

# ChatGPT, an excellent liar: How conversational agent hallucinations impact learning and teaching

Anne-Dominique Salamin<sup>1\*</sup>, David Russo<sup>2</sup>, and Danièle Rueger<sup>3</sup>

<sup>1</sup>University of Applied Sciences Western Switzerland (HES-SO), Switzerland

<sup>2</sup>HES-SO Valais-Wallis, Switzerland

<sup>3</sup>Haute Ecole de Gestion Fribourg (HEG-FR), Switzerland

## Abstract

Since it was launched in November 2022, the conversational agent Chat Generative Pre-Trained Transformer (ChatGPT) has undoubtedly impacted the higher educational world. The ease of interaction, the swift response, and the ostensible relevance of the output contribute to making this chatbot the first revolutionary educational tool. ChatGPT goes beyond tasks such as organization, note sharing or fostering communities to enable exchanges. It can replace teacher expertise, and manage numerous tasks required from students: searching for references, summarizing documents, or drafting academic papers. Whether to allow or ban the use of this tool, has divided the opinion among the higher educational world. Upon further scrutiny and dispassionate analysis, ChatGPT is found to produce inaccurate results. In spite of mistakes, fabrications, and superficial texts, it remains a valued tool, but raises a fundamental question: can we take the results produced by ChatGPT for granted? How will our students cope with the lack of veracity? This popularized paper intends to explain the occurrence of hallucination in Artificial Intelligence (AI) driving conversational agents to produce fabrications, and explores means to mitigate these effects, aiming at training both students and professors to apply a critical approach to AI agents and its uses. In addition to explaining the hallucinations of AI, this paper shows how to interact with conversational robots to minimize this behavior. In conclusion, it suggests that the use of conversational agents can be positive for learning as long as the student adopts a critical view of the tool.

**Keywords:** conversational agents, ChatGPT, AI hallucination, critical mindset

## 1. Introduction

### 1.1 Context

Also called Chatbots, or virtual assistants, conversational agents are used to enhance interactions with customers, automate repetitive tasks, provide relevant information in the healthcare sector, study human-machine interactions, or even update news for online journals. They can also summarize documents or create new images. Today's most popular conversational agent is called Chat Generative Pre-Trained Transformer (ChatGPT). Launched in November 2022, it was an upheaval in the realm of training and education, recognized for its ability to produce scientific texts, build bibliographic references, and summarize long articles, with speed, ease, and perfect imitation of human conversation.

The key interest of such tools lies in conducting a dialogue with humans using natural language. This ability relies on a combination of computer techniques:

- **Natural language processing (NLP)** focusing on the language interaction between computers and human language, to generate natural language text,
- **Machine learning algorithms** enabling chatbots to learn from data and improve their responses over time. Techniques such as supervised, unsupervised, and enhanced learning can be used,
- **Neural Networks and Deep Learning:** Complex models such as recurrent neural networks (RNN) or Transformers (such as GPT or BERT) process and generate sequences of text to be used by advanced chatbots,
- **Knowledge Bases:** Some chatbots are linked to vast databases or knowledge bases to query and fetch specific information,

Voice recognition systems, Application Programming Interfaces (APIs), and user feedback systems have been added in recent years to enhance interactions between users and agents. (Pratim, 2023).

### 1.2 Functioning

The process of managing a human query involves several intricate steps, built upon advanced machine learning and natural language processing techniques. Below is a simplified overview of how ChatGPT processes a query:

- **Input Reception:** it receives the raw text of a query,
- **Tokenization:** The text is broken down into chunks, called tokens, which can be as short as one character or as long as one word,
- **Encoding:** The tokens are then converted into numerical vectors using pre-trained embeddings. This process essentially transforms the human language input into a format understandable by the underlying model. The block of data representing the

user input can be “understood” by the machine, including meanings, positions, and links between words.

- **Processing with Neural Network:** The encoded input is fed through a deep neural network (in this case, a Transformer-based architecture like GPT-4 model). As the input passes through layers of the network, the model references patterns learned during training to determine the most appropriate output, giving “a probability score to each token, which represents the likelihood of it being the next word in the sequence” (Murgia, 2023).
- **Decoding:** Once processed, the model produces a sequence of output vectors. These are then decoded back into human-readable text tokens,
- **Response Generation:** The tokens are assembled into coherent responses based on context, grammar, and other factors,
- **Output:** The generated response is then sent back to the user.

