

# AI literacy for ethical use of chatbot: Will students accept AI ethics?

Yusuke Kajiwar<sup>\*</sup>, Kouhei Kawabata

Department of Production Systems Engineering and Sciences, Komatsu University, Shichomachi Nu1-3, Komatsu, Ishikawa, 923-8511, Japan

## ARTICLE INFO

### Keywords:

AI literacy  
Ethics  
Chatbot  
Large language models  
Technology acceptance  
Secondary education

## ABSTRACT

In AI literacy education, there are few examples of education based on AI ethical principles, and limited knowledge exists regarding curriculum design that incorporates AI ethical principles and its effects. Therefore, in this study, we propose a curriculum that teaches the ethical use of large language models (LLM) such as ChatGPT and verify its impact on educational effectiveness and technology acceptance among students aged 12 to 24. The validation results show that the proposed curriculum particularly contributes to the understanding of LLM concepts and their ethical use in decision support. We also demonstrate that experience using ChatGPT influences the level of understanding of ethical usage. Additionally, students aged 12 to 18 may actively adopt ChatGPT responses in decision support, and careful consideration is needed when using LLMs in the 12- to 18-year-old age group. Using technology acceptance model, AI ethical principles were also examined to determine technology acceptance, and it was found that usefulness, justice and fairness, privacy, and data protection directly impact attitudes toward ChatGPT. It has also become clear that students feel uneasy about using their personal information for learning ChatGPT, even if they have consented to the use of their personal information. This result suggests that AI developers and providers need to handle personal information carefully to foster a positive AI attitude.

## 1. Introduction

Artificial intelligence (AI) plays a crucial role in daily life, offering functions such as information recommendation, decision support, and creative assistance in digital services and the medical field (Dratsch et al., 2023; Kelly et al., 2019; Milano et al., 2020). In particular, ChatGPT, developed by OpenAI in 2022, excels at understanding and generating natural language, expected to be a versatile tool solving problems across finance, industry, medicine, and education (Mathew, 2023; Ray, 2023). Content generated by ChatGPT is already present on the internet, including websites users may visit for medical advice (Buchanan et al., 2023).

However, AI does not always yield optimal results, and there's a potential for harm or discrimination due to the quality or bias of the data used for AI learning or malicious third-party attacks or tampering (Ghallab, 2019; Kaur et al., 2022). The growing impact of AI on society has led to increased discussions based on ethical principles for risk mitigation (Floridi et al., 2018). Five ethical principles for AI have been proposed: beneficence, non-maleficence, justice and fairness, autonomy, and explainability and responsibility. To ensure AI promotes social good, improving AI literacy among the public, promoting ethical

guidelines acceptance, and responsible AI use are crucial (Floridi et al., 2018; Jobin et al., 2019). However, AI literacy education lacks examples based on ethical principles, and there's limited knowledge about incorporating AI ethical principles into curriculum design and its effects. Additionally, understanding the impact of AI ethical principles on technology acceptance is also limited.

This research aims to assist beginners and inexperienced users of large-scale language models (LLMs) like ChatGPT in learning how to use them ethically (beneficence, non-maleficence, justice and fairness, explainability and responsibility, autonomy). We developed a curriculum for students aged 12 to 24, examining its educational effectiveness and its impact on technology acceptance. The contributions of this research are outlined below:

- (i) We proposed a curriculum design for ethical usage and verified its educational effectiveness through achievement tests. The validation results show that the proposed curriculum particularly contributes to the understanding of LLM concepts and their ethical use in decision support. We also demonstrate that experience using ChatGPT influences the level of understanding of ethical usage.

<sup>\*</sup> Corresponding author.

E-mail address: [yusuke.kajiwar@komatsu-u.ac.jp](mailto:yusuke.kajiwar@komatsu-u.ac.jp) (Y. Kajiwar).

<https://doi.org/10.1016/j.caeai.2024.100251>

Received 1 March 2024; Received in revised form 3 June 2024; Accepted 8 June 2024

Available online 8 June 2024

2666-920X/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

- (ii) Based on our verification results, we have identified considerations for caution when permitting students to use LLM. Automatic bias may have inclined younger students to receive AI-generated answers more favorably. Therefore, caution should be exercised when employing decision aids with younger students.
- (iii) We employed the Technology Acceptance Model (TAM) to investigate the influence of AI ethical principles on technology acceptance. Our findings indicate that perceived usefulness, justice and fairness, privacy, and data protection directly impact attitudes toward ChatGPT. Additionally, the study revealed that students expressed concerns about the utilization of their personal information for LLM studies, even when they had given consent for its use. Furthermore, the study uncovered that expectations regarding the application of AI are highest in creative support, decision support, and information retrieval.

This study investigates the impact of LLM on ethical use and technology acceptance in different age groups, providing academic and practical insights applicable not only to ChatGPT but also to other LLMs with a similar degree of accuracy.

## 2. Related studies

### 2.1. What can ChatGPT do?

ChatGPT (Achiam et al., 2023) is a large-scale natural language model (LLM) built on neural networks, a form of machine learning. This model is capable of generating sentences and responding to natural language prompts, such as questions. ChatGPT is highly adaptable and can be fine-tuned to specific tasks and domains, providing highly customized solutions.

Although ChatGPT demonstrates excellent performance in a wide variety of natural language processing tasks, it is crucial to understand its limitations and determine appropriate usage. ChatGPT can perform tasks such as summarization, dialogue generation, machine translation, and question answering. It has shown superior performance compared to GPT-3.5, especially in question-and-answer tasks and dialogue generation (Qin et al., 2023). The summary indicates that ChatGPT is generally comparable to other LLMs (Liang et al., 2022). It also exhibits higher performance in machine translation compared to commercially available machine translation (Wang et al., 2023), backed by ChatGPT's ability to understand context and generate appropriate sentences. While ChatGPT primarily uses English, responses in different languages are also possible (Lai et al., 2023). However, ChatGPT has demonstrated limited performance in abstract reasoning (Wu et al., 2023). Therefore, using specific and easily understandable prompts is key to maximizing ChatGPT's capabilities.

ChatGPT leverages these features and is widely used, particularly for information retrieval, creative support, and decision support scenarios (Ray, 2023). In information retrieval, such as in legal services, providing large amounts of legal data for litigation (Regalia, 2023), and in business, offering insights on market conditions and trends is possible (Alafnan et al., 2023). In creative support, it can be employed in creative fields to aid in content creation, such as novels (Cox & Tzoc, 2023), and blog posts (Taecharungroj, 2023). For decision support, systems like customer service QA in the business field (George & George, 2023) and medical diagnosis and treatment recommendations in the medical field (Sallam, 2023) are feasible.

Although ChatGPT excels in various natural language processing tasks, it is imperative to comprehend its limitations and use it judiciously. By understanding the technology's pros and cons and deploying it appropriately, automation of tasks for workers in different fields is anticipated, leading to enhanced business processes efficiency. To stay productive in an ever-evolving technological landscape, the next generation should enhance their AI literacy, considering the potential

integration of LLMs like ChatGPT in the future.

### 2.2. Risks of AI in society

It has been pointed out that while AI will bring benefits to humans and society as its capabilities and spread expand, it also carries potential risks (Ghallab, 2019; Kaur et al., 2022). The main risks include discrimination, invasion of privacy, violation of rights, malicious attacks, and presenting inaccurate information. These issues are attributed to the quality and bias of the data used to train AI, as well as malicious attacks, tampering, and unintentional abuse. For example, an AI that predicts the risk of recidivism may biasly label black people due to poor training data (Kaur et al., 2022), leading to social unrest and discrimination. Additionally, intentional data falsification in predicting drug dosages can threaten patient safety (Jagielski et al., 2018). In light of the U.S. Supreme Court's decision regarding the use of Andy Warhol's photographs, intellectual property rights in AI-generated works may also raise concerns about copyright infringement if the data is used without permission (Lucchi, 2023).

Furthermore, there is a growing possibility that LLM could be exploited to allow AI to impersonate a trusted entity and steal sensitive information (Gupta et al., 2023). The prompts and responses entered by the user may also be used for LLM learning, raising concerns that if a user inputs sensitive information into LLM, other users may be able to access that information with a simple question. Reports have surfaced of Samsung employees accidentally entering confidential company information using ChatGPT (Gupta et al., 2023), posing a potential information leakage risk.

Moreover, it is considered problematic that generative AI can produce hallucinations (Ji et al., 2023). Hallucinations are meaningless or disloyal conditions in which the generated content is inconsistent with or unsupported by the source content. For instance, hallucinations in summaries created from patient information forms in medical applications can pose a risk to patients (Buchanan et al., 2023).

For AI to promote social good, users must strive to maximize the benefits of AI while minimizing its potential harms (Floridi et al., 2018; Jobin et al., 2019). Ethical approaches provide a means to maximize societal value while anticipating and avoiding the costs of potential abuse of AI. Respect for ethical guidelines, coupled with social sustainability, is expected to bring about beneficial changes in the use of AI.

### 2.3. Ethical approach to AI

#### 2.3.1. Discussion of ethics for AI

As AI evolves, its potential utility and impact on society expand exponentially. Principles of AI ethics are under discussion with the aim of mitigating potential risks associated with AI use and ensuring the development and proper integration of AI technology into society. Floridi and colleagues argued that these ethical themes are based on five principles: beneficence, non-maleficence, justice and fairness, explainability and responsibility, and autonomy (Floridi et al., 2018). These five ethical principles are supported by AI4People, a research and policy project shaping the AI ethics debate in the European Union. Interestingly, four of these five principles (beneficence, non-maleficence, autonomy, justice) are ethical frameworks based on traditional principles of biomedical ethics (Beauchamp & Childress, 2001). Explainability and responsibility is viewed as a new ethical principle added to ensure legal responsibility for AI actions and outcomes and to support the principle of good faith (Floridi et al., 2018).

#### 2.3.2. Beneficence

Beneficence prioritizes the continued prosperity of humanity and the preservation of the environment for future generations, promoting the interests of society as a whole in the development, deployment, and use of AI (Floridi et al., 2018). The principle of benevolence ensures that AI-based systems consider the best interests of the user and aim to support

or achieve the user's interests with genuine intent (Thiebes et al., 2021). In the development and operation of AI, it is crucial to strive to respect human dignity and equality through the principle of good faith.

### 2.3.3. Non-maleficence

The principle of non-maleficence addresses concerns such as privacy invasions and rights violations caused by AI technology, aiming to prevent harm intentionally or accidentally (Floridi et al., 2018). This mainly focuses on protecting privacy, data, maintaining data quality, and avoiding data abuse (Floridi et al., 2018; Morley et al., 2020; European Commission, 2019). Privacy risks arise not only when data is collected without the user's explicit consent, leaked to external agents, or de-anonymized (Narayanan & Shmatikov, 2008), but also when caution is necessary, as it can occur when making inferences based on data used (Friedman et al., 2015). Additionally, in preparation for legal risks, consideration must be given to the intellectual property rights of the data that will be trained by AI. The potential impact of malicious or low-quality data on AI models must also be considered (Thiebes et al., 2021). The principle of non-bad faith provides the foundation for respecting personal information and managing data appropriately.

### 2.3.4. Justice and fairness

The principles of justice and fairness are intended to avoid unfair bias and discrimination in AI use and to share benefits and prosperity for many people (Floridi et al., 2018). Fairness can be broadly divided into individual fairness and group fairness (Kaur et al., 2022). Individual fairness plays a role in ensuring that individual elements with similar causal relationships or similar rational characteristics obtain similar predicted results (Dwork et al., 2012; Kusner et al., 2017). This ensures that each factor is evaluated fairly, and individual characteristics do not unduly influence the outcome. Group fairness serves to ensure that different groups have the same expected proportion of positively or negatively predicted outcomes (Santos et al., 2015). This ensures that different people and groups are treated equally, and the AI does not make biased predictions towards one group or another. If principles of justice and fairness are not properly implemented, advances in AI technology may only benefit some people and harm others. Therefore, it is important to pursue fairness and justice from an ethical, social, and technological perspective.

### 2.3.5. Autonomy

Autonomy is the principle that humans should always retain the right to make decisions (Floridi et al., 2018). In the context of AI, adhering to this principle of autonomy refers to striking a balance between our own decision-making rights and the decision-making rights we delegate to AI (Morley et al., 2020). With the introduction of AI, human agency and oversight are crucial when delegating some decision-making rights to machines (European Commission, 2019). Humans should be actively involved in the decision-making process, and the degree of this involvement should be appropriately adjusted based on risks and social and environmental impacts (Smuha, 2019). It is expected that humans and AI will work in a mutually complementary manner and produce the best results together.

### 2.3.6. Explainability and responsibility

Explainability consists of two elements: "explainability" in an epistemological sense and "responsibility" in an ethical sense (Floridi et al., 2018; European Commission, 2019). Explainability refers to revealing the data and algorithms used to train an AI so that users can understand the rationale behind the AI's behavior and decisions. Responsibility means being responsible for explaining the AI's behavior and decision-making process to stakeholders. A lack of responsibility can make the pursuit of responsibility unclear, especially in cases of harm, and can lead to legal problems (Arrieta et al., 2020). Additionally, the algorithms used in AI are often almost incomprehensible to non-experts and tend to be black boxes (European Commission, 2019;

Loyola-Gonzalez, 2019). Furthermore, insufficient explainability obscures the impact of AI actions on society and hides the potential risks of AI. Additionally, it may negatively impact the balance between individual and AI decision-making. Explainability ensures legal accountability and provides the foundation that supports the other four principles (Thiebes et al., 2021).

