PACIS 2023 Proceedings

Pacific Asia Conference on Information Systems (PACIS)

7-8-2023

Toward an Objective Measurement of AI Literacy

Patrick Weber
Goethe University Frankfurt, weber@wiwi.uni-frankfurt.de

Marc Pinski
Technical University of Darmstadt, pinski@ise.tu-darmstadt.de

Lorenz Baum
Goethe University Frankfurt, baum@wiwi.uni-frankfurt.de

Toward an Objective Measurement of AI Literacy

Completed Research Paper

Patrick Weber
Goethe University Frankfurt
Theodor-W.-Adorno-Platz 4
60323 Frankfurt am Main, Germany
weber@wiwi.uni-frankfurt.de

Marc Pinski
Technical University of Darmstadt
Hochschulstraße 1
64289 Darmstadt, Germany
pinski@ise.tu-darmstadt.de

Lorenz Baum
Goethe University Frankfurt
Theodor-W.-Adorno-Platz 4
60323 Frankfurt am Main, Germany
baum@wiwi.uni-frankfurt.de

Abstract

Humans multitudinously interact with Artificial Intelligence (AI) as it permeates every aspect of contemporary professional and private life. The socio-technical competencies of humans, i.e., their AI literacy, shape human-AI interactions. While academia does explore AI literacy measurement, current literature exclusively approaches the topic from a subjective perspective. This study draws on a well-established scale development procedure employing ten expert interviews, two card-sorting rounds, and a between-subject comparison study with 88 participants in two groups to define, conceptualize, and empirically validate an objective measurement instrument for AI literacy. With 16 items, our developed instrument discriminates between an AI-literate test and a control group. Furthermore, the structure of our instrument allows us to distinctly assess AI literacy aspects. We contribute to IS education research by providing a new instrument and conceptualizing AI literacy, incorporating critical themes from the literature. Practitioners may employ our instrument to assess AI literacy in their organizations.

Keywords: AI Literacy, AI Competence, AI Skills, Objective Measurement Scale, Scale Development

Introduction

"Without universal AI literacy, AI will fail us."
World Economic Forum (2022)

Information Systems (IS) research and practice alike are increasingly engaging in the adoption and application of artificial intelligence (AI), dubbing its arrival as the "*dawn of a new age*" (Berente et al., 2021, p. 1433). For various decision-makers, AI-based systems evolved into essential parts (Fügener et al., 2022), thus offering companies opportunities to exploit their economic potential (Collins et al., 2021). AI is among the most important general-purpose technologies of the 21st century, and research and practice apply it in multiple areas, e.g., decision support systems, search engines, voice recognition, personalized training, spam detection, and autonomous driving (e.g., Brynjolfsson & McAfee, 2017; Russell et al., 2016; Abdel-Karim et al., 2021). In modern society, humans multitudinously interact with AI in their professional and private lives (Benbya et al., 2021). Exemplary interactions include developers creating new AI in their professional environment, e.g., technicians at Apple updating the voice assistant Siri, or lay users using AI in their private life, e.g., car drivers using Google Maps to navigate and circumvent traffic jams. Furthermore, as it permeates our society (Furman & Seamans, 2019), AI results are also increasingly being evaluated by individuals from third parties. These parties might include individuals from the government or individuals working in other regulatory institutions like the EU, in which individuals have to know how to handle AI in order to evaluate its compliance with regulations (European Commission, 2021). Evaluation interactions by individuals working in regulatory institutions are getting increasingly relevant in highly regulated application domains like healthcare, automotive, and finance, where regulatory authorities prescribe high levels of transparency due to high risks for involved humans (Weber et al., 2023).

We define AI literacy as a set of socio-technical competencies of humans that shape relevant types of human-AI interaction. Firms worldwide consider the lack of AI literacy the biggest challenge for digital transformation (Brock & von Wangenheim, 2019), thus highlighting AI literacy's relevance once more for IS education. Also, a proper understanding of AI is needed to implement AI in businesses successfully, e.g., as it works best for narrowly defined, data-rich, repeated tasks (Leprince-Ringuet, 2020).

All interactions between humans and AI are (heavily) shaped by technical features of the AI and the human's competencies, including prior experiences with AI (Amershi et al., 2019). While IS research has long been concerned with the technical features of AI, only recently are academics increasingly investigating the human counterpart – their competencies – to further improve human-AI interactions (Long & Magerko, 2020). Thereby, the competencies' importance increases for all stakeholders, not only technical ones (Barredo Arrieta et al., 2020). Human knowledge appears subjective and objective (Brucks, 1985), and their respective subjective and objective knowledge measurements can differ significantly (Carlson et al., 2009; Dunning, 2011). Literacy literature (Babiarz & Robb, 2014; Kiechle et al., 2015; Miller, 2016; T. H. Nguyen et al., 2017; Schaffner, 2005) provides strong evidence for the need of objective next to subjective literacy measurement, which would also benefit IS education. For example, T. H. Nguyen et al. (2017) explicitly compare subjective and objective health literacy measures, with the latter's central benefit being that "*there is inherent value to having empirically grounded data*" (T. H. Nguyen et al., 2017, p. 198). Further, Schaffner (2005) recommends objective literacy measures, when comparing them to their subjective counterparts in developing countries. Measuring perceptions of AI knowledge, like subjective AI literacy, has hitherto been approached with subjective scales (e.g., Wang et al., 2022); although a standard method in IS research, this may not serve as a good proxy for evaluating a person's AI literacy. The developed subjective instruments let humans judge their proficiency on Likert-type scales with statements relating to different aspects of AI literacy they can agree or disagree with to some degree. While conceptual and subjective measurement research on AI literacy steadily increases (Chiu et al., 2021; Wang et al., 2022), there is – to the best of our knowledge – no objective measurement instrument regarding AI literacy available so far. Following scientific calls (Laupichler et al., 2022; Weber, 2023), our paper offers a measurement instrument for AI literacy which is, contrary to prior instruments, objective, to measure individuals' AI-related competencies (T. H. Nguyen et al., 2017).

We followed an established scale development procedure for subjective measurements in IS research (MacKenzie et al., 2011) and adapted it – only where necessary – to obtain an objective measurement tool for AI literacy. After reviewing relevant literature, we defined AI literacy, i.e., our focal measurement construct, and conceptualized its dimensions based on established IS theory. Subsequently, we specified our instrument through n = 10 expert interviews, generating a set of n = 56 items in line with our definition and conceptualization and representing several aspects of AI literacy. We refined our items and selection through these interviews and twice inviting in total n = 12 peer IS researchers for a card-sorting exercise. After further item refinement and filtering, we conducted a between-subject comparison study with n = 88 participants outside academia: an AI-literate test group and a control group. We concluded with final instrument evaluation and refinement, ending with our final scale consisting of n = 16 items.

We contribute to research and practice alike. First, we provide a novel, objective measurement instrument for AI literacy research grounded in an established, rigorous scale development procedure (MacKenzie et al., 2011), thus, advancing the toolbox of IS researchers. Second, we provide a definition and conceptualization of AI literacy, incorporating multiple stakeholders, therefore delivering a concept other researchers may follow, adopt, validate, or extend. Third, we structure future IS education research on objective AI literacy measurement and invite future researchers to engage in the interplay of subjective and objective measurement of AI literacy or specialized objective AI literacy measurement scales for particularly interesting stakeholder groups. Practitioners may also benefit from the findings of our study. Organizations willing to investigate AI literacy may utilize our scale, e.g., a company may validate the effectiveness of IS education programs regarding AI by employing the scale before and after training or compare AI literacy levels among departments to identify needs for training.

Our research article is structured as follows. After this introductory section, we will elaborate on the theoretical background, thus presenting prior (AI) literacy literature next to the theoretical foundations of our conceptualization, namely the socio-technical perspective in IS (Sarker et al., 2019) and Bloom's taxonomy (Bloom et al., 1956; Krathwohl, 2002). We continue explaining our research design, highlighting the scale development procedure, including definition and conceptualization, instrument specification and item development, and instrument evaluation and refinement. We conclude with a discussion followed by limitations and future research.