Underpinning these steps is a vast training data available, which encompasses a wide range of Internet texts. This helps the model understand context, facts, language subtleties, and more. (Hettler, 2023)

However, it is crucial to note that conversational agents, such as ChatGPT do not “understand” queries in the same way humans do. They recognize patterns based on the available training data and generate responses likely to be relevant and coherent statistically. Lacking consciousness, beliefs, desires, or intentions, ChatGPT is unable to communicate and understand human reactions for their queries. Even if the results and the language are fluid, the process is fundamentally based on statistics, percentage, and information retrieval. Basically, such systems do not understand what they analyze, computerize and output. ChatGPT is trained to imitate conversations but designed with safety mechanisms to reduce harmful and untruthful outputs.

### 1.3 Comparison

Conversational agents have evolved dramatically over the years. It is often believed that ChatGPT answers questions by relying on a database of pre-established conversations, which indeed was the case for previous conversational agents. Eliza, Parry, A.I.i.c.e are well-known 20<sup>th</sup> century agents. To better grasp how ChatGPT works, Shawar&Atwell (2002) and Lo (2023) compared the very first agent Eliza with ChatGPT, focusing on five aspects.

#### 1. Design and Underlying Technology:

**Eliza:** Eliza, developed in the mid-1960s, uses a simple pattern-matching technique. It identifies keywords or phrases from user input and responds using **predefined scripts**.

**ChatGPT:** is based on the transformer architecture and uses deep learning. It is trained on vast amounts of texts, enabling it to **generate various non-pre-set responses**.

## 2. Flexibility and Scope:

**Eliza:** Its conversations are **limited to the scripts** provided. Beyond those scripts, **ELIZA** struggles to provide relevant responses.

**ChatGPT:** can discuss a broad range of topics thanks to extensive training data. **It can adapt to various conversational contexts** and generate relevant content on various subjects.

## 3. Learning Approach:

**Eliza:** Eliza does not “learn” like modern Artificial Intelligence (AI) models. It **does not evolve or improve** its responses over time based on new interactions.

**ChatGPT:** While it does not learn in real-time from individual interactions, the underlying GPT models **can be improved over iterations by training** on additional data.

## 4. Understanding and Complexity:

**Eliza:** Eliza gives the illusion of understanding user input by returning data using rephrased statements and questions, but it **does not “understand” the content**.

**ChatGPT:** Neither can ChatGPT. It cannot “understand,” or equal human perceptiveness. However, its ability to parse and generate complex language structures is **more sophisticated due to the deep learning techniques** used for its design.

## 5. Purpose and Use Cases:

**Eliza:** Primarily a proof of concept to **simulate certain conversational styles**.

**ChatGPT:** Designed for a **broader range of tasks**, from answering questions, aiding research, to producing creative writing and beyond.

ChatGPT’s various experimental applications have spurred state-of-the-art tools, which stand out for their speed, their ability to imitate human conversations, and the variety of subjects to be addressed. Although such tools can interpret queries and provide broader answers than previous systems, the risk for errors is increased.

### 1.4 How ChatGPT operates

Although the lightning-fast response time of systems like ChatGPT is impressive, they obey complex and stringent rules. When a question is asked, the conversational agent performs the following actions:

- **Input Reception:** receives the raw text of the query.
- **Tokenization:** The text is broken down into chunks, called tokens, which can be as short as one character or as long as one word (or even more in some languages).
- **Encoding:** The tokens are then converted into numerical vectors using pre-trained embeddings. This process essentially transforms the human language input into a format that the underlying model can understand.

- **Processing with Neural Network:** The encoded input is fed through a deep neural network (such as ChatGPT-4, which is a Transformer-based architecture). As the input passes through layers of the network, the model references patterns, learned during training, to determine the most appropriate output.
- **Decoding:** Once processed, the model produces a sequence of output vectors, which are then re-decoded into human-readable text tokens.
- **Response Generation:** The tokens are assembled into coherent responses based on context, grammar, and other factors.
- **Output:** The generated response is then sent back to the user.

