



Article

Development of an Assessment Scale for Measurement of Usability and User Experience Characteristics of Bing Chat Conversational AI

Goran Bubaš *, Antonela Čizmešija and Andreja Kovačić

Faculty of Organization and Informatics, University of Zagreb, 42000 Varaždin, Croatia; antonela.cizmesija@foi.unizg.hr (A.Č.); andreja.kovacic@foi.unizg.hr (A.K.)

* Correspondence: goran.bubas@foi.unizg.hr

Abstract: After the introduction of the ChatGPT conversational artificial intelligence (CAI) tool in November 2022, there has been a rapidly growing interest in the use of such tools in higher education. While the educational uses of some other information technology (IT) tools (including collaboration and communication tools, learning management systems, chatbots, and videoconferencing tools) have been frequently evaluated regarding technology acceptance and usability attributes of those technologies, similar evaluations of CAI tools and services like ChatGPT, Bing Chat, and Bard have only recently started to appear in the scholarly literature. In our study, we present a newly developed set of assessment scales that are related to the usability and user experiences of CAI tools when used by university students, as well as the results of evaluation of these assessment scales specifically regarding the CAI Bing Chat tool (i.e., Microsoft Copilot). The following scales were developed and evaluated using a convenience sample ($N = 126$) of higher education students: Perceived Usefulness, General Usability, Learnability, System Reliability, Visual Design and Navigation, Information Quality, Information Display, Cognitive Involvement, Design Appeal, Trust, Personification, Risk Perception, and Intention to Use. For most of the aforementioned scales, internal consistency (Cronbach alpha) was in the range from satisfactory to good, which implies their potential usefulness for further studies of related attributes of CAI tools. A stepwise linear regression revealed that the most influential predictors of Intention to Use Bing Chat (or ChatGPT) in the future were the usability variable Perceived Usefulness and two user experience variables—Trust and Design Appeal. Also, our study revealed that students' perceptions of various specific usability and user experience characteristics of Bing Chat were predominantly positive. The evaluated assessment scales could be beneficial in further research that would include other CAI tools like ChatGPT/GPT-4 and Bard.

Keywords: conversational artificial intelligence; higher education; Bing Chat; usability; user experience; survey



Citation: Bubaš, G.; Čizmešija, A.; Kovačić, A. Development of an Assessment Scale for Measurement of Usability and User Experience Characteristics of Bing Chat Conversational AI. *Future Internet* **2024**, *16*, 4. <https://doi.org/10.3390/fi16010004>

Academic Editors: Massimo Stella and Giulio Rossetti

Received: 4 December 2023

Revised: 18 December 2023

Accepted: 21 December 2023

Published: 23 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The term “conversational artificial intelligence” (CAI) was conceptualized concerning artificial intelligence (AI) tools with the ability to perform a dialogue or engage in a conversation with a human being. In the year 2023, several publicly available and free-to-use CAI tools gained worldwide popularity: ChatGPT, which was developed by OpenAI, Bing Chat, which was launched by Microsoft as an AI chat tool to complement the Bing search engine, and Bard, an experimental CAI chat service released by Google. The main goal of the study that is described in this paper is to present an assessment instrument and methodology for the evaluation of CAI tools regarding their use for teaching and learning activities in higher education from the perspective of usability and user experience.

The rate of the adoption of conversational artificial intelligence (CAI) tools in education is well illustrated by the number of attempts to produce review papers and meta-analyses in the one year after the introduction of ChatGPT in November 2022 (for instance, see: [1–8]),

including those focused on specific fields like medical education [9] and second language learning [10]. However, the adoption of CAI tools did not meet the early expectations that were driven by the initially rapid adoption of ChatGPT after its launch on 30 November 2023, when in its first week it attracted one million users, with an increase to over 100 million users in less than 2 months [11]. For instance, a survey performed by Ipsos [12] in April 2023 (N = 1008) revealed that only 16% of adults (aged 18+) in the USA reported having ever used a text-based or visual generative AI system (ChatGPT, DALL-E, Bard, Midjourney, Stable Diffusion, etc.). A similar rate of use was established by a Pew Research Center survey [13] performed in July 2023, in which only 18% of adults aged 18+ (N = 5057) stated that they had used ChatGPT, with most of the users (41%) belonging to the 18–29 age group. Another Pew Research Center survey [14], performed in the period from 26 September to 23 October 2023, revealed that 13% of teens aged 13–17 (N = 1453) had used ChatGPT for their schoolwork. The fact that a much larger adoption rate is to be expected in higher education was revealed by a large-scale survey performed among college students in Germany [15] from 15 May to 5 June 2023 (N = 6311 in the final sample). In this study, 63.2% of the students reported that they had used ChatGPT or other AI tools, but still, only 34.8% of them had used such tools “occasionally”, “frequently”, or “very often”. The last aforementioned study revealed that the most commonly used AI tool by the students was ChatGPT (as reported by 48.9% of the respondents), while the following were found to be the most frequent reasons for their use of AI tools: clarifying questions and having subject-specific concepts explained (56.5%), research and literature studies (45.4%), translation (42.2%), text analysis, processing, and creation (39.3%), problem-solving and decision making (35.1%), and exam preparation (20.3%). This level of adoption of ChatGPT and similar tools by students in higher education, which can be rated as at least moderate, confirms the importance of evaluation of such tools for teaching and learning in the academic environment, taking into account previously mentioned purposes and other potential educational uses of AI tools.

Before CAI tools like ChatGPT, Bing Chat, and Bard were introduced, the use of chatbots or conversational agents in various fields had already started to attract more interest from researchers, especially since 2001, with a rapid increase in 2017 and later [16]. However, in the specific field of education, chatbots attracted fairly small but continuous attention from researchers in the period from 2005 to 2019, which was followed by a considerable increase in the number of published scholarly papers beginning in 2018 (see: [17]). Regarding the use of chatbots in education, according to one review study, most of the scholarly papers were related to the pedagogical strategy of guided learning [18], while another review study revealed more specific uses of chatbots for learning activities (i.e., delivery of learning content), as well as for assessment, consultation/recommendation, and administration [19]. In fact, there are numerous potential uses of chatbots for recommendation and administration purposes regarding non-learning activities (class schedule, exam schedule, reminders, time management, organization of study plan, selection of courses, library information, various types of campus information, etc.) related to the study life of students (see: [20]). On the other side, a recent review study revealed that the most popular objective of the pedagogical use of chatbots was skill improvement [17]. The most frequently mentioned pedagogical design principles in related scholarly papers were personalized learning, experiential learning, social dialogue, and collaborative learning [21]. Finally, meta-analyses of empirical studies revealed a generally positive effect of the use of chatbots on learning [22] that is dependent on specific conditions like instruction method in the comparison group, as well as the chatbot type and tasks [23]. More specific positive learning outcomes were related to explicit reasoning, learning achievement, knowledge retention, and learning interest [24]. This brief overview of the studies on the educational use of chatbots indicates the potential fields of research regarding the use of CAI tools like ChatGPT, Bing Chat, and Bard for similar purposes in educational settings. However, due to the limited number of scholarly papers and the short time after the introduction of ChatGPT, Bing Chat, and Bard, quality reviews and meta-analyses regarding the pedagogical use of CAI tools in higher education are scarce.

According to Granić [25], the *adoption* of learning technology in higher education is a field of interest of researchers worldwide, with a focus on theoretical approaches like the *Technology Acceptance Model*—TAM [26] and *Unified Theory of Acceptance and Use of Technology*—UTAUT [27], sometimes combining TAM with *Innovation Diffusion Theory*—IDT [28] or the *Information Systems Success Model*—ISSM [29,30]. Theoretical approaches like TAM, UTAUT, and ISSM have been frequently used for the evaluation of various technology uses in teaching and learning, from the Moodle learning management system [31] to social media [32]. Educational technology can also be evaluated regarding *usability* and *user experience* aspects [33]. For instance, the *System Usability Scale* (SUS) and similar evaluation instruments have been used to evaluate the use of internet platforms, tutoring systems, mobile applications, and multimedia in education [34,35]. In our study, to evaluate the Bing Chat CAI tool in educational settings, the variables from the following aforementioned approaches for evaluation of learning technology were selected: TAM, ISSM, usability, and user experience.