Theoretical background

Prior technology literacy work and the emergence of AI literacy

The concept of *literacy* in the technology context has a long history in IS research (Bassellier et al., 2003; Wolfe, 1992). Historically, in the domain of Computer Science (CS), researchers conceptualized general computer literacy and IT competencies to guide the education of either IS professionals or students (Bassellier et al., 2003; Hindi et al., 2002; Wiesche et al., 2020). IS research and adjacent research fields have already applied the concept of literacy in a broader context and defined a multitude of *literacies*, such as digital literacy (Eshet-Alkalai, 2004; Gilster, 1997; A. Nguyen et al., 2020; Norman & Skinner, 2006), data literacy (Kerpedzhiev et al., 2021; Someh et al., 2019; Ongena, 2023), or statistical literacy (Gonda et al., 2022). With digital technology becoming ubiquitous in our economies and societies (Diaz Andrade & Techatassanasoontorn, 2021), the literacy concepts related to digital technology have also evolved beyond pure technical expert education. These concepts include more diverse stakeholders, such as lay users or domain experts who apply technology in their contexts. Hence, technology literacies are incorporating more and more competencies like critical thinking or ethical judgment, which are related to the social context of human-technology interaction (Feerrar, 2019; Pfeuffer et al., 2023).

The rise of AI significantly impacts our understanding of technology competence. AI invalidates core assumptions, such as functional consistency and transparency, which have guided IS research for decades (Schuetz & Venkatesh, 2020). Furthermore, the phenomenon of AI is expanding rapidly in scope across all human parts of life, such as mobility, health, economy, and others (Benbya et al., 2021; Berente et al., 2021). Given that AI is qualitatively different from traditional technology and relevant to many aspects of human life, IS research acknowledged that prior literature concepts are becoming insufficient. This urgent need formed a relatively young literature stream on AI literacy. Until today, research on AI literacy focuses primarily on exploring and conceptualizing its facets. Long and Magerko (2020) put the topic on the landscape and provide an initial definition of AI literacy, highlighting the competence to evaluate and use AI and to communicate and collaborate with it. They collocate a list of competencies that served as a starting point for further exploration. Heyder and Posegga (2021) built on this list and identified that apart from a *technical* part, there are *critical* and *sociocultural* aspects to AI literacy. Others distinguish technical aspects from non-technical, content-focused aspects, building on classifications like TPACK (Technological, Pedagogical, and Content Knowledge) (Ng et al., 2021). The implicit splitting of AI literacy further supports this segmentation into a technical and a social part, derived through a literature review (Cetindamar et al., 2022) or the analysis of job advertisements (Anton et al., 2020). In summary, the segmentation theme by *part* of the human-AI interaction in the literature distinguishes the AI itself (technical part) and the social context (social part), which comprises cultural, content-focused, and further non-technical aspects of AI. The *type* of interaction with the respective AI segments AI literacy further (Wang et al., 2022; Ng et al., 2021). For example, humans can use or develop AI, which constitutes two different types of interaction that necessitate entirely different sets of competencies.

Theoretical underpinnings of human-AI interaction dimensions

We identified two dominant segmentation approaches in the AI literacy literature, which compose dimensions of the human-AI interaction, i.e., *parts* and *types*. In the following section, we explore the theoretical foundations of these two segmentation approaches. The socio-technical perspective emphasizes the two core *parts* that constitute an IS artifact – the human and the artifact, which is the AI in this case (Sarker et al., 2019). An IS artifact can generally be defined as a "*system which assembles, stores, processes, and delivers information relevant to an organization (or to society), in such a way that the information is accessible and useful to those who wish to use it*" (Buckingham et al., 1986, p. 18). Hence, it is a system that mediates social and technical interaction (Ågerfalk, 2020). As a result, an IS artifact can, e.g., manipulate human communication (Stamper et al., 2000), enforce social norms in organizations (Aakhus et al., 2014), and express human values and beliefs (Goldkuhl & Agerfalk, 2005). Sarker et al. (2019) describe the equal recognition of the mutually interacting social and technical parts of such systems formed under the term *socio-technical perspective* as an axis of cohesion for the IS discipline. Adopting this socio-technical perspective when conceptualizing AI literacy implies that next to the technical aspects of AI training, such as development competencies, also social parts need to be included. Humans and AI interact in a socio-technical system, necessitating a literacy concept to cater to both parts.

Next to that, Bloom's taxonomy describes various competencies that relate to different *types* of interactions (Bloom et al., 1956). They developed the taxonomy to set up a common language for learning and to serve as a framework for determining the curriculum of a particular field of study (Krathwohl, 2002). Bloom et al. (1956) cluster competencies by the underlying educational objective independent of the field of study in question. Academia has been widely using this framework**,** and one of its creators extended it with a second dimension (Krathwohl, 2002). We leverage the revised version of the framework but focus on the original cognitive process dimension only since the second dimension adds complexity to the framework, which is unnecessary for our purposes. The revised cognitive process dimension distinguishes six consecutive categories of competencies: remembering, understanding, applying, analyzing, evaluating, and creating (Krathwohl, 2002). *Remembering* refers to the retrieval of relevant knowledge from memory, while *understanding* means the competence to determine meaning from instructional messages. *Applying* is carrying out a procedure in a given situation, whereas *analyzing* is decomposing material into its constituent parts. Lastly, *evaluating* includes the competence to judge based on specific criteria, and *creating* refers to putting elements together in a novel form. Each of these competence categories facilitates a different type of interaction. Adopting a segmentation by interaction type along Bloom's taxonomy when conceptualizing AI literacy implies the recognition of multiple stakeholders. For example, an AI developer needs to *create*, while an AI regulator needs to *evaluate*. Furthermore, an AI user needs to *remember* and *understand*.

Research design

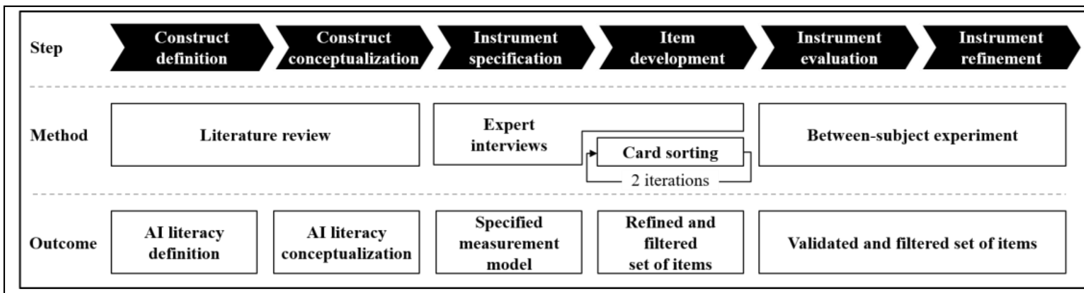

Figure 1. Scale development procedure adapted from MacKenzie et al. (2011)

This section elaborates on the process we followed throughout this study. Figure 1 provides an overview. To the best of our knowledge, there is no scale development procedure concerning objective measurement. Similar research papers concerning objective measurement followed no explicit procedure, or at least did not report so (e.g., Aggarwal et al., 2015; Vetter et al., 2011). Others simply adapted existing objective measurements, e.g., tests (Motta et al., 2018). However, in existing literature, there are similarities to our chosen procedure, e.g., employing experts for item generation (Ament, 2017). Hence, we based our approach on the procedure given by MacKenzie et al. (2011), adapting it for objective measurement and first extracted and synthesized a definition from the literature through a literature review. Afterward, we derived a conceptualization from our definition and specified our instrument through n = 10 interviews with experts from academia, generating the first set of items (n = 56). These interviews, together with peer IS researchers performing a first preliminary card-sorting (n_1 = 7), led us to refine our items and selection. Afterward, we performed a second round of card-sorting with different peer IS researchers (n_2 = 5) and further refined and filtered our item set. Integrating these insights led us to employ the scale in our between-subject comparison study (n = 88) to evaluate and refine our instrument, leaving us with a final validated and filtered set of items (n = 16).

Definition and conceptualization of the focal measurement construct

Definition. We start our scale development procedure by defining the focal measurement construct. First, we screened relevant literature from the IS and CS domain using multiple keywords (i.e., “ai competencies”, “ai abilities”, “ai knowledge”, “ai literacy”) and databases (i.e., ACM Digital Library, AIS electronic Library, Google Scholar, IEEE Xplore, Web of Science) to obtain valuable resources regarding the definition of the focal construct. We performed the search in July 2022. After removing duplicates, we screened titles and abstracts excluding irrelevant publications and further narrowing down our result set. Wherever necessary, we screened full texts to decide on the relevancy of publications. Finally, we performed forward- and backward search to derive the following set of n = 10 relevant publications containing AI literacy definitions.