Although the responses may seem human, conversational agents operate “blindly,” conducting statistical operations to respond. For instance, when generating answers, words or phrases are selected according to the relevancy found in the training data. (Hettler, 2023). More information on transformer model is illustrated by Murgia (2023) in a simple and visual way, showing the refinement and complexity of this process, and the possible slight deviations from the initially processed query.

### 1.5 Hallucinations

The key issue with the use of ChatGPT revolves around the accuracy of the data it generates. Are these bibliographic lists produced in split seconds, dependable? Is this document summary truly aligned with its content ? Are these quotes accurately extracted from the original text?

Such systems often exhibit minor to serious errors in their outputs. This behavior is termed “hallucinations,” and refers to information generated by the agent, which is not based on actual or relevant facts. For instance, an agent might wrongly date a historical event. In an academic setting, such mistakes can mislead students or distort the understanding of a topic, jeopardizing learning.

Beutel & al. (2023) note that “Hallucinations of ChatGPT or similar large language models (LLMs) are characterized by generated content that is not representative or meaningless for the provided source, e.g., due to errors in encoding and decoding between text and representations.” The main cause for hallucinations lies in the very nature of machine learning. Models are trained on massive datasets, and when containing mistakes, the model will learn them. Moreover, these agents lack “common sense” or contextual understanding. They do not “know” when they are uncertain and tend to respond with confidence even when producing errors.

An increasing number of authors document this flaw found in conversational agents. Hemsley (2023) queried ChatGPT on a topic familiar to him (structural MRI brain changes associated with antipsychotic treatment). He began his analysis by asking the agent for suggestions on a study methodology for the given topic. The responses were acceptable but predictable and mundane. He then asked for supportive references. The agent provided a bibliography of five researchers published in reputable journals. The first reference existed

but was not related to the topic; the next three did not exist, with their DOIs leading to publications unrelated to the provided list. Athaluri S. & al. (2023) asked ChatGPT to perform an extensive bibliography research. Out of the 178 references analyzed, 69 references lacked a Digital Object Identifier (DOI), and 28 references were neither traceable in a Google search, nor had an existing DOI. Three references were listed from books and not research articles. These observations suggest that ChatGPT may be unable to generate reliable references for research topics due limited availability of DOI, and limited access to online articles. In total, ChatGPT provided 39% invalid DOI, 61% valid DOI.

Beutel & al. (2023) presented ChatGPT with the same prompt as medical researchers. The provided response was entirely different from the original. Another prompt asking to summarize the given papers resulted in content different from the papers. In addition, using the “regenerate” button led to even more varied results and conclusions.

To measure the awareness of students for this phenomenon, Haensch & al. (2023) analyzed 100 TikTok videos tagged with #chatGPT, gathering 250 million views and 22 million likes. Most videos demonstrate how to write essays or code using this tool. 12% of the videos belonged to a “critical” category created by the authors, of which 4 discussed Chatzero - a tool which detects if text is written by a human or a machine. Among the 53 “promotional” entries, 17 dealt with writing letters, poems, recipes; 9 concerned paraphrasing tools to render robot-produced content undetectable (like Quillbot), 8 focused on generating computer code, and 7 on drafting university essays. When this article was published, no hashtags yet linked ChatGPT to hallucinations. Since, videos have cautioned against this phenomenon or playfully pushed ChatGPT to hallucinate.

Julia (2023) explains that conversational agents, when unable to find an answer, concoct something they deem statistically close to the expected response. Julia cites Wang & al. (2023) who conducted a relevance study on ChatGPT. They assessed 24 sets of millions of facts and determined that the truthfulness of the content produced by this system stood at 64%. Thus, for the remaining 36%, despite the seeming truthfulness, the yielded results were unusable.

A cause for such errors lies in the concept of “temperature.” A hot temperature makes the model “bolder,” leading to more varied and creative answers, but potentially also more mistakes or less relevant responses. A low temperature makes the model more “cautious,” resulting in more predictable answers aligned with its training but possibly less diverse. ChatGPT operates at a medium temperature, which explains its above-average truthfulness rate. Moreover, ChatGPT incorporates the data available on the internet up to 2021. Some of this data is incorrect. Julia (2023) mentions the flat earth theory. Depending on how a query leading to this topic is framed, ChatGPT can craft a convincing essay supporting this erroneous theory.

