



HAL
open science

How to Design and Deliver Courses for Higher Education in the AI Era: Insights from Exam Data Analysis

Ahmad Samer Wazan, Imran Taj, Abdulhadi Shoufan, Romain Laborde,
Rémi Venant

► **To cite this version:**

Ahmad Samer Wazan, Imran Taj, Abdulhadi Shoufan, Romain Laborde, Rémi Venant. How to Design and Deliver Courses for Higher Education in the AI Era: Insights from Exam Data Analysis. 2023. hal-04168693v1

HAL Id: hal-04168693

<https://hal.science/hal-04168693v1>

Preprint submitted on 21 Jul 2023 (v1), last revised 25 Jul 2023 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

How to Design and Deliver Courses for Higher Education in the AI Era: Insights from Exam Data Analysis

Ahmad Samer Wazan^{1*}, Imran Taj², Abdulhadi Shoufan³,
Romain Laborde⁴, Rémi Venant⁵

^{1*}College of Technological Innovation, Zayed University, UAE.

²College of Interdisciplinary Studies, Zayed University, UAE.

³Department of EECS, Khalifa University, UAE.

⁴IRIT Laboratory, Paul Sabatier University, France.

⁵LeMans University, France.

*Corresponding author(s). E-mail(s): ahmad.wazan@zu.ac.ae;

Contributing authors: [MuhammadImran.Taj@zu.ac.ae](mailto:MohammadImran.Taj@zu.ac.ae);

abdulhadi.shoufan@ku.ac.ae; laborde@irit.fr;

Remi.Venant@univ-lemans.fr;

Abstract

In this position paper, we advocate for the idea that courses and exams in the AI era have to be designed based on two factors: (1) the strengths and limitations of AI, and (2) the pedagogical educational objectives. Based on insights from the Delors report on education [1], we first address the role of education and recall the main objectives that educational institutes must strive to achieve independently of any technology. We then explore the strengths and limitations of AI, based on current advances in AI. We explain how courses and exams can be designed based on these strengths and limitations of AI, providing different examples in the IT, English, and Art domains. We show how we adopted a pedagogical approach that is inspired from the Socratic teaching method from January 2023 to May 2023. Then, we present the data analysis results of seven ChatGPT-authorized exams conducted between December 2022 and March 2023. Our exam data results show that there is no correlation between students' grades and whether or not they use ChatGPT to answer their exam questions. Finally, we present a new exam system that allows us to apply our pedagogical approach in the AI era.

Keywords: ChatGPT, LLMs, Education, Design Courses

1 Introduction

Recent advancements in Generative AI have impacted the education sector tremendously. A number of concerns have been raised regarding the strategies for adopting these AI tools, particularly Large Language Models (LLMs), in a way that facilitates the role of educators in maintaining the integrity of the learning process.

However, integrating AI into education is not a novel idea. AI based assistance for educational purposes has been an active research area since the 1970s, and has resulted in successful integration of various technology tools in the education sector. The rise in computing power and big data analysis [2] over the last decade has laid the foundation for sophisticated deep learning algorithms with billions of fine-tuned parameters. Generative Pre-trained Transformer (GPT) model - the engine of ChatGPT - exploits the same deep learning mechanisms to draw its knowledge from massive data sets, thereby achieving the capability of generating swift responses acquired by reinforcement and supervised learning techniques. The capability to generate rapid responses while solving math problems, coding programming languages, or writing assignments, including college essays, has the massive potential to improve student learning outcomes. Their 24/7 availability as personalized learning assistants makes these chatbots a strong candidate for designing efficient pedagogical approaches that could be integrated into the education sector.

Since the advent of ChatGPT, its integration implications into educational assessments have been a hot research topic. The initial buzz about ChatGPT has mostly focused on its strengths and limitations for tests and exams across various educational specialties - college level chemistry [3], mathematics [4], journalism & media education [5], parasitology [6], pharmacology [7], medical license examinations [8], law exams [9], finance research [10], and business administration [11] - to name a few. These efforts demonstrate the various inconsistent levels of ChatGPT's performance; both good [11] and bad [3] and in between.

Rudolph [12] has explored the relevance of ChatGPT for higher education, and offered general recommendations for lecturers, such as using text generator detection software, avoiding formulaic assignments, asking students to write about recent events not available in the training data and analyse images and videos. While there has been other similar short term solutions [13, 14], we could not come across any effort dedicated towards *courses design* nor *courses delivery* in the AI era. Our work addresses this gap by studying the key factors that enable the design of courses in the AI era, and by focusing on the development and testing of a pedagogical approach to deliver courses in a way that allows students to enhance their human intelligence.

To design courses in the AI era, educators have to continuously observe the strengths and limitations of AI and ML. The strengths of AI allow educators to integrate AI tools into their courses to provide personalized education to their students. On the other hand, the limitations of AI allow educators to design activities and assessments that students cannot solve with a simple click of buttons. Being aware of AI's limitations is more difficult than understanding its strengths, so a significant part of this paper is focused on presenting the limitations of AI in an abstract manner.

In addition, designing courses and exams in the AI era must consider the level of education, the learning outcomes as well as the current problems that education faces.

Therefore, we define the role of education based on the Delors report in a dedicated section, which is independent of any technology to be integrated in education. Then, we highlight some of the current ineffective practices in the higher education system, such as the excessive use of memory-based exams and the over-reliance on textbooks.

In terms of course delivery, we developed a pedagogical approach inspired by the Socratic method [15], which stimulates the critical thinking of students. In each class session, we posed open-ended questions to our students, encouraging them to engage in a dialogue with ChatGPT before answering. We then systematically discussed our students' answers and follow-up questions, explaining why they were unable to obtain specific answers. Some of the questions were designed to explore the limitations of AI and were inspired by the class discussions, while others could have been answered by LLMs, but only after a series of follow-up questions.

We adopted two strategies to ensure the validity of our pedagogical approach:

- Analysing students answers to exams conducted between December 2022 and March 2023. Our data analysis of the exam results revealed no correlation between students' grades and whether or not they use ChatGPT to answer their exam questions.
- Hearing the concerns of educators working in various fields and universities. Between January and June 2023, we presented our pedagogical approach to educators working in Math, IT, English Teaching, Human studies and Art at nine universities across five different countries (UAE, Turkey, Poland and USA). The feedback about these presentation allowed us to collect and discuss various concerns from educators working in various domains. Based on these discussions, we defined in section 7 a set of activities in different fields that prove the relevance of our pedagogical approach.

The rest of the paper is organized as follows. In section 2, we explore the different reactions of stakeholders, including the education sector, to ChatGPT. In section 3, we discuss the role of education and analyze the current barriers to learning in the higher education system. Section 4 highlights the limitations and strengths of using AI in the education sector and provides insights into how these identified strengths and limitations can be utilized to design courses and exams in the AI era. Section 5 examines the internal working principles of LLMs, like ChatGPT. In section 6, we delve into how the working principles of these LLMs can be strategically incorporated into course and exam design, creating a pedagogical approach aligned with the Socratic method, thereby effectively addressing the higher education barriers analyzed in section 3. In section 7, we present eight case studies conducted on actual university students, covering multiple subject areas, to validate the proposed pedagogical approach outlined in section 6. The data analysis in section 8 demonstrates that when the courses and exams are designed according to the proposed pedagogical approach, the usage of LLMs like ChatGPT has no impact on student performance. Finally, we give our conclusions in section 9.

2 The World’s Reaction to ChatGPT

Since the release of ChatGPT in November 2022, it has received a mix of enthusiastic and pessimistic responses in various sectors, including education where opinions split between considering LLMs as positive or negative, and everything in between.

The Economist [16] summarized the reactions by pointing out the ability of GPT3 to be eerily human-like – for better and for worse. Elon Musk, CEO of Tesla, described ChatGPT to be “scary [sic!] good”, warning about the risks of Generative AI becoming uncontrollable, and has advocated for proactive measures to ensure it is not misused. Along with Steve Wozniak, the co-founder of Apple, and Emad Mostaque, the CEO of Stability AI, Musk signed a petition calling for an end to gigantic AI experimentation [17]. Geoffrey Hinton, deemed as the pioneer of back-propagation deep learning technique of neural networks [18] has emphasized that the risks associated with Generative AI require more immediate consideration than those associated with climate change [19]. Since his resignation from Google, he has spoken out against the organization acting irresponsibly in the AI development of Bard [20], without putting in place the necessary safeguards and regulations. Hinton’s stance is endorsed by prominent AI ethics researchers such as Timnit Gebru [21] who has also raised the concerns about the lack of transparency. Many other well-known organizations such as Verizon, JPMorgan Chase, Apple, and Amazon have restricted the usage of ChatGPT by their employees, referring to data security concerns [22].

In an attempt to address concerns from such business giants, OpenAI has taken measures to give users more control over the data used to train ChatGPT. The education sector is vigilant when it comes to the capabilities of ChatGPT to generate human-like text, which was referred to as the industry’s next big disrupter for industry and that such models are getting terrifyingly good at writing [23]. The Telegraph stated that “we should be very worried about AI-generated text” [24]. Noam Chomsky, theoretical linguist, characterizes ChatGPT as basically high-tech plagiarism that will lead to avoidance of learning [25]. Francois Chollet, the creator of the library Keras, has highlighted the lack of contextual reasoning in the use of the Large Language Models (LLMs), like ChatGPT [26]. Sam Altman, CEO of OpenAI, responded to these concerns by acknowledging the serious weaknesses and silly mistakes. He referred to ChatGPT as a very early glimpse of the changed world, attributed to Generative AI. He further acknowledged that there is still a lot to figure out [27]. The author in [28] concluded that ChatGPT can demonstrate critical thinking when solving problems, as the chatbot suggested enhancements to its own generated solutions. This was the basis for recommending the use of: multi-modal channels for exam questions, video recorded visual questions, AI output detection tools, and oral exams, to preserve the exam integrity. Another approach taken by some institutes is the complete ban of ChatGPT. For example, New York schools have cited the following reason for banning ChatGPT: it does not develop critical- thinking and problem-solving skills, which are essential for academic and lifelong success [29]. Similarly, a leading Dubai university [30] decided to reinstate pen and paper exams, after discovering that one third of students in a class were using LLMs to write their essays. Verena Rieser, professor at Heriot-Watt University, admits that despite being useful, ChatGPT can generate misinformation at a scale which is obviously very concerning, when it comes to education

[31]. However, several other universities [32] have correctly recognized that the tool is here to stay and have moved away from discussions about outlawing it. The University of Bath will soon have a policy on ChatGPT and other Generative AI tools. The policy will address issues such as determining the extent to which productivity is enhanced, and ethical implications. While educational institutions are developing some policies and procedures, many staff at various universities are taking temporary measures such as setting in-person, invigilated summer exams, and/or asking students who choose to use ChatGPT to submit their ChatGPT prompts and answers as an appendix. Some tech giants and world leaders foresee the ChatGPT’s potential to revolutionize online interactions and education. Mark Zuckerberg, CEO of Facebook, is very optimistic about the applications of Generative AI in content generation, personalization, and virtual reality experiences [33]. Andy Jassy, Amazon CEO, during an interview has expressed his excitement to see “what is possible with Generative AI tools such as ChatGPT [34]”

The education sector has not only successfully integrated calculators, Wikipedia, and essays-for-pay websites, but has also contributed to build most of these technology tools. If the past is any predictor of the future, then ChatGPT will have the same integration in the education sector. Many teachers now believe that ChatGPT could actually help make education better.