Source	Definition of AI literacy
Cetindamar et al. (2022)	Four aspects of AI literacy: *Technology-related AI literacy* is about “*data, its collection and collation, analysis, and representation, the range of technologies for this, and their interaction with the world*”. *Work-related AI literacy* is “*understanding of human and AI capabilities within the domain, and interactions of capabilities across teams and disciplines, over time and complex decision environments*”. *Human-machine-related AI literacy* is “*skills to build and use intelligence augmentation recognising potential of human-robot-interaction in bolstering human capabilities*”. *Learning-related AI literacy* is “*skills that will allow the development of adaptive expertise to draw on self-learning and technological developments*”.
Mikalef and Gupta (2021)	AI capability “*is the ability of a firm to select, orchestrate, and leverage its AI-specific resources*”.
Hermann (2022)	AI literacy is “*individuals' basic understanding of (a) how and which data are gathered; (b) the way data are combined or compared to draw inferences, create, and disseminate content; c) the own capacity to decide, act, and object; (d) AI's susceptibility to biases and selectivity; and (e) AI's potential impact in the aggregate*”.
Kandlhofer and Steinbauer (2018)	AI literacy is “*sound knowledge about the principles of AI and its application*“.
Dai et al. (2020)	AI literacy is “*a student's ability to access and use AI-related knowledge and skills*”.
Chiu et al. (2021)	Chiu et al. (2021) focus (employee subjective) knowledge of AI, while subjective knowledge refers to “*consumer's perception of the amount of information they have stored in their memory*” (Flynn & Goldsmith, 1999).
Pinski and Benlian (2023)	Pinski and Benlian (2023) define general AI literacy as “*humans' socio-technical competence consisting of knowledge regarding human and AI actors in human-AI interaction, knowledge of the AI process steps, that is input, processing, and output, and experience in AI interaction*”.

Table 1. Further AI literacy definitions derived from literature review

From the literature, we extracted multiple definitions of which we want to introduce three as a subset highlighting relevant aspects of AI literacy for the scale development (see Table 1 for further definitions). Long and Magerko (2020, p. 2) define AI literacy as “*a set of competencies that enables individuals to evaluate AI technologies critically; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace*”. Hence, they highlight different competencies humans possess regarding AI, which they can deploy in various contexts. Next to that definition, Ng et al. (2021) identified four aspects of AI literacy, which they subsequently defined. The first three aspects correspond to Long and Magerko's definition of various competencies and were defined as follows (Ng et al., 2021, p. 4): Knowing and understanding AI is to “*[k]now the basic functions of AI and how to use AI applications”;* using and applying AI refers to “*[a]pplying AI knowledge, concepts, and applications in different scenarios”;* evaluating and creating AI is about “*[h]igher-order thinking skills (e.g., evaluate, appraise, predict, design) with AI applications*”. With the fourth aspect, Ng et al. (2021) add another facet of AI literacy, i.e., the ethical dimension, to the definition. They propose that AI ethics refers to "*[h]uman-centered considerations (e.g., fairness, accountability, transparency, ethics, safety)*" (Ng et al., 2021, p. 4). With this extension, they are in line with Wang et al. (2022, p. 1), who defined AI literacy as "*the ability to properly identify, use, and evaluate AI-related products under the premise of ethical standards*". We used these existing definitions to synthesize our definition for the scale development procedure, given as follows:

AI literacy is a set of socio-technical competencies of humans that shape relevant types of human-AI interaction.

With our definition, we highlight multiple aspects of AI literacy. Like other abilities, literacy, or competencies, AI literacy is something humans possess. For example, Engel et al. (1990, p. 281) define the knowledge artifact as "*the information stored within memory*". Similarly, AI literacy in the form of socio-technical competencies is imminent to humans, thus, something they hold within themselves. Next, we highlight the interplay between humans and AI in multiple ways. Accordingly, the application field of AI literacy refers to relevant types of interaction between humans and AI. Therefore, humans and AI must exist and interact with each other; i.e., the human has to start interacting with the AI or vice versa. These interactions may take place by, e.g., using or developing AI. Furthermore, influence shows that the socio-technical competencies of humans shape these types of human-AI interactions. Hence, socio-technical competencies have an actual (and measurable) influence on these types of interactions mentioned above. As a result, interactions may occur in different shapes and are malleable through the individual human's socio-technical competencies.

Conceptualization. Figure 2 provides an overview of how AI literacy influences the types of human-AI interaction derived from Bloom's taxonomy (Krathwohl, 2002). Human stakeholders may belong to different stakeholder classes: evaluator (e.g., an individual working in an regulatory authority), creator (e.g., a developer), or user (e.g., a lay or expert user). These stakeholder classes interact in diverse ways with AI. These interactions include the human stakeholder and the AI affecting each other. For example, a user uses a chatbot (AI) by engaging with it. This interaction is two-sided, as not only does the user manipulate the AI, but the AI's answers also affect the user. Hence, the interaction's influence is, as well, two-sided.

We argue that these interactions are shaped by humans' AI literacy, i.e., socio-technical competencies. As mentioned above, humans possess these socio-technical competencies, and these competencies manifest themselves in different human-AI interaction types. For example, regarding relatively low complexity tasks, like using an AI, a user's AI literacy shapes the interaction between the focal user and the AI, thus possibly determining the effectiveness of the engagement with the AI. A creator interacting with an AI by creating it may face a rather technical AI development, in which their AI literacy shapes the interaction by varyingly accelerating the development process. Similarly, an evaluator interacting with an AI by interpreting its implementation and outputs may be successful to varying degrees, depending on their AI literacy.

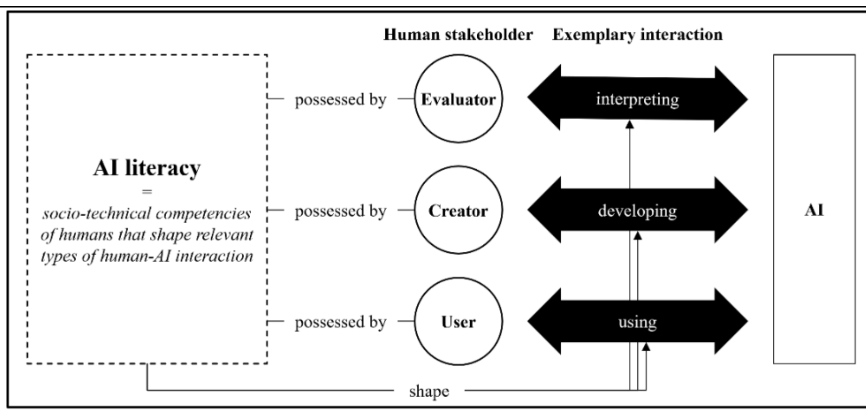

Figure 2. AI literacy's influence in the human-AI interaction landscape

Table 2 gives an overview of the dimensions of human-AI interaction we identified. We distinguish between different *parts* and *types* of human-AI interaction. The two parts of human-AI interaction we identified are

Socio AI Literacy and *Technical AI Literacy*. We derived this dimension from the well-established socio-technical perspective of IS (Sarker et al., 2019). As human-AI interactions happen in a socio-technical environment, requiring socio-centric next to techno-centric competencies, we adopted this scheme to our construct of AI literacy. Accordingly, we define Socio AI Literacy as *competencies regarding social components or social contextualization of human-AI interaction* and Technical AI Literacy as *competencies regarding technical components of human-AI interaction*.