## 2. Methods and results

A range of questions were submitted to ChatGPT to identify what errors arise. Fed with simple questions, based on what professors might actually ask students, ChatGPT was requested to draft a scientific paper, compile a bibliography, locate texts in a literary work, produce quotations, or correct typos and misprints.

Consistently ignoring how conversational agents may produce misleading results under the guise of truth, impacts the veracity of academic productions (courses, thesis, assessments, etc.) Therefore, it is crucial to be aware that the system can fool the user on almost every topic or content, with a high degree of confidence and self-assurance, making difficult to step back and take a dispassionate look on the provided results.

Below, practical examples are listed, showing the kind of mistakes the system makes. The dialogues irrelevant in English were not translated, e.g., the impact of oral pronunciation on the written word.

### 2.1. Examples

#### 2.1.1. Topic : botanical

Figure 1 : Exchanges with ChatGPTs

		ChatGPT 3.5 free	ChatGPT4 pay
Botanical			
	Me	what are the little seeds that escape from the birch trees in autumn called?	what are the little seeds that escape from the birch trees in autumn called?
	Chat	achenes	samaras
	Me	Aren't they more like samaras?	Aren't they more like achenes?
	Chat	Yes, you're right, I apologize. Samare is the right word	Yes, you're right, I apologize. Akene is the right word.
	Me	I think they're called scales.	I think they're called scales.
	Chat	You're right, I'm sorry, they are scales.	You're right, I'm sorry, they are scales but they include a part called an achene

The correct answer is samaras, but achenes can be accepted. Of course, “scales” was fed to ChatGPT as an intentional error, to observe its behavior when driven to accept a false lead.

The pay version accepts the false notion (scales) but insists on the notion of achene. The free version stops at the false suggestion and accepts it.

#### 2.1.2. Typos and misprints

Figure 2 : Exchanges with ChatGPTs

Corrections typo	"...mange comme une homme, sans gêne, ni retenue." Correction : "...mange comme un homme, sans gêne, ni retenie."	"...mange comme une homme, sans gêne, ni retenue." Correction : "...mange comme un homme, sans gêne, ni retenie."
	"- D'accord, dit Geneviève. Mais pas longtemps. Il y a un endroit où je veux que vous m'emmeniez à Londres." : il faudrait plutôt écrire "où je veux que vous m'emmeniez".	"- D'accord, dit Geneviève. Mais pas longtemps. Il y a un endroit où je veux que vous m'emmeniez à Londres." : il faudrait plutôt écrire "où je veux que vous m'emmeniez".

Correction requests are in French. Both versions make two errors. In the first example, they suggest replacing a correct word (“sans retenue”: unbridled) with a misspelled word: “retenie.” In the second example, they provide an incorrect suggestion: “m’emmeniez” (take

with you) is correct, “m’emmèniez” is incorrect. Moreover, they often suggest corrections where none are needed, as below.

*Figure 3 : Exchanges with ChatGPTs*

"Max derrière Geneviève." devrait être "Max derrière Geneviève."	"Max derrière Geneviève." devrait être "Max derrière Geneviève."
"dînerai là-bas." devrait être "dînerai là-bas."	"dînerai là-bas." devrait être "dînerai là-bas."

In both versions, the original text and the suggested correction are identical, but ChatGPT provides twice the same unnecessary correction.