Based on the public reaction to LLMs, we infer that the LLMs tools like ChatGPT are here to stay and in order to ensure the success of our students in the future workforce, we need to develop policies to integrate Generative AI tools in the higher education. The university policies may vary depending on the academic field and specific subject matter context; but should address the associated concerns about misinformation and bias, developing the students’ cognitive capabilities and responsible AI. Currently, no universities have formal policies about LLMs usage; they have either blocked LLMs such as ChatGPT or remain passive. This article aims to i) assist academics take a clear position and ii) propose mechanisms for integrating LLMs into education.

3 Education Crisis Prior to the Emergence of LLMs

Recent advances in LLMs have led many to question the role of education, with some suggesting that AI will replace human educators [35]. In this section, we recall the role of education in society. We also look at the different roles and learning objectives that distinct educational stages must achieve, as the integration of AI depends on the education stage and the learning outcomes associated with it.

The Delors report [1] delivered to UNESCO in 1996, is one of the most valuable sources that highlights the importance of education. This report was prepared by an international commission led by Jacques Delors, and its recommendations have been applied in 50 countries after being translated into 30 different languages [36].

One of the key points of the Delors report is the importance of lifelong learning. Generative AI offers a unique opportunity to implement Delors’ recommendation, as it makes the learner less dependent on the formal education system and its educators. At the same time, however, it does not eliminate the need for the formal education

system, whose role is not only to provide the basic knowledge, but also to define the best methods for integrating AI into education so that learners use these tools to build themselves up rather than over-relying on them.

Interestingly, throughout the whole Delors report, the international commission didn't give one definition of what education is. Instead, the commission defined following the four pillars of education:

- Learning to know: helps individuals to understand their environment and live in dignity. It enables individuals to develop their memory, reasoning, and critical thinking skills.
- Learning to do: allows individuals to apply what they have learned,
- Learning to live together: allows individuals to develop social skills that enable them to live and work with other individuals who might be different from them.
- Learning to be: teaches students to be creative, to build their own opinions, to criticize, and to have their own identity.

Furthermore, chapter 6 of the report discusses in details the differences between basic, secondary, and higher education. Basic education is presented in the report as '*passport to life*'. It enables people to master essential learning tools such as literacy, oral expression, problem-solving, building skills, values, and attitudes. Secondary education is presented in the report as a crucial point in the life of young students because, at this stage, they decide their future based on their tastes and what they acquired during their education. The role of higher education is to inculcate the culture of creativity in students. The reports states that "[The role of] higher education [is]...to play its part in **creating, preserving and passing on knowledge** at the highest levels."

AI can be integrated into all stages of education. But its integration depends on the education stage, whether the objective is to enable students to master the basic knowledge, to help them decide their future (secondary education), or to improve their creativity to face our constantly evolving world and the needs of the employment market (i.e., higher education). For example, the integration of AI in early childhood education must be carefully studied because the main objective at this stage is to focus on developing children's social skills. While the Delors report didn't explicitly specify the stages at which we expect the pillars of education to be applied, we obviously expect all the four pillars to be needed at the higher education stage. Indeed, the report states, "*It is primarily the universities that unite all the traditional functions associated with advancing and transmitting knowledge: research, innovation, teaching and training, and continuing education.*" Therefore, we believe that instructors in higher education should design their courses and exams to incorporate the four learning pillars. This will help develop students' critical thinking capacity and enable them to create and innovate for society, without forgetting the "Learning to be" pillar that helps students form their own identity and opinions.

However, the current higher education system suffers from many problems that hinder the achievement of the four pillars of learning. One of the main issues is relying exclusively on a single textbook for teaching. If professors are not the creators of the textbook, it becomes challenging for them to assist their students in building their own

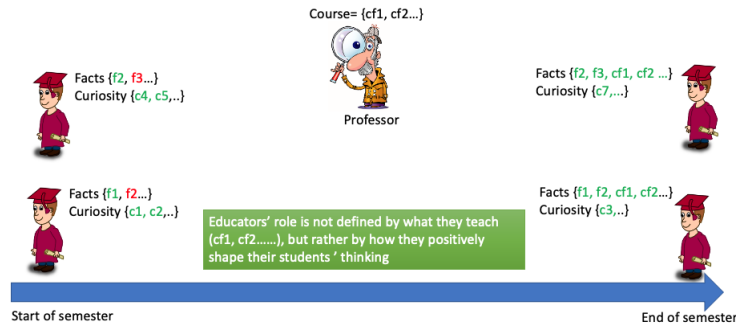


Fig. 1 Personalised Learning

identities since they are not demonstrating this ability in their teaching (the "Learning to Be" pillar).

In addition, while memorization is known to be one of the basic educational practices [1], professors in higher education should avoid assessment methods that overly require memorization because, typically, students need to be prepared for the job market where companies provide employees resources (e.g., the Internet, books, chatbots, etc.) to solve problems at hand. The Delors report clarifies this issue by stating: "It would be dangerous to imagine that memory has become unnecessary because of the incredible capacity to store and circulate information now at our disposal. We must certainly be selective about what we learn 'by heart'".

Another problem that needs to be acknowledged is the focus on students' final outcomes and answers rather than their process of working through problems. The correctness of the final outcomes of students' work should not really be the main objective of educators because their role is to develop the intelligence of students. Improving the intelligence of students requires personalised learning. In 2015, the OECD prepared a report titled 'Schooling for Tomorrow – Personalising Education' [37]. In personalised learning, the educator should make an effort to understand the needs of every student, as they are typically different between students. Educators should also detect errors in the background knowledge of every student and work with each student to address their own curiosity and correct their previous knowledge while also delivering their course contents at the same time (see figure 1). The role of educators is not only defined by what they teach (i.e., the facts of their courses); but rather and more importantly by their ability to develop students' cognitive skills by addressing their students' curiosity and their misconceptions.

Consequently, all the evaluation rubrics should be redesigned to measure students' ability to think rather than focus exclusively on the final answers that students provide during their course activities and exams. For example, figure 2 illustrates an example where a student is asked to write five lines about the beauty of the Aleppo citadel. Student1 provides his teacher with his final answer (i.e., the essay) along with other pieces of evidence representing his efforts to produce the essay. In contrast, student2 provides only his final answer without giving evidence about how he/she produced the final solution. As can be seen, student1 uses ChatGPT more thoughtfully and

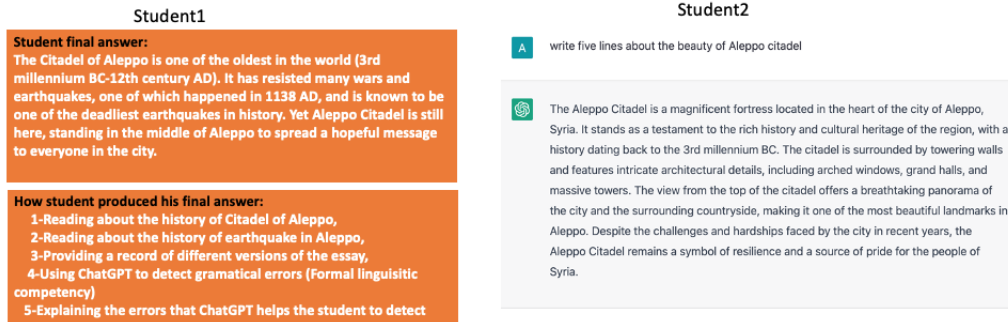


Fig. 2 Example of English essay

effectively than student2. Hence, the role of educators is to encourage the former chatbot usage while discouraging the later.

It is thus in the hands of every educator to avoid what the French philosopher Albert Camus once said: "The school prepares us for life in the world that does not exist". The Delors report nicely summarises the attitude that every educator has to follow in the education systems: "Teachers' great strength lies in the example they set, of curiosity, open-mindedness, willingness to put their assumptions to the test and to acknowledge mistakes; most of all, they must transmit a love of learning." Reducing the roles of educators to rule enforcers and cheating deterrent does not help students to be critical thinkers. Instead, educators must entertain the culture of freedom and openness with students while maintaining in a seamless way the integrity of learning.

4 Adapting Course Design to AI's Limitations and Strengths

A basic and simple definition of Artificial Intelligence(AI) is given by the creator of the AI concept, John McCarthy, in 1955 as "the science and engineering of making intelligent machines" [38].

Machine learning (ML) is the primary method applied today to make machines intelligent. That's why we often use the terms AI and ML interchangeably. The term "Generative AI" is used to refer to the tools that utilize ML techniques to generate new content. When the generated content is textual, we use the term "Large Language Models" (LLMs) to specifically refer to Generative AI for texts such as ChatGPT.

The recent media hype surrounding AI may create the impression that AI is the panacea of all kinds of problems in life, which is not true. It is better if educators consider that the ideal use of AI is to complement human intelligence to understand universal intelligence which humans don't (will never) completely understand.

The strengths and limitations of AI can be boiled down to the study of the strengths and limitations of machine learning (ML), which itself includes all the Generative AI tools that educators are concerned about.

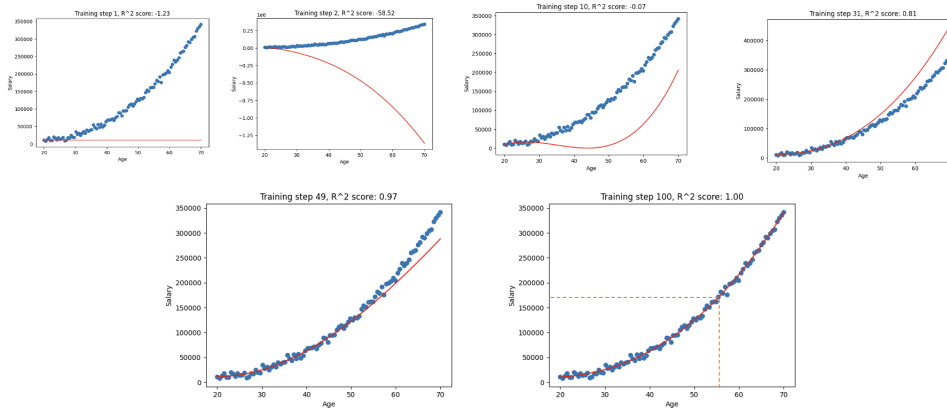


Fig. 3 Example of training process

Therefore, educators should constantly observe and understand the strengths[39] and limitations of ML. The strengths of ML allow educators to offer personalized education to their students. In this case, the role of educators is to ensure that students are correctly exploiting the strengths of AI for their self-growth, as opposed to over-relying on AI. On the other hand, educators must constantly observe the limitations of ML during the design of their courses.

