko, B. (2020). What is AI Literacy? Competencies and Design Considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–16. https://doi.org/10.1145/3313831.3376727

Ng, W. (2012). Can we teach digital natives digital literacy? *Computers & Education*, *59*(3), 1065–1078. https://doi.org/10.1016/j.compedu.2012.04.016

Nomura, T., Kanda, T., Suzuki, T., & Kato, K. (2008). Prediction of Human Behavior in Human–Robot Interaction Using Psychological Scales for Anxiety and Negative Attitudes Toward Robots. *IEEE Transactions on Robotics*, *24*(2), 442–451. https://doi.org/10.1109/TRO.2007.914004

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

Wang, P., Fang, W., & Guo, B. (2017). A colored petri nets based workload evaluation model and its validation through Multi-Attribute Task Battery-II. *Applied Ergonomics*, *60*, 260–274. https://doi.org/10.1016/j.apergo.2016.11.013

Wienrich, C., & Carolus, A. (2021). Development of an Instrument to Measure Conceptualizations and Competencies About Conversational Agents on the Example of Smart Speakers. *Frontiers in Computer Science*, *3*. https://www.frontiersin.org/articles/10.3389/fcomp.2021.685277

Wilson, M., Scalise, K., & Gochyyev, P. (2015). Rethinking ICT literacy: From computer skills to social network settings. *Thinking Skills and Creativity*, *18*, 65–80. https://doi.org/10.1016/j.tsc.2015.05.001

Zhang, H., Lee, I., Ali, S., DiPaola, D., Cheng, Y., & Breazeal, C. (2022). Integrating Ethics and Career Futures with Technical Learning to Promote AI Literacy for Middle School Students: An Exploratory Study. *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-022-00293-3