The main intention of the study that is reported in this paper was to develop and evaluate assessment scales for measurement of usability and user experience characteristics of CAI tools with a specific focus on Bing Chat, which was introduced by Microsoft in February 2023 and upgraded to GPT-4 OpenAI large language model (LLM) in March 2023 [36]. By November 2023, Microsoft had introduced multiple upgrades of its products with AI, for instance, by launching Bing Chat Enterprise for organizations and upgrading its 365 platform with Copilot (a version of Bing Chat) [37], as well as enabling the integration of Copilot into the Windows 11 operating system [38]. Finally, in November 2023, the Bing Chat CAI was, without significant technological alterations, renamed Microsoft Copilot, or briefly, Copilot. This inclusion of CAI functionalities of Bing Chat/Copilot into diverse products and services is placing an even greater emphasis on the importance of evaluation of their usability and user experience characteristics.

With a simple assumption that educational technology that is used by teachers to create learning activities has to be (previously) evaluated, the focus of our study was to develop and test a set of measurement instruments (assessment scales) for that purpose, with a specific reference to Bing Chat/Copilot.

2. Methodology

Apart from the studies that can be found in pre-print archives like arXiv or SSRN, published quality scholarly papers that investigated Bing Chat (or Google Bard) from the theoretical standpoint of technology adoption, usability, or user experience are difficult to find. However, the quantity of published scholarly research on ChatGPT or GPT-4 is growing rapidly, and as of November 2023, several representative papers have been found that used TAM/UTAUT or a similar theoretical background for the design of empirical research in educational settings (for instance, see: [39–42]). Unfortunately, quality studies regarding usability and user experience assessment of ChatGPT/GPT-4 (or other recently introduced CAI tools) were scarce at the time of our study among reviewed papers that are published in the scholarly literature. Therefore, the focus of our study was to develop and evaluate assessment scales for the measurement of selected usability and user experience variables of the Bing Chat CAI tool keeping in mind its educational use in higher education.

2.1. Goals and Research Questions

CAI tools like ChatGPT/GPT-4, Bing Chat, and Bard are designed to meet the interests of diverse users (in private, corporate, and educational settings). As mentioned before, Microsoft has been rapidly integrating Copilot (Bing Chat) with its services and applications, especially Microsoft 365, and Google has been pursuing a similar effort by integrating Bard in Google applications like Gmail, Docs, Drive, Google Maps, and Google Flights [43]. Even though there is currently a broad list of settings in which CAI tools can be evaluated, in this study, the focus is on the educational environment and the potential use of such tools by teachers and students in higher education. So far related studies have revealed

that it is not only traditional technology adoption variables that influence one's intention to use such tools in education, but also the variables associated with hedonic [39] and intrinsic motivation [40].

The *first goal* of our study was to create and perform an initial evaluation of an assessment instrument for CAI tools from a much broader perspective than technology adoption itself. Having in mind this first goal, assessment scales were created to measure students' perceptions of various characteristics of Bing Chat that are broadly related to usability (*Perceived Usefulness, General Usability, Learnability, System Reliability, Visual Design and Navigation, Information Quality, Information Display*) and user experience (*Cognitive Involvement, Design Appeal, Trust, Personification, Risk Perception*), as well as to the criterion (dependent) variable frequently found in TAM studies labeled *Intention to Use*. This was done by adapting the existing assessment scales, as well as by developing several new ones. The evaluation of those scales was performed empirically by collecting survey data and using Cronbach's alpha coefficient as an indicator of the internal consistency of an assessment scale.

The *second goal* of our study was to investigate how specific usability and user experience characteristics of Bing Chat are evaluated by students in higher education, as well as to identify the most influential usability and user experience characteristics, as *independent or predictor variables*, in relation to *Intention to Use* Bing Chat, as a dependent or criterion variable.

According to the goals of our study, the following three research questions were formulated:

RQ1—What are the internal reliabilities (Cronbach's alpha coefficients) of the assessment scales that were used in our study to measure selected characteristics of Bing Chat?

RQ2—How do higher education students perceive specific usability and user experience characteristics of Bing Chat?

RQ3—Which usability and user experience variables are the best predictors of *Intention to Use* Bing Chat (or ChatGPT/GPT-4) in the future?

2.2. Instrument

The assessment scales that were used in our study were developed based on previous research on the educational use of a videoconferencing tool by the authors [44] and were also partly evaluated using two small convenience samples in an earlier pilot study on the use of Bing Chat for learning activities in higher education [45].

The survey that was used in our study consisted of demographic questions and questions related to respondents' previous use of LLMs and other AI-based systems like ChatGPT and GPT-4, as well as of assessment scales (with 4–7 items each) that were constructed to measure the variables associated with students' perceptions of usability and user experience characteristics of Bing Chat. The participants responded to each of the items in the assessment scales using the following 1–5 Likert-type scale: “1—I totally disagree”; “2—I disagree”; “3—I neither agree nor disagree”; “4—I agree”; “5—I totally agree”. The variables (constructs) for which the assessment scales were developed are briefly explained (with the number of items and a sample item in brackets) in the continuation of this section. The complete assessment scales with all of their items can be found in the Appendix A.

Perceived Usefulness is a construct introduced in 1989 by Davis [26] as one of the main variables of the Technology Acceptance Model (TAM), and as mentioned earlier, is frequently found in contemporary research that uses TAM/UTAUT as a theoretical background (5 items; an example item: “I have determined that I can make good use of Bing Chat.”).

General Usability is a construct related to the System Usability Scale (SUS) [46], a popular instrument for the measurement of perceived usability that was developed by Brooke in 1986. However, the structure of SUS comprises both usability and learnability [47], while in our study, this scale is used to specifically measure general usability. It must be noted that this scale is still very popular with numerous recently published studies that report its use for the evaluation of diverse technologies in various environments, including higher educa-

tion, for instance, (a) concerning the use of the chat application *Differ* for communication with students [48], as well as (b) the evaluation of text generation by ChatGPT [49] (5 items; a sample item: “Bing Chat responds to my queries/commands as I expect it to.”).

Learnability is often considered an important component of the usability of technological systems and is also one of the two factors measured by the widely used System Usability Scale [46]. For instance, in educational settings, it was a component of usability that was investigated concerning the Blackboard learning management system [50] (5 items; representative item: “One can quickly learn the basics of working with the Bing Chat tool.”).

System Reliability is one of the components of the DeLone and McLean Information System Success Model (ISSM) [29,30] and an important aspect of information system quality (4 items; illustrative item: “Bing Chat worked fast enough and reliably.”).

Visual Design and Navigation is a construct related to the attributes of computerized display of information, which was first mentioned in the scholarly literature about hypertext (see: [51]). It must be noted that the components of visual design and navigation are interlinked, ensuring both usability and the desired user experience of websites [52] (6 items; exemplary item: “Functionalities on the Bing Chat interface are well organized and easily accessible, e.g., menus, copying, etc.”).

Information Quality, as a component of the DeLone and McLean ISSM [29,30], commonly includes attributes like [53]: accuracy, relevance, timeliness, understandability, completeness, and usefulness (6 items; sample item: “The use of the Bing Chat service enabled the collection of accurate information.”).

Information Display is an assessment scale specifically designed in this study for Bing Chat to investigate how users perceive the way the text of the dialogues is presented, as well as the form in which information is provided to them after they write their prompt for Bing Chat (7 items; illustrative item: “The data obtained by the Bing Chat service is in a suitable and easy-to-use format for further use”).

Cognitive Involvement, as conceptualized for our assessment scale, has similarities to the concepts of cognitive absorption [54] (i.e., deep mental involvement characterized by distortion in the perception of time flow, immersion in mental activity, and enjoyment) and cognitive engagement [55] (which can range from simple memorization to deep understanding and include one’s effort to comprehend complex ideas and learn difficult skills) (7 items; representative item: “When I use the Bing Chat application, I feel as if I am immersed in the communication process and the information I receive”).

Design Appeal as a construct can be associated with the hedonic aspects of using technology (e.g., with attributes like “enjoyable”, “exciting”, “pleasant”, and “interesting”) [56] (6 items; exemplary item: “I feel satisfied and fulfilled during and after using the Bing Chat service”).

Trust has been a subject of research associated with information systems [57], including AI systems like ChatGPT [58]. For instance, it was established that trust is related to user satisfaction (and loyalty) in mobile commerce [59] (6 items; sample item: “I have the same trust in the Bing Chat service as I do in other internet services, such as social networks, etc.”).