Dimension	Categorization	Definition of AI literacy quadrants
Part of human-AI interaction	Socio AI Literacy	Competencies regarding social components (e.g., human, society) or social contextualization (e.g., algorithmic biases) of human-AI interaction
	Technical AI Literacy	Competencies regarding technical components of human-AI interaction
Type of human-AI interaction	Creator/Evaluator AI Literacy	Competencies regarding analysis, evaluation, and creation of human-AI systems
	User AI Literacy	Competencies regarding recalling, understanding, and applying AI knowledge

Table 2. Dimensions of human-AI interaction and definitions of corresponding quadrants

Next to the part of the human-AI interaction, we inspected the type of human-AI interaction. We distinguish between *User AI Literacy* and *Creator/Evaluator AI Literacy*. We adapted Bloom's taxonomy (Krathwohl, 2002) and divided it into two halves for simplicity. We also follow prior AI literacy research by doing so, like Ng et al. (2021). We group the lower half of Bloom's taxonomy, i.e., remembering, understanding, and applying, and label them User AI Literacy, representing relatively low complexity tasks. The upper half of Bloom's taxonomy, i.e., analyzing, evaluating, and creating, forms our so-called Creator/Evaluator AI Literacy representing relatively high-complexity tasks. Subsequently, we define User AI Literacy as *competencies regarding recalling, understanding, and applying AI knowledge* and Creator/Evaluator AI Literacy as *competencies regarding analysis, evaluation, and creation of human-AI systems*.

Application of definition and conceptualization. We conclude our definition and conceptualization phase by illustrating our emerging concept of AI literacy in Figure 3. In this figure, we combine both dimensions: part and type of AI literacy. The horizontal axis shows the part of AI literacy with Socio AI Literacy (S) on the left and Technical AI Literacy (T) on the right. The vertical axis refers to the type of AI literacy, with User AI Literacy (U) in the lower half and Creator/Evaluator AI Literacy (CE) in the upper half. Combining these two dimensions with two manifestations each leads to the below shown 2×2 matrix with four quadrants of AI literacy. As our dimensions, part of human-AI interaction with manifestations socio and technical and type of human-AI interaction with manifestations user and creator/evaluator, are derived from holistic scales (Sarker et al., 2019; Krathwohl, 2002), our matrix's quadrants are collectively exhaustive. Thus, our matrix reflects all possible interactions and competencies. We use these four quadrants to generate and sort items (see the following subsection).

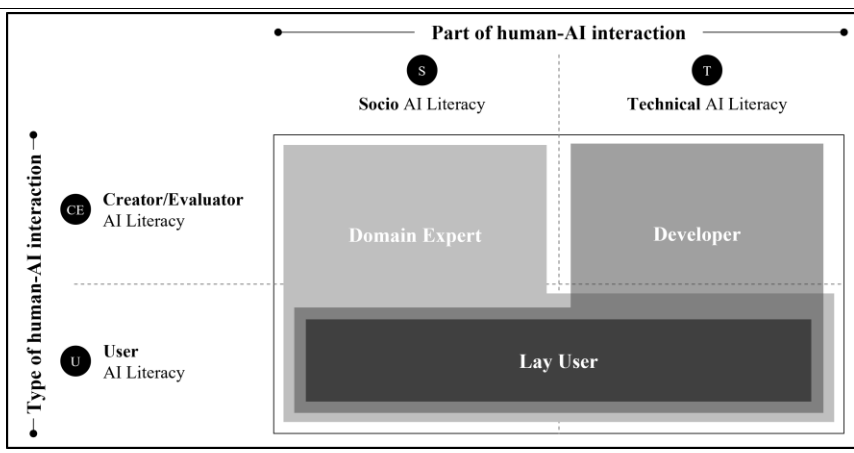

Figure 3. Conceptualization of AI literacy, including exemplary personas

To illustrate our suggested concept of AI literacy, we added exemplary personas to Figure 3. A lay user is an instance of the stakeholder class user (Figure 2). This lay user requires low complexity, i.e., User AI Literacy in Socio AI Literacy and Technical AI Literacy, to successfully use the AI (S-U and T-U). Similarly, a developer is one possible instance of the stakeholder class creator (Figure 2) next to another possible instance being the domain expert, e.g., a doctor detecting cancer or a product manager responsible for an AI-based customer support system. A developer's AI literacy is a superset to a lay user's AI literacy, i.e., the developer has S-U and T-U just as a lay user (as mentioned, the categories of competencies given by Bloom's taxonomy are consecutive (Krathwohl, 2002)). Additionally, as the developer is concerned with the technical development of AI, they are required to have high complexity, i.e., Creator/Evaluator AI Literacy for the technical (T-CE) but not for the socio part. Eventually, domain experts represent another instance of the stakeholder class creator (Figure 2). Furthermore, a domain expert's AI literacy is a superset to a lay user's AI literacy, possessing the same S-U and T-U as a lay user. A domain expert supports the creation process of AI by incorporating domain specifics and AI's non-technical implications for that domain, thus, requiring a relatively high complexity level of AI literacy on the socio part, i.e., Socio Creator/Evaluator AI Literacy (S-CE). Note that this list of exemplary personas is not exhaustive, i.e., one could consider other personas, such as the expert user as an instance of the stakeholder class user or regulator as an instance of the stakeholder class evaluator (Figure 2).

Instrument specification and item development

After we defined AI literacy and developed our focal measurement construct, we continued our scale development procedure by specifying the instrument and developing its items (Figure 1). Here, as a first step, we specified the measurement model. The item development incorporates the results of the previous section, especially the literature review, which already revealed a collection of competencies AI literacy should entail. Additionally, expert interviews served as a qualitative source to the construct and item development, and two rounds of card-sorting helped verify the items' mapping to the parts and types of human-AI interaction, i.e., the assignment to the four quadrants of AI literacy (Figure 3).

There are various ways to measure AI literacy. Ng et al. (2021) cluster and evaluate 30 articles regarding AI literacy in their overview work. While conceptual and subjective research on AI literacy steadily increases (Chiu et al., 2021; Wang et al., 2022), no study investigated an objective measurement instrument regarding AI literacy. As knowledge, e.g., AI literacy, may be present in the forms of objective knowledge and subjective knowledge, we must distinguish between the two. While objective knowledge is what a person actually knows about an artifact, subjective knowledge describes what a person perceives to know (Brucks, 1985). These two are somewhat correlated (Carlson et al., 2009), but research shows that subjective and objective measurements can differ significantly (Dunning, 2011). Humans also tend to overestimate their competencies in general (Moore & Healy, 2008). Therefore, subjective measurements may not serve as a good proxy for evaluating a person's AI literacy leading us to the necessity for developing an objective measurement tool.

As described above, AI literacy consists of a collection of competencies (Table 2) shaping relevant human-AI interaction types. These competencies are not necessarily interchangeable and collectively constitute AI literacy (Jarvis et al., 2003). These properties of AI literacy led to the conclusion that AI literacy can only be measured objectively as a formative construct through the assemblage of its competencies, i.e., proxied by the questions in our construct (Coltman et al., 2008). Furthermore, an increase in one quadrant (Figure 3) does not necessarily increase another, suggesting a formative rather than a reflective measurement model, which would propose such a relationship between the skills (MacKenzie et al., 2011). With the two dimensions, i.e., parts and types of human-AI interaction, forming AI literacy, each quadrant of our measurement construct is also formed by several competencies.

The envisioned tool should measure the subject's AI literacy objectively. To do so, we decided to move away from inherently subjective Likert-type questions and instead construct our instrument with items, i.e., questions, that offer four distinct answer choices, of which only one is correct in each case. Those items are proxies for the competencies AI literacy entails. Presenting four answers per question item represents a trade-off between, on the one hand, possibly overwhelming the subjects with too many answer options and, on the other, low test-specificity (true negative rate) when random guesses have a high chance of being correct (e.g., 50% chance of randomly picking the correct answer with two answers presented). This way, by collecting the list of questions a subject answered correctly (and the list of incorrectly answered questions), the instrument can objectively determine what a subject actually knows and does not know (Brucks, 1985). Weighting the items equally, the proportion of correctly answered items gives an estimate of the subject's AI literacy. Given the four quadrants of AI literacy (Figure 3), we subdivide the overall AI literacy into these quadrants, i.e., S-U, T-U, S-CE, and T-CE. For example, a developer might entail a higher T-CE than S-CE while being equally literate regarding the two parts of User AI Literacy, i.e., S-U and T-U. Another level of abstraction emerges when looking at only one of the two dimensions, i.e., part and type of human-AI interaction. This viewpoint allows testimonies about subjects regarding their User and Creator/Evaluator AI Literacy and their Socio and Technical AI Literacy level. For example, a lay user might have lower Creator/Evaluator AI Literacy while having higher User AI Literacy.

After having specified the measurement model, we developed the construct items. For this, we crafted an initial set of n = 56 items based on the literature review results to completely cover all conceptualized dimensions of human-AI interaction (e.g., Long & Magerko, 2020). Given those, we also sorted the items into the four quadrants of our measurement construct.