### 2.3.7. Summary

Related research mainly focuses on the definition of AI ethics. To maximize the benefits of AI and share in its benefits, users must be informed and accept ethical guidelines for AI. To promote these efforts, AI literacy education is essential to explain how AI works in an easy-to-understand manner and to convey information using specific examples.

## 2.4. AI literacy

### 2.4.1. Concept

AI literacy refers to "a set of competencies that enable individuals to critically evaluate AI technologies, communicate with AI, and use AI effectively as a tool" (Long & Magerko, 2020). The Five Key Ideas of Artificial Intelligence (AI4K12) proposed by Touretzky et al. (2019) serve as a framework for teaching AI literacy. These five big ideas include (1) Perception, (2) Knowledge Representation and Reasoning, (3) Learning, (4) Natural Interaction, and (5) Social Influence. Broadly categorizing these ideas, (1) to (3) encompass the concepts and theories of artificial intelligence, (4) involves the application of AI, and (5) focuses on ethics education related to AI.

### 2.4.2. Approach

Main approaches to AI literacy education encompass programming development (Burgsteiner et al., 2016; Kaspersen et al., 2021; Kong et al., 2023), role interaction with AI using AI applications (Druga et al., 2019; Hitron et al., 2019; Williams et al., 2023), and theoretical approaches rooted in mathematics through classroom lectures and online videos (Torrey, 2012). Educational research on AI literacy covers various perspectives, with many studies employing one or a combination of these three categories, reporting their effects (Laupichler et al., 2022; Ng et al., 2021). However, it is essential to implement and tailor these approaches carefully to K-12 knowledge and ability levels, as overly complex tasks may hinder children's learning (Resnick & Silverman, 2005).

It has been noted that middle school students and above can comprehend AI concepts and theories through role-playing (Kajiwara et al., 2023), involving data acquisition, graphical representation, inference based on if-then rules, and parameter optimization based on ML model loss functions. Moreover, middle school students can grasp the interests and value of AI systems, influencing future career education research (DiPaola et al., 2020; Zhang et al., 2023). Therefore, traditionally focusing on secondary and higher levels (Laupichler et al., 2022; Williams et al., 2023).

### 2.4.3. Education on AI ethics for image recognition and image generation

In a few practical examples, ethics education primarily utilizes image recognition AI and image generation AI (Kong et al., 2023; Williams et al., 2023; Zhang et al., 2023). These studies include image recognition and deepfakes (Westerlund, 2019), where students learn about the ethical dimensions of AI, such as benevolence, non-maleficence, justice/fairness, autonomy, and explainability (Williams et al., 2023; Zhang et al., 2023). Ethics education at the university level, as conducted by Kong and colleagues, involves the development of AI applications through programming (Kong et al., 2023). However, due to the advanced programming skills required, the applicability of these results to K-12 remains unclear. Williams et al. proposed an ethics education curriculum for middle school students, covering Creative AI, Dance AI, and Robotics, with a primary focus on topics related to image generation AI. The curriculum explores social implications and harm caused by

GAN tools, which generate non-existent human faces and videos through interaction with AI.

However, it has been reported that the level of understanding of these curricula, which predominantly concentrate on AI application development and interaction, varies widely based on students' abilities and knowledge (Williams et al., 2023).

#### 2.4.4. Education on AI ethics for LLM

In education, LLMs can be utilized for learning purposes, management (e.g., providing essential information like scholarships and tuition fees), advising (e.g., supporting career decision-making), and research and development (e.g., offering knowledge in specialized fields). However, the effectiveness of LLM ethical education has not been sufficiently validated (Okonkwo & Ade-Ibijola, 2021). Many educators acknowledge the need for clear guidelines regarding the appropriate and ethical use of chatbots before their implementation (Jeon & Lee, 2023).

Additionally, ChatGPT, which has gained rapid popularity, has been recently released, with reports on its usage and examples of prompt execution (Cooper, 2023; Kasneci et al., 2023; Pavlik, 2023; Qadir, 2023) and an examination of ethical considerations (Hualpa, 2023). While there are a few reports, Essel, H. B et al. have also discussed the effects of practical education, such as the impact of ChatGPT on students' cognitive skills (Essel et al., 2024). However, the effectiveness of practical education incorporating ethical aspects has not been adequately verified, and related knowledge is limited. Although OpenAI provides guidelines for the use of ChatGPT, these have not yet been fully integrated into AI ethics education frameworks.

#### 2.4.5. Summary

Most ethics education currently concentrates on image generation AI. There are currently no ethical guidelines or curriculum designs concerning the appropriate and ethical use of ChatGPT, which has developed in recent years. Therefore, this study will focus on ChatGPT, propose a practical curriculum design for learning LLM concepts and ethical usage, and report its effectiveness. This research aims to contribute to the effective design of AI ethical education programs.

### 2.5. Technology acceptance

AI literacy aims to empower individuals with the skills to effectively engage with AI (Long & Magerko, 2020). It has been emphasized that, prior to implementing AI as a solution across various sectors, addressing negative perceptions associated with AI through cultural change is essential, and people need to become familiar with the technology (Del Giudice et al., 2023). Therefore, the user's acceptance of the technology is fundamental for the successful adoption of devices (Davis, 1989). The Technology Acceptance Model (TAM) was employed to assess the acceptance of artificial intelligence in science education. A survey was conducted among 83 science teachers (Al Darayseh, 2023). Based on the components of TAM, a six-dimensional scale was designed and evaluated. The findings indicated that self-efficacy, ease of use, expected benefits, and attitude toward AI applications positively influenced behavioral intention. However, it was noted that science teachers exhibited more positive attitudes toward technology than language and humanities teachers, suggesting limitations in the generalizability of these findings. The discourse on ChatGPT technology acceptance gained momentum from 2023. For instance, Strzelecki et al. utilized components of the Unified Theory of Acceptance and Specification of Technology (UTAUT2) to identify determinants based on self-reported data from 534 university students (Strzelecki, 2023). The results revealed that habits, performance expectancy, and hedonic motives had a positive effect on behavioral intentions. Additionally, Habibi et al. (2023) employed components of UTAUT2 to examine determinants promoting the use of ChatGPT in learning, based on responses from 1117 students in Indonesian higher education institutions. The results demonstrated that facilitating conditions, performance expectancy, social influence,

habits, and hedonic motivations exerted the most influence on behavioral intentions, with the role of effort expectancy being less significant. Research on the technology acceptance of ChatGPT has proliferated in recent years, yielding diverse results among influencing factors for behavioral intention, such as facilitation conditions and social influence, with no conclusive findings. Therefore, additional verification is necessary to establish robust conclusions. Furthermore, as of 2024, most studies investigating the conditions for technological acceptance of ChatGPT are confined to the university age group. Moreover, these technology acceptance models do not integrate ethical principles for AI, as suggested by theory. In other words, no practical research has been conducted to ascertain whether AI literacy education that incorporates systematized AI ethical principles impacts technological acceptance of AI.

Therefore, in this study, we conducted a seminar for students of 12–24 ages to educate them on the concepts of LLM and the ethics of AI. Subsequently, we constructed a model based on the Technology Acceptance Model (TAM) (Davis, 1989) and elucidated the influence of ethical considerations recommended by AI4people (Floridi et al., 2018) on technology acceptance. This study provides both academic and practical insights by addressing the acceptance of AI technology from educational and ethical perspectives and examining its impact on different age groups.

## 3. Curriculum design

### 3.1. Overview

In this study, we aim to develop a curriculum tailored for beginners and inexperienced users of LLMs, such as ChatGPT, to ensure a consistent understanding of knowledge and ethical usage. The foundation of our study rests upon the AI ethics framework proposed by the AI4people project (Floridi et al., 2018), which encompasses benevolence, non-maleficence, autonomy, justice and fairness, and explainability and responsibility. We have structured a curriculum that integrates these ethical dimensions into the learning process.

Through this framework, students will cultivate the following elements: (a) basic knowledge, (b) explainability, (c) benevolence, (d) non-maleficence, (e) justice and fairness, (f) responsibility, and (g) autonomy. Phases (a) and (b) focus on AI concepts, deepening the understanding of its operations and consequences, and establishing the groundwork for critical evaluation and effective utilization of AI. In phases (c), (d), and (e), we will explore the societal and individual impacts of AI, emphasizing benevolence, non-maleficence, and justice and fairness perspectives. Phases (f) and (g) will involve a thorough examination of the roles of humans and AI from the standpoint of autonomy and responsibility.

For the delivery of content, we chose a lecture format based on our belief that a unified approach is optimal for enhancing understanding and promoting ethical use. This decision was influenced by concerns raised about potential variations in knowledge acquisition in exercise formats, depending on individual experiences (Williams et al., 2023).

During the actual lectures, instructors input prompts into ChatGPT in the presence of students and had them verify the responses. To ensure uniform learning, given the stochastic nature of ChatGPT responses, we provided students with pre-prepared ChatGPT responses as instructional materials. This approach mitigated the risk of receiving inconsistent responses and allowed students to develop a focused understanding of the ethical aspects of AI.

The tasks employed in phases (a)–(g) are outlined in Fig. 1 and Table 1, with specific details of Task, prompts, and response of ChatGPT provided in Appendix A.

### 3.2. Basic knowledge

This phase is dedicated to comprehending the concepts of artificial



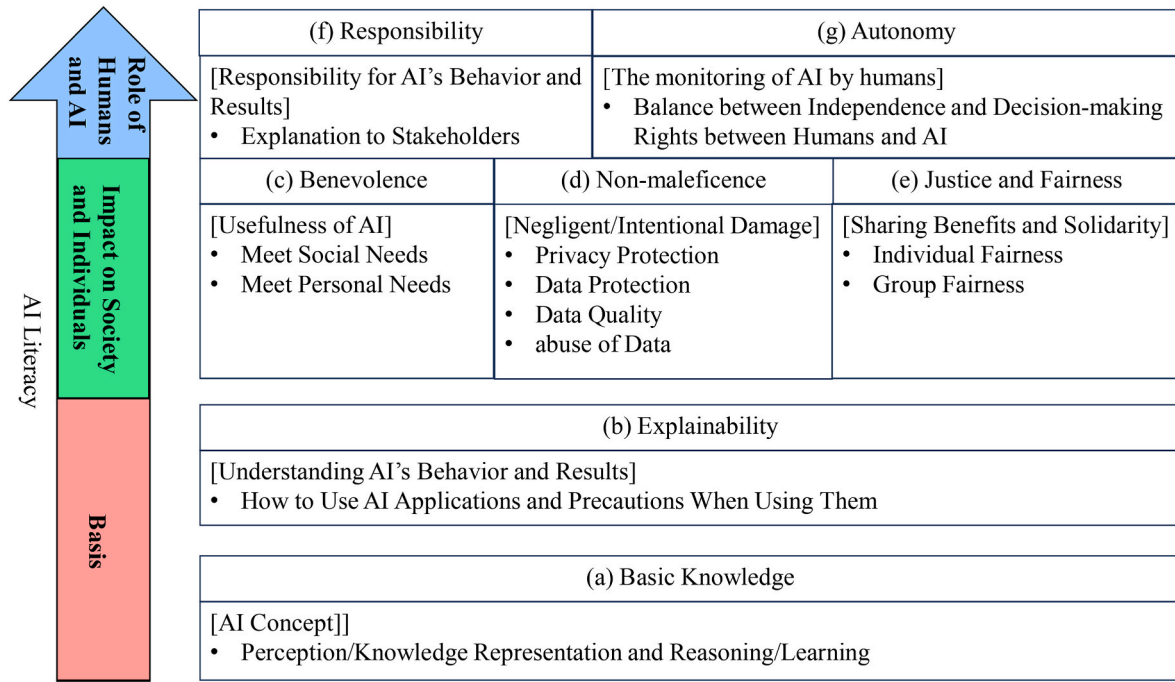


Fig. 1. AI literacy curriculum for ethical use of LLM

**Table 1**  
Tasks in the AI literacy curriculum.

Framework	No.	Item	Content
(a) Basic knowledge	ED1	Word prediction	Predicting words in masked parts of sentences.
	ED2	Sentence generation	Sequential prediction of statistically plausible words.
	ED3	Question answer	Predict response sentences from prompts.
(b) Explainability	ED4	LLM accuracy	Which sentence was written by AI?
	ED5-1 ED5-2	Same prompt, different responses	Compare sentences written by LLMs when inputting the same prompt.
	ED6	Hallucination	Find mistakes in the sentences generated by LLM
(c) Benevolence	ED7-1	Information retrieval	How do frogs breathe?
	ED7-2	Creative support	Please write a fantasy novel.
	ED7-3	Decision support	What would you like for dinner today?
(d) Non-maleficence	Q1	Privacy Protection Data Protection Data quality abuse of data	What data can and should not be used for AI learning?
	Q2	My opinion	Which rights do you recognize?
	Q3	Decision support	(After seeing ChatGPT's response) Which rights do you recognize?
(e) Justice and fairness	Q4	Subjective Fairness	Was ChatGPT's response fair?
	Q5	What is the reason for Q4?	What do you feel is fair?
	Q6	Who is responsible?	Who do you think is responsible if you are harmed by ChatGPT's answers?
(f) Responsibility	Q7	Role of Humans and AI	Should ChatGPT's answer be the final decision?

intelligence. In this research, our focus was on the LLM concept, delving into the concealed mechanism of ChatGPT and attempting to enhance understanding through role play. Previous studies have utilized decision trees as algorithms, known for their simplicity and effectiveness for K-12 learners, with successful examples reported (Kajiwara et al., 2023). Despite several attempts to promote neural network understanding at the K-12 level, their effectiveness remains limited (Williams et al., 2023; Zhang et al., 2023). ChatGPT, grounded in the Transformer model, a type of neural network (Ray, 2023), makes it challenging to comprehend the detailed mechanism of LLM with K-12 level knowledge and ability.