Personification is the label for a new scale specifically designed for this study that denotes the level to which a user experiences the dialogue with a CAI tool to be human-like, i.e., similar to a conversational interaction with a real person (6 items; illustrative item: “My conversations with Bing Chat resembled an exchange of online text messages with an informed person”).

The Risk Perception assessment scale was included in our study since it represents an important factor in technology adoption [60] (6 items; sample item: “I am sure that there is no danger or potential threat to Bing Chat users”).

Intention to Use (or Behavioral Intention) is a frequent variable in the research on information systems and IT-based services (for instance in TAM [26], UTAUT [27], and ISSM [29,30] (6 items; representative item: “I expect that I will need to use Bing Chat and similar services such as GPT for a long time to come”).

2.3. Subjects

The participants in our study were higher education students from a large university in Croatia. They were enrolled in several different study programs in the economics of entrepreneurship, implementation of IT in business, and information systems. The courses in which they had to use Bing Chat for learning activities before the survey was conducted were English Language 1, Business English Language, Communication and Virtual Teams in the Organization, and Digital Communication and Media. A total of 126 students participated in our study, 67 of whom were of male and 57 of female gender (2 students did not indicate their gender in the survey). The age structure of the respondents was as follows: 16.7% aged 18–19, 34.1% aged 20–21, 16.7% aged 22–23, 22.2% aged 24–25, and 10.3% aged 26 or above. The students were in their first (67.5%), second (11.1%), or third (21.4%) year of undergraduate study. All of the students had at least some previous experience with ChatGPT, Bing Chat, or similar CAI tools before performing learning activities with Bing Chat in their university courses and participating in the survey.

2.4. Procedure

Before the survey was conducted, the higher education students enrolled in two undergraduate study programs in the implementation of IT in business and information systems, respectively, were assigned to prepare oral presentations on predefined topics using the information provided by Bing Chat, as well as to compare it with the information acquired from online sources like Google Search and Google Books. They were also asked to evaluate the relevance of the links that Bing Chat supplemented to its responses. In the second assignment, the students performed an asynchronous translation activity using the Bing Chat service, which included checking the translations generated by Bing Chat for plagiarism and checking whether the answers to their questions that were acquired from Bing Chat could be recognized as being created by AI by using websites like Copyleaks and other.

Somewhat different learning activities were performed by the students enrolled in two English language courses, who were given the following assignments: (a) searching for content related to their English language course with Bing Chat and checking the accuracy of the collected information using the Google search engine; (b) performing a dialogue-based learning activity with Bing Chat about phrasal verbs, including their definition and examples; (c) correcting grammatically incorrect sentences in English using Bing Chat; (d) creating multiple-choice questions for grammar practice; (e) learning about a typical conversation structure on an example of a telephone dialogue, etc.

After the learning activities with Bing Chat were completed by the students, they participated in a paper and pencil survey. The survey was conducted at the end of the summer semester of the 2022/2023 academic year. This survey was voluntary and anonymous for all students in our convenience sample. The use of the survey was previously approved by the Ethics Committee of the higher education institution.

3. Results

The data analyses were performed with IBM SPSS Statistics software. The results of data analyses are presented according to the research questions (RQ1–RQ3).

3.1. Internal Consistencies of Assessment Scales

As can be concluded from the data presented in Table 1, Cronbach's alpha coefficient, as a measure of internal consistency, was below the most frequently accepted minimal level of 0.70 only in the case of the *Learnability* assessment scale, which can still be considered as acceptable for an early stage of research. Having in mind the qualifications of Cronbach's alpha coefficients in the literature [61] and the *first research question* (RQ1) in our study, it can be concluded that the internal reliabilities of the assessment scales that were used in our study to measure the selected characteristics of Bing Chat can be categorized as "adequate" and "satisfactory".

Table 1. Scale labels, number of items, and internal consistency of assessment scales (N = 126).

Scale Label	Number of Items	Cronbach's Alpha
Perceived Usefulness	5	0.77
General Usability	5	0.76
Learnability	5	0.67
System Reliability *	4	0.79
Visual Design and Navigation	6	0.78
Information Quality	6	0.82
Information Display	7	0.79
Cognitive Involvement	7	0.88
Design Appeal	6	0.81
Trust	6	0.82
Personification	6	0.81
Risk Perception	6	0.86
Intention to Use	6	0.90

* One item was excluded from the scale because of redundancy.

3.2. Perceptions of Usability and User Experience Characteristics of Bing Chat

To confirm the categorization of the assessment scales that were used in our study into the broad categories of (a) *usability* measures and (b) *user experience* measures, a forced factor analysis was performed with two fixed factors in a principal components analysis and varimax rotation with Kaiser normalization. The common recommendation is that the minimal number of subjects/cases (N) for factor analysis is 100, and that the ratio of the number of participants (N) to the number of variables (p) is at least 5:1 (see: [62,63]). Since the number of subjects in our convenience sample is 126 and the total number of usability and user experience variables in our study is 12, the N:p ratio amounts to 10.5:1, which means that the minimal prerequisites for the use of factor analysis have been met. The projections of variables on the two factors (F1 and F2), according to the results of the factor analysis, are presented in Table 2 and they indicate that the broad labels of “usability” for components of F1 (*Perceived Usefulness*, *General Usability*, *Learnability*, *System Reliability*, *Visual Design and Navigation*, *Information Quality*, *Information Display*) and “user experience” for components of F2 (*Cognitive Involvement*, *Design Appeal*, *Trust*, *Personification*, and *Risk Perception*) manifest at least some correspondence with their theoretical classification. According to the data presented in Table 2, in the results of this forced factor analysis, the common criterion was met that for the loading of an item to a factor to be considered relevant it needs a primary loading on one factor with the value of at least 0.60, with no secondary loading on some other factor above 0.40.

The summarized percentages of *positive responses* “4—I agree” and “5—I totally agree” (N = 126 for all assessment scales; the percentages were rounded to the nearest whole number) for *all the items* of usability and user experience assessment scales are provided in the Appendix A. The following data were selected to represent the percentages of positive responses that are related to the *second research question* (RQ2), which addresses the students’ perception of specific *usability* and *user experience* characteristics of Bing Chat.

Having in mind the items of the *usability* scales, *Perceived Usefulness* attributes were positively evaluated in the range from 60% (“By using Bing Chat I can do whatever I want”) to 87% (“Bing Chat can be used for many different things”). In the case of *General Usability* attributes, positive evaluations ranged from 53% (“I did not notice an inconsistency in the operation of Bing Chat”) to 88% (“Bing Chat is not too complex for everyday use”). Regarding the *Learnability* attributes, positive responses ranged from 68% (“I remember well what I learned to do with Bing Chat”) to 94% (“One can quickly learn the basics of working

with the Bing Chat tool”). With *System Reliability* characteristics, the percentages of positive evaluations were from 72% (“There were no unexpected interruptions in the operation of Bing Chat during its use”) to 81% (“Bing Chat worked fast enough and reliably”). In relation to *Visual Design and Navigation*, the positive responses ranged from 75% (“The way Bing Chat displays the discussion and its responses is visually attractive and engaging”) to 91% (“The use of the interface with Bing Chat is logical and intuitive (easily understandable) regarding the functionalities that I use”). The *Information Quality* attributes received positive responses ranging from 73% (“Verification of information obtained from Bing Chat shows that one can have confidence in its correctness”) to 94% (“The data provided by the Bing Chat service was clear and easy to understand”). Finally, the characteristics related to *Information Display* were positively evaluated in the range from 52% (“The information provided in a conversation with Bing Chat can also be obtained after a few days/weeks”) to 90% (“The data obtained by the Bing Chat service is in a suitable and easy-to-use format for further use”).

Table 2. Results of forced factor analysis of *usability* and *user experience* variables with two fixed factors (N = 126; principal components analysis; varimax rotation).