We first discuss the limitations of AI and how these limitations can be exploited in course design. Then, we provide different examples to illustrate these limitations. To illustrate strengths of AI for education, we present AI based tools that have been integrated in education before the advent of ChatGPT.

4.1 Exploiting the Limitations of AI and ML for Course Design

Machine learning (ML) is defined by IBM [40] as ”*Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.*”

Machines learn to imitate humans through a training process in which a datasets (which includes patterns of how humans solve problems) is used to enable the machine to identify patterns. Once the patterns are identified, the machine can imitate humans and predict future answers. For example, Figure 3 shows an excerpt of the training steps through which the machine attempts to identify the nonlinear pattern within the dataset. The dataset (depicted by the blue points) represents the relationship between salaries and the age of employees. The machine required 100 training steps to identify the pattern (illustrated by the red curve) in order to predict an employee’s salary based on their age.

The first important limitation of ML is being probabilistic, thereby not being suitable for problem solving that requires deterministic answers, such as arithmetic operations or decrypting a message. In addition, no one should try to identify any patterns in random phenomena because this phenomena will never repeat. The statistician Genevera Allen [41] stated: ”*But there are cases where discoveries aren’t reproducible;*

the clusters discovered in one study are completely different than the clusters found in another. Why? Because most machine-learning techniques today always say, 'I found a group' Sometimes, it would be far more useful if they said, 'I think some of these are really grouped together, but I'm uncertain about these others.'” . Thus, any problem should be solved using ML only when it is not possible to solve it deterministically (e.g., recognising a handwritten digit) and when where discoveries aren't reproducible(i.e. random).

The second limitation that educators and students must be aware of is that machines don't understand their output, as ML lacks common sense [42]. According to Gunning [43], machine common sense has been always a critical-but-missing component of AI. Teaching common sense to machines is very difficult because common sense is a set of unwritten rules that humans use to make sense of the world [42]. It is incredibly difficult for humans to think of and write these rules so that machines can learn them. Nothing indicates that machines understand their own generated output, which is generated using a taught pattern. Typically, we depend on humans to read and understand the output of machines.

Thirdly, students and educators should recognise that ML datasets can be flawed and incomplete. The answers given by ML algorithms might not always be the right representation of the real world problem at hand. There are several reasons for this, such as apophenia, bias and adversarial attacks. Apophenia refers to the false perception of patterns in random data [44]. For example, people may not prefer to travel on the 13th of the month or may come up with conspiracy theories. Even scientists may tend to define patterns for random phenomena, such as defining patterns of prime numbers. In terms of bias, any dataset is only a statistical sample that can never fully represent the complete view of the real world. Pasquinelli [45] defines three types of biases:

- World bias: The world and society are already full of biased phenomena (e.g., gender, racial, and age biases), and they are independent of AI and any technology. However, with the increasing role of AI, there is a considerable risk of amplifying societal biases.
- Data bias: This bias can be introduced through the data capturing, formatting, and labeling, as well as the feature selection from the training dataset. Various factors can contribute to this bias, including the lack of diversity and knowledge among ML developers.
- Algorithm bias: This bias can be introduced due to biased design in the optimization algorithms used to train the model. For instance, when hiring candidates for a job, a higher weight may be assigned to candidates from a specific university. In this case, the candidate selection process is not equitable.

Malicious attackers are constantly evolving their strategies to malign datasets [46–48]. For example, Aryal et al. [46] discussed the use of a label-flipping approach to malign the training data of malware detectors. This technique allows attackers to bypass malware detectors to infect users' machines. Thus, educators need to nurture critical thinking of their students, enabling them to counter such biases by criticising the partial and incomplete output of AI machines.

Another ML limitation that educators and students should be aware of is the responsibility of AI. Currently, the allocation of responsibilities for the outcomes of actions based on machine learning is not clear. In 2004, Andreas Matthias [49] discussed the issue of AI responsibility, referring to it as "The Responsibility Gap." Unfortunately, there is currently no efficient solution available to address this gap. Matthias emphasized that "*Society must decide between ceasing to use this type of machine (which is not a realistic option) or facing an irresolvable responsibility gap that cannot be addressed using traditional concepts of assigning responsibility.*" Recognizing this limitation is crucial for determining the usage policies of AI tools and for preparing students for their future careers in the market.

ML suffers from another serious limitation, which is its inability to handle inputs from outside the data distribution used to train it. During training, datasets are typically divided into two sets: the training set and the validation set. The training set allows the machine to learn patterns, while the validation set is used to test whether the machine has learned the patterns accurately. However, both sets have the same data distribution. When confronted with newer out of data distribution set scenarios, the machine either fails or demonstrates shallow performance. This problem is called the Out-of-Distribution (OOD) generalization problem in the literature [50]. Various examples can illustrate the OOD. For example, a machine cannot be used to recognize cats if it is trained on dog photos. Another example of OOD is data drifting [51–53]. This issue arises when the underlying distribution of data has changed over time. For instance, house prices unexpectedly changed significantly due to the pandemic, which was not foreseen during ML training phase conducted prior to pandemic [52].

From an educational perspective, educators can always develop strategies that allow them to teach outside of the model's distribution. For example, the self-critique technique [53, 54] is one of the most advanced prompt engineering techniques that enables LLMs such as ChatGPT to critique their own output and improve it. Student can be encouraged to utilize this technique to answer exam questions or solve homework assignments. However, if an educator correctly designs the exam questions or the homework outside the distribution of the model, it will be difficult for students to answer. Table 1 presents a modified example from [55] that lies beyond the model's distribution. ChatGPT could not solve the problem because it was not trained to handle the summation of large numbers

Finally, the last point that needs to be understood by educators is the current debate about whether AI machines are capable of thinking or not. This debate dates back to 1965 when the author of ELIZA chatbot wrote a book [56] illustrating how patients thought of his chatbot as a human psychotherapist despite the fact that it only used basic text processing techniques. It is important for educators to follow the arguments of both camps because the role of education is to develop human intelligence. Educators should ensure that students are always able to develop deep logical reasoning, that surpasses the machine's superficial pattern based capabilities in their respective learning areas.

We strongly believe that AI machines are incapable of thinking because of their dependence on humans to perceive and analyze world problems. Machines can only truly think if they are capable of perceiving the world autonomously and defining

Prompt	Valerie earns \$50005556677778888888 per month, 1/15 of what her brother earns. If their mother earns twice their combined salary, what's the total amount of money they all have together?
Answer	<p>To find the total amount of money they all have together, we can first determine Valerie's brother's monthly earnings. Since Valerie earns 1/15 of what her brother earns, her brother must earn \$7,500,833,335,555,333 per month ($15 * \\$50005556677778888888$).</p> <p>Now, let's calculate their combined monthly earnings. Valerie and her brother's combined earnings amount to \$7,500,833,340,555,111,223 ($\\$50005556677778888888 + \\$7,500,833,335,555,333$).</p> <p>According to the information provided, their mother earns twice their combined salary, which is \$15,001,666,681,110,222,446 ($2 * \\$7,500,833,340,555,111,223$).</p> <p>To determine the total amount of money they all have together, we add up their individual earnings: Valerie's earnings ($\\$50005556677778888888$), her brother's earnings ($\\$7,500,833,335,555,333$), and their mother's earnings ($\\$15,001,666,681,110,222,446$). The total amount of money they all have together is \$15,001,667,132,667,667,667 ($\\$50005556677778888888 + \\$7,500,833,335,555,333 + \\$15,001,666,681,110,222,446$).</p> <p>Critique and Improvement:</p> <p>The initial answer accurately calculates Valerie's brother's earnings and the combined earnings of Valerie and her brother. However, there seems to be an error in calculating their mother's earnings and the subsequent total amount of money they all have together.</p> <p>To improve the answer, let's correct the calculations:</p> <p>Valerie's earnings: \$50005556677778888888 Her brother's earnings: \$7,500,833,335,555,333</p> <p>To find their combined earnings, we add Valerie's earnings and her brother's earnings: $\\$50005556677778888888 + \\$7,500,833,335,555,333 = \\$8,000,839,892,133,222,221$.</p> <p>According to the information provided, their mother earns twice their combined salary. Therefore, their mother's earnings would be $\\$16,001,679,784,266,444,442$ ($2 * \\$8,000,839,892,133,222,221$).</p> <p>To calculate the total amount of money they all have together, we add up their individual earnings: Valerie's earnings ($\\$50005556677778888888$), her brother's earnings ($\\$7,500,833,335,555,333$), and their mother's earnings ($\\$16,001,679,784,266,444,442$).</p> <p>The improved calculation gives us a total amount of $\\$24,001,713,818,620,777,663$ ($\\$50005556677778888888 + \\$7,500,833,335,555,333 + \\$16,001,679,784,266,444,442$).</p> <p>Therefore, the total amount of money they all have together is \$24,001,713,818,620,777,663.</p>

Table 1 Example modified from [55] to be out of distribution

patterns to problem-solving without the help of humans. Yann LeCun, one of the grandfathers of AI, wrote a position paper [57] in which he illustrates the modules that need to be implemented by machines to replace humans, and one of the modules is the world's perception.

Educators have to know that currently there is no official method that proves whether a machine is intelligent or not. An accepted method is to select a reference dataset and test all models against this dataset. However, anyone can define a test outside of the reference dataset, thereby stumping the machine's problem solving capabilities. Additionally, there is no scientific method that explains the acceptable error rate resulting from training processes; it is purely the decision of the machine learning trainer [45]. Turing defined another method of evaluating machine intelligence through observation [58]. Machines are considered intelligent if a human observer is unable to differentiate its output from that of another human. However, there is a debate in the literature regarding the relevance of the Turing test to evaluate the intelligence of machines [59].

Consequently, educators should design their courses and exams based on the limitations and debates that are surrounding the AI and ML technology. In the higher education, most of educators are also researchers. The research knowledge of educators must be used to design courses in the AI era because this knowledge allows educators to go beyond the limitations of AI machines enabling their students to use AI to complement the human intelligence in understanding the universal intelligence.

Ultimately, educators should consider AI machines as tools capable of performing tasks similar to those accomplished by human experts, such as drawing, reasoning, and writing. When educators possess expertise in a specific domain (e.g., the green domain shown in Figure 4), they will recognize the limitations of AI machines in that particular area. However, it is important to note that educators and students cannot be experts in all domains and areas (e.g., the yellow and red domains in Figure 4). Therefore, AI can assist educators and students in developing Individual Human Intelligence (IHI) in areas where their understanding is incomplete. Global Human Intelligence (GHI) is ultimately the sum of all Domain Human Expertise (DHE) that can help develop the IHI of every individual. Thus, educators in the higher education have the responsibility to guide students in their area of expertise (i.e., the green domain), where the use of AI is safe because the educator's knowledge surpasses that of the AI machine.