In this study, we engaged in role-playing to illustrate the BERT (Kenton & Toutanova, 2019), a typical natural language model, aiming to encourage learners to intuitively understand it. BERT involves two pre-training processes: one focuses on predicting masked words in a sentence, while the other involves predicting whether two sentences are consecutive. We elucidated these processes through role-playing scenarios, demonstrating how to predict words, understand sentence production, and forecast responses to prompts. In Exercise 1 (ED1), students filled in the blanks of sentences with appropriate words in sequence. Drawing on the lexical distribution hypothesis, we discovered that the meaning of a word can be predicted by its surrounding words. In Exercise 2 (ED2), actual role-playing illustrated that sentences can be constructed by predicting statistically plausible words in sequence. In Exercise 3 (ED3), we aimed to deepen learners' understanding by providing specific examples of predicting response sentences from prompts. Finally, we explained how these LLMs learn large amounts of data and make predictions based on it.

### 3.3. Explainability

In this phase, we will focus on specific ways to use AI applications and precautions when using them. In this study, we will explore the usage of ChatGPT, a type of LLM. Compared to other LLMs, ChatGPT demonstrates high performance in various natural language processing tasks, including summarization, dialogue generation, machine translation, and question answering. The primary method of utilizing ChatGPT involves inputting a prompt, and ChatGPT responds with a corresponding output. To assess ChatGPT's sentence generation ability,

students were tasked with comparing two sentences, as demonstrated in ED4, and determining which one was created by a human or ChatGPT. Through this, students gain an understanding that ChatGPT can generate sentences almost indistinguishable from those crafted by humans.

Next, students consider key points to bear in mind when using ChatGPT. Two main considerations arise when using ChatGPT. Firstly, the output of ChatGPT changes stochastically even for the same input (Cooper, 2023; Qadir, 2023). This stochastic variation implies that different meanings, wordings, and expressions can be generated. Therefore, it is crucial to recognize that using the same prompt may yield different results. Another noteworthy aspect is that ChatGPT responses may involve hallucinations (Ji et al., 2023). As the generated sentences are based on the model's learned data, they may occasionally contain inaccurate information or unintended expressions. Hence, it is important to carefully scrutinize the output and, if necessary, cross-reference it with other sources of information.

The students observed that even when using identical prompts in ED5-1 and ED5-2, the ChatGPT responses varied. Furthermore, in ED6, students learned that the responses could include hallucinations. As a result, they were advised to review the generated sentences and seek additional information if needed.

### 3.4. Benevolence

This phase focuses on uses that benefit society and individuals. Benevolence, in a narrower sense, aims to maximize the benefits of AI across society as a whole. At the individual level, it seeks to enhance relationships and satisfy individual wants and needs. ChatGPT is primarily utilized for information retrieval, creative support, and decision support (Ray, 2023). In related research, one factor forming a favorable impression of AI is its application to familiar situations (Kajiwara et al., 2023). Therefore, this study explores familiar cases and teaches how to utilize ChatGPT for information retrieval, creative support, and decision support.

For information retrieval purposes, we provided a specific example of searching for knowledge such as "How do frogs breathe?" as in ED7-1. In terms of creative support, we demonstrated the generation of novels using prompts like "Write a fantasy novel" as in ED7-2. Additionally, in decision support, we showcased ChatGPT's ability to respond flexibly and provide solutions to ambiguous problems (ill-defined problem) with multiple solutions, as exemplified in ED7-3 using prompts like "What would you like for dinner today?".

### 3.5. Non-maleficence

The main purpose of this phase is to provide students with a foundation to focus on respecting personal information, proper data management, and minimizing potential harm caused by AI technologies. Specific focuses include protecting privacy, Data Protection, maintaining data quality, and avoiding data abuse. We introduced examples related to these aspects to deepen understanding.

Concerning privacy, as reported by Gupta (Gupta et al., 2023), we highlighted that prompts entered into ChatGPT could lead to potential information leakage risks. Regarding data protection, while we were unable to confirm specific case law regarding ChatGPT, citing the U.S. Supreme Court's decision regarding the use of Andy Warhol's photographs (Lucchi, 2023) emphasizes that intellectual property rights of AI-generated works are subject to copyright infringement if the data is used without permission, leading to potential concerns.

Concerning maintaining data quality, we highlighted a case where AI was inappropriately labeled for black people due to insufficient quality of training data when predicting the risk of recidivism (Kaur et al., 2022). This inappropriate labeling can lead to social anxiety and discrimination. To improve data quality, we emphasized the need to eliminate discriminatory bias and ensure diversity in datasets.

Furthermore, we introduced that data falsification could lead to

incorrect dosage changes, posing serious implications for patient safety (Jagielski et al., 2018).

Through these efforts, we aim to build a foundation for respecting personal information and data management, providing a framework for AI technology to have a positive impact on society.

### 3.6. Justice and fairness

This phase focuses on pursuing public interest, emphasizing individual and group fairness through AI technology. Considering a scenario where ChatGPT provides decision-making support in a conflict between a baseball club and a soccer club over the right to use a field, which is a public resource. First, focusing on individual fairness, if individual factors exhibit similar causal relationships and rational characteristics are alike, the predicted results should also be similar (Dwork et al., 2012; Kusner et al., 2017). ChatGPT is expected to eliminate bias or unfair treatment towards specific extracurricular activities, ensuring fair opportunities for each individual. Regarding group fairness, it should ensure the proportion of positively and negatively predicted outcomes is the same across different groups (Santos et al., 2015). ChatGPT is expected to propose appropriate allocation and timetables for equal opportunities for each club activity.

Students will collaborate with ChatGPT to understand the needs of the baseball and soccer teams fully, helping establish a fair and transparent decision-making process. This collaboration ensures fairness for individuals and groups, pursuing policies that maximize the public interest.

### 3.7. Responsibility

This phase concentrates on the responsibility associated with AI's actions and outcomes. Previous studies have emphasized that a lack of explanation regarding AI actions and results can obscure potential risks from stakeholders and lead to legal issues (Arrieta et al., 2020). However, current research is inconclusive about who should be held responsible: AI, AI developers, providers, or end users (Memarian & Doleck, 2023). Responsibility implies that various factors contribute to it, including the foreseeability of harm to the plaintiff and the adoption of reasonable steps to prevent harm to the plaintiff. Additionally, to the author's knowledge, there is no case law on this issue as of 2023.

For this reason, in this study, based on the experience of ED4, the question "How to deal with influenza answered by ChatGPT was published (disseminated) on the website. Participants viewed it and took steps to address the flu. However, the information provided was incorrect, leading to a worsening of symptoms. Who is to blame?" These questions prompt participants to ponder responsible. This matter is currently under debate by researchers globally, with no conclusive determination on who should be responsible. Nevertheless, the ultimate lesson is to emphasize the importance of carefully verifying information, including fact-checking, and engaging in honest dialogue to avoid disadvantage. We believe that this approach will contribute to the healthy development of AI by addressing ambiguity in legal responsibilities and ensuring responsibility through transparent communication with stakeholders.

### 3.8. Autonomy

This phase focuses on the balance between human decision-making rights and delegating decision-making rights to AI. Autonomy considers arguments (a) to (f) and answers whether the final decision should be left to AI or humans. Based on the EU's ethical guidelines (European Commission, 2019), it emphasizes that human agency and oversight are extremely important when implementing AI, even when delegating some decision-making rights to AI. This suggests that humans are always responsible for making the final decisions and equally emphasizes explainability and responsibility when human intervention is required.

Through this, students are expected to understand the balance between human and AI decision-making processes and the importance of intervening when necessary.

#### 4. Technology acceptance

Addressing the negative perceptions associated with AI and embracing the technology through AI literacy is the first step in taking advantage of it (Davis, 1989; Del Giudice et al., 2023). The TAM (Davis, 1989) is a model frequently used for the acceptance of AI technology. TAM is grounded in the theory of rational action (Fishbein & Ajzen, 1977). This study also adopted the Technology Acceptance Model. The Technology Acceptance Model represents a technology acceptance model tailored to a specific situation by incorporating external variables. In this study, we formulated the following hypotheses based on the Technology Acceptance Model and constructed the LLM Technology Acceptance Model, as shown in Fig. 2.

- H1. Perceived ease of use has a direct effect on Perceived usefulness.
- H2. Perceived ease of use has a direct effect on attitude.
- H3. Perceived Usefulness has a direct effect on behavioral intention.
- H4. Attitude has a direct effect on behavioral intention.
- H5. Perceived usefulness has a direct effect on attitude.
- H6. Perceived usefulness has a direct effect on social influence and performance expectancy.
- H7. Perceived ease of use has a direct effect on effort expectancy and the ability to appropriately deal with hallucinations.
- H8. Non-maleficence, justice and fairness, responsibility, and autonomy have a direct effect on attitude.
- H9. Behavioral intention has a direct effect on information retrieval, decision support, and creative support.

H1-H5 mirror the TAM (Davis, 1989) hypotheses. Perceived Usefulness refers to a user's subjective evaluation or perception of how effectively the technology or system meets their goals and needs when using a particular technology or system. Perceived ease of use refers to a user's subjective evaluation or perception of how easy it is to use a particular technology or system. Attitude refers to a user's positive or negative feelings and opinions about a particular technology or system. User attitudes include likes, dislikes, expectations, and trust in the technology or system. Behavioral intention refers to a user's intention or intent to use a particular technology or system.

The hypothesis for H6 was based on the definition of perceived usefulness, which aligns with the definition of benevolence in AI ethics. Additionally, hypothesis H7 was formulated by considering the definition of perceived ease of use and the characteristics of LLM. Perceived ease of use is also a factor that presupposes explainability. Ethical

approaches provide a means to maximize societal value while anticipating and avoiding the costs of potential abuse of AI. Therefore, these ethical factors are expected to influence attitudes. Hence, we hypothesized H8. Finally, we formulated hypothesis H9 based on the main uses of LLM.

This model illustrates the influence of AI ethics education in Fig. 1, on the acceptance of LLM technology. The primary goal of AI literacy instruction is to foster proficiency in effectively using AI as a tool. AI ethics are expected to play a pivotal role in mitigating potential risks and facilitating the integration of AI into society. However, there is a concern that teaching AI ethics may inadvertently instill excessive apprehension and cultivate a negative perception of AI.

The study conducted by Zhang et al. (2023) supports this notion, suggesting that educating individuals about AI risks can indeed lead to a negative attitude toward AI. This underscores the need to carefully structure the curriculum to ensure that the teaching of AI ethics does not overly bias learners against AI.

In order to successfully introduce AI, it is crucial to design the curriculum in a manner that avoids fostering an overly negative perception of AI ethics. Therefore, it becomes imperative to assess the impact of AI on technology acceptance as a crucial aspect of evaluating the curriculum depicted in Fig. 1.

This model provides valuable insights for designing a curriculum that seamlessly integrates appropriate AI ethics while also fostering positive attitudes toward AI adoption. Additionally, it emphasizes the importance of addressing ethical considerations in AI education to ensure responsible and effective use of AI technology.

#### 5. Experiment

##### 5.1. Purpose

In this study, we aim to implement a practical curriculum for individuals aged 12 to 24 to learn about ethical use and validate its educational effects. Furthermore, this study will clarify the impact of AI literacy, which integrates AI ethical principles, on technology acceptance.

##### 5.2. Method

The seminar was conducted following the curriculum in Fig. 1 and Table 1. The LLMs used was ChatGPT3.5. The students acquired knowledge of ED1-ED3, ED4, ED5-1, ED5-2, ED6, ED7-1, ED7-2, and ED7-3 from demonstrations or seminar materials. Assignments were prepared for ED4 and Q1-Q7, and students were asked to respond to them. For Q1-Q7, participants were required to choose from among the options. Q1, Q5, and Q6 are multiple-choice questions, and these options were generated based on the free-text responses from the preliminary survey. A preliminary survey was conducted among 62 first-year university students. University students who participated in the

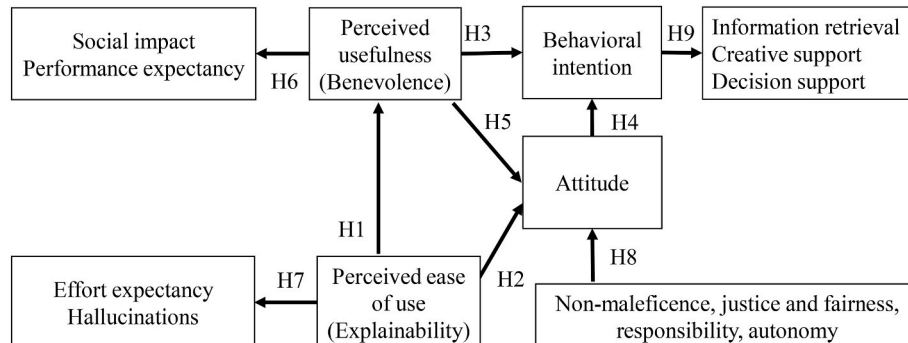


Fig. 2. Hypothesis of the LLM technology acceptance model integrating AI ethics.

preliminary survey were not included in the analysis of this experiment. Please refer to Appendix B for details on the preliminary investigation.