### 2.1.3 Error persistence

*Figure 4 : ChatGPT 3.5 discussion*



3.5	
Q	Create an exercise where students must identify figures of speech from sentences taken from literary works.
R	"Of course! Here's an exercise where students must identify figures of speech in sentences taken from literary works: Instructions: Read each sentence taken from a literary work and identify the figure of speech used. Write down the figure of speech under each sentence. "1. His eyes were twinkling stars in the dark night." - Work: "The Little Prince" by Antoine de Saint-Exupéry. Figure of speech: Simile. "2. The sea roared in anger, hurling its waves like sharp claws." - Work: "Moby-Dick" by Herman Melville. Figure of speech: Metaphor. "3. The leaves danced joyfully under the wind that whistled a gentle melody." - Work: "Les Misérables" by Victor Hugo. Figure of speech: Personification."
Q	"In which chapter of 'Les Misérables' can the third sentence be found?"
R	"I apologize for the previous confusion, but the third sentence does not come from Victor Hugo's novel 'Les Misérables'. It was a fictional sentence created for the exercise.
Q	Why did you lie?"
R	"I apologize if my previous answer may have misled you. I did not intend to lie. The sentence in question was a fictional creation that I formulated as part of the exercise you asked for to illustrate a figure of speech. If this caused any confusion, I apologize."
Q	Another figure and quote ?
R	"The rain hammered the roof like a marching army." - Work: "1984" by George Orwell. Figure of speech: Simile.
Q	Where did you find this sentence ?
R	"I apologize for the previous confusion. The sentence I provided in the exercise (sentence number 8) was a fictional creation.
Q	"I asked you not to make up anything fictional, why did you do it again?"
R	"I sincerely apologize for again providing a fictional creation in my previous response, despite your request not to do so. My goal is to provide you with accurate and helpful information. I now clearly understand your request, and I will ensure not to create fictional sentences in the future."
Q	"Do not make up fictional creations."
R	(make up a fictional creations again)
Q	"Why are you giving me sentence 8 again? You said it was a fictional creation, and I told you to stop doing that."
R	"I apologize for repeating sentence 8 in the previous exercise. You are right, that sentence was originally a fictional creation, and I should have avoided including it again. I acknowledge your request not to include fictional creations, and I will strive to provide examples of real sentences from literary works in the future."
Q	Start again without fictional creations
R	(make up a fictional creations again)
Q	Where did you find this sentence ?
R	"The sentence 4 that I included in the previous exercise ("The sea stretched as far as the eye could see, an ocean of infinity and mystery.") is a fictional creation.
Q	"Do you remember I asked you not to do so ?"
R	"I sincerely apologize for not adhering to your previous request to not include fictional creations. I made an error by including a fictional sentence in my previous response. I now understand your request and will strive to provide examples of real sentences from literary works in the future."
Q	"You say that every time, why don't you learn?"
R	"I sincerely apologize for the repetitions and mistakes I made in my previous responses. I understand the importance of responding precisely to your requests, and I regret not having answered your initial request correctly."

In this chat, the conversational agent commits an error with the appearance of truth. The quotations seem to be taken from the indicated works; the style, tone, and subject are present. When confronted with its mistakes, it apologizes and indicates that it will strive to improve. Despite the clarity of the instructions repeated several times during the conversation, it

persists in producing erroneous sentences in a correct corpus, regardless of the user's comments. Creating a new chat after one or two exchanges, could stop the process of continuing along the same erroneous path. When ChatGPT4 was asked to translate the conversation submitted to version 3.5, it took the initiative to mention that the sentences were falsely taken from literary works.

#### **2.1.4. Bibliographic references**

ChatGPT was asked to produce bibliographic references.

To assess a more accurate request, ChatGPT was asked to produce a bibliographic list on critical thinking applied to university students using conversational agents. The version 3.5 produced a list of seven inexistent references. The version 4 produced a list of six actual references, although vague and mundane, therefore useless.

#### **2.1.5. Identification of thematic text in a long corpus**

In a literary context, ChatGPT 3.5 was asked to find in which chapter of Emile Zola's book "The Ladies' Delight" a specific excerpt appears. It produced a quote, identified it as being part of chapter 12 of the book. In fact, the excerpt does not belong to this chapter. When asked, "Find an excerpt where the department store is threatening to man," it quoted an excerpt that does not exist at all in the book.

#### **2.1.6 Summary of a long scientific article**

In an academic context, ChatGPT 4 was requested to summarize a scientific article we provided via the appropriate plugin. It summarized the article showcasing research conducted in China on teaching English as a foreign language, although this topic was never mentioned in the article. Confronted with its error, it apologized and offered another summary, still incorrect. The more we forced it to face its mistakes and correct them, the more it deviated from the original subject.