Next, we conducted n = 10 expert interviews with researchers from IS, CS, and Management disciplines, each scheduled for 60 minutes, from August to October 2022. Table 3 gives an overview of the interview partners showing the different disciplines and qualification levels, including their level of AI (literacy) engagement and the AI stakeholder perspective they represent. By interviewing not only AI researchers, but AI *literacy* researchers, we ensured incorporating multiple perspectives, especially regarding AI creators and AI evaluators. Additionally, we explicitly interviewed AI users, i.e., AI appliers. We collected thorough qualitative feedback on the focal measurement construct during the interviews using a semi-structured interview approach. We started with general questions about the experts' view on AI literacy and human-AI interaction and asked for feedback on our conceptualization of AI literacy, thereby confirming the results of the first steps of the scale development procedure (Figure 1). Afterward, getting more specific, we presented the items, i.e., the questions and respective answers, in batches to the experts and asked for feedback on the items and their assignment to the four quadrants of AI literacy, including their relevance for and ability to measure the focal construct. Also, the interviews asked whether items seemed inappropriate, questions or answers were not understandable or ambiguous, or crucial topics/competencies were missing in the question set.[1] We aggregated the expert interview results to develop a refined list of items. This refinement included changing the wording of questions and answers, changing the items' assortment to the quadrants of AI literacy, and removing items from the set that were rated low by the experts. After this, a subset of 35 question items remained for further investigation.

Scientific qualification	Discipline	AI engagement	U	C	E	Count (n)
Senior researcher (Professor)	IS	AI literacy researcher	x	x	x	4
	CS	AI literacy researcher	x	x	x	
	Management	AI literacy researcher	x	x	x	
	Management	AI researcher		x	x	
Researcher (Postdoc)	IS	AI researcher		x	x	2
	Management	AI researcher		x	x	
Junior researcher (Ph.D. graduates)	IS	AI researcher		x	x	4
	IS	AI researcher		x	x	
	CS	AI applier	x			
	CS	AI applier	x			
Total (n)						**10**
Note: U: AI User, C: AI Creator, E: AI Evaluator						

Table 3. Overview of Expert Interview Partners (IS: n = 4, CS: n = 3, Management: n = 3; AI literacy researcher: n = 3, AI researcher: n = 5, AI applier: n = 2)

[1] For example, the following questions were dropped *(correct answer in italics):* 1. What distinguishes Deep Learning (DL) from Artificial neural networks (ANNs)? *A. ANNs are shallow models compared to DL models.* B. ANNs can better model the human brain. C. DL delivers higher accuracy measures than ANNs. D. DL lacks behind ANNs in terms of data type flexibility. Question was too complicated. 2. Humans are relevant in the following AI tasks: *A. Programming, choosing models, and fine-tuning AI systems* B. Calculating F1-scores by hand, maintaining readability, and updating C. Manually separating testing data, and cross-validating them D. Skimming procedures and improving processes. Question was too easy.

The next step in the scale development procedure is card-sorting, which aims to confirm the categorization of items. We performed two card-sorting rounds using an online survey tool ($n_1 = 7$ and $n_2 = 5$). Each time, we asked a separate set of peer IS researchers to sort each question item into one of the four quadrants of AI literacy to check the inter-rater agreement. We first presented our definition and conceptualization of AI literacy by explaining the dimensions and categorizations. Then, we presented the question items, including their answers. Afterward, the participants needed to assign the items to one of the four quadrants of the measurement instrument (S-U, S-CE, T-U, T-CE). The results of the first card-sorting were somewhat promising, with a 54.1% hit ratio between allocated and theoretical assignment and Cohen's kappa coefficient $\kappa_1 = 0.25$, although some items seemed to have problems. Such problems were either the participants sorting items into a different quadrant than we did or, worse, the participants' answers diverged by selecting multiple quadrants (low inter-rater agreement). For this reason, we decided to refine the items again and perform another card-sorting study. This item refinement mainly included changing the wording of the questions and answers next to the reassignment of items to the quadrants. After the second card-sorting study, we narrowed the number of items down to a set of n = 25 items by looking at the card-sorting and expert interview results to keep a balanced item set per quadrant.

Table 4a shows the results of the second card-sorting for the final set of n = 16 items (after the evaluation and refinement in the next section). This crosstabulation puts the actual categorization on the left in relation to the participants' answers on the top. The numbers represent the number of allocations. Overall, participants agreed with us regarding 77.5% of the assignments to the four quadrants of AI literacy (hit ratio). Following Hinkin (1998), this indicates an acceptable agreement between allocated and theoretical groups. Also, Cohen's kappa κ indicates a moderate strength of agreement among respondents, with $\kappa_2 = 0.46$ (Landis & Koch, 1977). Table 4b shows the same results grouped by parts and types of AI literacy only. When looking at the assignment of items solely based on the parts (Socio and Technical), participants agree with us in 92.5% of the assignments. Focusing on types of AI literacy (User and Creator/Evaluator), participants agree with us in 83.8% of the assignments.

		Allocated group				**Hit ratios**	
		S-U	S-CE	T-U	T-CE		
Theoretical group	S-U	**13**	3	3	1	0.65	**0.775**
	S-CE	6	**13**	0	1	0.65	
	T-U	1	0	**17**	2	0.85	
	T-CE	0	0	1	**19**	0.95	

Note: S = Socio; T = Technical; U = User; CE = Creator/Evaluator.

Table 4a. Results of second card-sorting (n_2 = 5) for the final items (n = 16)

			Allocated group				**Hit ratios**
			S	T	U	CE	
Theoretical group	Part	S	**35**	4			**0.925**
		T	1	**39**			
	Type	U			**34**	6	**0.838**
		CE			7	**33**	

Note: S = Socio; T = Technical; U = User; CE = Creator/Evaluator.

Table 4b. Results of second card-sorting regarding parts and types of AI literacy for the final items

Instrument evaluation and refinement

The final steps in the scale development procedure include the evaluation of the instrument and the final refinement of the items (Figure 1). For this, we deployed our preliminary construct of AI literacy to an online survey tool. The survey, therefore, included our set of 25 items emerging from the second card-sorting study. The order of the question items and their answers was randomized to avoid order effects. Finally, we included standard demographic questions like age, gender, ethnicity, nationality, level of education, and employment status and asked for experience with AI using a scale with four Likert-type questions (Pinski & Benlian, 2023). To increase the quality of the results, we included four instructional manipulation checks, i.e., attention checks, in the survey process (Oppenheimer et al., 2009).

This study aimed to evaluate the "*convergent, discriminant, and nomological validity*" (MacKenzie et al., 2011, p. 310) of our focal measurement construct for AI literacy by employing a between-subject comparison. We decided to take educational background together with the weekly device usage (mobile phone, tablet, laptop, desktop) to discriminate the groups, as we assume that these are proxies for AI literacy levels (overall, not toward the quadrants). We expect both educational background and weekly device usage to be proxies of how frequently subjects interact with and thus learn by experience about AI. The group we assume to have higher AI literacy should have an educational background in "Computer Science" or "Computing (IT)" and use such devices at least every day or multiple times every day (group A).

We defined the other group with lower expected AI literacy (group B) to come from all remaining educational backgrounds and with lower weekly device usage, i.e., never, once a week, 2-6 times a week.

In cooperation with a market research company through which we acquired our two samples, we ran the experiment in October 2022 and repeated it in May 2023 to increase the sample size. The two samples (groups A and B) are balanced regarding gender, and people from all countries were allowed to participate. In total, we collected $n = 88$ ($n_A = 44$; $n_B = 44$) successful participations by people not failing the attention checks and completing the survey. Table 5 summarizes the demographics and other statistics for both groups. These results confirm a high diversity of the sample, e.g., regarding the level of education and ethnicity, but also some flaws regarding group differences in employment. These can probably be explained by the group selection criteria, especially the subject of education, and differences between groups in the level of education. For the sample size, these differences are within the expected ranges; therefore, we consider the sample sufficient for our evaluation, as, e.g., the education level of group B would indicate higher AI literacy, which was not the case. One aspect to note is that the degree of AI experience measured on an abstract scale is higher for group A, those with CS backgrounds, compared to group B (3.24 > 2.63). This fact increases the robustness of the study, as we are not only comparing AI literacy based on CS backgrounds but also based on stated AI experience.