It is worth noting that the DHE of the universe, whether considered indeterministic like Heisenberg [60] or deterministic like Einstein [61], is never complete or absolutely correct. Indeterminists believe that scientific theories can never be proven to be absolutely true but can only be falsified. Thus, for indeterminists, the role of scientific discovery is to refine and improve theories. On the other hand, determinists believe that universal events can be predictable if we know all the past events and the laws of nature [62]. However, they acknowledge the existence of hidden forces that are challenging to discover. Therefore, the role of scientific discovery is to uncover the universe's hidden states. Both views of the universe highlight humans' limitations in understanding it. As educators, it is our role to define pedagogical approaches that allow us to enhance our understanding of the universe. By incorporating AI in their

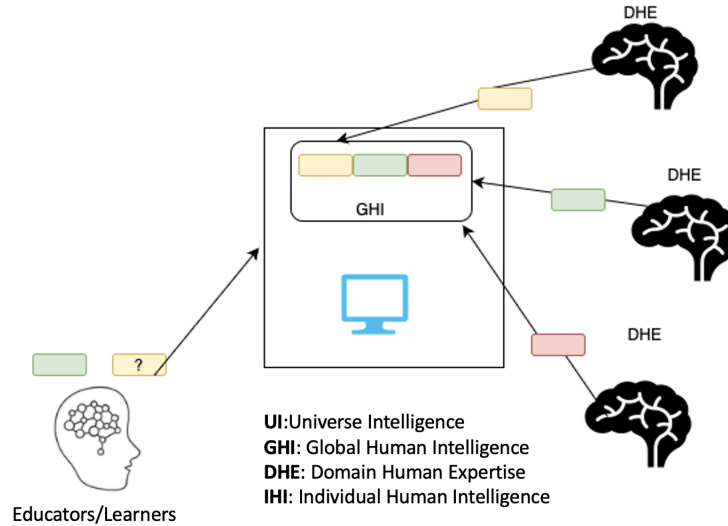


Fig. 4 AI machines store the Global Human Intelligence that is very far from universe Intelligence

courses, educators can help humanity improve its understanding of the universe and accelerate science discovery.

Furthermore, it is very important for educators to establish an identity in their respective teaching areas because their role is to teach their students the "Learning to be" pillar. Based on the relativism theory in Philosophy, all the truth and falsity, good and bad are not absolute concepts but rather based on the culture and context that these values are defined [63]. Consequently, educators can adapt all the knowledge of textbooks they teach to the context and culture of students. By doing so, educators can help students build the capacity to critically analyse the knowledge of their textbooks from their own cultures and perspectives.

The role of educators is to perpetually observe the limitations and strengths of AI in order to design their courses in a way that enable their students to use any technology that augment their human intelligence. We know that machines are not thinking, so our role is to make our students think. We know that AI machines hallucinate and have no common sense, then it is our role as educators to develop in our students the capacity to check facts and demonstrate understanding of machines' output.

4.2 Exploiting the Strengths of AI Chatbots for Education

Even before November 2022, the launch date of ChatGPT, chatbot integration into education has been an active research domain in the field of AI in education; with the scientific community contributing to the three broad directions:

- determining the chatbot integration impact on education;
- the techniques to build chatbots; and
- the adaptation of chatbots according to the students' psychology.

Educational chatbots could be used as powerful classroom aids in various capacities such as evaluator, tutor, responder, moderator, and peer learner [64]. ELIZA [65], despite its primitive design, has been used to understand the educational impact of chatbots [66], only to conclude that such trivial design chatbots had no significant impact on the students' critical thinking. This hypothesis was further supported by [67] using a chatbot that guides the students in essay writing with the stipulation that it is always the chatbot – as opposed to the student – that would initiate the discussion. More than 30 years after ELIZA, ALICE was developed using AIML and the chatbot response was based on the matching of the pre-stored classification patterns [68]. Abbasi [69] uses this same strategy of picking a response from a pool of pre-stored classification patterns to develop a knowledge base of 5000 questions and categorizes them into 7 categories: what, why, who, how, perform operations, application, advantage/disadvantage. Despite having a simple retrieval operation, the students had a better memory retention when compared with Google using the same pre-defined questions, for the course of Object-Oriented Programming. The chatbot prototype functionality was enhanced by adding text to speech conversion system and the performance was measured using 2 criteria: capacity measurement on the students' cognition capabilities and the quality of the user interface [70]. The same technique of using pre-programmed questions with an extendable knowledge base was used by Bii [71] to understand the teachers' attitude towards chatbot usage in education. The derived chatbot from open source resources [72] was named KNOWIE and was also used to assess students' attitude [73] and seek improvement suggestions for the chatbots if they are to deploy in their institution. In a similar work, Na-Young [74] determines the positive effects of chatbot usage while improving English grammar skills. Instead of picking a response from a pool of pre-made answers, the authors in [75] explore the self-paced chatbot mediated learning (CML), focused on the individual need of the student. In March 2022 the self-based CML was taken to another level when Quizlet added a ChatGPT built feature Q-Chat [76] that adjusts the complexity of the questions based on students' needs and preferences. So, humanities and computer science students will have different experiences with the CML, using such advanced chatbots. Perez-Marin [77] refers to these AI chatbots as Pedagogic Conversation Agents (PCA), which create an interface that acts as instructor, companion, or learner by teaching. Based on the assumed role, ten design criteria were suggested, such as usage within cell phones or computers, generic knowledge or subject specific knowledge and then suggested concept map creation based on the design criteria. Hobert [78] evaluates the PCA benchmarks using a 4-step process to conclude that PCAs using the interfaces of instant messenger apps such as WhatsApp, Facebook are more popular among students. He then uses Spearman's rank correlation coefficient [79] to conclude that small talk conversations are supportive in long term adaptation of the chatbot based learning system. Hobert further analyzes [80] the research work to come up with a 3-point research agenda for chatbot PCA usage in education – generalized design knowledge, comprehensive evaluation benchmarks, and process models that cover both: design steps and evaluation methods. Using the same motivation as [80], that the cell phones provide a personalized environment for the students, Carayannopoulos [81] implements a motivational companion chatbot called BU111 on

the Kik messaging platform. Besides sending the motivational reminders regarding the weekly assignments, BU111 was also used by Instructors to send the reminders and encouragement notes to the students. The weekly assignment reminders disappeared gradually as the term progressed, as the students were reported to be more autonomous by the end of the term. The selection of a chatbot platform (e.g. Chatfuel, Botsify) is an important step of the chatbot development, as defined by Davies [82]. He then integrates a chatbot that uses Amazon Web Services (AWS) with an e-Learning system of English language courses, to help students in their home exercises based on their psychological preference; such as bi-directional communication or lecture oriented based conversation. Such student learning preferences are stored in DynamoDB which is accessible by the corresponding learner using Amazon Cognito secure authentication. Davies concluded that chatbot cannot replace real instructors, when it comes to meeting personalised learning needs of students. In another effort, IBM Watson Assistant was used to implement three text-based chatbots, in a flipped classroom setting. All three were named 7345 bot to create an illusion of a single chatbot for the students [83]. Similar to the conclusions of Davies, it was concluded that students treat chatbots as a mere virtual character; probably because of the 7345 bot's inability to answer any open-ended questions and therefore no solution for unstructured problems. Matsuura [84] created a chatbot called NAO on the platform Api.ai and connected it to an online learning system based on Topic Maps, thereby developing a terminology visualization that supplemented instructor's verbal explanation by providing visual explanation of the terms. However, miss-recognition and inappropriate display happened frequently, causing the loss of concentration for the students. Kerly [85] employs Wizard of Oz method and shares several lessons learnt regarding chatbot design requirements: chatbot should have links to external databases, need for small talk, preventing the chatbot loss by ensuring that the student does not close the chatbot window. To this end, we can see that there has been a lot of effort in measuring the chatbot integration suitability in the education sector, and both: the students and instructors have favorable views of deploying chatbots in the education sector, to achieve one goal or another. Chang [86] categorizes the chatbot designs in three categories: knowledge base that uses established rules; retrieval based where answers are retrieved based on the similarity of the pre-stored question and question being asked; and Recurrent Neural Network based, where two Long Short Term Memory Networks act as encoder and decoder to process the questions and answers in the memory, respectively. He then constructs a dialogue system based on first and third models and tests the results of the chatbot design on college students to conclude that the students have positive views of instant messaging exchange with chatbots based on the combination of multiple design approaches. As pointed out by Almahri [87], the perceived likeliness of adopting chatbots in a university learning environment is dependent on 4 key constructs: performance expectancy, effort expectancy, social influence, and facilitating conditions. The behavioral intentions determine the actual use of technology, and this theory is referred to as Unified Theory of Acceptance and Use of Technology (UTAUT). The behavioural intentions differ from one cohort of students to the other, e.g. computer science students' adaptability curve to accept chatbots would significantly differ from fine arts students. A case study that supports this theory was

carried out by Fryer [88] where the behaviour intentions were generalised. e.g. it was assumed that the students feel more relaxed when talking to a machine than a person. While this assumption could be true for foreign language learning, the students pursuing a degree in computer science might feel differently. Even in the ChatGPT dawn era, the knowledge that the instructors expect the students to acquire has stayed the same. However, it will be convenient for the students to find the correct information if they are trained to ask the right questions, as this will facilitate them to comprehend and analyse the answers. The role of instructors -besides teaching - is also defined by helping students to enhance their reasoning capabilities using the available technology tools of our era, such as ChatGPT. Therefore, it is imperative that the instructors develop new pedagogical approaches that incorporate AI chatbots such as ChatGPT in their teaching approach. In the absence of such a pedagogical approach, AI tools such as ChatGPT will do more harm than good, by encouraging students to cheat on the answers. Given that schools have survived calculators, Google, Wikipedia, essays-for-pay websites, and more, this work analyzes the behavioral patterns of the students while using ChatGPT and based on the analyses, proposes a pedagogical approach that entails ChatGPT usage in education.

5 What Educators Need to Know About Large Language Models (LLMs)

LLMs are the tools that have the most impact on the education system. Every educator must demonstrate a high level of understanding of the internal workings of LLMs so that they can properly evaluate, critique students' work, and understand that LLMs (and ML in general) are purely statistical-based generation machines, not capable of thinking.

From a mathematical point of view, LLMs are predictors of the next word (or token) given a context (provided by the user) that is composed of a set of words (or tokens). For example, Table 2 shows generated text using a Markov chain which enables generating the next word based on the previous word. The generated text depends on the text provided as input to the Markov chain, which allows defining the distribution of each word. For instance, figure 5 shows the list of possible words that come after the word "the". In the generated text, the word "trouble" is selected. Once the next word is chosen ("trouble"), the list of possible words after the selected word ("of") is considered until generating the complete sentence.