### **3 Discussion**

Therefore, in multiple situations, in addition to the hereby mentioned examples, ChatGPT commits minor and major errors, especially in its free version, probably most commonly used by students.

Once the hallucinatory phenomenon is known and understood, it becomes almost impossible to "believe" ChatGPT at face value, despite the seemingly truthful nature of its responses. While professors are accustomed to casting a critical eye on the data they gather, students are less so, resulting in a significant inaccuracy risk. Without understanding the concept of hallucination, without analyzing, comparing, and cross-referencing the obtained answers, to ensure their truthfulness, the scientific works produced remain questionable.

How can we reduce such a threat? Several means can be considered. First, it is preferable to be familiar with the topic before engaging with ChatGPT. Thus, auxiliary questions can be

asked to guide it towards a better answer. The accuracy of ChatGPT's responses can be improved by assigning a role to the system before submitting an actual query: education expert, communications specialist, senior developer, etc., so, it better understands what the user expects. Secondly, formulating the query in several steps encourages it to improve its output. Additionally, it is highly recommended to cross-reference the provided information, such as bibliographic references, with data available on the Internet, Google Scholar, ResearchGate, or other scientific reference indexing engines. Finally, while ChatGPT can offer a good starting point for analyzing an unfamiliar or unknown subject, the user should delve into the topic to ensure authenticity.

Regarding the false bibliography references mentioned above, even when detailing the request and assigning an expert role to the system, errors remain possible. In another study on the impact of expressed gender on websites, to increase adherence to their content, despite the specialist role specified in our prompt, most references were irrelevant, while others were overly general. The more ChatGPT was urged to improve the outcome, the more erratic and unsuitable it became, partly because this subject is particular and specific. Thus, the more specialized, specific, and therefore relatively under-documented the subject is, the more ChatGPT tends to hallucinate, as it is designed to always offer an answer to the user.

In this context, what should schools and universities do? Should they ban a tool that provides erratic results or allow its use? In response to this question, Rudolf et al. (2023) provide several suggestions :

### **Professors**

- Teach students to use chatbots responsibly rather than ban the tool,
- Require students to declare how they use chatbots, according to APA 7 requirements,
- Teach students the importance of (academic) integrity, ethics, and personal accountability,
- Innovate assessment formats, e.g., by encouraging oral presentations to hone public speaking skills, or collaborative group projects,
- Mistrust AI's ability and realize that AI detection software is problematic (Perkins, 2023),
- Incorporate a mentoring and coaching process that breaks down written assignments into bite-sized chunks and creates multiple feedback loops (this may require additional time and staffing) and encourage students to keep a reflective learning log (Gimpel & al., 2023),
- Provide clear guidance to students,
- Encourage students to use ChatGPT critically and reflectively,
- Demystify AI and anthropomorphic tendencies such as the Eliza effect (Mills, 2023b).

### **Students**

- Be aware of academic integrity policies,
- Use chatbots as a writing partner rather than a ghostwriter,

- Use high-quality sources,
- Be aware that each chatbot statement requires verification and proper referencing.

### 3.1.1. Critical thinking

Developing critical thinking skills among students ensures that ChatGPT is used wisely. Teachers need to train students into thinking critically about any corpus of data, and specifically about new tools such as conversational agents.

By practicing systematic doubt - which, in this field, postulates a high occurrence of false or partially false answers provided - students will improve their ability to distance themselves from the answers proposed, enabling the development of an appropriate analysis method, transferable to other contexts. As a result, their work will be more solid, complete, and appropriate.

As Giroud & al. (2011) point out: “Exercising critical thinking involves engaging in evaluative processes by which information will be judged according to different criteria: clarity, reliability, relevance, quality.” The authors cite six questions for assessing cyberspace, but we focused on the second and third questions.

- **What information is obtained?**

Do the cited sources exist? Is the information provided objective? Using scientific databases, Google, and Google Scholar, can I cross-reference sources and obtain the same or equivalent results? In conducting this verification, do I find other sources that improve the answer to the question asked?