Scale Label	F1	F2
Perceived Usefulness	0.70	0.38
General Usability	0.81	0.22
Learnability	0.67	0.23
System Reliability	0.64	0.10
Visual Design and Navigation	0.68	0.13
Information Quality	0.78	0.28
Information Display	0.67	0.40
Cognitive Involvement	0.17	0.87
Design Appeal	0.09	0.77
Trust	0.37	0.62
Personification	0.24	0.72
Risk Perception	0.33	0.51

Regarding the *user experience* scales, the attributes of *Cognitive Involvement* were positively rated from 44% (“Time seems to pass quickly while I am using Bing Chat”) to 71% (“I feel like I control what happens while working with Bing Chat because I use it as I want and get what I want”). The evaluations of characteristics associated with *Design Appeal* were slightly higher on average and ranged from 54% (“I feel satisfied and fulfilled during and after using the Bing Chat service”) to 79% (“The visible representations of the content of the computer screen during Bing Chat use are modern and enjoyable to use”). Regarding *Trust*, the positive evaluations were in the range from 52% (“I believe I can rely more on Bing Chat than on most other sources of information, knowledge, and advice”) to 84% (“I believe that the Bing Chat service is designed with the goal of helping the widest possible number of people”). The items of the *Personification* assessment scale received on average the lowest level of positive evaluations, from 36% (“It would suit me if Bing Chat and similar services, e.g., GPT, could react as much as possible like a human being”) to 59% (“My conversations with Bing Chat resembled an exchange of online text messages with an informed person”), but it must be noted that a high level of human-like interaction can in practice be both favored by some and disliked by other users. Finally, the aspects of Bing Chat that are associated with *Risk Perception* were positively perceived in the range from 40% (“I am sure that my privacy isn’t under any threat by my use of the Bing Chat service”) to 70% (“The privacy and security of Bing Chat users are not lower than, for example, those of users of social networks and similar services”).

To briefly summarize the findings related to the *second research question* (RQ2) in this study, having in mind the responses of students in our convenience sample to individual items of assessment scales in the *usability* category, the attributes of Bing Chat addressed in 19 out of 31 items (61% of them) in those scales (see the Appendix A) received positive evaluations (i.e., “4—I agree” and “5—I totally agree”) by 75% or more of the respondents. In the category of *usability*, it is important to emphasize that only the item related to *not noticing an inconsistency in the operation of Bing Chat* received less than 60% of positive confirmations. Another similarly evaluated item, with the *statement that the information provided in a conversation with Bing Chat can also be obtained after a few days/weeks*, refers to an issue that has been resolved in the newer adaptations of Bing Chat.

After analyzing the responses to the items in the *user experience* category, it can be concluded that the confirmations of positive experiences attributed to *Cognitive Involvement* in using Bing Chat were only moderately frequent and mostly in connection to interest, fun, attention, and control. Similar results were obtained concerning *Design Appeal*, where technical aspects of Bing Chat were, on average, evaluated more favorably than its encouragement of the user’s innovativeness and creativity or the user’s feeling of being satisfied and fulfilled when using Bing Chat. The level of *Trust* was rather high regarding the users’ assessment of the intention of its designers that Bing Chat would provide help and benefit its users, but lower when compared to the perception of possible reliance on other sources of information. Finally, *Risk Perception* attributes were less favorably evaluated regarding privacy concerns, threats to computer and data security, as well as potential danger in general for Bing Chat users.

The analyses of responses to individual items of usability and user experience assessment scales indicate that a more detailed inspection of the characteristics of tools like Bing Chat is opportune when planning to use them in educational settings. It must be noted that the percentage of *negative* (“1—I totally disagree”; “2—I disagree”) or *neutral* evaluations (“3—I neither agree nor disagree”) should not be disregarded in favor of the (pre)dominantly *positive* evaluations (“4—I agree” and “5—I totally agree”) when it comes to introducing technology in education, especially for minors in K-12. Finally, since CAI technology is constantly evolving and its adoption rate is on the increase, the current or future perception of its attributes is bound to be different than at the time the survey in our study was performed.

3.3. Regression Analysis of the Predictors of the Intention to Use Bing Chat (or GPT) in the Future

The rather high percentage of *positive* statements (“4—I agree” and “5—I totally agree”) of participants in our study regarding the individual items of the *Intention to Use* assessment scale (see the Appendix A), which was in the range from 75% to 79%, slightly exceeds the findings of a large-scale survey that was performed among college students in Germany [15] in May and June 2023, in which 63.2% of the respondents reported using ChatGPT or other AI tools. The data analysis regarding the *correlation* between *usability* and *user experience* variables, on the one side, and *Intention to Use* Bing Chat, on the other, revealed that the highest correlations (Pearson; statistically significant at $p < 0.001$) were obtained for *Trust* ($r = 0.55$), *Perceived Usefulness* ($r = 0.42$), *Personification* ($r = 0.40$), *Cognitive Involvement* ($r = 0.37$), *Design Appeal* ($r = 0.37$), *Information Quality* ($r = 0.35$), *Risk Perception* ($r = 0.34$), and *Information Display* ($r = 0.33$), respectively, on the one side, and *Intention to Use* Bing Chat on the other. Lower correlation coefficients (statistically significant at $p < 0.01$) were obtained for *General Usability* ($r = 0.29$) and *Learnability* ($r = 0.25$), respectively, on the one side, and *Intention to Use* Bing Chat on the other. Interestingly, no statistically significant correlation ($r < 0.1$) was found for *System Reliability* and *Visual Design and Navigation* variables, on the one side, and *Intention to Use* Bing Chat on the other.

To provide a more detailed investigation and better meet the requirement of a minimal number of subjects to perform a regression analysis (see: [64]), two separate regression analyses were performed (see Tables 3 and 4)—one for the variables in (a) the *usability* category and one for those in (b) the *user experience* category.

Table 3. Results of stepwise regression analysis with *usability* variables as predictors of *Intention to Use* (N = 126).

Regression Summary					
Model	R	R Square		Adjusted R Square	Std. Error of the Estimate
1	0.417 ^a	0.174		0.167	3.919
Model Summary ^{b,c}					
Model		Beta	Standard Error	t	Sig
1	Perceived Usefulness	0.588	0.115	5.11	<0.001

^a Predictors: (Constant), Perceived Usefulness; ^b Dependent variable: Intention to Use; ^c Excluded variables: General Usability, Learnability, System Reliability, Visual Design and Navigation, Information Quality, Information Display.

Table 4. Results of stepwise regression analysis with *user experience* variables as predictors of *Intention to Use* (N = 126).

Regression Summary					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	
1	0.549 ^a	0.301	0.296	3.604	
2	0.578 ^b	0.335	0.324	3.532	
Model Summary ^{c,d}					
Model		Beta	Standard Error	t	Sig
1	Trust	0.578	0.079	7.31	<0.001
2	Trust	0.505	0.83	6.10	<0.001
	Design Appeal	0.206	0.83	2.48	0.015

^a Predictors: (Constant), Trust; ^b Predictors: (Constant), Trust, Design Appeal. ^c Dependent variable: Intention to Use; ^d Excluded variables in Model 2: Cognitive Involvement, Risk Perception, Personification.

The data presented in Table 3 reveal that, according to the results of stepwise regression analysis, the only predictor of the dependent variable *Intention to Use* Bing Chat among the *usability* variables was *Perceived Usefulness*. For comparison, it must be noted that the same result was obtained using regression analysis with the “Enter” method. However, in the presented stepwise regression analysis, the explanatory power (common variance with the dependent variable) of *Perceived Usefulness* as a predictor is considered *weak* with R^2 of only 0.174.

When only the *user experience* variables are used as predictors of *Intention to Use* Bing Chat, results of stepwise regression analysis that are presented in Table 4 reveal two predictor variables: *Trust* and *Design Appeal* (these two variables also appear as the only predictors when stepwise regression is used on joint sets of *usability* and *user experience* variables). Again, for comparison, it must be noted that, among the *user experience* variables, *Trust* was the only predictor obtained with the use of the “Enter” method. In the results of the stepwise regression analysis that are presented in Table 4, the explanatory power (common variance with the dependent variable) of *Trust* and *Design Appeal* as predictors is considered *moderate* with R^2 of 0.335.

The results of the regression analyses interpreted above (also, see Tables 3 and 4) respond to the *third research question* (RQ3) of our study: “Which usability and user experience variables are best predictors of *Intention to Use* Bing Chat (or ChatGPT/GPT-4) in the future?” Regarding the *usability* variables, this was *Perceived Usefulness*, while in

relation to *user experience* variables, the obtained predictors were *Trust* and *Design Appeal*. The results presented in Tables 3 and 4 emphasize the importance of including diverse constructs/variables in technology acceptance studies of CAI tools like Bing Chat/Copilot, ChatGPT/GPT-4, or Bard, especially when empirical research is performed in educational settings. These findings can contribute to a scholarly discussion that, alongside the perceived *usefulness* of introducing CAI systems, the level of *trust* and *hedonic experience* of their users should also be taken into consideration.