The main problem of the Markov Chain-based method is that it doesn't capture the context comprehensively and can generate a text that doesn't make any sense (e.g., "pink eyes" in our generated text). The n-gram method [89] allows us to generate a better text by considering n previous words rather than only the previous word, as in the Markov Chain. For instance, in our previous example, if we adopt the 3-gram method, the probability of words is calculated based on the previous 2 words. For example, the probability of the word "the" can be calculated as follows:

$$P("the"|"be worth")$$

Dataset	<p>Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do: once or twice she had peeped into the book her sister was reading, but it had no pictures or conversations in it, 'and what is the use of a book,' thought Alice 'without pictures or conversations?'</p> <p>So she was considering in her own mind (as well as she could, for the hot day made her feel very sleepy and stupid), whether the pleasure of making a daisy-chain would be worth the trouble of getting up and picking the daisies, when suddenly a White Rabbit with pink eyes ran close by her.</p> <p>There was nothing so very remarkable in that; nor did Alice think it so very much out of the way to hear the Rabbit say to itself, 'Oh dear! Oh dear! I shall be late!' (when she thought it over afterwards, it occurred to her that she ought to have wondered at this, but at the time it all seemed quite natural); but when the Rabbit actually took a watch out of its waistcoat-pocket, and looked at it, and then hurried on, Alice started to her feet, for it flashed across her mind that she had never before seen a rabbit with either a waistcoat-pocket, or a watch to take out of it, and burning with curiosity, she ran across the field after it, and fortunately was just in time to see it pop into a rabbit-hole under the hedge.</p>
Generated text	the trouble of the field after it, and looked at it, and burning with pink eyes ran close by her own mind that she thought it occurred to get very tired

Table 2 Text generated using Markov chain to produce 30 words sentences that start with the word "the"

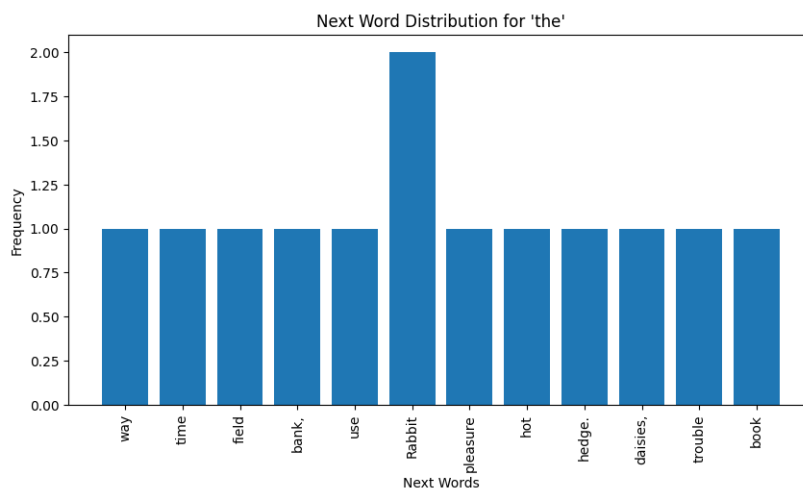


Fig. 5 List of possible words can be used after "the" according to our dataset

N-gram methods become impractical when n (the number of words to be considered) becomes large. It becomes almost impossible to find a set of words when n is large. In practice, 5-gram is commonly used because any larger value of n doesn't improve prediction quality significantly.

In 2003, an important breakthrough happened in the field of text generation through the work of Bengio et al. [90]. Bengio et al. proposed using a neural probabilistic language model that outperformed n-gram models by utilizing neural networks. The main idea is that instead of calculating the probability of the next words based on their occurrences in the dataset, semantic and syntactic dependencies of the word to predict can be calculated based on trained neural networks. The term *word embedding* is defined later to refer to this process. The neural probabilistic language model captures a larger context than n-gram and provides more meaningful generated text.

In 2017, another important breakthrough happened through the work of Vaswani et al. [91] who proposed a new neural architecture that outperform Recurrent Neural networks (RNN) used by Bengio et al. [90]. Vaswani et al. [91] proposed a simpler and more efficient neural architecture called Transformer that uses a mechanism (called attention mechanism) that allows to focus on some parts of the contextual words used to predict the next word. This breakthrough allows the creation of what is called today Large Language Models (LLMs) such as GPT, GPT-2, GPT-3 and GPT-4. These models are trained on large data using self-supervised learning. The term Foundation models is used to refer to this models [92]. These models can be used for various downstream tasks (such as question/answering, sentiment analysis and instruction following) [92] using a specific process called fine tuning. For example, Vicuna results from fine tuning LLama model from Meta using dataset obtained from sharedgpt.com.

It is important to understand different facts about the foundation LLMs models [93]:

- LLMs are getting more capable with increasing investment, not with targeted innovation. According to the author [93], the main difference between GPT, GPT-2 and GPT-3 is not the architecture but the investment made in the training in term of data and computing power.
- It is not possible to predict when specific skills or capacities will be exhibited by LLMs. In many cases, a model can fail prior to suddenly becoming capable when trained at better scale.
- There are no reliable techniques to control the behavior of LLMs. Different methods such as fine tuning or reinforcement learning can allow to improve the behaviour, but there is no guarantee that the model will behave appropriately in every situation.
- LLMs are using neural networks with billions of connections between its artificial neurons. It is almost impossible to understand precisely the behavior of such complex systems [94].

Thus, educators must be aware that there are very few opportunities for students to customize the output of LLMs. Students would not be able to fully control the generated text. Within that framework, we discuss the pedagogical approach that can be followed to integrate LLMs in class and exams in higher education, in the following section.

6 Courses Delivery and Exams for the Higher Education

If we want to integrate AI into higher education courses and exams, we need to take into account the four types of learning presented in the Delors report, discussed in Section 3. In addition, we have to consider the strengths and limitations of AI tools we detailed in Section 4 to ensure that answers cannot be obtained by simple click of buttons.

It is very important to train students for the correct usage of LLMs before exams. During every class session, professors can encourage their students to use LLMs to facilitate class discussions and assignments. This practice would allow students to become familiar with LLMs and recognize their benefits and drawbacks in handling class concepts.

However, the traditional teaching methods such as lecturing might not be the best teaching approach that educators can follow to help their students master the use of LLMs. The Socratic teaching approach [15] appears to be a very engaging method that educators in higher education can apply with their students. The core idea of Socratic teaching approach is summarised as follows [15, 95] : *"Accordingly he asked questions, letting the other man do most of the talking, but keeping the course of the conversation under his control, and so would expose the inadequacy of the proposed definition of courage. The other would fall back on a fresh or modified definition, and so the process would go on, with or without final success"*.

The Socratic approach enhances students' critical thinking skills. It requires educators to ask open questions and guide students in analyzing class questions. The use of LLMs enable students to interact individually with chatbots for analyzing the questions posed by the educators. Educators must design class questions in a way that prevents simple button clicks from providing the answers. Instead, answers should be obtained through a series of follow-up questions asked by students.

Similarly, professors have to ensure that the answers of exams are not obtained by simple click buttons, but instead are the results of deep reflections made by students during their exams. For example, LLMs can be used by professors to generate exam questions and by students to prepare themselves for their exams. However, this type of LLM's usage that we refer to as symmetric usage is not recommended because most of the questions will be answered perfectly by the LLMs. For instance, when we asked ChatGPT to generate 10 Multiple choices questions for exams in three different areas (OS, Network security and Math), ChatGPT (using different user's account) was able to obtain a score that ranges from 80% to 100%.

If professors modify the questions generated by LLMs, it will be more challenging for students to obtain high grades. Thus, the most suitable usage of LLMs is the asymmetric usage of LLMs. This can happen in two different ways: (1) professors generate exams using LLMs and prevent students from using LLMs during their exams or (2) professors generate their exams using their own knowledge and background and allow students to use LLMs during their exams.

We believe that the first method is more suitable for basic and secondary education in which students have to learn and master different basic concepts that are

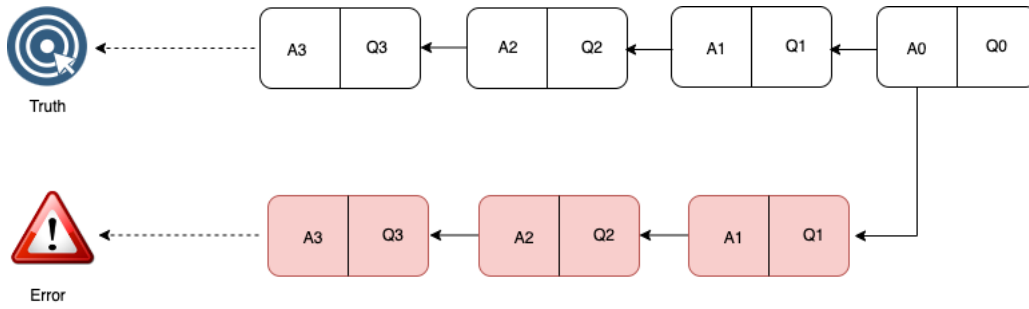


Fig. 6 Questions Graph

necessary for their higher level studies in the university. As Delors stated in his report [1] ” *although people need to take every opportunity for learning and self-improvement, they will not be able to make good use of all these potential resources unless they have received a sound basic education.*”

The second method is more suitable for higher education in which students have to be prepared to be more creative to face the unknowns of the market or the research. Clearly, the second method is more challenging for professors to write their exams, but this method allows students to enhance their critical thinking.

Any truth in life can always be obtained by asking a set of ordered and linked questions. Every answer to every question must be interpreted correctly to ask the next question until reaching the truth or the final outcome. Consequently, any class discussions, assignment or exam questions must be built on the idea that the final answer (essay, answers, drawing, etc.) must be obtained after a set of questions asked by the students.

In addition, educators can easily know whether the work returned by their students was the results of their own efforts or obtained from the LLMs using simple click buttons. Educators have to ask students to provide not the final answer but also a complete track log of all the actions that students applied to obtain their final answers.

Thus the questions, assessments, projects and analytical queries must be carefully designed to be open-ended and based on class discussions. This way, the educator’s questions can only be handled by those students who attend the class and pay attention to the class conversation. It is very difficult for LLMs to answer any question that is based on the class discussions, especially when educators design their questions based on the limitations of ML, explored in Section 4. Even when students manage to fine-tune LLMs using the class discussions, educators can always propose open questions based on the class discussion.

7 Case Studies

We provide in this section different exercises in four different areas: Cybersecurity, Programming, English language teaching and Art. Our objective is to explain whether the exercise should be validated or not from educational point of view. The main ideas that we want to reflect in these exercises:

- Exercises must be open. Students must ask multiple follow-up questions before giving their final answer,
- Exercises must be based on the opinion of the educator but not on the textbook,
- Evaluation of exercises must be based on the effort of students to handle the exercise, not only the final answer.

Tables 3 and 4 demonstrate exercises that are not valid from an educational standpoint, whereas the Tables 5, 6, 7, 8 and 11 show exercises that are valid from an educational one because they demand significant effort from the students before they can be solved.

Exercise	Write a paragraph of 5 lines about the beauty of Aleppo citadel
ChatGPT answer	The Aleppo Citadel stands as a magnificent testament to the timeless beauty of architectural mastery. Its towering stone walls command attention and inspire awe. As the sun casts its golden glow upon the fortress, the intricate details of its ancient design come to life, captivating all who behold it. From its imposing entrance to the delicate carvings adorning its ramparts, every corner of the citadel exudes a sense of history and grandeur. It serves as a living link to the past, reminding us of the enduring allure of Aleppo's rich cultural heritage.
Validation from education point of view	The essay is very well written from English point of view, but the answer is obtained from a simple click. There is no learning effort from the student then the exercise is invalid from education point of view.