Statistics of the sample		Group A	Group B	Total
Gender	Female	46%	48%	47%
	Male	52%	52%	52%
	Other	2%	0%	1%
Age*	Mean (years)	24.4	25.0	24.7
Level of education**	Lower education	8%	1%	9%
	High school diploma	27%	23%	25%
	University/College degree	36%	64%	50%
Field of education	With CS background	100%	11%	56%
	Without CS background	0%	89%	44%
Employment	Working (full-time/part-time)	34%	59%	47%
	Not working (unemployed/retired)	39%	32%	35%
	Other	27%	9%	18%
AI experience	Mean***	3.42	2.21	2.82
Correct answers (25 items)	Mean (proportion) (significant with $p < 0.1$)	31.3%	27.5%	29.4%
Correct answers (16 items)	**Mean (proportion) (significant with $p < 0.1$)**	**43.8%**	**38.2%**	**41.0%**

** One person did not report their age. ** Excluding "other" option. Note: This refers to the highest level of education obtained and does not prevent subjects from having a current CS background. *** Scale from 1 "completely disagree" to 7 "completely agree".*

Table 5. Statistics of the between-subject evaluation study sample (n = 88)

The between-subject comparison study with 25 question items asked already revealed a clear difference between groups A and B regarding the portion of correctly answered questions (Table 5). This difference is significant (Mann-Whitney $U = 757$, $n_A = n_B = 44$, $p = 0.076$ two-tailed)[2] and also points in the expected direction, as people from the field of CS or Computing obtained higher levels of AI literacy than people from other fields (means: 31.3% > 27.6%; medians: 32% > 28%)[2]. By combining the results of this study with the results of previous steps in the development procedure, we filtered the set of items (1) to deliver a concise construct that is of appropriate size for practice use and (2) to remove possibly problematic items which, for example, had a relatively high or low rate of correct answers.

From the final set of 16 questions (see Appendix A), group A correctly answered 43.8% (mean; median: 43.8%), while group B correctly answered 38.2% (mean; median: 37.5%). This difference in AI literacy levels between the groups is still significant (Mann-Whitney $U = 757.5$, $n_A = n_B = 44$, $p = 0.076$ two-tailed).

[2] We use the non-parametric Mann-Whitney U test due to a Kolmogorov-Smirnov test rejecting a normal distribution for the AI literacy levels.

Ultimately, with half of the question items correctly answered, the measurement construct also is at a good level of difficulty (neither too easy nor too hard) to successfully discriminate between subjects with different AI literacy levels.

Discussion

This paper sets out to develop an objective scale for AI literacy. Therefore, we reviewed the current literature to derive an actionable definition of AI literacy for objective measurement, conceptualized the construct along two theoretically underpinned dimensions, and subsequently developed, refined, and validated objective measurement items. Our final instrument consists of n = 16 items and successfully discriminates between an AI-literate test and a control group.

Our conceptualization of AI literacy and the operationalized measurement instrument contribute to IS theory in three ways. First, we provide a new, validated instrument for IS academics enabling the measurement of an additional perspective on AI literacy. Prior research aimed to measure AI literacy from a subjective perspective (Wang et al., 2022). While subjective measurement has substantial advantages, such as simplicity in the measurement process, it also has significant drawbacks. Research shows that subjective and objective knowledge is, at best, mildly correlated (Carlson et al., 2009; Dunning, 2011). We extend prior IS research with an instrument that enables efficient, objective AI literacy measurement. While the instrument certainly is not without limitations, its operationalization allows the inclusion into, e.g., survey or experimental studies and therefore constitutes a usable tool that research and practice can easily deploy. Second, we conceptualize AI literacy that incorporates critical themes discussed in prior AI literacy literature. Prior exploratory and conceptual work on AI literacy (Long & Magerko, 2020; Ng et al., 2021; Wang et al., 2022) identifies relevant aspects of AI literacy by explaining diverse competencies that humans can possess. We extend AI literacy literature by conceptualizing all these potential competencies along theoretically grounded, well-established dimensions (socio-technical perspective and Bloom's taxonomy). The conceptualization allows us to identify and assess further competencies that research might not have identified yet. Third, we structure future IS education research within the AI literacy field. Our conceptualization, which other researchers may follow, adopt, validate, or extend, can guide IS academics in choosing promising future research endeavors, e.g., incorporating multiple human stakeholders.

The developed measurement instrument also offers multiple relevant applications for practitioners. Organizations might leverage it, e.g., to assess the status quo of AI literacy within different departments, thus reducing information asymmetry between employer and employee. They could define thresholds of required AI literacy within the different dimensions for specific departments, such as a Technical AI Literacy requirement among developers or a Socio AI Literacy requirement among product managers – or vice versa. An assessment with the tool could yield valuable information on necessary AI training to, e.g., improve inter-departmental cooperation in AI-related projects. Furthermore, organizations could leverage the tool to assess the effectiveness of their IS education programs regarding AI or, if not in place yet, to guide the development of training programs. The segmentation of AI literacy into parts and types of human-AI interaction can help organizations to develop a curriculum for their employees that covers different potential use cases of AI in the organizational context. Regarding the software development lifecycle, organizations may employ our tool at all stages, from development to use.

Limitations, future research, and conclusion

The developed AI literacy measurement instrument has certain limitations that pose interesting avenues for future research: First, a quantitative multiple-choice assessment is, by design, a proxy for the actual underlying competence and hence a simplification of the real world. Therefore, we had to make necessary simplifications, such as focusing on Bloom's cognitive process dimensions to segment our included questions. The multiple-choice format can only test a limited number of competencies and does not allow feedback for the participants during the test and therefore necessitates that the participants understand each question on their own. Furthermore, participants can also randomly get answers right. Future research could develop objective measurement instruments that follow a different format and complement a multiple-choice test. Researchers and practitioners might also be interested in a highly reliable but even shorter scale, increasing applicability in diverse research or organizational settings. Following the rise of generative AI[3], future scholars may enhance our scale with questions regarding prompt engineering.

Second, we purposefully developed a broad instrument to achieve wider applicability, e.g., for different departments within a firm. However, the gain in breath naturally limits the depth of the measurement. Furthermore, we evaluated our measurement in a single study at a single point in time. Also, as AI, like other digital technologies, is constantly evolving, a frequent revisit of the measurements questions is inevitable. Future research might specify the instrument with regards to particularly interesting stakeholder groups, such as a manager AI literacy scale or employee AI literacy scale (Heyder & Posegga, 2021), and develop a refined and more targeted version of the instrument which allows for measuring stakeholder-specific AI literacy objectively in greater depth. For example, IS researchers may focus on one quadrant of our provided matrix and refine its measurement. Additionally, future researchers should replicate our study in other national and organizational contexts, e.g., they could research four groups derived from our conceptualization (see Figure 3), thereby enhancing our employed between-subject comparison study. Finally, our conceptualization itself may be challenged, e.g., the aggregation of creators and evaluators regarding the distinction of different types of human-AI interaction.

Lastly, future IS education research could contextualize the objective measurement of AI literacy regarding other key IS constructs. Studies might investigate, e.g., the relationship between subjective and objective AI literacy or the impact of objective AI literacy on outcomes of human-AI interaction, such as attitude toward AI (Nomura et al., 2006; Sindermann et al., 2021), trust in AI (Gillath et al., 2021) or intention to use AI (Bhattacherjee, 2001). Furthermore, AI literacy alone will not be sufficient to drive successful AI within organizations. Future research might also investigate which complementary capabilities, such as purposeful data culture or providing the necessary infrastructure (e.g., Jöhnk et al., 2021), exist next to AI literacy and how they interact or interfere with AI literacy within the organizational context.

In conclusion, with this paper, we conceptualized and empirically validated an objective measurement instrument for AI literacy. We showed that our developed instrument successfully discriminates an AI-literate test group from a control group. Furthermore, the structure of our instruments allows us to assess distinct aspects of AI literacy. We encourage future research to further investigate our categorization of AI literacy to enable more specific and accurate measurements.

References

Aakhus, M., Ågerfalk, P. J., Lyytinen, K., & Te'eni, D. (2014). Symbolic Action Research in Information Systems: Introduction to the Special Issue. *MIS Quarterly*, *38*(4), 1187–1200.