4. Discussion

The *first research question* (RQ1) of our study was related to the investigation of the *internal consistency* of the assessment scales that were used in our study to measure selected characteristics of Bing Chat. From the results of data analyses that are presented in Table 1, it can be concluded that, for most of the assessment scales, the internal reliabilities (Cronbach's alpha coefficients) were above the 0.70 threshold criterion or, in other words, "adequate" and "satisfactory". Even though some of the assessment scales that were used in our study need further refinement, most of them can be, with some customization and improvement, used for further studies of CAI tools in educational and other settings.

The collected survey data regarding the second research question (RQ2—"How do higher education students perceive specific usability and user experience characteristics of Bing Chat?") were analyzed with much detail in the Section 3.2 "Perceptions of usability and user experience characteristics of Bing Chat" (see also the percentages presented in the Appendix A). To make the presentation of these findings more concise, only the sums of positive statements ("4—I agree" and "5—I totally agree") in response to the items of assessment scales were provided. It can be concluded that most of the characteristics of Bing Chat received a positive evaluation. However, the percentages of positive responses are not sufficient to conclude that there are no potential issues pertaining to its use in teaching and learning activities in higher education. This is especially true regarding the (a) correctness of retrieved information and (b) privacy concerns. In this section, a forced factor analysis was performed (see Table 2) that indirectly justified the broad categorization of variables in our study to those related to (a) usability and (b) user experience. Due to the rather small convenience sample ($N = 126$) in our study, a more detailed exploratory factor analysis was not performed.

The *third research question* (RQ3) was directed toward the identification of *usability* and *user experience* constructs as independent variables that are best predictors of *Intention to Use Bing Chat* (or ChatGPT/GPT-4) in the future as a dependent variable. Because of the availability of similar (and competing) CAI tools like ChatGPT/GPT-4 and Bard, the items in the assessment scale for the measurement of *Intention to Use* addressed not only Bing Chat but also included the mention of "GPT", which denotes ChatGPT, GPT-3.5, GPT-4, and the like. It must be emphasized that this creates some uncertainty in the interpretation of the results of regression analyses, but not in a way that significantly questions the findings. The variable/construct labeled *Trust* was both in *highest correlation* ($r = 0.55$) with *Intention to Use* and was also found to be its *strongest predictor* in regression analyses. From the set of *user experience* variables, *Trust* together with *Design Appeal* explained 33.5% of the variance of the dependent variable *Intention to Use*. These findings are important and relevant regarding the higher education environment and students' use of Bing Chat or other CAI tools. In several other studies, trust was found to be a critical variable for the adoption of ChatGPT by adults [58], professionals [65], and students [66]. Also, hedonic motivation, which *Design Appeal* can be considered a component of, was in some other studies also found to affect students' intention to use ChatGPT [39,41,67,68]. One study also revealed that, among technical variables, only *Perceived Usefulness* successfully predicted ChatGPT usage [69], as was found in our regression analysis with predictor variables related to *usability* and *Intention to Use Bing Chat* as a dependent variable (see Table 3).

The findings that are reported in the Section 3.3 "Regression analysis of the predictors of the Intention to Use Bing Chat (or GPT) in the future" place additional emphasis on

the importance of the use of a broad range of instruments for in-depth evaluation of CAI tools beyond sole reliance on the TAM/UTAUT sets of variables. For this purpose, the assessment scales labeled Personalization, Cognitive Involvement, and Design Appeal were constructed or adapted for use in our study. For instance, in a related study [70], a construct similar to Personalization that was labeled perceived humanness was found to influence effort expectancy and positively correlate with willingness to accept the use of ChatGPT.

The first goal of our study was to create and perform an initial evaluation of an assessment instrument for CAI tools from a much broader perspective than that presented in related studies of ChatGPT that utilized TAM or UTAUT theoretical models [39–42], i.e., with variables broadly associated with the concepts of usability and user experience. This goal was achieved in our study and the assessment scales are presented in the Appendix A (see also the responses to RQ1).

The second goal of our study was to investigate how specific characteristics of Bing Chat are evaluated by students in higher education, as well as to identify the most influential usability and user experience characteristics, as independent or predictor variables, in relation to Intention to Use Bing Chat, as a dependent or criterion variable. This goal is related to the second and third research questions (RQ2 and RQ3) and was also achieved.

The main limitation of this study concerns the relatively small convenience sample of respondents ($N = 126$) that was used for data collection. Also, results could vary depending on the study year of students, their major and minor, and the course in which the activities with Bing Chat were performed, as well as on the learning activities that were used before the application of the survey. The timing of the study is also relevant—CAI tools are constantly being advanced and expanded with additional features like voice and picture recognition, text-to-speech, and others, as well as integrated into common applications and services (Windows 11, Microsoft 365, Google applications, etc.), alongside their more wide-spread presence and longer experience of subjects in research with their use, all of which may considerably affect the perceptions and evaluations of their characteristics.

This research was focused on the use of one tool—Bing Chat (i.e., Microsoft Copilot). One possible future direction for the authors is to perform comparative studies across various CAI tools (Bing Chat/Copilot, ChatGPT, Bard) that would include similar usability and user experience characteristics of this technology and a much larger sample of subjects.

Since the technological evolution in the field of CAI is progressive—in terms of the development of their new functionalities and use of more complex LLMs—it is recommended that the assessment scales that were used in our study are not only improved and refined but also adapted to the novel characteristics of such tools/services and having in mind their use in specific educational and other contexts.

Possible other future directions for research in this area could include the exploration of various pedagogical applications of CAI tools for learning in educational activities in diverse courses (e.g., STEM and social studies).

5. Conclusions

The introduction of CAI tools based on LLMs has been exerting an important impact on the higher education environment. Our study intended to provide early findings and convenient tools in the form of assessment scales for the in-depth evaluation of potentially important characteristics not only of Bing Chat (i.e., Microsoft Copilot), but also of ChatGPT/GPT-4, Bard, and other CAI tools. It is a strong belief of the authors of this paper that instructors in higher education should evaluate CAI tools that they intend to use for teaching and learning activities or be able to find such evaluations in the reviewed and published (preferably open access) scholarly literature. Numerous studies and their meta-analyses have revealed a generally positive effect of the use of chatbots and CAI tools on learning [22,23], which places considerable additional importance on the investigation and evaluation of their usability and user experience characteristics.

Author Contributions: Conceptualization, G.B., A.Č. and A.K.; methodology, G.B., A.Č. and A.K.; investigation, G.B., A.Č. and A.K.; resources, G.B., A.Č. and A.K.; data analyses: G.B. and A.Č.; writing—original draft preparation, G.B. and A.Č.; writing—review and editing, G.B., A.Č. and A.K. All authors have read and agreed to the published version of the manuscript.

Funding: This study was performed as part of the standard research obligation of the authors with no external funding outside the regular support of their higher education institution that is provided to all of its employees.

Data Availability Statement: The survey data for this study are available upon request after the completion of the study. A signed agreement between research institutions is required, as well as the approval of the Ethics committee.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Assessment scales (with the percentage of positive responses to individual items)

The responses to the items of the assessment scales were on a 1–5 Likert-type scale (“1—I totally disagree”; “2—I disagree”; “3—I neither agree nor disagree”; “4—I agree”; “5—I totally agree”). The percentage in the brackets next to each item represents the sum of positive responses “4—I agree” and “5—I totally agree” (N = 126 for all assessment scales; the percentages were rounded to the nearest whole number). The original items were in the Croatian language and their translation into English was performed by a human expert. However, the wording of some of the items needed to be adapted for the use of the survey in the English language.

Perceived Usefulness

1. Bing Chat is useful for my needs (71%).
2. By using Bing Chat, I can do whatever I want (60%).
3. Bing Chat can be used for many different things (87%).
4. I will easily find new ways of using Bing Chat (79%).
5. I have determined that I can make good use of Bing Chat (86%).

General Usability

1. Bing Chat is not too complex for everyday use (88%).
2. The diverse functionalities of Bing Chat are well integrated (78%).
3. I did not notice an inconsistency in the operation of Bing Chat (53%).
4. I am successful at making Bing Chat do what I want (76%).
5. Bing Chat responds to my queries/commands as I expect it to (69%).