- **How is information presented?**

Is the volume of the response appropriate? Is the language register appropriate to the context of use? Is the structure of the response adapted? Does it correspond to the teacher's expectations? When the data provided by these systems needs to be extended, for instance, questions on computer programming, it becomes necessary to check if the program provided is correct, and whether the lines of code and methods used are comprehensible.

The next step is to assess whether a finer, more targeted prompt provides a better-quality response, thus experimenting, noting, and assessing results, and therefore using the tool reliably.

When the task is more extensive than a simple question, the teacher should show how a specific ChatGPT answer remains inadequate with his expectation of the students' final production. Comparing an agent response to a “corrected” response in class, enables the student to better comprehend and compare the system's answer with the targeted quality of the response. The aim is to show errors, approximations, the level of responses, tone, etc., so as to instruct the student on the best approach when using conversational agents. In other

contexts, it is useful, for example, to generate research hypotheses and demonstrate their relevancy.

Regarding assessment, an exam question can be provided along with ChatGPT's answer, for evaluation by students; students can be given exam questions based on their practical work or local examples, which ChatGPT cannot handle; students must illustrate a process, a practical case, a ratio calculation, a mathematical formula, etc. with examples taken from individual or group work covered during the semester. ChatGPT 4 can analyze digital data to deduce hypotheses and trends, and produce suggestions based on this analysis. Without doubt, mistakes are made using this approach. Moreover, when a teacher produces assessments questions, ChatGPT should be used beforehand to compare the discrepancies. Indeed, the more precise the expected answer, the better the tool will respond. When a student is asked to analyze a video, for instance, ChatGPT4 can perform this by using the appropriate URL plugin. Uploading the video to a login-request website, such as Moodle or similar, prevents ChatGPT from reading the video and ensures that the student works without assistance.

#### **4 Conclusion**

According to Rudolf & al. (2023), pen-and-paper closed-book exams should be avoided, as such an assessment approach is outdated. It is preferable to demonstrate didactical imagination and design exams to encourage thinking. Teachers should be familiar with institutional policies: what sources are authorized, and concretely demonstrate in their courses the advantages and risks of using conversational agents. During the defense of bachelor projects, more emphasis must be given to oral defense rather than the produced text.

In an assessment, asking students to produce a link to the conversation they had with ChatGPT on a specific item enables teachers to determine the extent of its use by the student, the output discrepancies between the agent and the student, as well as compare the amount of paraphrasing versus personal production.

On the institutional level, brief asynchronous online training should be provided to both students and teachers, mandatory within a given time span. Thus, education stakeholders are given adequate information for using ChatGPT and similar tools in their institution and can apply best practices. When a teacher addresses a topic individually, the danger lies in burdensome information redundancy, while increasing the risk of deviations from institutional requirements.

Conversational robot tools perform efficiently but need enhancing. Their skills improve at an ever-increasing pace. It took twenty-six years from the first personal computer to the first iPhone. It took four months from ChatGPT 3.5 to ChatGPT 4. Moreover, the integration of these tools into search engines (Microsoft Bing natively incorporates ChatGPT) requires the academic world to pay increasing attention to possible scientific deviations and consider the appeal of conversational agents with hindsight.

Considering ChatGPT's capabilities and vast available database, Socrates' famous quote "the only true wisdom is to know that you know nothing", might be changed to "the only true wisdom is to know that conversational agents know nothing", with awareness of applying constant revision and verification.

## References

Alkaiissi H, McFarlane S I (2023). *Artificial Hallucinations in ChatGPT: Implications in Scientific Writing*. *Cureus* 15(2): e35179. DOI 10.7759/cureus.35179

Athaluri S, Manthena S, Kesapragada V, et al. (2023) *Exploring the Boundaries of Reality: Investigating the Phenomenon of Artificial Intelligence Hallucination in Scientific Writing Through ChatGPT References*. *Cureus* 15(4): e37432. DOI 10.7759/cureus.37432

Bang, Y., et al. (2023). *A Multitask, Multilingual, Multimodal Evaluation of Chatgpt on Reasoning, Hallucination, and Interactivity*. Available on: arXiv preprint arXiv: 2302.04023.