Abdel-Karim, B. M., Pfeuffer, N., & Hinz, O. (2021). Machine Learning in Information Systems-a Bibliographic Review and Open Research Issues. *Electronic Markets*, *31*(3), 643–670.

Ågerfalk, P. J. (2020). Artificial intelligence as digital agency. *European Journal of Information Systems*, *29*(1), Article 1.

Aggarwal, R., Kryscynski, D., Midha, V., & Singh, H. (2015). Early to Adopt and Early to Discontinue: The Impact of Self-Perceived and Actual IT Knowledge on Technology Use Behaviors of End Users. *Information Systems Research*, *26*(1), 127–144. https://doi.org/10.1287/isre.2014.0564

Ament, C. (2017). *The Ubiquitous Security Expert: Overconfidence in Information Security*.

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. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13.

Anton, E., Behne, A., & Teuteberg, F. (2020). The Humans Behind Artificial Intelligence–An Operationalisation of AI Competencies. *Proceedings of the 28th European Conference on Information Systems (ECIS)*. An Online AIS Conference, June 15-17, 2020. https://aisel.aisnet.org/ecis2020_rp/141

Babiarz, P., & Robb, C. A. (2014). Financial Literacy and Emergency Saving. *Journal of Family and Economic Issues*, *35*(1), 40–50. https://doi.org/10.1007/s10834-013-9369-9

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial

[3] See https://www.jmis-web.org/cfps/JMIS_SI_CfP_Generative_AI.pdf (last access May 19, 2023).

Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115.

Bassellier, G., Benbasat, I., & Reich, B. H. (2003). The influence of business managers' IT competence on championing IT. *Information Systems Research*, *14*(4), 317–336.

Benbya, H., Pachidi, S., & Jarvenpaa, S. (2021). Special Issue Editorial: Artificial Intelligence in Organizations: Implications for Information Systems Research. *Journal of the Association for Information Systems*, *22*(2). https://doi.org/10.17705/1jais.00662

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly*, *45*(3), 1433–1450.

Bhattacherjee, A. (2001). Understanding information systems continuance: An expectation-confirmation model. *MIS Quarterly*, 351–370.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). Taxonomy of educational objectives: The classification of educational goals. *New York: David McKay*.

Brock, J. K.-U., & von Wangenheim, F. (2019). Demystifying AI: What digital transformation leaders can teach you about realistic artificial intelligence. *California Management Review*, *61*(4), 110–134.

Brucks, M. (1985). The effects of product class knowledge on information search behavior. *Journal of Consumer Research*, *12*(1), 1–16.

Brynjolfsson, E., & McAfee, A. (2017). The Business of Artificial Intelligence. What it can – and cannot – do for your organization. *Harvard Business Review*. https://starlab-alliance.com/wp-content/uploads/2017/09/The-Business-of-Artificial-Intelligence.pdf

Buckingham, R. A., Hirschheim, R. A., Land, F. F., & Tully, C. J. (Eds.). (1986). *Information systems education: Recommendations and implementation*. Cambridge University Press.

Carlson, J. P., Vincent, L. H., Hardesty, D. M., & Bearden, W. O. (2009). Objective and subjective knowledge relationships: A quantitative analysis of consumer research findings. *Journal of Consumer Research*, *35*(5), 864–876.

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*.

Chiu, Y.-T., Zhu, Y.-Q., & Corbett, J. (2021). In the hearts and minds of employees: A model of pre-adoptive appraisal toward artificial intelligence in organizations. *International Journal of Information Management*, *60*, 102379.

Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. *International Journal of Information Management*, *60*, 102383.

Coltman, T., Devinney, T. M., Midgley, D. F., & Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. *Journal of Business Research*, *61*(12), 1250–1262.

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), Article 16.

Diaz Andrade, A., & Techatassanasoontorn, A. A. (2021). Digital enforcement: Rethinking the pursuit of a digitally-enabled society. *Information Systems Journal*, *31*(1), 184–197.

Dunning, D. (2011). The Dunning–Kruger effect: On being ignorant of one's own ignorance. In *Advances in Experimental Social Psychology* (Vol. 44, pp. 247–296). Elsevier. https://doi.org/10.1016/B978-0-12-385522-0.00005-6

Engel, J. F., Blackwell, R. D., & Miniard, P. W. (1990). *Consumer Behavior*. Dryden Press.

Eshet-Alkalai, Y. (2004). Digital literacy: A conceptual framework for survival skills in the digital era. *Journal of Educational Multimedia and Hypermedia*, *13*(1), 93–106.

European Commission. (2021). *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

Feerrar, J. (2019). Development of a framework for digital literacy. *Reference Services Review*.

Flynn, L. R., & Goldsmith, R. E. (1999). A short, reliable measure of subjective knowledge. *Journal of Business Research*, *46*(1), 57–66.

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2022). Cognitive challenges in Human–Artificial Intelligence Collaboration: Investigating the path toward productive delegation. *Information Systems Research*, *33*(2), 678–696.

Furman, J., & Seamans, R. (2019). AI and the Economy. *Innovation Policy and the Economy*, *19*, 161–191. https://doi.org/10.1086/699936

Gillath, O., Ai, T., Branicky, M. S., Keshmiri, S., Davison, R. B., & Spaulding, R. (2021). Attachment and trust in artificial intelligence. *Computers in Human Behavior*, *115*, 106607.

Gilster, P. (1997). *Digital literacy*. Wiley Computer Pub. New York.

Goldkuhl, G., & Agerfalk, P. J. (2005). IT Artifacts as Socio-Pragmatic Instruments: Reconciling the Pragmatic, Semiotic, and Technical. *International Journal of Technology and Human Interaction (IJTHI)*, *1*(3), 29–43.

Gonda, D., Pavlovičová, G., Ďuriš, V., & Tirpáková, A. (2022). Implementation of Pedagogical Research into Statistical Courses to Develop Students' Statistical Literacy. *Mathematics*, *10*(11), 1793.

Hermann, E. (2022). Artificial intelligence and mass personalization of communication content—An ethical and literacy perspective. *New Media & Society*, *24*(5), 1258–1277.

Heyder, T., & Posegga, O. (2021). Extending the foundations of AI literacy. *ICIS 2021 Proceedings*. https://aisel.aisnet.org/icis2021/is_future_work/is_future_work/9

Hindi, N. M., Miller, D., & Wenger, J. (2002). Computer literacy: Implications for teaching a college-level course. *Journal of Information Systems Education*, *13*(2), 143.

Hinkin, T. R. (1998). A Brief Tutorial on the Development of Measures for Use in Survey Questionnaires. *Organizational Research Methods*, *1*(1), 104–121.

Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research. *Journal of Consumer Research*, *30*(2), 199–218.

Jöhnk, J., Weißert, M., & Wyrtki, K. (2021). Ready or Not, AI Comes—An Interview Study of Organizational AI Readiness Factors. *Business & Information Systems Engineering*, *63*(1), 5–20.

Kandlhofer, M., & Steinbauer, G. (2018). A driving license for intelligent systems. *Proceedings of the AAAI Conference on Artificial Intelligence*, *32*(1).

Kerpedzhiev, G. D., König, U. M., Röglinger, M., & Rosemann, M. (2021). An exploration into future business process management capabilities in view of digitalization. *Business & Information Systems Engineering*, *63*(2), 83–96.

Kiechle, E. S., Bailey, S. C., Hedlund, L. A., Viera, A. J., & Sheridan, S. L. (2015). Different Measures, Different Outcomes? A Systematic Review of Performance-Based versus Self-Reported Measures of Health Literacy and Numeracy. *Journal of General Internal Medicine*, *30*(10), 1538–1546. https://doi.org/10.1007/s11606-015-3288-4

Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice*, *41*(4), 212–218.

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159–174.

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*, 100101. https://doi.org/10.1016/j.caeai.2022.100101

Leprince-Ringuet, D. (2020, April 1). *AI for business: What's going wrong, and how to get it right*. ZDNET. https://www.zdnet.com/article/ai-for-business-whats-going-wrong-and-how-to-get-it-right/

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

MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly*, 293–334.

Mikalef, P., & Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Information & Management*, *58*(3), 103434.

Miller, T. A. (2016). Health literacy and adherence to medical treatment in chronic and acute illness: A meta-analysis. *Patient Education and Counseling*, *99*(7), 1079–1086. https://doi.org/10.1016/j.pec.2016.01.020

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, *115*(2), 502.