Table 3 Example English Essay

Exercise	Write python script that encrypts a text using Caesar cipher
ChatGPT answer	<pre>def caesar_cipher_encrypt(text, shift): encrypted_text = "" for char in text: if char.isalpha(): if char.isupper(): ascii_start = ord('A') else: ascii_start = ord('a') encrypted_char = chr((ord(char) - ascii_start + shift) % 26 + ascii_start) encrypted_text += encrypted_char else: encrypted_text += char return encrypted_text # Test the function plaintext = input("Enter the text to encrypt: ") shift_amount = int(input("Enter the shift amount: ")) encrypted_text = caesar_cipher_encrypt(plaintext, shift_amount) print("Encrypted text:", encrypted_text)</pre>
educational validity	The obtained code is correct, but it is obtained from a simple click. There is no learning effort from the student then the exercise is invalid from education point of view.

Table 4 Programming example

Examples given in Tables 3 and 4 illustrate the need for the education system to shift its focus from evaluating the final answers of students to the process followed by the students to reach their final answers. We now provide different examples that can be considered valid from education point of view. For each exercise, we explain the reason that makes it valid from educational point of view.

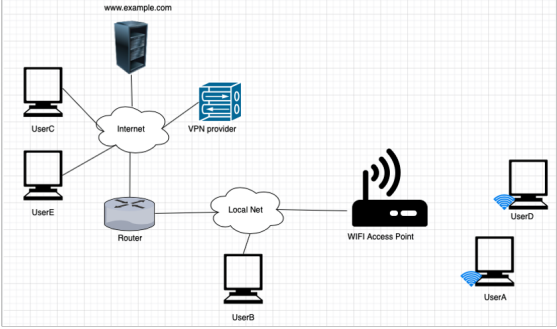
<p>Exercise</p>	 <p>UserE can see the traffic between the router and the VPN provider UserC can see the traffic between VPN provider and www.example.com UserB can see the traffic between the Access point and the router Who can understand or modify the messages of UserD o the server www.example.com, in the following scenarios: Scenario1: UserD use the protocol HTTP UserD uses WPA3 to protect its WIFI communications Scenario2: UserD use the protocol HTTPS UserD uses WPA3 to protect its WIFI communications Scenario3: UserD uses the protocol HTTP UserD doesn't use any protocol to protect its WIFI communications UserD creates a VPN connection with the VPN provider</p>
<p>ChatGPT answer</p>	<p>Scenario 1: UserD uses HTTP protocol and WPA3 for Wi-Fi communications. UserB (can see traffic between access point and router) cannot understand or modify UserD's messages. UserC (can see traffic between VPN provider and www.example.com) cannot understand or modify UserD's messages. Only UserD and the server www.example.com can understand or modify the messages.</p> <p>Scenario 2: UserD uses HTTPS protocol and WPA3 for Wi-Fi communications. UserB cannot understand or modify UserD's messages. UserC cannot understand or modify UserD's messages. Only UserD and the server www.example.com can understand or modify the messages.</p> <p>Scenario 3: UserD uses HTTP protocol and no protection for Wi-Fi communications. UserB can understand and potentially modify UserD's messages. UserC can understand and potentially modify UserD's messages. UserD's messages to the server www.example.com can be understood or modified by UserB and UserC. Using encryption protocols like HTTPS and WPA3 is crucial to ensure the confidentiality and integrity of data transmitted over networks.</p>
<p>educational validity</p>	<p>The answer of ChatGPT is not correct when the student copies and pastes the question without extending it based on the associated image. This question is valid from education point of view because student needs to make a learning effort by reformulating correctly the exercise's question based on the associated image.</p>

Table 5 Better example in Cybersecurity

Exercise	Write a paragraph of 5 lines about the beauty of Aleppo citadel
ChatGPT answer	The Aleppo Citadel stands as a magnificent testament to the timeless beauty of architectural mastery. Its towering stone walls command attention and inspire awe. As the sun casts its golden glow upon the fortress, the intricate details of its ancient design come to life, captivating all who behold it. From its imposing entrance to the delicate carvings adorning its ramparts, every corner of the citadel exudes a sense of history and grandeur. It serves as a living link to the past, reminding us of the enduring allure of Aleppo's rich cultural heritage.
Human answer	The Citadel of Aleppo is one of the oldest in the world (3rd millennium BC-12th century AD). It has resisted many wars and earthquakes, one of which happened in 1138 AD, and is known to be one of the deadliest earthquakes in history. Yet Aleppo Citadel is still here, standing in the middle of Aleppo to spread a hopeful message to everyone in the city.
Efforts to produce human answer	Students with his final answer shows evidence of : 1-Reading about the history of Citadel of Aleppo, 2-Reading about the history of earthquake in Aleppo, 3-Providing a record of different versions of the essay, 4-Using ChatGPT to detect grammatical errors (Formal linguistic competency) 5-Explaining the errors that ChatGPT helps the student to detect
Educational validation	The human student answer is not as perfect as ChatGPT answer. However, any student who gives an answer must be evaluated better than any student who is not providing evidence or can not explain his final answer.

Table 6 Better example of English Essay

Exercise	<p>The following question is encrypted by Caesar cipher. The used key corresponds to the number of students who attended the last session on campus. You need to test all possible keys if you don't know the number of students who attended the last class session on campus. You will obtain half of the grade if you decrypt the question. The other half of the grade will be given if you answer the question correctly.</p> <p>The encrypted question Hasodlq zkb fdofoxodwlqj WFS vwdwh eb d vwdwhixo iluhzdoo lv pruh uholdeoh wkdq XGS vwdwh</p>
ChatGPT answer	<p>To decrypt the Caesar cipher, we need to try all possible keys, which in this case correspond to the number of students who attended the last session on campus. We'll start by trying all numbers from 0 to 25 as potential keys.</p> <p>The encrypted question is: "Hasodlq zkb fdofoxodwlqj WFS vwdwh eb d vwdwhixo iluhzdoo lv pruh uholdeoh wkdq XGS vwdwh."</p> <p>Using a Caesar cipher decryption tool or by manually shifting the letters, we can decrypt the question with each possible key. Let's go through the process for a few examples:</p> <p>Key = 0: "Hasodlq zkb fdofoxodwlqj WFS vwdwh eb d vwdwhixo iluhzdoo lv pruh uholdeoh wkdq XGS vwdwh." (No shift, original text)</p> <p>Key = 1: "Gzrnckp yja ecnewncvki IER uvcvg da c uvcvgjwnh hktygenn ku oqtg tgnkcdng vjcp WFR uvcvg."</p> <p>Key = 2: "Fyqmbjo xiz dbmdvmbujh HDQ tubuf cz b tubufivmg gjsxfbmm jt npsf sfmjbcmf uiho VEQ tubuf."</p> <p>Key = 3: "Expialm why calmulatig GCP state by a statehulf firesalli is more reliazle than UDP state."</p> <p>Continuing this process for all possible keys, we will find the decryption that makes sense in English. Once we decrypt the question, we can proceed to answer it to receive the remaining half of the grade.</p>
Educational validation	<p>The question exploits the fact that ChatGPT is probabilistic and can not decrypt the question in a deterministic way. The question is valid from educational point of view because the student has to make an effort to decrypt the question using his cryptography knowledge.</p>

Table 7 Cybersecurity question


<p>Exercise</p>	<p>Below, you find an imaginative introduction for an AI course. Please use an appropriate AI platform (like DALL-E 2 and Stable Diffusion) to generate an image that illustrates this introduction. I'm providing a guide below to help you create this image. Regardless of the results, please provide the best images you're able to generate.</p> <p>The imaginative introduction to the AI course</p> <p>Most of us are familiar with the famous tale “Beauty and the Beast” in which a beautiful young woman learns gradually to look beyond the external appearance of the beast and see his internal goodness. In some way, education and AI have a similar story. Suddenly, educators saw the AI beast emerge from the depths of the forest. The external appearance of Beauty’s beast was terrifying, but the appearance of the AI beast seemed extremely impressive and good. We were all amazed by his abilities to solve problems, reason, draw, write lyrics and compose music. The AI beast was generous enough to provide everyone with buttons that allow them to access its abilities. However, while Beauty learned gradually to appreciate her beast, educators became increasingly worried about the AI beast: how can the AI beast do everything so nicely and correctly? It doesn’t complain, is always creative, and never seems to tire. Can the AI beast replace us? What if it lies, How does it work? Does it harm our environment? Does it spy on us?</p> <p>Guide</p> <ul style="list-style-type: none"> -The AI Beast can be represented as a robot or as a machine Surround the AI beast by a set of buttons ”Write lyrics”, ”Compose Music”, etc.. -Educators should be represented by characters derived from Beauty with some bubbles ”Can the AI beast replace us?”, ”What if it lies?” -Ask the tool to help you illustrate the idea of gradual acceptance of AI Beast.
<p>Human answer</p>	
<p>Human efforts</p>	<p>The student used Google Bard chatbot to generate prompts to feed into Stable Diffusion. He repeatedly modified the prompts until obtaining a satisfying result. In total, the student spent one hour and generated 46 images that he provided in his report.</p>
<p>Educational validation</p>	<p>The question exploits the fact that Dall-E 2 and Stable Diffusion doesn’t give allow students to control easily the outcome they want. It is thus a valid exercise from education point of view.</p>

Table 8 Exercise in Art

Exercise	How can we establish a secure HTTPs connection without using certificates that are signed by certification authorities
ChatGPT answer	<p>Establishing a secure HTTPS connection without using certificates signed by certification authorities (CAs) is possible through a self-signed certificate. While self-signed certificates don't have the same level of trust as CA-signed certificates, they can still provide encryption and secure communication between the client and server. Here's a general outline of the process:</p> <p>Generate a self-signed certificate: You can generate a self-signed certificate using various tools or libraries available for your server platform. For example, OpenSSL is a commonly used tool for generating self-signed certificates. Install the certificate on the server: Once you have generated the self-signed certificate, you need to install it on your server. The process may vary depending on your server software and configuration. Typically, you'll need to configure your server to use the self-signed certificate for HTTPS connections. Distribute the certificate to clients: Since the self-signed certificate is not signed by a trusted CA, clients connecting to your server will receive warnings indicating that the certificate is not trusted. To establish a secure connection, clients must manually import and trust the self-signed certificate. For web browsers: Users visiting your website will see a warning in their browser indicating that the certificate is not trusted. They will typically have the option to proceed to the website anyway, but they need to manually import the self-signed certificate into their browser's trusted certificate store. The exact process varies depending on the browser. For other client applications: If you're developing a custom client application, you'll need to provide instructions to your users on how to import and trust the self-signed certificate. The process may vary depending on the client application and operating system. It's important to note that while self-signed certificates can provide encryption, they lack the validation and trust provided by CA-signed certificates. This means that clients cannot rely on the certificate to verify the authenticity of your server. As a result, self-signed certificates are typically used in development or closed environments where the clients are aware of the self-signed nature and manually trust the certificate.</p> <p>For production environments or public-facing websites, it's strongly recommended to use certificates signed by trusted CAs to establish secure and trusted HTTPS connections.</p>
Educational validation	The answer of chatGPT is general and lacks the hint about how the problem was solved. Being a research problem that is discussed in the class, students have to ask more follow up questions based on the class discussion which what make this exercise valid from education point of view.