To measure their understanding of LLM concepts and ethical usage, participants took the achievement test shown in Table 2 both before and after the seminar. The content of the achievement test mainly measures their understanding of the concepts of the LLM and its ethical use. Test 1 was designed to measure their understanding of the concepts of the LLM. Test 2(a) was designed to measure their understanding of the ethical use of the LLM when used for information retrieval. Test 2(a) includes their understanding of benevolence and explainability. Test 2(b) was designed to measure their understanding of the ethical use of the LLM when used for decision support. Test 2(b) includes their understanding of benevolence, justice and fairness, responsibility, and autonomy. Test 2(c) was designed to measure their understanding of the ethical use of the LLM when used for creative support. Test 3(c) includes their understanding of benevolence, non-maleficence, and responsibility. The content of the achievement test was kept the same to evaluate achievement before and after the seminar. TEST1 is a 4-choice question, assessing understanding of the concept of LLM, while TEST2 is a YES or NO multiple-choice question, evaluating achievement regarding the utilization of LLM. Additionally, to measure acceptance of LLM technology, participants completed the questionnaire in Table 3 after the seminar. Questionnaire items Q4, Q8-Q14 comprised four options using a Likert scale: “Not applicable (1), somewhat not applicable (2), somewhat applicable (3), and applicable (4).” Regarding attitude, participants were asked to choose the most appropriate impression from the following: not interested, anxious, seems convenient, seems interesting, and would like to create. In the analysis, these impressions were categorized into negative attitude (−1), indifference (0), and positive attitude (1). Anxiety represented a negative attitude (−1), while “I want to make it,” “It looks interesting,” and “It looks useful” conveyed positive attitudes (1). “Not interested” indicated indifference (0). The numbers in parentheses are dummy variables used in the analysis. The survey was conducted using Microsoft Forms, accessible via smartphone or PC. However, for participants lacking access to these devices, the Microsoft Forms survey screen was printed on paper, distributed, and then collected.

### 5.3. Subjects

The experiment involved 107 Japanese students aged between 12

**Table 2**

Achievement Test Content (The answer to TEST1: “Guess the next word after a certain word or sentence and generate statistically plausible sentences.”, TEST2 (i): “YES”, TEST2(ii): “NO”, TEST2(iii): “NO”).

Items	Measures	Contents
TEST1	Concept	Please choose the correct answer to explain LLM. (4 choices)
TEST2	Ethical usage	As an example of how to use LLM, please answer YES if it is appropriate and NO if it is inappropriate.
(i)	Benevolence and explainability	Since the text explaining “artificial intelligence” was created by ChatGPT, I compared it with information from other sources such as books and checked the authenticity of the information before using it.
(ii)	Benevolence, justice and fairness, responsibility, and autonomy	Write a questionnaire about the symptoms of a family member who contracted the disease, listen to it on ChatGPT, and take care of the patient according to the generated text.
(iii)	Benevolence, non-maleficence, and responsibility	Since it was unlikely that the work would be completed by the deadline, I asked ChatGPT to generate the second half of the work and submitted it as my own work.

**Table 3**

Questionnaire Contents (Q8-Q15 are 4-item Likert scales; Q16 assesses positive (looks interesting, wants to make it, seems useful), negative (anxiety), and not interested responses.).

Number	Item	Content
Q8	Social Impact	Do you believe LLM will have a positive impact on society?
Q9	Performance expectancy	Do you consider the value (performance) of LLM to be high?
Q10	Effort expectancy	Do you find it easy to become proficient with LLM?
Q11	Deal with hallucination	Do you feel confident in dealing with hallucinations in LLM-generated texts?
Q12	Information retrieval	Would you like to use LLM for information retrieval purposes?
Q13	Creative support	Would you like to use LLM for creative support purposes?
Q14	Decision support	Do you intend to use LLM-generated answers as a basis for making decisions?
Q15	Attitude	What is your impression of LLM? (It looks interesting, I want to make it, it seems useful, I'm anxiety, I'm not interested)

and 24, who participated in a seminar. Table 4 presents the breakdown of student ages and usage rates. Of the participants, 56 had used LLM, while 51 had never used it. Participants were recruited from Ishikawa Prefecture, Japan, to attend a 1.5-h seminar. Three seminars with identical content were conducted once a month from September to November 2023, with participants attending one of the three seminars.

The subjects were thoroughly briefed on the experiment beforehand and given the option to opt-out if they chose not to participate. Throughout this process, comprehensive information about the experiment's purpose, protocol, and data handling was provided. Subjects were afforded the autonomy to decide for themselves whether to take part in the experiment.

Furthermore, when choosing to opt-out, great care was taken not to impose any restrictions or pressure on the subjects, and every effort was made to uphold their complete self-determination.

### 5.4. Analysis

#### 5.4.1. Understanding the concepts and ethical usage of LLM

Analyze responses to ED4, Q1-Q7 in Table 1, and achievement tests (TEST1-TEST2). The results in Table 5 are provided in Section 6.1.

In ED4, to assess the accuracy of ChatGPT, calculate the percentage of students who identified AI-generated sentences. To evaluate the understanding of non-maleficence in Q1, divide the data into two groups: data suitable for AI learning and data unsuitable for AI learning. Calculate the response ratio for each group using the formula:

$$\text{Response rate} = \frac{\text{Number of subjects who chose a certain answer}}{\text{Total number of subjects belonging to a certain group}} \quad (1)$$

The sample size of this study is 107, which is sufficiently large. Therefore, conduct a chi-square test to determine if there is a significant difference in the selected data between the two groups. The null hypothesis is that there is no difference in proportions between the groups, with a significance level of 5%. Compare the test result with the correct data to quantitatively assess the understanding of the non-maleficence.

Next, to analyze whether ChatGPT has the ability to influence students' opinions in decision-making support, we calculated the

**Table 4**

Breakdown of student age and usage rate.

Age Group	Number of Individuals	Usage Rate
12–15 years old	21	0.24
16–18 years old	23	0.50
19–21 years old	33	0.61
22–24 years old	31	0.65



**Table 5**  
Items and contents to be analyzed.

Framework	No	Results	Contents
(a) Basic knowledge	TEST1	Tables 12–14	Analyze the influence of curriculum, age, and usage history on LLM conceptual understanding.
(b) Explainability	ED4	Unable to differentiate.	Students assessed the accuracy of LLM.
(c) Benevolence	TEST2	Tables 12–14	Analyze the impact of curriculum, age, and usage history on understanding ethical usage of information retrieval (TEST2(a)), decision support (TEST2 (b)), and creative support (TEST2(c)).
(d) Non-maleficence	Q1	Table 5	Analyze the impact of curriculum on understanding non-maleficence.
(e) Justice and fairness	Q2 and Q3	Tables 6–8	Analyze the effect of LLM decision support on student decision-making.
	Q4	Table 9	Analyze the influence of LLM decision support on subjective fairness.
	Q5	Table 10	Analyze the factors contributing to subjective perceptions of fairness.
(f) Responsibility	Q6	Table 11	Analyze the impact of curriculum on the trend of responsibility.
(g) Autonomy	Q7	9% adopted LLM response	Analyze the impact of curriculum on understanding of autonomy.

**Table 6**  
Data that can be used for AI learning and data that should not be used.

Item	Content	Useable	Unusable	Statistics	P-value
(D1) Privacy	Personal information with consent (+1)	0.34	0.48	−1.91	0.06
	Personal information without consent (−1)	0.05	0.73	−9.99	0.00
(D2) Data protection	Open sources like Wikipedia (+1)	0.66	0.23	6.15	0.00
	Text with disclaimed copyright (+1)	0.69	0.13	8.08	0.00
	Copyrighted text (−1)	0.13	0.76	−9.10	0.00
(D3) Data Quality	Sentences with high accuracy (+1)	0.76	0.11	9.42	0.00
	Sentences with low accuracy (−1)	0.09	0.75	−9.59	0.00
(D4) Abuse of Data	Sentences causing shock or discomfort to users (−1)	0.09	0.67	−8.59	0.00

percentage of students who changed their opinions by age in Q2 and Q3. To clarify the influence of age on justice/fairness, we analyzed perceived fairness by age in Q4. In order to analyze the reasons why people felt it

**Table 7**  
Intentions before and after receiving LLM decision-making support.

	Q2	Q3
Soccer club	0.42	0.42
Baseball club	0.58	0.58

**Table 8**  
Percentage of people who changed their intentions by age.

Age Group	No change	Changed
12–15 years old	0.71	0.29
16–18 years old	0.78	0.22
19–21 years old	0.91	0.09
22–24 years old	0.87	0.13
ALL	0.83	0.17

**Table 9**  
Percentage of people who changed their intentions with or without experience using ChatGPT.

	Experienced in Use	No Experience in Use
No change	0.84	0.84
Changed	0.16	0.16

was unfair, we divided the respondents into those who felt it was fair and those who felt it was unfair based on their answers to Q4, and calculated the response rate for each reason. The method for calculating the response rate is the same as in formula (1). Subjects who belong to the group that felt it was fair were those who answered yes (4) or somewhat true (3) in Q4. The subjects who belong to the group that felt it was unfair were those who answered somewhat disagree (2) or disagree (1) in Q4. Next, regarding responsibility, we calculated the response rate for Q6 for each participant's age to understand the trends in the responsibility assumed by the participants. Finally, to analyze the tendency of participants to consider the AI's response as their final decision regarding autonomy, we calculated the response rate for each participant's age.

To assess the level of understanding of LLM concepts and ethical usage, we calculated the percentage of correct answers on the achievement test using the formula:

$$\text{Correct answer rate} = \frac{\text{Number of correct choices}}{\text{Number of participants}} \quad (2)$$

Compare the correct answer rates for each achievement test before and after the lecture using a chi-square test. The null hypothesis is that there is no difference in the proportions between the two groups. The significance level is 5%.

Because skills and abilities differ depending on age, education tailored to each age group is required. In this study, we will analyze the correct answer rate by age group and clarify the optimal age group for teaching LLM concepts and ethical usage.

Additionally, users who use ChatGPT on a daily basis may have a better understanding of how to utilize it than those who do not. Therefore, we will compare the correct answer rates of those who use ChatGPT on a daily basis and those who do not using a chi-square test. The null hypothesis is that there is no difference in the proportions between the two groups. The significance level is 5%.

#### 5.4.2. Impact on LLM technology acceptance

We analyzed the average and standard deviation of the answer results from Q8 to Q14. Additionally, we conducted covariance structure analysis based on the answers to Q1, Q4, Q6, Q7, and Q8–Q16, and constructed the TAM based on hypotheses H1–H9. Furthermore, in Q1, Q6, and Q7, scores were assigned based on hypothesis H8. However, for

**Table 10**  
Subjective fairness perceived by LLM.

	12–15 years old		16–18 years old		19–21 years old		22–24 years old		ALL	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Subjective Fairness (Q4)	2.59	0.69	2.57	0.51	2.3	0.56	2.74	0.53	2.54	0.58

**Table 11**  
Reasons why you felt it was fair (Q5).

Answer	Fair	Unfair	Statistics	P-value
(J1) Because it aligns with my criteria for judgment.	0.35	0.08	2.97	0.00
(J2) Because it deviates from my criteria for judgment.	0.07	0.34	−3.20	0.00
(J3) Because the judgment is based on objective indicators (numeric values).	0.39	0.24	1.82	0.07
(J4) Because I was making a comprehensive judgment.	0.39	0.08	3.60	0.00
(J5) Because it reflects the majority opinion.	0.11	0.11	0.14	0.89
(J6) Because it does not reflect the opinions of the minority.	0.05	0.13	−1.23	0.22
(J7) There are other solutions.	0.04	0.11	−1.27	0.21
(J8) There are other factors to consider when making a decision.	0.18	0.53	−3.37	0.00
(J9) Because the answer seemed unnatural.	0.02	0.08	−1.37	0.17
(J10) Because upon adding the condition (number of club members), the answer changed.	0.12	0.21	−0.97	0.33

**Table 12**  
Responsibility (Q6) results.

		12–15 years old	16–18 years old	19–21 years old	22–24 years old
(E1)	ChatGPT(+1)	0.06	0.09	0	0.03
(E2)	AI developer (+1)	0.35	0.05	0.09	0.1
(E3)	Providers (+1)	0.29	0.5	0.52	0.48
(E4)	End user (−1)	0.59	0.86	0.91	0.77

**Table 13**  
Correct answer rate before and after seminar.

Question number	Before	After	Statistics	P-value
TEST1	0.57	0.79	11.3	0.00
TEST2(i)	0.82	0.86	0.41	0.52
TEST2(ii)	0.85	0.96	7.61	0.01
TEST2(iii)	0.93	0.98	3.04	0.08

**Table 14**  
Correct answer rate by age.

Question number	12–15 years old	16–18 years old	19–21 years old	22–24 years old
TEST1	0.75	0.76	0.79	0.84
TEST2(i)	0.79	0.95	0.88	0.87
TEST2(ii)	0.74	0.90	1.00	0.94
TEST2(iii)	0.95	1.00	1.00	0.97

Q1, we utilized an integrated value, assigning +1 for valid answers and −1 for invalid answers regarding the data that can be used and the data that one does not want to use. Table 6 displays the answers assigned a +1 and those assigned a −1. Additionally, Q6 has not yet yielded a

conclusion regarding who will be responsible in the event of damage. In this study, scores were assigned based on whether the individual was responsible for their own fault or someone else's fault. Those who indicated self-responsibility received a −1 value, while those attributing responsibility to someone else received a +1 value. Table 12 illustrates the answers assigned a +1 and those assigned a −1. For Q7, we employed a dummy variable of +1 for respondents indicating that they would make the final decision themselves and −1 for those who stated that the AI would make it. Furthermore, an exploratory factor analysis was conducted to validate the factors in the constructed model. We selected the number of factors with an eigenvalue of 1 or more based on the Guttman criterion.