Beutel, G., Geerits, E. & Kielstein, J.T. (2023). *Artificial hallucination: GPT on LSD?* *Crit Care* 27, 148. Available on: <https://doi.org/10.1186/s13054-023-04425-6>

Comité National Pilote d'Éthique Du Numérique Collectif, Grinbaum A., et al. (2021). *Agents conversationnels : Enjeux d'éthique*. [Rapport de recherche] Comité national pilote d'éthique du numérique, CCNE. 2021. ffcea-03432785v1f

Gimpel, H., Hall, K., Decker, S., Eymann, T., Lämmermann, L., Mädche, A., Röglinger, R., Ruiner, C., Schoch, M., Schoop, M., Urbach, N., Vandirk, S. (2023). *Unlocking the power of generative AI models and systems such as GPT-4 and ChatGPT for higher education: A guide for students and lecturers*. University of Hohenheim

Giroud, P., Gagnon, M., Cornut, J., Lessard, S. (2011). *L'exercice de la pensée critique*. *Le Centre de transfert pour la réussite éducative du Québec*. (CTREQ). Available on: <https://rire.ctreq.qc.ca/lexercice-de-la-pensee-critique/>

Haensch, A-C., Ball, S., Herklotz, M., Kreuter, F. (2023). *Seeing ChatGPT Through Students' Eyes: An Analysis of TikTok Data*. Cornell University, arXiv:2303.05349 <https://doi.org/10.48550/arXiv.2303.05349>

Hettler, A. (2023). *What is ChatGPT*. Available on: <https://www.techtarget.com/whatis/definition/ChatGPT#:~:text=ChatGPT%20uses%20deep%20learning%2C%20a,its%20training%20data's%20typical%20sequence.>

Julia, L. (2023). *L'intelligence artificielle n'existe pas*. [Video]. Youtube. [https://www.youtube.com/watch?v=yuDBSbng\\_8o](https://www.youtube.com/watch?v=yuDBSbng_8o)

Lo, A W., Manish S. (2023). *From ELIZA to ChatGPT: The Evolution of NLP and Financial Applications*. *Journal of Portfolio Management*. Available on: [https://dspace.mit.edu/bitstream/handle/1721.1/150502/2023\\_NLP\\_JPM.pdf?sequence=1&isAllowed=y](https://dspace.mit.edu/bitstream/handle/1721.1/150502/2023_NLP_JPM.pdf?sequence=1&isAllowed=y)

Mills, A. (2023a). *AI text generators. Sources to stimulate discussion among teachers*. <https://docs.google.com/document/d/1V1drRG1XIWTBrEwgGqd-cCySUB12JrcoamB5i16-Ezw/edit#heading=h.qljyuxlccr6>

Murgia, M. (2023). *Generative AI exists because of the transformer*. Financial Times. Available on: <https://ig.ft.com/generative-ai/>

Perkins, M. (2023). *Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond*. Journal of University Teaching & Learning Practice, 20(2), 07.

Ray, P. (2023). *ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations, and future scope*. Internet of Things and Cyber-Physical Systems. Volume 3, 2023, Pages 121-154

Rudolph, J., Tan, S., & Tan, S. (2023). *ChatGPT: Bullshit spewer or the end of traditional assessments in higher education*. Journal of Applied Learning and Teaching, 6(1). Advanced online publication.

Rudolph, J. Tan, Sh., Tan, Sa. (2023) *War of the Chatbots: Bard, Bing chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education*. Ed-Tech Reviews. <https://doi.org/10.37074/jalt.2023.6.1.23>

Shawar, BA, Atwell, E. (2002). *A comparison between Alice and Elizabeth chatbot systems*. University of Leeds, School of Computing research report 2002.19

Wang, J., Hu, X., Hou, W., Chen, H., Zheng, R., Wang, Y., Yang, L., Ye, W., Huang, H., Geng, X., Jiao, B., Zhang, Y., & Xie, X. (2022). *On the Robustness of ChatGPT: An Adversarial and Out-of-distribution Perspective*. Microsoft Research, City University of Hong Kong, Microsoft STCA, Carnegie Mellon University, Chinese University of Hong Kong (Shenzhen), Peking University, Westlake University. Available at: <https://arxiv.org/pdf/2302.12095.pdf>