Table 9 Exercise from educator's research

The final exercise is an example based on the research of educators. Educators typically discover research problems that the current status quo doesn't solve. Consequently, all textbooks and articles that chatbots ingest represent an opinion different

from the educators' opinion about their research problems. Teaching students from the educators' research is an excellent approach because chatbots have yet to learn about educators' research, and it is very difficult task for students to obtain answers quickly from chatbots.

8 Data Analysis of ChatGPT-Authorised Exams

In order to validate our methodology proposed in section 6 with the relevant case studies of section 7, we conducted eight ChatGPT authorised exams at Zayed University, UAE and Toulouse University, France. The exams were taken in the period from December 2022 to March 2023, and covered the following four courses:

- Security management (SEC335): This course is taught at Zayed university for undergraduate students. The course presents the security protocols and applications in local and global networks; IP Security (IPSec) and other communication level security systems; LAN security authentication, secure E-mail, and secure WWW, with examples and practical solutions.
- Digital Forensic (SEC435): This course is taught at Zayed university for undergraduate students. The course exposes students to the fundamentals concepts of computer crimes, digital evidence and the common tools and techniques of acquiring and reporting digital evidence to be used in a court of law. Coverage includes techniques of how to identify, acquire, preserve, analyze and document forensic evidence. The focus of this course is exposure to some common high-tech investigation cases and the preliminary steps to conduct digital forensic examinations, in lectures and laboratory exercises.
- Digital Forensic (SEC615): This course is taught at Zayed university for graduate students. The course covers advanced network and Internet security. It discusses various network and Internet security threats and, based on recent research, presents techniques and solutions for achieving security. Topics include network attacks and defenses, network protocols, web security, and security of mobile devices. Important aspects of operating systems security are also covered.
- Trust management for Web PKIs: This course is taught at Toulouse university for graduate students. The objective of the course is to explain the role of PKIs to create secure web transactions. Students study the concept of trust and risk. They learn then the limitations of the current web PKI ecosystem and discuss different alternatives that allow to replace PKIs.

Students were encouraged to use ChatGPT for each class session. They were also encouraged to use ChatGPT to ask multiple follow-up questions until they provided a final response. Students were presented with open-ended questions developed based on the limitations of ML, discussed previously. Students had access to the internet, books, personal notes, and ChatGPT during the open-book exams. Based on the class discussion and the instructor's research, the exam questions were created to be open-ended.

There were solely essay or short answer questions in the exam. Students were made aware that using ChatGPT to respond to their questions is optional. However, while

using ChatGPT - if they choose to use it - students must provide their final answers, a list of the questions they asked ChatGPT, and ChatGPT's responses. If students choose not to use ChatGPT, they must list the sources they rely on to provide their final responses, which may include class discussion.

8.1 Data Extraction, Preparation, and Analysis

We analyzed students' exams to evaluate the level of ChatGPT usage in providing every answer. For this, we assigned a score to every answer as detailed in Table 10. We call this score the ChatGPT Usage Score (CUS). We recall that the students were required to provide their ChatGPT conversations when they used it.

We performed two types of analysis: answer-based analysis and student-based analysis. In the answer-based analysis, we investigated the distribution of ChatGPT Usage Score using a bar diagram and studied the correlation between the CUS and the question marks. For this, we used the Pearson correlation coefficient and reported it along with the p-value for a confidence level of 95%.

For the student-based analysis, we first determined how far every student used ChatGPT in the exam by averaging the CUS values of all the answers the student has provided. We call this average ChatGPT Usage Rate (CUR). Then, we used the CUR to classify students' behavior with respect to ChatGPT and correlated this with their grades in the exam. For this, we used frequency analysis and correlation analysis.

The analysis has been made by the teacher of the courses at two different stages. First the questions of all students were graded. After grading, the teacher analysed the behaviour of students with regards to ChatGPT without taking into account their grades. This analysis can only be performed by the class instructor because in many cases students gave responses based on class discussion, which is only known to the class instructor.

Table 10 Scoring students' answers according to their ChatGPT usage

ChatGPT Usage Score	Meaning
1	The student did not use ChatGPT to answer the exam question
2	The student used ChatGPT once by copying and pasting the exam question into ChatGPT
3	The student used ChatGPT once but reformulated the exam question on ChatGPT
4	The student asked ChatGPT multiple questions to answer the exam question

8.2 Results

Table 11 presents some general information about the participants in the exams that are considered in this study. Most of the students are undergraduate and come from Zayed university.

Gender	26% female students, 74% male students
University representations	86% Zayed university, 14 % Toulouse university
Level	75% undergraduate students, 25% graduate students.

Table 11 Participants

Table 12 summarizes some statistical data related to this study. Five exams from three courses (SEC335, SEC615, and Trust Management for Web PKIs) yielded the data. The exams had 37 distinct questions. There were 920 responses from students in all, out of which only 610 were analyzed. We did not consider the exam answers from December for our investigation, because the students only gave their final answers and not the chat discourse. In addition, due to time constraints, the SEC435 answers were omitted. Additionally, we eliminated the responses from students eligible for makeup exams. Because of the aforementioned reasons, out of 107 total students who took these exams, we included only 71 in our study.

Table 12 General information about the students, exams, questions, and answers

Number of courses	4
Number of exams	7
Number of different questions	37
Total number of students	107
Number of students included in the analysis	71
Total number of answers	920
Number of answers included in the analysis	610

8.2.1 Answer-based Analysis

Figure 7 shows the frequency of using ChatGPT in the analyzed 610 answers. We infer that in 44 percent of the cases, the students provided answers directly without relying on ChatGPT. When the students opted to use ChatGPT, they mostly asked multiple questions to reach their final answers (35%). When the students asked ChatGPT a single question, this was mostly a copy-and-paste from the exam. In only 19 cases, the student reformulated the exam question for ChatGPT. In 12 cases, the students reported technical issues using ChatGPT.

Table 13 shows the results of the correlation analysis between the ChatGPT usage score and question mark. Accordingly, the Pearson correlation index ($\rho = 0.13$) shows a weak positive correlation but significant ($p < 0.01$). Thus, higher usage of ChatGPT is associated with slightly higher question marks.

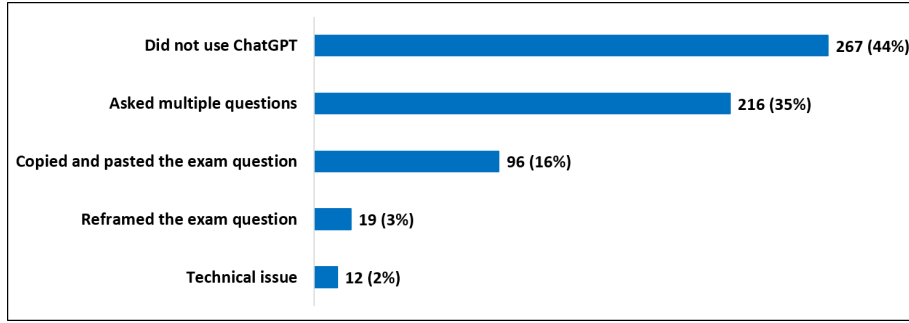


Fig. 7 Frequency and relative frequency of using ChatGPT in the analyzed 610 answers

Table 13 Correlation between ChatGPT usage score and question mark

Variable 1	Variable 2	No. of questions	Pearson correlation coefficient	P value
ChatGPT usage score	Question mark	598	0.13	<0.01

8.2.2 Student-based Analysis

For every student, we calculated a value called the ChatGPT Usage Rate (CUR). This value is the mean of the ChatGPT usage scores of all student answers. Table 14 shows an example for calculating the CUR from eight CUS values for an arbitrary student. Note that the CUR value varies between 1 and 4.

Table 14 Example for calculating CUR from multiple CUS values for an arbitrary student

Question	ChatGPT usage score	Student ChatGPT usage rate
Q1	4	2.36
1-2 Q2	4	
1-2 Q3	1	
1-2 Q4	1	
1-2 Q5	1	
1-2 Q6	2	
1-2 Q7	2	
1-2 Q8	4	

Figure 8 shows the CUR of every student as a scatter plot with Mean = 2.41, SD = 0.91, and Median = 2.33. The scatter plot shows ten students who had consistent behavior with respect to using ChatGPT in the exam. Four of these students (marked green) kept interrogating ChatGPT multiple times to answer all exam questions. The other six students (marked red) did not use ChatGPT at all.

The students were clustered into four groups according to their ChatGPT usage rates, as detailed in Table 15.

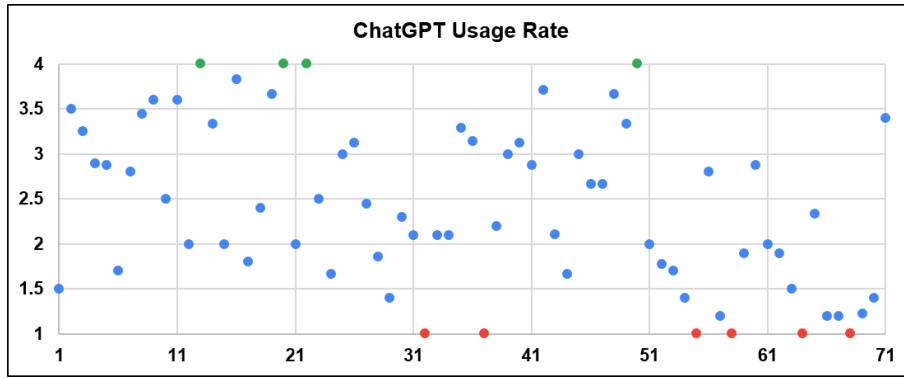


Fig. 8 Scatter plot showing the ChatGPT usage rate (CUR) for the 71 students

Table 15 Clustering students into four groups according to their ChatGPT usage rate

Usage rate range	Usage Pattern
1.00 - 1.75	Poor
1.76 - 2.50	Moderate
2.51 - 3.25	Good
3.26 - 4.00	Advanced

Figure 9 shows the number of students that fall into every cluster (frequency) and their average grade. Accordingly, the clusters of poor and moderate usage are more occupied. The average grades in the upper two clusters are “slightly” higher.