Learnability

1. One can quickly learn the basics of working with the Bing Chat tool (94%).
2. Using Bing Chat does not require technical foreknowledge (84%).
3. It is easy to learn to use Bing Chat as an aid to studying (91%).
4. I can easily comprehend how to make Bing Chat do what I want (89%).
5. I remember well what I learned to do with Bing Chat (68%).

System Reliability

1. There were no unexpected interruptions in the operation of Bing Chat during its use (72%).
2. Bing Chat worked fast enough and reliably (81%).
3. There was no loss of data obtained from Bing Chat (75%).
4. The user interface of the Bing Chat tool worked flawlessly (79%).

Visual Design and Navigation

1. The way Bing Chat displays the discussion and its responses is visually attractive and engaging (75%).

2. Functionalities on the Bing Chat interface are well organized and easily accessible, e.g., menus, copying, etc. (86%).
3. The choice of text and background color, as well as the size and positioning of content on the screen and the design of icons are refined and appealing (77%).
4. The use of the interface with Bing Chat is logical and intuitive (easily understandable) regarding the functionalities that I use (91%).
5. Greetings and other directional messages by Bing Chat that are not part of the conversation are understandable and appropriate (83%).
6. The textual content of the interface (“Ask me anything”, “New topic”, “Recent activity” and similar) is clear and not confusing (83%).

Information Quality

1. The use of the Bing Chat service enabled the collection of accurate information (78%).
2. Verification of information obtained from Bing Chat shows that one can have confidence in its correctness (73%).
3. The information provided by the Bing Chat service is useful and satisfactory for my needs (86%).
4. The information obtained by the Bing Chat service was as a rule sufficient for me concerning the reasons for its use (74%).
5. The data provided by the Bing Chat service were clear and easy to understand (94%).
6. The information provided by the Bing Chat service was up-to-date, i.e., not obsolete (80%).

Information Display

1. The data obtained by the Bing Chat service are in a suitable and easy-to-use format for further use (90%).
2. The way information is displayed in Bing Chat’s responses is clear and well structured (90%).
3. I was able to easily share the information from Bing Chat with others (deliver it to others) (84%).
4. It was easy for me to connect the information provided in different responses during a longer conversation with Bing Chat (71%).
5. The information provided in a conversation with Bing Chat can also be obtained after a few days/weeks (52%).
6. Bing Chat displays the requested information quickly and without much waiting on my part (77%).
7. By using the Bing Chat service, I obtain the requested information without asking many questions (76%).

Cognitive Involvement

1. When using Bing Chat, I can retain attention and interest in this activity longer than when using other information search systems (60%).
2. When I use the Bing Chat application, I feel as if I am immersed in the communication process and the information I receive (52%).
3. Time seems to pass quickly while I am using Bing Chat (44%).
4. During the use of Bing Chat, my attention and focus will be difficult to reduce by other potentially distracting things and external distractors (47%).
5. Using the Bing Chat service to search for information is interesting and fun for me (68%).
6. I feel like I control what happens while working with Bing Chat because I use it as I want and get what I want (71%).
7. While working with Bing Chat, I can forget about other things that are not related to my interaction with that service (50%).

Design Appeal

1. The visual design of the Bing Chat application is impressive and very attractive (67%).

2. The visible representations of the content of the computer screen during Bing Chat use are modern and enjoyable to use (79%).
3. The technical aspects of interaction and ways of working with Bing Chat are very interesting to me (69%).
4. I am very interested in the practical and technical capabilities of Bing Chat which are still unknown to me and unexplored (64%).
5. The use of Bing Chat encourages me to innovate more and be even more creative (56%).
6. I feel satisfied and fulfilled during and after using the Bing Chat service (54%).

Trust

1. I believe that the Bing Chat service is designed with the goal of helping the widest possible number of people (84%).
2. I believe that using the Bing Chat service will bring much more benefits than potential harm (71%).
3. I have the same trust in the Bing Chat service as I do in other Internet services, such as social networks, etc. (57%).
4. The more I used Bing Chat, the more I felt that I could rely on this tool if I needed it (65%).
5. I am sure Bing Chat will work just as well in the future as it did when I needed it earlier (75%).
6. I believe I can rely more on Bing Chat than on most other sources of information, knowledge, and advice (52%).

Personification

1. My conversations with Bing Chat resembled an exchange of online text messages with an informed person (59%).
2. At some points during my discussions with Bing Chat, it seemed to me like Bing Chat had “human” traits (48%).
3. In some interactions with Bing Chat, I thought something along the lines of “What if this were a living being?” (48%).
4. It happened to me that I was having a conversation with Bing Chat without thinking it was an artificial system (38%).
5. Some restrictions set on the format/shape of discussion in Bing Chat make it seem less human (54%).
6. It would suit me if Bing Chat and similar services, e.g., GPT, could react as much as possible like a human being (36%).

Risk Perception

1. I am sure that my privacy is not under any threat by my use of the Bing Chat service (40%).
2. I do not feel uneasy about using the Bing Chat (or GPT) service that is based on artificial intelligence (68%).
3. I believe that the security of my computer and the data on it are not compromised when I use Bing Chat (62%).
4. I generally felt relaxed and safe when using Bing Chat (67%).
5. I am sure that there is no danger or potential threat to Bing Chat users (56%).
6. The privacy and security of Bing Chat users are not lower than, for example, those of users of social networks and similar services (70%).

Intention to Use

1. I plan to use Bing Chat or similar services (e.g., GPT) whenever I get the chance (78%).
2. The decision to use Bing Chat or GPT for a particular reason is in my case accompanied by positive feelings (75%).
3. I hope that in the future I will be able to use Bing Chat, GPT, and similar services as much as possible (79%).

4. I expect that I will need to use Bing Chat and similar services such as GPT for a long time to come (75%).
5. I believe that I will complete my future jobs and tasks faster and better if I use Bing Chat or GPT in the process (77%).
6. I will certainly find reasons and will not miss the opportunity to use Bing Chat, GPT, or a similar tool in the future (79%).