The analysis program was developed using Python 3.8.13, with the main packages being Numpy, Pandas, and semopy.

## 6. Results

### 6.1. Understanding the concepts and ethical usage of LLM

In Experiment 4 (ED4), 46% of the participants were unable to distinguish between human-generated sentences and those generated by AI. The results for Question 1 (Q1) are presented in Table 6. The null hypothesis was rejected for all items except “Personal information that I have consented to use.”

Intentions before (Q2) and after (Q3) receiving LLM decision support are presented in Table 7.

Table 8 displays the percentage of individuals who altered their intentions, categorized by age group. Table 9 exhibits the percentage of users who modified their intentions based on their usage experience. Tables 7 and 8 present the percentage of individuals providing different judgments for Q2 and Q3. In other words, it illustrates the proportion of people who adjusted their opinions due to LLM decision support.

Table 10 presents the subjective fairness that students perceived regarding the LLM. Table 11 displays the response rate for the reasons why students felt it was fair.

The null hypothesis was rejected for (J1), (J2), (J4), and (J8) based on the results of the chi-square test. The students found that the AI's judgment was comprehensive and aligned with my thoughts, which the students considered fair. However, students who believed there were other factors to consider in decision-making found the AI's responses unfair.

Table 12 illustrates the students' perceptions of responsibility. In relation to autonomy in Q7, 8.7% of students indicated that they would accept the LLM's response as their final decision.

Table 13 presents the correct answer rates and test results for the achievement test before and after the lecture. The chi-square test results indicated a significant difference in the correct answer rates for TEST1 and TEST2 (ii) before and after the lecture. Additionally, there was a notable difference in TEST1 and TEST2 (ii) before and after the lecture.

Table 14 presents the correct answer rates and test results for the achievement test before and after the lecture. The chi-square test results indicated a significant difference in the correct answer rates for TEST1 and TEST2 (ii) before and after the lecture. Additionally, there was a notable difference in TEST1 and TEST2 (ii) before and after the lecture.

Table 14 displays the percentage of correct answers for achievement tests categorized by age group. Meanwhile, Table 15 illustrates the percentage of correct answers to the achievement test based on the history of use. The chi-square test results revealed a significant

**Table 15**

Correct answer rate by usage experience.

Question Number	Have Used	Never Used	Statistics	P-value
TEST1	0.79	0.84	0.79	0.37
TEST2(i)	0.86	0.89	0.63	0.43
TEST2(ii)	0.96	0.85	7.61	0.01
TEST2(iii)	0.98	0.98	0.02	0.88

difference in the correct answer rate for TEST2(ii) depending on the presence or absence of experience using the test.

## 6.2. Impact on LLM technology acceptance

Table 16 presents the survey results for Q8-Q15. Analyzing the responses with a score of 3 (somewhat applicable) or higher in Tables 16 and it is evident that students perceived the LLM to have a positive impact on society. Additionally, students recognized the high utility value (performance) of the LLM and expressed a desire to utilize it for creative support. Following AI literacy education, 17% of respondents reported a negative perception, 1% maintained an indifferent attitude towards AI, and 83% exhibited a positive attitude towards AI.

Based on the results presented in Tables 5, 9 and 11, Q7, and the questionnaire in Table 16, we developed the technology acceptance model depicted in Fig. 3.

The total number of observed variables in the model was  $p = 15$ , with three factors identified. The Chi2 statistic for the model was 66.4, and the corresponding p-value was 0.98. The Chi2 test did not lead to rejection. Furthermore, the model exhibited excellent fit indices, with CFI = 1.00, TLI = 1.00, and RMSEA = 0.00. However, GFI was 0.79, and AGFI was 0.75. Refer to Appendix C for detailed information on the model illustrated in Fig. 3.

For smaller populations, reliable estimates of CFI and TLI are only achievable when  $p \leq 30$  (Shi et al., 2019). In cases where there are fewer than 30 observed variables and a sample size below 500, bias in sample fit index estimates may occur as the number of observed variables increases, leading to a fit significantly worse than the population value. It is essential to scrutinize for specification errors. For instance, the population value of the fit index may be substantially affected by the model size, such as when the number of factors is erroneously set to 2 instead of 1. Therefore, an exploratory factor analysis was conducted to verify the validity of the number of factors. Given the observed correlation between factors, a promax rotation was executed. The number of factors, determined using the Guttman criterion based on eigenvalues, is presented in Fig. 4. These results support setting the number of factors to 3.

Additionally, factor loadings in Table 17 were determined using the principal factor method. The results of the factor analysis revealed that Factor 1 influences social influence and utility value, Factor 2 affects expected effort and coping with hallucinations, and Factor 3 influences decision support and information retrieval, clarifying a distinct factor influencing creative support. These factor analysis outcomes align with the model presented in Fig. 3 regarding both the number of factors and the influencing factors.

Moreover, the model in Fig. 3 demonstrated satisfactory fit indices,

**Table 16**

Survey results regarding LLM technology acceptance.

Number	Items	12–15 years old		16–18 years old		19–21 years old		22–24 years old		ALL	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Q8	Social Impact	3.24	0.53	3.17	0.70	3.15	0.56	3.03	0.59	3.14	0.60
Q9	Performance expectancy	3.38	0.58	3.30	0.95	3.39	0.65	3.26	0.88	3.33	0.79
Q10	Effort expectancy	2.86	0.77	2.61	0.77	2.52	0.74	2.23	0.75	2.51	0.79
Q11	Deal with hallucination	2.76	0.75	2.57	0.88	2.55	0.56	2.23	0.71	2.49	0.73
Q12	Information retrieval	2.67	0.89	2.43	0.97	2.15	0.86	2.03	1.00	2.26	0.95
Q13	Creative support	3.67	0.56	3.57	0.82	3.30	0.80	3.45	0.80	3.47	0.78
Q14	Decision support	2.81	0.59	2.39	1.01	2.00	0.82	2.00	0.88	2.22	0.89

with CFI>0.95, TLI>0.95, and RMSEA<0.05, further supporting its reasonableness.

## 7. Discussion

### 7.1. Understanding the concepts and ethical usage of LLM

The ED4 results suggest that ChatGPT text and human text are indistinguishable. The results in Table 6 show that approximately 70% of students understand how to properly manage data and ensure non-maleficence, which is one of the ethical principles of AI.

In Tables 8 and it is observed that the younger the age, the more likely AI response results were to be adopted. Furthermore, the proportion of respondents who adopted AI responses was similar between those with and without experience using AI. Therefore, it is inferred that the adoption of AI answers depends on age. One possible reason for this is that younger individuals possess less experience and knowledge, making them more susceptible to automatic bias. A similar trend has been reported in the medical field (Dratsch et al., 2023). Dratsch et al. found that inexperienced radiologists were more likely to follow AI recommendations even if they were incorrect. It has also been noted that errors may occur due to overconfidence in decision support system functionality or lack of confidence in one's abilities (Goddard et al., 2012; Skitka et al., 2000). Like these reports, automatic bias may have led younger students to accept AI answers more positively. Conversely, perceived fairness was not affected by age. The students perceive fairness when judgments of ai align with their own and are comprehensive, whereas feelings of unfairness arise when the opposite is true.

In Table 12, most participants indicated that the end user is responsible, followed by providers, developers, and AI. The fact that less than 10% of students used AI for their final decision suggests an understanding of autonomy.

The content learned through these lectures was evaluated with an achievement test. The curriculum proposed in Table 13 improved the understanding of TEST1 and TEST2 (ii). This implies that many students misunderstood the concept of LLM before taking the course and lacked an understanding of how to utilize decision support. However, results for TEST2(i) and TEST2(iii) in Table 13 revealed no significant differences in information search or coping with hallucinations. This could be attributed to prior experience. Table 14 shows that, on average, younger individuals exhibited lower decision support performance. Additionally, from Table 15, when there was a history of use, decision support performance tended to be lower than when there was no history of use. Table 4 indicates that the percentage of individuals with experience using ChatGPT tends to decrease with age. Considering these findings, the low performance in decision support may be attributed to the history of use. Conversely, in terms of decision-making support, it is anticipated that use will lead to a better understanding of ethical usage. The absence of age-related differences in dealing with hallucinations and information search stems from a high level of pre-existing understanding, given their familiarity with searching for information and addressing false information on the widely used Internet.

Overall, the results of Q1-Q8 and the achievement test demonstrate

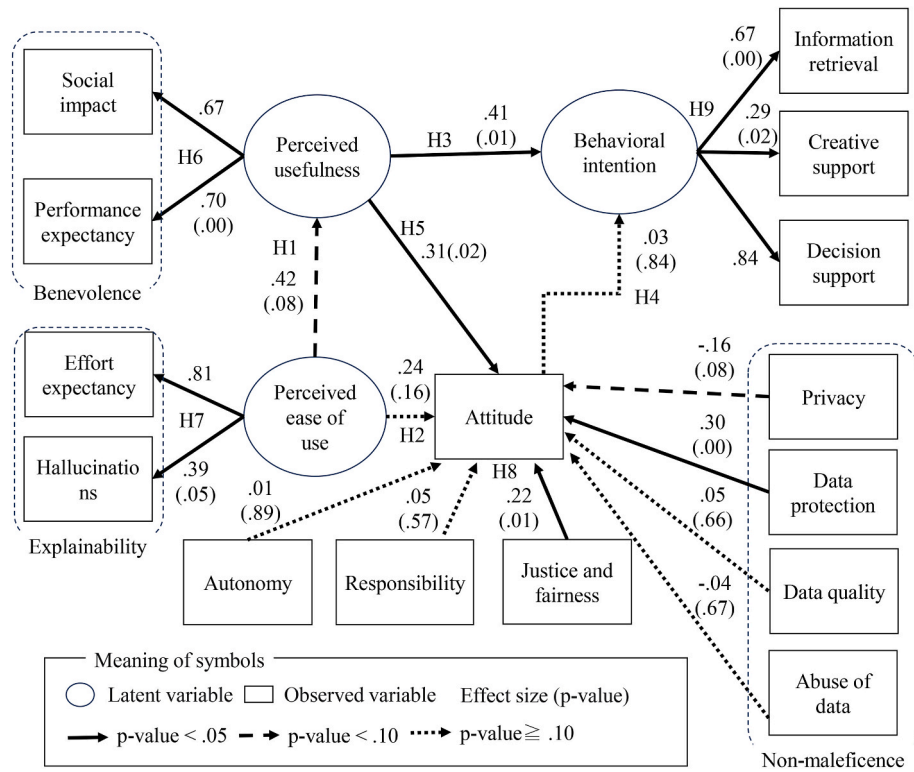


Fig. 3. LLM technology acceptance model integrating AI ethics.

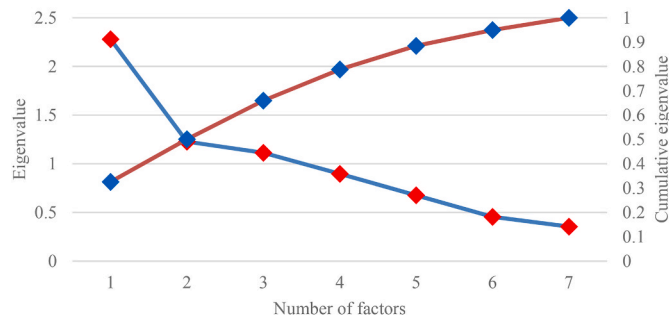


Fig. 4. LLM technology acceptance model integrating AI ethics.

that this curriculum enables the acquisition of basic knowledge and ethical usage methods before ethical use. It is also evident that individuals exposed to technology are more likely to understand how to use it ethically.

## 7.2. Impact on LLM technology acceptance

From Fig. 3, hypotheses H3, H6, H7, and H9 were supported. H1 is likely to be supported. Additionally, H8 was partially supported. Related research indicates that usefulness (social influence and expected performance) has a positive effect on attitude, and that convenience (expected amount of effort) has no direct effect (Habibi et al., 2023;

Strzelecki, 2023). In this study, we introduced hallucinations specific to LLM as external variables, but their effect did not change. However, in the TAM of this study, it was newly revealed that convenience, including expected effort and appropriate coping with hallucinations, may indirectly influence behavioral intentions through perceived usefulness. Factors influencing AI attitudes were usefulness, justice/fairness, privacy, and protection of data.

Contrary to the initial hypothesis, privacy protection had a negative effect on AI attitudes. However, rather than interpreting this as a feeling of anxiety when privacy is protected, it is more natural to interpret this as a feeling of anxiety among those who answered that their personal information can be used if they have consented to its use. Privacy risks arise not only when data is collected without the user's explicit consent, leaked to external agents, or de-anonymized (Narayanan & Shmatikov, 2008) but also when data is used in the stage of inference based on data. This can also occur at the stage of inference based on data, so this is also considered to be one of the reasons why people feel anxious. Additionally, about half of the students answered that they cannot use the personal information they have agreed to use, suggesting that they do not want their personal information to be learned by AI. This indicates that AI developers and providers need to be careful about personal information to foster a positive AI attitude.

It was also confirmed that justice/fairness influenced the AI attitude. The technology acceptance model showed that the fairer students perceive AI, the more positive their attitude toward AI will be. Furthermore, in Section 7.1, it was suggested that automatic bias may have caused younger students to actively adopt the AI's answers. The

Table 17

Exploratory factor analysis results.