Motta, M., Callaghan, T., & Sylvester, S. (2018). Knowing less but presuming more: Dunning-Kruger effects and the endorsement of anti-vaccine policy attitudes. *Social Science & Medicine*, *211*, 274–281. https://doi.org/10.1016/j.socscimed.2018.06.032

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, *2*, 100041.

Nguyen, A., Hong, Y., & Gardner, L. A. (2020). A taxonomy of digital learning activities for digital inclusion. *ECIS 2020 Research Papers*. https://aisel.aisnet.org/ecis2020_rp/135

Nguyen, T. H., Paasche-Orlow, M. K., & McCormack, L. A. (2017). The state of the science of health literacy measurement. *Information Services & Use*, *37*(2), 189–203. https://doi.org/10.3233/ISU-170827

Nomura, T., Suzuki, T., Kanda, T., & Kato, K. (2006). Measurement of negative attitudes toward robots. *Interaction Studies*, *7*(3), 437–454.

Norman, C. D., & Skinner, H. A. (2006). eHEALS: The eHealth Literacy Scale. *Journal of Medical Internet Research*, *8*(4), e507. https://doi.org/10.2196/jmir.8.4.e27

Ongena, G. (2023). Data literacy for improving governmental performance: A competence-based approach and multidimensional operationalization. *Digital Business*, *3*(1), 100050. https://doi.org/10.1016/j.digbus.2022.100050

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867–872.

Pfeuffer, N., Baum, L., Stammer, W., Abdel-Karim, B. M., Schramowski, P., Bucher, A. M., Hügel, C., Rohde, G., Kersting, K., & Hinz, O. (2023). Explanatory Interactive Machine Learning. *Business & Information Systems Engineering*. https://doi.org/10.1007/s12599-023-00806-x

Pinski, M., & Benlian, A. (2023). AI Literacy-Towards Measuring Human Competency in Artificial Intelligence. *Proceedings of the 56th Hawaii International Conference on System Sciences | 2023*.

Russell, S. J., Norvig, P., & Davis, E. (2016). *Artificial Intelligence: A Modern Approach*. Pearson.

Sarker, S., Chatterjee, S., Xiao, X., & Elbanna, A. (2019). The sociotechnical axis of cohesion for the IS discipline: Its historical legacy and its continued relevance. *MIS Quarterly*, *43*(3), 695–720.

Schaffner, J. (2005). Subjective and objective measures of literacy: Implications for current results-oriented development initiatives. *International Journal of Educational Development*, *25*(6), 652–657.

Schuetz, S., & Venkatesh, V. (2020). Research Perspectives: The Rise of Human Machines: How Cognitive Computing Systems Challenge Assumptions of User-System Interaction. *Journal of the Association for Information Systems*, *21*(2). https://doi.org/10.17705/1jais.00608

Sindermann, C., Sha, P., Zhou, M., Wernicke, J., Schmitt, H. S., Li, M., Sariyska, R., Stavrou, M., Becker, B., & 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. https://doi.org/10.1007/s13218-020-00689-0

Someh, I., Shanks, G., & Davern, M. (2019). Reconceptualizing synergy to explain the value of business analytics systems. *Journal of Information Technology*, *34*(4), Article 4.

Stamper, R., Liu, K., Hafkamp, M., & Ades, Y. (2000). Understanding the roles of signs and norms in organizations—A semiotic approach to information systems design. *Behaviour & Information Technology*, *19*(1), 15–27.

Vetter, J., Benlian, A., & Hess, T. (2011). Overconfidence in IT Investment Decisions: Why Knowledge can be a Boon and Bane at the same Time. *ICIS 2011 Proceedings*. https://aisel.aisnet.org/icis2011/proceedings/generaltopics/4

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.

Weber, P. (2023). Unrealistic Optimism Regarding Artificial Intelligence Opportunities in Human Resource Management. *International Journal of Knowledge Management (IJKM)*, *19*(1), 1–19. https://doi.org/10.4018/IJKM.317217

Weber, P., Carl, K. V., & Hinz, O. (2023). Applications of Explainable Artificial Intelligence in Finance—A systematic review of Finance, Information Systems, and Computer Science literature. *Management Review Quarterly*. https://doi.org/10.1007/s11301-023-00320-0

Wiesche, M., Joseph, D., Thatcher, J., Gu, B., & Krcmar, H. (2020). IT workforce research curation. *MIS Quarterly Research Curation Series*. https://open.bu.edu/handle/2144/44097

Wolfe, H. (1992). Computer Literacy for the 1990's. *Journal of Information Systems Education*, *4*(1), 18–22.

World Economic Forum: Firth-Butterfield, K., Toplic, L., Anthony, A., and Reid, E. (2022, March 17). *Artificial Intelligence. Without universal AI literacy, AI will fail us*. https://www.weforum.org/agenda/2022/03/without-universal-ai-literacy-ai-will-fail-us/

Acknowledgements

This work has been funded by the German Research Foundation (DFG) within the Collaborative Research Center (CRC) 1053 MAKI.

Appendix A: Objective measurement scale

Item (n = 16)	**Answer Options** *(correct answer in italics – order randomized in test)*
Socio User AI Literacy	
AI was first mentioned in…	*A. …the 1950s.* B. …the 2000s. C. …the 1880s. D. …the 1980s.
How are human and artificial intelligence related?	*A. They are different, each has its own strengths and weaknesses.* B. Their strengths and weaknesses converge. C. They predict each other. D. They are the same, concerning strengths and weaknesses.
AI research…	*A. …happens in an interdisciplinary field including multiple technologies.* B. …is only fiction at this point in time. C. …revolves predominantly around optimization. D. …refers to one specific AI technology.
What is a possible risk for humans of AI technology?	*A. Deep fakes render videos unattributable.* B. Digital assistants take over self-driving cars. C. Image generators break the rules of art. D. Voice generators make people unlearn natural languages.
Socio Creator/Evaluator AI Literacy	
What is not an objective of current AI regulation?	*A. Enforcing a 'no-bias policy' to ban all potential biases that can arise from AI.* B. Ensuring that AI systems placed on the market are safe and respect existing law on fundamental rights. C. Facilitating the development of a market for lawful, safe and trustworthy AI. D. Ensuring legal certainty to facilitate investment and innovation in AI.
Which is a major regulation that has been passed specifically for artificial intelligence?	*A. European Commission's Act for Artificial Intelligence* B. European Regulation for Responsible AI C. United Nations' Framework for the Ethical Use of AI D. American Regulation on the Usage of AI
Which potential consequence can working with an AI have on humans that interact with it?	*A. Shift of tasks performed by humans* B. Shift of evaluation periods C. Debiasing of human literacy D. Debiasing of result interpretation
Key ethical issues surrounding AI include…	*A. …diversity, bias, and transparency* B. …ANN, GA, and Simulated Annealing C. …future predictions and past overfitting D. …cold start problem, omitted variable trap, and sunk cost fallacy
Technical User AI Literacy	
What is the central distinction between supervised and unsupervised learning?	*A. Supervised learning uses labeled datasets.* B. Unsupervised learning may happen anytime. C. Supervised learning is performed by supervised personnel. D. Supervised learning supersedes unsupervised learning.
Which of the following statements is true?	*A. ML is a part of AI* B. ML and AI are mutually exclusive C. AI is a part of ML D. AI and ML are the same
What is a typical application of an AI at which it is usually better than non-AI?	*A. Image recognition* B. Creating annual reports C. Undefined processes D. Hardware space analysis
Running the same request with the same data on the same AI…	*A. …could give different results.* B. …increase the computing speed. C. …never give different results. D. …double the computing time.
Technical Creator/Evaluator AI Literacy	
What always distinguishes decision trees from support vector machines?	*A. Decision trees are more interpretable.* B. Decision trees are more implicit. C. Decision trees generate more predictions. D. Decision trees are trained faster.
What is a typical split of testing and training data for development purposes?	*A. 80% Training and 20% Testing* B. 40% Training, 40% Testing, 20% Test-Training together C. 95% Testing and 5% Training D. It does not matter
What is not a strictly necessary part of a single AI system's development process?	*A. Benchmarking* B. Training/Learning C. Data preprocessing D. Model definition
What is not part of an ANN?	*A. User Layer* B. Input Layer C. Output Layer D. Hidden Layer