References

1. Imran, M.; Almusharraf, N. Analyzing the Role of ChatGPT as a Writing Assistant at Higher Education Level: A Systematic Review of the Literature. *Contemp. Educ. Technol.* **2023**, *15*, 464. [CrossRef] [PubMed]
2. İpek, Z.H.; Gözü, A.İ.C.; Papadakis, S.; Kallogiannakis, M. Educational Applications of the ChatGPT AI System: A Systematic Review Research. *Educ. Process Int. J.* **2023**, *12*, 26–55. [CrossRef]
3. Jahic, I.; Ebner, M.; Schön, S. Harnessing the Power of Artificial Intelligence and ChatGPT in Education—A First Rapid Literature Review. In Proceedings of the EdMedia + Innovate Learning, Vienna, Austria, 10 July 2023; Bastiaens, T., Ed.; Association for the Advancement of Computing in Education (AACE): Vienna, Austria, 2023; pp. 1489–1497. Available online: <https://www.learntechlib.org/primary/p/222670/> (accessed on 20 November 2023).
4. Lo, C.K. What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Educ. Sci.* **2023**, *13*, 410. [CrossRef]
5. Montenegro-Rueda, M.; Fernández-Cerero, J.; Fernández-Batanero, J.M.; López-Meneses, E. Impact of the Implementation of ChatGPT in Education: A Systematic Review. *Computers* **2023**, *12*, 153. [CrossRef]
6. Perera, P.; Lankathilaka, M. AI in Higher Education: A Literature Review of ChatGPT and Guidelines for Responsible Implementation. *Int. J. Res. Innov. Soc. Sci.* **2023**, *7*, 306–314. [CrossRef]
7. Pradana, M.; Elisa, H.P.; Syarifuddin, S. Discussing ChatGPT in Education: A Literature Review and Bibliometric Analysis. *Cogent Educ.* **2023**, *10*, 2243134. [CrossRef]
8. Vargas-Murillo, A.R.; de la Asuncion Pari-Bedoya, I.N.M.; de Jesús Guevara-Soto, F. Challenges and Opportunities of AI-Assisted Learning: A Systematic Literature Review on the Impact of ChatGPT Usage in Higher Education. *Int. J. Learn. Teach. Educ. Res.* **2023**, *22*, 122–135. [CrossRef]
9. Sallam, M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare* **2023**, *11*, 887. [CrossRef]
10. Ghafar, Z.N. ChatGPT: A New Tool to Improve Teaching and Evaluation of Second and Foreign Languages A Review of ChatGPT: The Future of Education. *Int. J. Appl. Res. Sustain. Sci.* **2023**, *1*, 73–86.
11. Trust, T.; Whalen, J.; Mouza, C. Editorial: ChatGPT: Challenges, Opportunities, and Implications for Teacher Education. *Contemp. Issues Tech. Teach. Educ.* **2023**, *23*, 1–23. Available online: <https://citejournal.org/volume-23/issue-1-23/editorial/editorial-chatgpt-challenges-opportunities-and-implications-for-teacher-education> (accessed on 20 November 2023).
12. Ipsos. *Americans Hold Mixed Opinions on AI and Fear Its Potential to Disrupt Society, Drive Misinformation*; Tech and Society Survey; Ipsos: Washington, DC, USA, 2023. Available online: <https://www.ipsos.com/en-us/americans-hold-mixed-opinions-ai-and-fear-its-potential-disrupt-society-drive-misinformation> (accessed on 20 November 2023).
13. Park, E.; Gelles-Watnick, R. *Most Americans Haven't Used ChatGPT; Few Think It Will Have a Major Impact on Their Job*; Pew Research Center: Washington, DC, USA, 2023. Available online: <https://www.pewresearch.org/short-reads/2023/08/28/most-americans-havent-used-chatgpt-few-think-it-will-have-a-major-impact-on-their-job/> (accessed on 20 November 2023).
14. Sidoti, O.; Gottfried, J. *About 1 in 5 U.S. Teens Who've Heard of ChatGPT Have Used It for Schoolwork*; Pew Research Center: Washington, DC, USA, 2023. Available online: <https://www.pewresearch.org/short-reads/2023/11/16/about-1-in-5-us-teens-who-heard-of-chatgpt-have-used-it-for-schoolwork/> (accessed on 20 November 2023).
15. von Garrel, J.; Mayer, J. Artificial Intelligence in Studies—Use of ChatGPT and Ai-Based Tools Among Students in Germany. *Humanit. Soc. Sci. Commun.* **2023**, *10*, 799. [CrossRef]
16. Adamopoulou, E.; Moussiades, L. Chatbots: History, Technology, and Applications. *Mach. Learn. Appl.* **2020**, *2*, 100006. [CrossRef]
17. Wollny, S.; Schneider, J.; Di Mitri, D.; Weidlich, J.; Rittberger, M.; Drachler, H. Are We There Yet?—A Systematic Literature Review on Chatbots in Education. *Front. Artif. Intell.* **2023**, *4*, 654924. [CrossRef] [PubMed]
18. Hwang, G.-J.; Chang, C.Y. A Review of Opportunities and Challenges of Chatbots in Education. *Interact. Learn. Environ.* **2021**, *31*, 4099–4112. [CrossRef]
19. Ibna Riza, A.N.; Hidayah, I.; Santosa, P.I. Use of Chatbots in E-Learning Context: A Systematic Review. In Proceedings of the 2023 IEEE World AI IoT Congress (AIoT), Seattle, WA, USA, 7–10 June 2023; pp. 819–824. [CrossRef]
20. Ramandanis, D.; Xinogalos, S. Investigating the Support Provided by Chatbots to Educational Institutions and Their Students: A Systematic Literature Review. *Multimodal Technol. Interact.* **2023**, *7*, 103. [CrossRef]
21. Kuhail, M.A.; Alturki, N.; Alamlawi, S.; Alhejori, K. Interacting with Educational Chatbots: A Systematic Review. *Educ. Inf. Technol.* **2023**, *28*, 973–1018. [CrossRef]
22. Wu, R.; Yu, Z. Do AI Chatbots Improve Student's Learning Outcomes? Evidence from a Meta-Analysis. *Br. J. Educ. Technol.* **2023**. [CrossRef]
23. Alemdag, E. The Effect of Chatbots on Learning: A Meta-Analysis of Empirical Research. *J. Res. Technol. Educ.* **2023**. [CrossRef]

24. Deng, X.; Yu, Z. A Meta-Analysis and Systematic Review of the Effect of Chatbot Technology Use in Sustainable Education. *Sustainability* **2023**, *15*, 2940. [\[CrossRef\]](#)
25. Granić, A. Educational Technology Adoption: A Systematic Review. *Educ. Inf. Technol.* **2022**, *27*, 9725–9744. [\[CrossRef\]](#)
26. Davis, F.D. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Q.* **1989**, *13*, 319–340. [\[CrossRef\]](#)
27. Venkatesh, V.; Morris, M.G.; Davis, G.B.; Davis, F.D. User Acceptance of Information Technology: Toward a Unified View. *MIS Q.* **2003**, *27*, 425–478. [\[CrossRef\]](#)
28. Rogers, E. *Diffusion of Innovations*, 5th ed.; Free Press/Simon and Schuster: New York, NY, USA, 2003.
29. DeLone, W.H.; McLean, E.R. Information Systems Success: The Quest for the Dependent Variable. *Inf. Syst. Res.* **1992**, *3*, 60–95. [\[CrossRef\]](#)
30. DeLone, W.H.; McLean, E.R. The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. *J. Manag. Inf. Syst.* **2003**, *19*, 9–30.
31. Gamage, S.H.P.W.; Ayres, J.R.; Behrend, M.B. A Systematic Review on Trends in Using Moodle for Teaching and Learning. *Int. J. STEM Educ.* **2022**, *9*, 9. [\[CrossRef\]](#)
32. Al-Qaysi, N.; Mohamad-Nordin, N.; Al-Emran, M. A Systematic Review of Social Media Acceptance from the Perspective of Educational and Information Systems Theories and Models. *J. Educ. Comput. Res.* **2020**, *57*, 2085–2109. [\[CrossRef\]](#)
33. Lu, J.; Schmidt, M.; Lee, M.; Huang, R. Usability Research in Educational Technology: A State-Of-The-Art Systematic Review. *Educ. Technol. Res. Dev.* **2022**, *70*, 1951–1992. [\[CrossRef\]](#)
34. Vlachogianni, P.; Tselios, T. Perceived Usability Evaluation of Educational Technology Using the System Usability Scale (SUS): A Systematic Review. *J. Res. Technol. Educ.* **2022**, *54*, 392–409. [\[CrossRef\]](#)
35. Vlachogianni, P.; Tselios, N. Perceived Usability Evaluation of Educational Technology Using the Post-Study System Usability Questionnaire (PSSUQ): A Systematic Review. *Sustainability* **2023**, *15*, 12954. [\[CrossRef\]](#)
36. Microsoft. *The New Bing: Our Approach to Responsible AI*; Microsoft Corporation: Redmond, WA, USA, 2023. Available online: <https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2023/04/RAI-for-the-new-Bing-April-2023.pdf> (accessed on 20 November 2023).
37. Stallbaumer, C. *Introducing Bing Chat Enterprise, Microsoft 365 Copilot Pricing, and Microsoft Sales Copilot*; Microsoft Corporation: Redmond, WA, USA, 2023. Available online: <https://www.microsoft.com/en-us/microsoft-365/blog/2023/07/18/introducing-bing-chat-enterprise-microsoft-365-copilot-pricing-and-microsoft-sales-copilot/> (accessed on 20 November 2023).
38. Mehdi, Y. *Announcing Microsoft Copilot, Your Everyday AI Companion*; Microsoft Corporation: Redmond, WA, USA, 2023. Available online: <https://blogs.microsoft.com/blog/2023/09/21/announcing-microsoft-copilot-your-everyday-ai-companion/> (accessed on 20 November 2023).
39. Strzelecki, A. To Use or Not to Use ChatGPT in Higher Education? A Study of Students' Acceptance and Use of Technology. *Interact. Learn. Environ.* **2023**, 1–14. [\[CrossRef\]](#)
40. Lai, C.Y.; Cheung, K.Y.; Chan, C.S. Exploring the Role of Intrinsic Motivation in ChatGPT Adoption to Support Active Learning: An Extension of the Technology Acceptance Model. *Comput. Educ. Artif. Intell.* **2023**, *5*, 100178. [\[CrossRef\]](#)
41. Tiwari, C.K.; Bhat, M.A.; Khan, S.T.; Subramaniam, R.; Khan, M.A.I. What Drives Students Toward ChatGPT? An Investigation of the Factors Influencing Adoption and Usage of ChatGPT. *Interact. Technol. Smart Educ.* **2023**, ahead of print. [\[CrossRef\]](#)
42. Saxena, A.; Doleck, T. A Structural Model of Student Continuance Intentions in ChatGPT Adoption. *EURASIA J. Math. Sci. Tech.* **2023**, *19*, em2366. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Pinsky, Y. *Bard Can Now Connect to Your Google Apps and Services*; Google LLC: Mountain View, CA, USA, 2023. Available online: <https://blog.google/products/bard/google-bard-new-features-update-sept-2023/> (accessed on 20 November 2023).
44. Bubaš, G.; Čižmešija, A. Measuring Video Conferencing System Success in Higher Education: Scale Development and Evaluation. *Int. J. Emerg. Technol. Learn.* **2023**, *18*, 227–254. [\[CrossRef\]](#)
45. Bubaš, G.; Babić, S.; Čižmešija, A. Usability and User Experience Related Perceptions of University Students Regarding the Use of Bing Chat Search Engine and AI Chatbot: Preliminary Evaluation of Assessment Scales. In Proceedings of the SISY 2023, IEEE 21st International Symposium on Intelligent Systems and Informatics, Pula, Croatia, 21–23 September 2023; pp. 607–612.
46. Brooke, J.B. SUS: A Retrospective. *J. Usability Stud.* **2013**, *8*, 29–40. Available online: https://uxpajournal.org/wp-content/uploads/sites/7/pdf/JUS_Brooke_February_2013.pdf (accessed on 20 November 2023).
47. Lewis, J.R. Usability: Lessons Learned and Yet to Be Learned. *Int. J. Hum. Comp. Inter.* **2014**, *30*, 663–684. [\[CrossRef\]](#)
48. Plantak Vukovac, D.; Horvat, A.; Čižmešija, A. Usability and User Experience of a Chat Application with Integrated Educational Chatbot Functionalities. In *Learning and Collaboration Technologies: Games and Virtual Environments for Learning, Proceedings of the HCII 2021, Online, 24–29 July 2021*; Zaphiris, P., Ioannou, A., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2021; Volume 12785, pp. 216–229. [\[CrossRef\]](#)
49. Mulia, A.P.; Piri, P.R.; Tho, C. Usability Analysis of Text Generation by ChatGPT OpenAI Using System Usability Scale Method. *Procedia Comput. Sci.* **2023**, *227*, 381–388. [\[CrossRef\]](#)
50. Salman, H.; Mohsin, E.A.; Al Rawi, A.; Shatnawi, S. Investigating HCI of the LMS Blackboard Ultra Using WAMMI during COVID-19: Usability and Design Interactivity. In Proceedings of the 2022 International Conference on Innovation and Intelligence for Informatics Computing and Technologies (3ICT), Sakheer, Bahrain, 20–21 November 2022; pp. 519–525. [\[CrossRef\]](#)
51. Nielsen, J. The Art of Navigating Through Hypertext. *Commun. ACM* **1990**, *33*, 296–310. [\[CrossRef\]](#)