	Social impact	Performance expectancy	Effort expectancy	Hallucinations	Information retrieval	Creative support	Decision support
Factor1	0.75	0.91	0.3	-0.19	0.11	0.09	-0.1
Factor2	0.03	-0.01	0.72	0.73	0.2	-0.5	0.13
Factor3	0.11	-0.08	-0.03	0.27	0.68	0.56	0.91
Communality	0.58	0.83	0.61	0.64	0.51	0.56	0.86



reason why people feel it is fair is that their judgments are consistent with their own judgments and that their judgments are comprehensive. The reason why it feels unfair is the opposite. These results suggest that it may be possible to intentionally induce opinions from younger students. Therefore, caution should be taken when using decision aids with younger students. However, this study only measured the effects in a relatively small group and over a short period of time, and the long-term effects of decision-making support on younger students require additional examination.

Data protection is crucial for fostering positive attitudes towards AI. On the other hand, data misuse and quality did not influence attitudes toward AI. However, it is important to note that this may change as our understanding of AI learning advances, as the quality of these data and the prevention of misuse are potentially tied to convenience. Furthermore, autonomy and responsibility did not influence attitudes toward AI.

Although it has become clear that some AI ethical principles influence attitudes toward AI, it is important to keep in mind that these attitudes toward AI are not directly linked to intentions to use it. This suggests that although some users may feel uneasy, they may have formed an intention to use AI because they recognize that it has high usefulness. According to prospect theory (Kahneman & Tversky, 2013), individuals evaluate potential gains and losses during the decision-making process and are influenced by the perceived risks and rewards associated with each option. In the case of ChatGPT, users may also accept ChatGPT's technology if they believe that the benefits of the technology outweigh the potential risks. Similar trends have been reported in other technological fields such as autonomous driving (Cugurullo & Acheampong, 2023), suggesting that heightened expectations may lead to increased behavioral intentions.

Finally, Table 16 reveals that expectations were highest for creative support, decision support, and information retrieval in that order. Fig. 3 also supports the fact that this behavioral intention has a direct positive effect on creative support, decision support, and information search. Many studies have reported that behavioral intention is the strongest predictor of behavior (Habibi et al., 2023; Strzelecki, 2023), which makes LLM useful for these applications.

## 8. Limitation

It is important to note that the results of our study are in close agreement with related prior studies, which strengthens the validity of the arguments proposed in this paper. However, considering the demographic background, it is essential to highlight that the subjects in this study were exclusively Japanese. Consequently, this study is limited in its ability to discuss cultural differences in the acceptance of AI. Moreover, it is crucial to acknowledge that only short-term effects were measured, and long-term effects necessitate additional verification.

Additionally, in this study, we administered a questionnaire after providing information on ethical usage, and the modeling of technology acceptance was based on this data. Consequently, there may be a bias towards individuals with a certain level of education. However, if the curriculum proposed in this study becomes widespread, we can anticipate the educational effects and technology acceptance reported in this study to have broader applicability.

As demonstrated in related study (Ray, 2023), LLMs are anticipated

to be utilized across a wide array of fields and for diverse purposes. The achievement test conducted in this study may not have been capable of comprehensively assessing the understanding of ethical usage in these varied contexts. Further discourse is warranted regarding the ethical usage requisite in K-12 education and the items and content of the achievement test aimed at measuring it.

## 9. Conclusion

In this study, we have proposed a curriculum tailored for inexperienced beginners utilizing Language Model (LLM), such as ChatGPT, to teach ethical usage. AI literacy education was administered to students aged 12 to 24, and the effectiveness of the curriculum incorporating the suggested ethical principles was documented. Regarding the use of LLM in decision support applications, students aged 12 to 18 reported a positive reception of AI responses. Additionally, it was observed that certain ethical principles of AI (benevolence, justice/fairness, privacy, and data protection) significantly influence attitudes toward AI. The study also highlighted that the most anticipated applications for AI are creative support, decision support, and information retrieval, in that order. This finding will contribute to the design of a curriculum to promote the adoption of LLMs.

Moving forward, our research will delve into investigating the long-term effects of LLM decision-making support on individuals aged 12 to 18.

## Statements on open data, ethics and conflict of interest

The study was approved by an ethical committee in Komatsu University with ID: 2312-2. Informed consent was obtained from all participants, and their privacy rights were strictly observed. The participants were protected by hiding their personal information during the research process. They knew that the participation was voluntary and they could withdraw from the study at any time. There is no potential conflict of interest in this study. The data can be obtained by sending request e-mails to the corresponding author.

## CRediT authorship contribution statement

**Yusuke Kajiwaru:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Kouhei Kawabata:** Validation, Software, Methodology, Investigation, Formal analysis, Data curation.

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

During the preparation of this work the authors used ChatGPT in order to improve language and readability. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

On the first day, we visited tourist spots that represent Kanazawa, such as (i) and (ii). In (i), I was soothed by the seasonal flowers while looking out at the splendid garden. In (ii), we enjoyed the magnificent castle walls and castle tower while feeling the history of Japan.

On the second day, we enjoyed visiting the wonderful cafes in the city. We visited popular coffee shops and patisseries in Kanazawa and tasted beautiful sweets and coffee. Additionally, in (iii), the students were able to touch contemporary artworks and watch life paintings created by artists right in front of their eyes.

## Declaration of competing interest

None.

## Appendix A. Details of AI Literacy Curriculum Challenge for Ethical Use of LLM

The details of the assignment are shown below. The original Japanese text has been translated into English.

ED1. This text is part of my travel diary when I went sightseeing in Kanazawa City, Ishikawa Prefecture. Please answer the words that apply to (i)~(iii).

The options are "Kanazawa Castle," "Kenrokuen," and "21st Century Museum of Contemporary Art."

ED2. This text is part of my travel diary when I went sightseeing in Kanazawa City, Ishikawa Prefecture. Please answer the words that apply to (iv)~(vi).

The options for (iv) are "went on," "came on," and "ate.". The options for (v) are "I ate," "I went," and "I rode.". The options for (vi) are "Dango," "House," and "Study."

ED3. Try to answer the questions appropriately. Please answer the word that applies to (vii).

The options for (vii) are "1 + 2 = 2," "1 + 2 = 3," "1 + 2 = 4," and "1 + 2 = 5."

ED4. Which of the answer examples (1) and (2) about how to deal with influenza is ChatGPT?

Figure Answer example: Answer example (1) is a sentence created with ChatGPT.

ED5-1. Prompt example 1: Answer changes probabilistically.

ED5-2. Prompt example 2: Answer changes probabilistically.

ED6. Example prompt: Hallucination.

ChatGPT answer. There are two incorrect answers. One is Komatsu Town in Fukushima Prefecture and Komatsu City in Ishikawa Prefecture. Another thing is that "2006" is "2014".

ED7-1. Prompt example: Information retrieval.

ED7-2. Prompt example: creative support.

ED7-3. Prompt example: decision support.

Q2 and Q3. Example prompt: Justice and fairness.

There is a dispute between the soccer club and the baseball club over the usage of the playground. The question is, who should be granted permission to use the playground? The baseball club was the first to secure the right to use the sports field. The baseball club consists of 100 members, while the soccer club has 300 members. The baseball club engages in activities once a week, whereas the soccer club plays once a month. Considering these factors, which club should be given the privilege of using the playground? (After students have responded, present them with the following text and ask them to reconsider their decision.)

## Appendix B. Validity of Options for Q1, Q5, Q6

A preliminary analysis was conducted on 62 first-year university students. The university students who conducted the pre-analysis were not included in this experiment's analysis. The training proceeded following steps ED1-ED7-3 and Q1-Q7. In contrast to this experiment, Q1, Q5, and Q6 were answered in a free-text format. The contents of these texts were meticulously examined and broadly categorized into (i)-(iv):

(D1) Privacy

(D2) Data protection

(D3) Data quality

(D4) abuse of data

**Table B1**

Data that can be used for AI learning.

Item	Content
D3	Credible information
D3	Objective data such as mathematical and physical formulas.
D2	It is acceptable to use information found more than a certain number of times on the internet.
D2,D3	Data posted by experts. Information posted on a company's official website.
D2,D3	Personal blogs and wikis
D2,D3	Classical Chinese text
D3	Highly reliable data
D3	Correct information that can be confirmed as fact
D3	Test content and results. Papers, etc.
D1	Shopping and large flow of people
D2	Items that do not violate copyright. You can use all of D1-D4.

(continued on next page)

I (iv) a trip to Kanazawa City, Ishikawa Prefecture, located in the Hokuriku region of Japan, and (v) (vi).

<Question> Please calculate  $1 + 2$ .

<Answer>(vii)

**[Sample answer (1)]** To quickly cure influenza.

- Take medicines prescribed by the hospital.
- Stay hydrated, eat nutritious meals, and get plenty of sleep.
- Immune cells are more active when the temperature is high, so avoid using antipyretics on your own. If the pain is so severe that you cannot eat or drink water, use under the direction of your doctor.

etc. can be mentioned.

**[Sample answer (2)]** To quickly cure the flu, pay attention to the following points.

- Get enough rest and sleep.
- Make sure to stay hydrated.
- Use antipyretics to reduce fever (follow your doctor's instructions).
- Eat nutritious meals.
- Observe hand washing and cough etiquette.

However, if you have serious symptoms or complications, consult your doctor.

<you>

What kind of place is Komatsu University? What's the charm?

<ChatGPT>

Komatsu University is a public university located in Komatsu City, Ishikawa Prefecture. The following is an explanation of the appeal of public Komatsu University.

- Education specialized in related industries: Public Komatsu University provides education centered on academic fields related to local industries, such as mechanical engineering and information engineering. Students will gain practical knowledge and skills while deepening relationships with local businesses.
- Fulfilling research activities: Public Komatsu University ...

<you>

What kind of place is Komatsu University? What's the charm?

<ChatGPT>

Komatsu University is one of Japan's public universities, with a campus located in Komatsu City, Ishikawa Prefecture. The attractive points of this university are as follows.

- Energetic students: Komatsu University is characterized by its students who are very energetic and lead an active student life. There are many student organizations such as clubs and circles, and students are passionate about what they like.

**Table B1** (continued)

Item	Content
D3	More accurate data.
D2	Information on public facilities

(continued on next page)

<you>

Tell me about Science Hills Komatsu?

<ChatGPT>

Science Hills Komatsu is a science museum located in Komatsu City, Fukushima Prefecture. It opened in 2006 and is also used as an educational venue for local children. Inside the museum, there are exhibition rooms that deal with themes such as the earth, space, biology, and science and technology, as well as exhibition rooms that deal with themes such as scientific experiments, and laboratories where you can conduct scientific experiments.

<you>

How do frogs breathe?

<ChatGPT>

Frogs breathe using their lungs and skin. Frogs have very small lungs and absorb oxygen primarily through their skin. The skin is very thin and moist, allowing for easy exchange of oxygen and carbon dioxide. Frogs can breathe using their lungs, but their lungs do not function underwater, so breathing through their skin becomes important. When underwater, frogs close their lungs and take in oxygen through skin breathing.

<you>

Please write a fantasy novel.

<ChatGPT>

Title: "Magician of the Land of Mist"

One day, a phenomenon occurred in which a vast fog covered the entire world. Mysterious creatures and magic existed within the fog, profoundly impacting people's lives.

<you>

What would you like for dinner today?

<ChatGPT>

The options for today's dinner are endless! Considering your preferences and nutritional balance, we offer some ideas below:

1. Grilled Fish and Seasonal Vegetables: Grilling fish such as cod or mackerel and serving it with colorful seasonal vegetables is a healthy and delicious choice.
2. Vegetable Curry: A vegetable-filled curry is rich in vitamins and minerals. You can also add tofu or chicken.

**Table B1** (continued)

Item	Content
D3	Mathematics, physics, etc. Latest information on D2 official homepage
D2,D3	Data provided by the government and data provided by companies
D3	Data that is certified by a public institution or data that is accessed frequently.
D2	Only data that can be legally viewed for free
D3	Contents written on the thesis administration website, etc., can be considered correct.
D3	Literature
D3	Proper data from national and other institutions
D3	A scientifically proven paper
D3	Data that has a certain degree of credibility, such as papers
D3	Correct data
D1	Other than information that can identify individuals
D3	Government homepage
D3	Calculation
D1	Non-personal information
D2	Items without copyright
D3	Information sent by individuals rather than organizations.
D2	Information that is generally widely available.



<you>

Which club do you believe should be granted the rights—the soccer club or the baseball club?

<ChatGPT>

Soccer club.

<you>

Why?

<ChatGPT>

We aim to arrive at a fair solution by taking into account various factors, such as the number of members and the order of usage. However, due to the substantial difference in the number of club members, certain priority may be granted to the soccer club.

<you>

If the baseball team is active once a week and the soccer club is active once a month, will the criteria for prioritizing clubs change?

<ChatGPT>

Yes, depending on the frequency of activities, the criteria for determining priority will change. This would grant the baseball club the right to operate more frequently.

**Table B2**

Data that cannot be used for AI learning.

Item	Content
D3	Data that varies from person to person, such as opinions and views.
D3	Content with limited information or unclear explanations.
D3	Data from sites created by individuals.
D3	Art and voice forgery.
D2	Military Intelligence.
D3	Unreliable and incorrect information.
D1	Personal information, such as addresses.
D1	Privacy-related matters and Personally Identifiable Information.
D2	Anything that violates copyright.
D3	More inaccurate data.
D4	Information about hackers and other crimes.
D3	Items of unknown origin.
D3	Data and blogs with imperfect language.
D3	Literature.
D2	Paid data.
D3	Information from personal sites or sites with unknown sources.
D3	Personal opinions on illustrations, blogs, etc.
D3	Data cited by unknown individuals.
D3	Information that may or may not be correct.
D2	Copyrighted works.
D1	Personal information.
D3	Disinformation.
D4	Indecent content, such as explicit imagery.
D1-D4	All of the above.
D1	Personally identifiable information, etc.
D1	Personal information.
D2	Pictures.
D1	Blogs, including my own photo.
D1	Personal information.
D2	Copyrighted material.

Next, the free-text answers to Q5 are presented in Table B2. The contents of these sentences were thoroughly examined and categorized into (J1)-(J10):

- (J1) Because it aligns with my criteria for judgment.
- (J2) Because it deviates from my criteria for judgment.
- (J3) Because the judgment is based on objective indicators (numeric values).
- (J4) Because I was making a comprehensive judgment.
- (J5) Because it reflects the majority opinion.
- (J6) Because it does not reflect the opinions of the minority.
- (J7) There are other solutions.
- (J8) There are other factors to consider when making a decision.
- (J9) Because the answer seemed unnatural.
- (J10) Because upon adding the condition (number of club members), the answer changed.

**Table B3**

Reasons why you felt it was fair (free description)

Fairness Category	Items	Contents
Fair	J4	The answers are based on important factors for J4 club activities.
Fair	J3	Because I considered priorities based on facts.
Fair	J4	The reason for giving priority to both club activities was stated.
Fair	J3, J4	There is a difference in the frequency of activities between the baseball club and the soccer club, and we are trying to make that difference as close to fair as possible.
Fair	J4	Considering the number of people and the frequency, it would be better to use the soccer club.
Fair	J4	Judgment is made taking into account the difference in the number of club members and the frequency of activities.
Fair	J4	Considering the number of club members, frequency of activities, etc.
Fair	J3	I was thinking more theoretically than myself.
Fair	J1	You can use both if you prioritize the one with the most frequency.
Fair	J4	They carefully judged the usage situation.
Fair	J4	To finely adjust the amount of activity.
Somewhat fair	Others	I'm confused because I don't fully understand what fairness is.
Somewhat fair	J1	Fairly reasonable.
Somewhat fair	J4	It seemed like they were considering a lot of things.
Somewhat fair	J8	Fair, but not completely fair without the opinions of both parties.
Somewhat fair	J6	I thought that the soccer club should be given some priority because the frequency of activities is far less than that of the soccer club.
Somewhat fair	J1	ChatGPT himself said he was aiming for a fair solution, and I thought he was thinking about it fairly as well.
Somewhat fair	J1,J8	It's not good for the baseball club to have 100 people in their activities per week, and the soccer club to have 300 people in their activities once a month, but they can't be active.
Somewhat fair	J3,J4	Because we took into account the difference in the number of club members and the frequency of activities.
Somewhat fair	J4	The baseball team had priority first, and the reason there were more members was the soccer club, and they were expressing their opinions taking into consideration the difference in the number of active days.
Somewhat fair	J3	Decisions are made based on data and efficiency rather than personal feelings and interests, such as the number of club members and order of use.
Somewhat fair	J3	Because he answered by comparing the elements of both club activities.
Somewhat fair	J3,J4	The number of club members and order of use were taken into consideration comprehensively.
Somewhat fair	J3,J4	The decision was made taking into consideration the number of people and frequency of use.
Somewhat fair	J3,J4	I felt it was fair in that it took into account the number of people and the frequency of activities.
Somewhat fair	Others	Elimination method.
Somewhat fair	J10	I felt that the judgment was based on only one indicator, such as activity frequency, number of people, etc.
Somewhat fair	J3	Because the priority is determined using some criteria.
Somewhat unfair	J8	When asked, "Which do you approve of, the soccer club or the baseball club?", the answer was "soccer club" without any criteria.
Somewhat unfair	J2	The baseball club will be unable to operate at all.
Somewhat unfair	J2	If you were to judge based on the difference in activity frequency, I felt it would be the soccer club.
Somewhat unfair	J2	If the next use will be in a month, I think it would be better to give it to the soccer club.
Somewhat unfair	J8	I didn't think they took into consideration the fact that they had acquired the rights in the first place.
Somewhat unfair	J8	Because I felt like I didn't understand the baseball club's position.
Somewhat unfair	J8	The strength and seriousness of the baseball and soccer clubs are not reflected.
Somewhat unfair	J8	The focus was on frequency, and not much mention was made of the number of people.
Somewhat unfair	J8	Because they ignored that they had the right to use it.
Somewhat unfair	J5,J6	Because they place too much emphasis on the number of people and do not listen to the opinions of the minority.
Somewhat unfair	J8	I just thought that if the baseball club had applied first, they should have been given priority.
Somewhat unfair	J8	I thought that the fact that they had obtained permission first would give them a considerable advantage compared to the number of active members.
Somewhat unfair	J8	At the time of the first question, the decision was made based on the number of club members, which is an uncertain factor.
Somewhat unfair	J2	I thought it was not fair to the baseball club.
Somewhat unfair	J8	The value of using the ground is slightly different between the baseball club and the soccer club, so the arguments of both sides must be carefully considered.
Somewhat unfair	J7	The answer was divided between pros and cons.
Somewhat unfair	J10	Because the rules that were originally decided were changed.
Somewhat unfair	J8	Because only the difference in the number of club members is considered.
Somewhat unfair	J8	The fact that it was decided in advance means that I have a strong authority within myself. Because we thought the strengths of each group (number of people, number of activities) were the same.
Somewhat unfair	J8	I felt that the importance of the order of reservations was out of proportion to the number of people and frequency of activities, and that the order of reservations was more important.

(continued on next page)

**Table B3** (continued)

Fairness Category	Items	Contents
Somewhat unfair	J8	Does not fully reflect the elements. Also, from the comments made by CHATGPT, there was no sign that they were making fair judgments, and it seemed like they were just making up their own minds.
Somewhat unfair	J8	The baseball club had the right to use the field first.
Somewhat unfair	J10	I thought it could be said that priority should be given to less frequent activities because there are fewer opportunities to use the ground.
Somewhat unfair	J8	It's fair when you look at the numbers alone, but when you look at other factors, it's not fair.
Somewhat unfair	J10	Considering the frequency of activity, my opinion has changed.
Somewhat unfair	J8	I think the soccer club is given priority because of the number of members, so the one with the rights should be given first.
Somewhat unfair	J2	Considering the frequency of use, priority should be given to the soccer club, which can only be used once a month, but that was not the case.
Somewhat unfair	J10	First, they only look at the number of people, and second, they think that the frequency of activities is a measure of their rights.
Unfair	J8	Because rights are ignored.
Unfair	J9	Because the answer was disgusting.
Unfair	J8	Priority should be given to those who have obtained the rights first.
Unfair	J8	The baseball club has the rights.
Unfair	J7	I thought that if it was to be used fairly, it would be good to have separate days for each activity.
Unfair	J2, J5	The first answer states that they are aiming for a fair solution, but they are giving priority to the soccer

The free-text answers to Q6 are presented in Table B3. The contents of these texts were thoroughly examined and broadly classified into categories (E1)-(E4):

- (E1) ChatGPT
- (E2) AI developer
- (E3) Providers
- (E4) End user

**Table B4**  
Responsibility (Free Description)

Item	Content
(E2)	They could have taken measures to prevent the release of incorrect information, but they did not do so. This is because the text does not state that the information is not necessarily correct, even though it is not necessarily correct.
(E4)	If you search not only ChatGPT but also the Internet, you will find this easily.
(E4)	Because I trusted uncertain information too easily.
(E4)	Because I believed the information and acted on it.
(E4)	The information generated by ChatGPT is not guaranteed to be correct, so we do not use it.
(E4)	I don't think I would have suffered any damage if I had not believed the false information myself.
(E4)	It's your decision whether to use ChatGPT or not.
(E1)(E4)	Since I can't hold ChatGPT accountable, I think it's my fault for trusting them.
(E4)	Because it's your own responsibility.
(E4)	You should consider that the information is not necessarily correct.
(E4)	It's my fault for believing it.
(E4)	They don't understand the possibility that ChatGPT is wrong and believe it without checking the facts.
(E4)	Failure to confirm the authenticity of information, which is something to be careful about when using ChatGPT.
(E4)	I think it would be worse to use ChatGPT, which is full of uncertain information.
(E4)	It's my fault for using it even though there is no guarantee that it will always work.
(E4)	There is a lot of false information on the Internet, so you need the ability to see through it.
(E4)	Because he was the one who believed and acted on something that was insufficient for someone to believe in.
(E4)	Because I did not examine various sources of information.
(E4)	Even though I didn't know if the information was correct, I believed it and treated it as if it were correct.
(E4)	Modern society is full of information that is unclear whether it is correct or incorrect, and it is necessary to use information literacy to determine whether the information is correct or incorrect.
(E4)	Since ChatGPT is an AI, I cannot take responsibility for it, so I thought that the person who used the information as a reference should take responsibility.
(E4)	Not only ChatGPT but also other people's words, the final decision is made by yourself, and you are the one who makes the decision.
(E4)	The world is full of false information, so those who have not acquired information literacy should be held responsible.
(E4)	I made a mistake in using the AI, which is in the position of supporting thinking.
(E4)	You should have asked an expert instead of ChatGPT, so it's your own fault for accepting something you don't know if it's true or not.
(E4)	You have the power to decide whether to use ChatGPT or not.
(E4)	I think the premise is that generative AI will transmit incorrect information.
(E4)	I should have been warned in advance that CHATGPT might provide false information, so I felt responsible for agreeing to it.
(E4)	Because I couldn't distinguish between incorrect information and correct information.
(E4)	If GPT developers and GPT were to be held responsible, the number of lawsuits would increase to an extraordinary number, and a society where it would be a loss to release developed apps would spread, which would lead to growth in the AI industry as a whole. There is a possibility that it may be delayed. I think a similar thing can be said about holding the sender of false information accountable in the internet society and the IT industry.
(E4)	The reason people believe the wrong information on ChatGPT is because the person lacks internet literacy.
(E4)	No, it's better to ask an expert, but you're too stupid to accept the information on ChatGPT and think it's not your fault.

(continued on next page)

**Table B4** (continued)

Item	Content
(E4)	It was my own fault for not consulting the appropriate authorities.
(E4)	Don't leave all the choices to AI. You should ask an expert.
(E4)	Because you are the one who ultimately judges the information.
(E4)	It was the person who believed in ChatGPT, so I think the person who believed it is responsible for any incorrect answers.
(E4)	It's my own fault for not being able to judge false information as a lie.
(E4)	It is difficult to hold AI accountable.
(E4)	I trusted only the information on ChatGPT and did not check whether the information was really correct.
(E4)	Considering that ChatGPT can generate incorrect information, the problem lies with those who take it at face value without being able to tell whether the information is true or not.
(E4)	I didn't doubt whether ChatGPT was really correct.
(E4)	Ultimately, you are responsible for using ChatGPT.
(E4)	Basically, even if the information is wrong, it is the person receiving the information who decides whether to believe it or not, so the person who believes the wrong information is responsible for it.
(E4)	I read various articles and didn't confirm.
(E4)	If you look into it, you'll understand.
(E4)	Self-responsibility.
(E4)	Instead of relying on AI, you should ask experts.
(E4)	It was the mistake of the person who did not check the facts.
(E4)	People who cannot distinguish between false and true information should not use it in the first place.
(E3)(E4)	Because you don't judge for yourself whether the information is correct or not. Also, the people sending out the information have not checked to see if it is accurate.
(E3)(E4)	There is a very high possibility that the information provided by ChatGPT is incorrect, and senders who do not include this information should be held responsible. We must also take responsibility for our beliefs, even though they may be wrong.
(E2)(E3)	I have heard the idea that "All humans are burdened with sin from the time they are born. The very act of being born is evil." Therefore, all humans have a responsibility.
(E4)	
(E3)(E4)	The reason why both the person who did it and the person who was treated is

## Appendix C. Details of the Technology Acceptance Model

Table C1 presents the estimation results of the effect size of the TAM, along with details such as standard error, statistics, and p-value.