52. Cuddihy, E.; Spyridakis, J.H. The Effect of Visual Design and Placement of Intra-Article Navigation Schemes on Reading Comprehension and Website User Perceptions. *Comput. Hum. Behav.* **2012**, *28*, 1399–1409. [\[CrossRef\]](#)
53. Petter, S.; McLean, E.R. A Meta-Analytic Assessment of the Delone and McLean Is Success Model: An Examination of Is Success at the Individual Level. *Inf. Manag.* **2009**, *46*, 159–166. [\[CrossRef\]](#)
54. Saadé, R.; Bahli, B. The Impact of Cognitive Absorption on Perceived Usefulness and Perceived Ease of Use in On-Line Learning: An Extension of the Technology Acceptance Model. *Inf. Manag.* **2005**, *42*, 317–327. [\[CrossRef\]](#)
55. Fredricks, J.A.; Blumenfeld, P.C.; Paris, A.H. School Engagement: Potential of the Concept, State of the Evidence. *Rev. Educ. Res.* **2004**, *74*, 59–109. [\[CrossRef\]](#)
56. van der Heijden, H. User Acceptance of Hedonic Information Systems. *MIS Q.* **2004**, *28*, 695–704. [\[CrossRef\]](#)
57. Söllner, M.; Leimeister, J.M. What we really know about antecedents of trust: A critical review of the empirical information systems literature on trust. In *Psychology of Trust: New Research*; Gefen, D., Ed.; Nova Science Publishers: Hauppauge, NY, USA, 2013; pp. 127–155. Available online: <https://ssrn.com/abstract=2475385> (accessed on 15 November 2023).
58. Choudhury, A.; Shamszare, H. Investigating the Impact of User Trust on the Adoption and Use of ChatGPT: Survey Analysis. *J. Med. Internet Res.* **2023**, *25*, e47184. [\[CrossRef\]](#)
59. Sarkar, S.; Chauhan, S.; Khare, A. A Meta-Analysis of Antecedents and Consequences of Trust in Mobile Commerce. *Int. J. Inf. Manag.* **2020**, *50*, 286–301. [\[CrossRef\]](#)
60. Im, I.; Kim, Y.; Han, H.-J. The Effects of Perceived Risk and Technology Type on Users' Acceptance of Technologies. *Inf. Manag.* **2008**, *45*, 1–9. [\[CrossRef\]](#)
61. Taber, K.S. The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Res. Sci. Educ.* **2018**, *48*, 1273–1296. [\[CrossRef\]](#)
62. Kyriazos, T.A. Applied Psychometrics: Sample Size and Sample Power Considerations in Factor Analysis (EFA, CFA) and SEM in General. *Psychology* **2018**, *9*, 2207–2230. [\[CrossRef\]](#)
63. Mundfrom, D.J.; Shaw, D.G.; Ke, T.L. Minimum Sample Size Recommendations for Conducting Factor Analyses. *Int. J. Test.* **2005**, *5*, 159–168. [\[CrossRef\]](#)
64. Green, S.B. How Many Subjects Does It Take to Do a Regression Analysis. *Multivar. Behav. Res.* **1991**, *26*, 499–510. [\[CrossRef\]](#)
65. Hasan Emon, M.M.; Hassan, F.; Hoque Nahid, M.; Rattanawiboonsom, V. Predicting Adoption Intention of Artificial Intelligence. *AIUB J. Sci. Eng.* **2023**, *22*, 189–199. [\[CrossRef\]](#)
66. Jo, H. Decoding the ChatGPT Mystery: A Comprehensive Exploration of Factors Driving AI Language Model Adoption. *Inform. Dev.* **2023**, 02666669231202764. [\[CrossRef\]](#)
67. Foroughi, B.; Senali, M.G.; Iranmanesh, M.; Khanfar, A.; Ghobakhloo, M.; Annamalai, N.; Naghmeh-Abbaspour, B. Determinants of Intention to Use ChatGPT for Educational Purposes: Findings from PLS-SEM and fsQCA. *Int. J. Hum.-Comput. Interact.* **2023**, ahead of print. [\[CrossRef\]](#)
68. Romero-Rodríguez, J.; Ramírez-Montoya, M.; Buenestado-Fernández, M.; Lara-Lara, F. Use of ChatGPT at University as a Tool for Complex Thinking: Students' Perceived Usefulness. *J. New Approaches Educ. Res.* **2023**, *12*, 323–339. [\[CrossRef\]](#)
69. Faruk, L.I.D.; Rohan, R.; Ninrutsirikun, U.; Pal, D. University Students' Acceptance and Usage of Generative AI (ChatGPT) from a Psycho-Technical Perspective. In Proceedings of the 13th International Conference on Advances in Information Technology (IAIT '23), Bangkok, Thailand, 6–9 December 2023; Association for Computing Machinery: New York, NY, USA, 2023; p. 15. [\[CrossRef\]](#)
70. Ma, X.; Huo, Y. Are Users Willing to Embrace ChatGPT? Exploring the Factors on the Acceptance of Chatbots from the Perspective of AIDU Framework. *Technol. Soc.* **2023**, *75*, 102362. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Copyright of Future Internet is the property of MDPI and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.