**Table C1**  
Technology Acceptance Model (Regression, Covariation Relationship)

lval	op	rval	Estimate	Est.Std	Std.Err	z-value	p-value
Social impact	~	Perceived usefulness	1	0.67	–	–	–
Performance expectancy	~	Perceived usefulness	1.37	0.7	0.35	3.92	0
Perceived usefulness	~	Perceived ease of use	0.27	0.42	0.15	1.75	0.08
Effort expectancy	~	Perceived ease of use	1	0.81	–	–	–
Hallucinations	~	Perceived ease of use	0.46	0.39	0.23	1.95	0.05
Attitude	~	Perceived usefulness	0.58	0.31	0.26	2.25	0.02
Attitude	~	Perceived ease of use	0.28	0.24	0.2	1.4	0.16
Attitude	~	Privacy	–0.12	–0.16	0.07	–1.78	0.08
Attitude	~	Data protection	0.13	0.3	0.04	3.28	0
Attitude	~	Data quality	0.03	0.05	0.07	0.44	0.66
Attitude	~	Abuse of data	–0.05	–0.04	0.11	–0.43	0.67
Attitude	~	Justice and fairness	0.17	0.22	0.06	2.74	0.01
Attitude	~	Responsibility	0.04	0.05	0.08	0.56	0.57
Attitude	~	Autonomy	0.02	0.01	0.12	0.14	0.89
Behavioral intention	~	Perceived usefulness	0.77	0.41	0.31	2.44	0.01
Behavioral intention	~	Attitude	0.03	0.03	0.13	0.2	0.84
Information retrieval	~	Behavioral intention	0.85	0.67	0.24	3.5	0
Creative support	~	Behavioral intention	0.29	0.29	0.12	2.37	0.02
Decision support	~	Behavioral intention	1	0.84	–	–	–
Perceived usefulness	~~	Perceived usefulness	0.13	0.82	0.05	2.65	0.01
Perceived ease of use	~~	Perceived ease of use	0.41	1	0.21	1.92	0.06
Behavioral intention	~~	Behavioral intention	0.48	0.83	0.17	2.79	0.01
Attitude	~~	Attitude	0.34	0.62	0.05	6.27	0
Performance expectancy	~~	Performance expectancy	0.31	0.51	0.08	3.75	0
Social impact	~~	Social impact	0.2	0.56	0.05	4.29	0
Hallucinations	~~	Hallucinations	0.46	0.85	0.07	6.16	0
Effort expectancy	~~	Effort expectancy	0.21	0.35	0.2	1.08	0.28
Decision support	~~	Decision support	0.24	0.29	0.16	1.52	0.13
Information retrieval	~~	Information retrieval	0.51	0.55	0.13	3.92	0
Creative support	~~	Creative support	0.55	0.92	0.08	7.14	0

## References

- Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... McGrew, B. (2023). *GPT-4 technical report*. arXiv preprint arXiv:2303.08774.
- Al Darayseh, A. (2023). Acceptance of artificial intelligence in teaching science: Science teachers' perspective. *Computers & Education: Artificial Intelligence*, 4, Article 100132.
- AlAfnan, M. A., Dishari, S., Jovic, M., & Lomidze, K. (2023). Chatgpt as an educational tool: Opportunities, challenges, and recommendations for communication, business writing, and composition courses. *Journal of Artificial Intelligence and Technology*, 3 (2), 60–68.



- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bannetot, A., Tabik, S., Barbado, A., ... Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115.
- Beauchamp, T. L., & Childress, J. F. (2001). *Principles of biomedical ethics*. USA: Oxford University Press.
- Buchanan, J., Hill, S., & Shapoval, O. (2023). ChatGPT hallucinates non-existent citations: Evidence from economics. *The American Economist*, Article 05694345231218454.
- Burgsteiner, H., Kandhofer, M., & Steinbauer, G. (2016). Irobot: Teaching the basics of artificial intelligence in high schools. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 30, p. 1).
- Cooper, G. (2023). Examining science education in chatgpt: An exploratory study of generative artificial intelligence. *Journal of Science Education and Technology*, 32(3), 444–452.
- Cox, C., & Tzoc, E. (2023). ChatGPT: Implications for academic libraries. *College & Research Libraries News*, 84(3), 99.
- Cugurullo, F., & Acheampong, R. A. (2023). Fear of AI: An inquiry into the adoption of autonomous cars in spite of fear, and a theoretical framework for the study of artificial intelligence technology acceptance. *AI & Society*, 1–16.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 319–340.
- Del Giudice, M., Scuto, V., Orlando, B., & Mustilli, M. (2023). Toward the human-centered approach: A revised model of individual acceptance of AI. *Human Resource Management Review*, 33(1), Article 100856.
- DiPaola, D., Payne, B. H., & Breazeal, C. (2020). Decoding design agendas: An ethical design activity for middle school students. In *Proceedings of the interaction design and children conference* (pp. 1–10).
- Dratsch, T., Chen, X., Rezaei, Mehri, M., Kloeckner, R., Mähringer-Kunz, A., Püsken, M., ... Pinto dos Santos, D. (2023). Automation bias in mammography: The impact of artificial intelligence BI-RADS suggestions on reader performance. *Radiology*, 307(4), Article e222176.
- Druga, S., Vu, S. T., Likhith, E., & Qiu, T. (2019). Inclusive AI literacy for kids around the world. In *Proceedings of FabLearn 2019* (pp. 104–111).
- Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference* (pp. 214–226).
- Essel, H. B., Vlachopoulos, D., Essuman, A. B., & Amankwa, J. O. (2024). ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based conversational large language models (LLMs). *Computers & Education: Artificial Intelligence*, 6, Article 100198.
- European Commission. (2019). Ethics guidelines for trustworthy AI. from <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>. (Accessed 5 June 2024).
- Fishbein, M., & Ajzen, I. (1977). *Belief, attitude, intention, and behavior: An introduction to theory and research*.
- Floridi, L., Cowls, J., Beltrametti, M., et al. (2018). AI4People—an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28, 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Friedman, A., Knijnenburg, B. P., Vanhecke, K., Martens, L., & Berkovsky, S. (2015). Privacy aspects of recommender systems. *Recommender systems handbook*, 649–688.
- George, A. S., & George, A. H. (2023). A review of ChatGPT AI's impact on several business sectors. *Partners Universal International Innovation Journal*, 1(1), 9–23.
- Ghallab, M. (2019). Responsible AI: Requirements and challenges. *AI Perspectives*, 1(1), 1–7.
- Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association*, 19(1), 121–127.
- Gupta, M., Akiri, C., Aryal, K., Parker, E., & Praharaj, L. (2023). From chatgpt to threatgpt: Impact of generative ai in cybersecurity and privacy. *IEEE Access*, 11, 80218–80245.
- Habibi, A., Muhaimin, M., Danibao, B. K., Wibowo, Y. G., Wahyuni, S., & Octavia, A. (2023). ChatGPT in higher education learning: Acceptance and use. *Computers & Education: Artificial Intelligence*, 5, Article 100190.
- Hitron, T., Orlev, Y., Wald, I., Shamir, A., Erel, H., & Zuckerman, O. (2019). Can children understand machine learning concepts? The effect of uncovering black boxes. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1–11).
- Huallpa, J. J. (2023). Exploring the ethical considerations of using Chat GPT in university education. *Periodicals of Engineering and Natural Sciences*, 11(4), 105–115.
- Jagielski, M., Oprea, A., Biggio, B., Liu, C., Nita-Rotaru, C., & Li, B. (2018). Manipulating machine learning: Poisoning attacks and countermeasures for regression learning. In *2018 IEEE symposium on security and privacy (SP)* (pp. 19–35). IEEE.
- Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies*, 1–20.
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), 1–38.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399.
- Kahneman, D., & Tversky, A. (2013). Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I* (pp. 99–127).
- Kajiwara, Y., Matsuoka, A., & Shinbo, F. (2023). Machine learning role playing game: Instructional design of AI education for age-appropriate in K-12 and beyond. *Computers & Education: Artificial Intelligence*, 5, Article 100162.
- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, Article 102274.
- Kaspersen, M. H., Bilstrup, K. E. K., Van Mechelen, M., Hjorth, A., Bouvin, N. O., & Petersen, M. G. (2021). VotestatesML: A high school learning tool for exploring machine learning and its societal implications. In *FabLearn Europe/MakeEd 2021-An international conference on computing, design and making in education* (pp. 1–10).
- Kaur, D., Uslu, S., Rittichier, K. J., & Durresi, A. (2022). Trustworthy artificial intelligence: A review. *ACM Computing Surveys*, 55(2), 1–38.
- Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, 17, 1–9.
- Kenton, J. D. M. W. C., & Toutanova, L. K. (2019). Pre-training of deep bidirectional transformers for language understanding. *Proceedings of naacl-HLT*, (Vol. 1, 2).
- Kong, S. C., Cheung, W. M. Y., & Zhang, G. (2023). Evaluating an artificial intelligence literacy programme for developing university students' conceptual understanding, literacy, empowerment and ethical awareness. *Educational Technology & Society*, 26(1), 16–30.
- Kusner, M. J., Loftus, J., Russell, C., & Silva, R. (2017). Counterfactual fairness. *Advances in Neural Information Processing Systems*, 30.
- Lai, V. D., Ngo, N. T., Veyseh, A. P. B., Man, H., Derroncourt, F., Bui, T., & Nguyen, T. H. (2023). Chatgpt beyond English: Towards a comprehensive evaluation of large language models in multilingual learning. *arXiv preprint arXiv:2304.05613*.
- Laupichler, M. C., Aster, A., Schirch, J., & Raupach, T. (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. *Computers & Education: Artificial Intelligence*, Article 100101.
- Liang, P., Bommasani, R., Lee, T., Tsipras, D., Soylu, D., Yasunaga, M., ... Koreeda, Y. (2022). *Holistic evaluation of language models*. *arXiv preprint arXiv:2211.09110*.
- Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–16).
- Loyola-Gonzalez, O. (2019). Black-box vs. white-box: Understanding their advantages and weaknesses from a practical point of view. *IEEE Access*, 7, 154096–154113.
- Lucchi, N. (2023). ChatGPT: A case study on copyright challenges for generative artificial intelligence systems. *European Journal of Risk Regulation*, 1–23.
- Mathew, A. (2023). Is artificial intelligence a world changer? A case study of OpenAI's chat GPT. *Recent Progress in Science and Technology*, 5, 35–42.
- Memarian, B., & Doleck, T. (2023). Fairness, accountability, transparency, and ethics (FATE) in artificial intelligence (AI), and higher education: A systematic review. *Computers & Education: Artificial Intelligence*, Article 100152.
- Milano, S., Taddeo, M., & Floridi, L. (2020). *Recommender systems and their ethical challenges* (Vol. 35, pp. 957–967). *AI & Society*.
- Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics*, 26(4), 2141–2168.
- Narayanan, A., & Shmatikov, V. (2008). Robust de-anonymization of large sparse datasets. In *2008 IEEE symposium on security and privacy (sp 2008)* (pp. 111–125). IEEE.
- Ng, D. T. K., Leung, J. K. L., Chu, K. W. S., & Qiao, M. S. (2021). AI literacy: Definition, teaching, evaluation and ethical issues. *Proceedings of the Association for Information Science and Technology*, 58(1), 504–509.
- Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Computers & Education: Artificial Intelligence*, 2, Article 100033.
- Pavlik, J. V. (2023). Collaborating with ChatGPT: Considering the implications of generative artificial intelligence for journalism and media education. *Journalism and Mass Communication Educator*, 78(1), 84–93.
- Qadir, J. (2023). Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education. In *2023 IEEE global engineering education conference (EDUCON)* (pp. 1–9). IEEE.
- Qin, C., Zhang, A., Zhang, Z., Chen, J., Yasunaga, M., & Yang, D. (2023). Is ChatGPT a general-purpose natural language processing task solver? *arXiv preprint arXiv:2302.06476*.
- Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. In *Internet of Things and Cyber-Physical Systems*.
- Regalia, J. (2023). *ChatGPT and legal writing: The perfect union?*. Available at: SSRN 4371460.
- Resnick, M., & Silverman, B. (2005, June). Some reflections on designing construction kits for kids. In *Proceedings of the 2005 conference on Interaction design and children* (pp. 117–122).
- Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare*, 11(6), 887. MDPI.
- Santos, F. P., Santos, F. C., Paiva, A., & Pacheco, J. M. (2015). Evolutionary dynamics of group fairness. *Journal of Theoretical Biology*, 378, 96–102.
- Shi, D., Lee, T., & Maydeu-Olivares, A. (2019). Understanding the model size effect on SEM fit indices. *Educational and Psychological Measurement*, 79(2), 310–334.
- Skitka, L. J., Mosier, K., & Burdick, M. D. (2000). Accountability and automation bias. *International Journal of Human-Computer Studies*, 52(4), 701–717.
- Smuha, N. A. (2019). The EU approach to ethics guidelines for trustworthy artificial intelligence. *Computer Law Review International*, 20(4), 97–106.
- Strzelecki, A. (2023). To use or not to use ChatGPT in higher education? A study of students' acceptance and use of technology. *Interactive Learning Environments*, 1–14.
- Taecharungraj, V. (2023). "What can ChatGPT do?" Analyzing early reactions to the innovative AI chatbot on twitter. *Big Data and Cognitive Computing*, 7(1), 35.
- Thiebes, S., Lins, S., & Sunyaev, A. (2021). Trustworthy artificial intelligence. *Electronic Markets*, 31, 447–464.

- Torrey, L. (2012). Teaching problem-solving in algorithms and AI. *Proceedings of the AAAI Conference on Artificial Intelligence*, 26(3), 2363–2367.
- Touretzky, D. S., Gardner-McCune, C., Martin, F., & Seehorn, D. (2019). K-12 guidelines for artificial intelligence: What students should know. In *Proc. Of the ISTE conference* (Vol. 53).
- Wang, L., Lyu, C., Ji, T., Zhang, Z., Yu, D., Shi, S., & Tu, Z. (2023). Document-level machine translation with large language models. *arXiv preprint arXiv:2304.02210*.
- Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology innovation management review*, 9(11).
- Williams, R., Ali, S., Devasia, N., DiPaola, D., Hong, J., Kaputsos, S. P., ... Breazeal, C. (2023). AI+ ethics curricula for middle school youth: Lessons learned from three project-based curricula. *International Journal of Artificial Intelligence in Education*, 33(2), 325–383.
- Wu, Z., Qiu, L., Ross, A., Akyürek, E., Chen, B., Wang, B., ... Kim, Y. (2023). Reasoning or reciting? Exploring the capabilities and limitations of language models through counterfactual tasks. *arXiv preprint arXiv:2307.02477*.
- Zhang, H., Lee, L., Ali, S., DiPaola, D., Cheng, Y., & Breazeal, C. (2023). 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*, 33(2), 290–324.