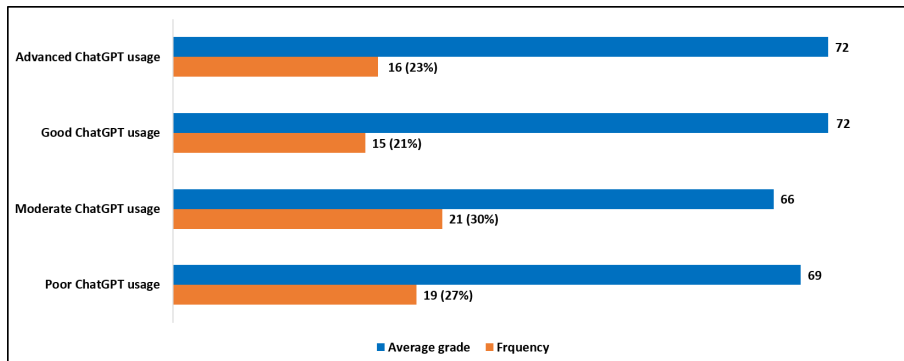


Fig. 9 Frequency of the ChatGPT clusters and average grade in every cluster

Table 16 shows the results of the correlation analysis between the ChatGPT usage rate and the exam grade. Accordingly, the Pearson correlation index ($\rho = 0.14$) shows a weak positive correlation. However, this correlation is insignificant since $p > 0.05$. Thus, students’ behaviour in using ChatGPT is not correlated with their performance in the exam according to these results.

Table 16 Correlation between ChatGPT usage rate and exam grade

Variable 1	Variable 2	No. of students	Pearson correlation coefficient	P value
ChatGPT usage rate	Exam grade	71	0.14	= 0.24

Consequently, our results show when questions of exams are correctly designed, the use of chatbots will not help students obtain high grades.

It should be noted that we identified different behaviours with regards to exam questions, we list here some of them:

- Even after training on the use of ChatGPT, a considerable number of final answers were paraphrased from the very initial answer of ChatGPT to the exam question (A0),
- Some students obtained good answers from ChatGPT but they misinterpreted them and they gave wrong final answers,
- Many students were not able to analyze images associated with exam questions to ask a better question to ChatGPT,
- One student went beyond what was requested by providing his answer and the evolution of his answer after reading the responses of ChatGPT,
- Students prefer giving final answers based on class discussions, but sometimes they misunderstand the class discussion, leading to wrong answers,
- In some cases, students were asking good questions but ChatGPT did not provide specific correct answers,
- Some students used external resources (e.g. Google).

These observations illustrate the need to prepare our students for enhanced and correct chatbots usage. Our role as educators is to ensure that our students are using these tools in a way that helps them in self-growth by developing their critical thinking capabilities.

9 Conclusions and Future Works

In this paper, we propose to design courses based on the strengths and limitations of AI. We believe that educators need to recognise that science is not and will not fully understand the universe. However, AI can expedite our understanding of the universe. Therefore, educators have to continuously observe the progress of AI technology and update their courses accordingly.

To help educators achieve this objective, we explored the working principles, strengths, and limitations of ML and LLMs. These steps enabled us to strategically build upon the LLMs strengths and limitations, while designing courses and exams for the modern AI era. The designed courses and exams showed strong potential to efficiently address the education crisis. Thus, this research suggests that integrating LLMs - like ChatGPT - in education hold the promise to implement the recommendations of Delors Report [1], eventually fostering students' self-growth by strengthening their analytical thought process.

Regarding course delivery, we proposed a pedagogical approach that is inspired from the Socratic based approach. While LLMs can provide eloquent answers to the

questions, the responsibility of inculcating the progressive thinking pattern in students rests with the educators.

To validate our approach, we present eight case studies conducted on actual university students covering multiple subject areas. We evaluate and analyze our methodology using the benchmark scores of CUS and CUR. The data analysis of our case studies conducted between December 2022 and March 2023 demonstrate that there is a very low correlation between students' grades and their choice of using ChatGPT (or otherwise) when courses and exams are designed using our proposed pedagogical approach.

The students' behaviors observed during our case studies and data analysis motivated us to design *Transparency Exam System*, which is an examination tool to facilitate our pedagogical approach implementation.

The current exam tools such as Moodle[96] are incapable of integrating our pedagogical approaches and have several issues, including but not limited to:

- Requiring students to manually copy and paste their conversations with chatbots,
- Creating difficulties for educators to grade students' responses, as each student follows a different answering format,
- Manual reading of a lot of text by the educators to grade the students,
- Leaving educators with little choice but to assume that students have provided the complete conversation log with the consulted chatbot.

To address the above identified issues, our future work will design and implement Transparency Examination System that will run inside a lockdown browser (see figure 10), enabling the tool to log and present the students actions during examinations in a chronological way, to the examiner. We are currently working on the mentioned examination tool, with our initial prototype being tested. The complete functional exam system once implemented will be fully integrated with external resources (chatbot, books, slides, and search engines); in order to record and present a log of complete student activity; and employ sophisticated AI techniques requiring minimal examiner intervention, to grade students based on their answers. Transparency Examination System will be fully compatible with the proposed pedagogical approach and modern exam requirements.

We have implemented the first version of the exam system and published the code source of our exam system here under GPL3 license: <https://github.com/theDartagnan/isorceit>. We show in figures 11 and 12 how every question is presented to the evaluator. The evaluator can check clearly how the student was handling the exam question and what was his thinking approach to handle the exam question.

We are working to extend our exam system to add more features that allow students to take advantage of AI tools to handle their courses problems that are designed based on the strengths and limitations of AI tools.

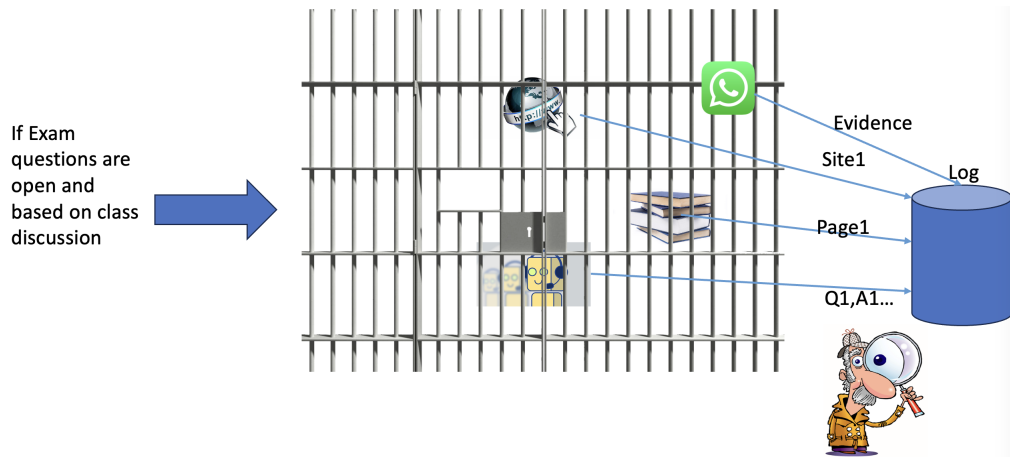


Fig. 10 Action logging in the transparency exam system

Start the exam

09/05/2023, 11:31:57

(Q1) Wrote initial answer

I forgot how AND cipher work, so i need to refresh my memory

09/05/2023, 11:33:09

(Q1) Ask Chat AI

Prompt: can you explain how AND cipher works? with an example?
Example: Certainly! An AND cipher is a simple encryption method that involves using a binary "AND" operation to combine the binary representation of a message with a secret key.

Here's an example:

Let's say we want to encrypt the message "HELLO" using the secret key "1101". First, we need to convert each letter into its binary representation using ASCII encoding:

H = 01001000
 E = 01000101
 L = 01001100
 L = 01001100
 O = 01001111

Next, we perform the AND operation between each binary digit of the message and the corresponding binary digit of the key.

For example, the first letter "H" would be encrypted as follows:

```

01001000 (H) AND
1101 (Key)
-----
01000000 (Result)

```

We repeat this process for each letter in the message, using the same key. The resulting encrypted message would be:

01000000 01000100 01000000 01000000 01001001

To decrypt the message, you would need to perform the same AND operation between each binary digit of the encrypted message and the corresponding binary digit of the key.

Note that this encryption method is not very secure, as the key is easily guessable and the encryption process is reversible. It is more commonly used as a bitwise operation in computer programming than as an encryption method.

Fig. 11 Chronological presentation of student's action to one exam question

Assess: To encrypt the string "10101100" using the key "10111101" with the AND cipher, we need to perform the AND operation between each bit of the string and the corresponding bit of the key.

Here are the steps:

```
String: 1 0 1 1 0 1 1 0 1
Key:   1 0 1 1 1 1 0 0 1
-----
Encrypted: 0 1 1 0 1 1 0 0 0
```

Therefore, the encrypted string would be "01101100".

09/05/2023, 11:40:13

(Q1) Write initial answer

I forgot how AND cipher work, so I need to refresh my memory. From what I understood, we get 1 when the value is 1 and 0 when its not. The encryption result will be 10101100 and I checked my answer with chatgpt.

09/05/2023, 11:40:30

(Q1) Ask Chat AI

Prompt: is AND cipher a perfect secure cipher?

Assess: No, the AND cipher is not a perfect secure cipher.

One reason is that the key used in the AND cipher must be at least as long as the message being encrypted. If the key is shorter than the message, it is possible to use frequency analysis or other methods to break the encryption.

Additionally, the AND cipher is vulnerable to a known-plaintext attack. If an attacker knows part of the plaintext and the corresponding ciphertext, they can use this information to deduce the key and decrypt the rest of the message.

Overall, the AND cipher is a very simple encryption method that is not suitable for secure communication. It is more commonly used as a bitwise operation in computer programming than as an encryption method.

09/05/2023, 11:41:55

(Q1) Write final answer

1. The encrypted string is 10101100. 2. Based on my understand, any perfect secure cipher must use a key that is the same or bigger than the plaintext. Also, the AND cipher is known that any string value of 1 is 1 otherwise it will be 0. This can help the attacker to perform frequency analysis attack or even worse known-plaintext attack.

09/05/2023, 11:42:15

(Q1) Add resource

Title: www.baikidung.com
Description: https://www.baikidung.com/cek-dan-cek-pasokan-known-plaintext-attack-how-does-known-plaintext-attack
Type: url | Browsers: no

09/05/2023, 11:46:09

Submit exam

09/05/2023, 11:46:57

Fig. 12 Chronological presentation of student's action to one exam question

References

- [1] Twenty-first Century, I.C., Delors, J., (Eds.), U.: Learning, the Treasure Within: Report to UNESCO of the International Commission on Education for the Twenty-first Century. Unesco Pub., ??? (1996)
- [2] Taj, I., Jhanjhi, N.: Towards industrial revolution 5.0 and explainable artificial intelligence: Challenges and opportunities. International Journal of Computing and Digital Systems **12**(1), 295–320 (2022) <https://doi.org/10.12785/ijcds/120128>
- [3] Moore, S., Nguyen, H.A., Bier, N., Domadia, T., Stamper, J.: Assessing the quality of student-generated short answer questions using gpt-3. In: European Conference on Technology Enhanced Learning, pp. 243–257 (2022). Springer
- [4] Frieder, S., Pinchetti, L., Griffiths, R.-R., Salvatori, T., Lukasiewicz, T., Petersen, P.C., Chevalier, A., Berner, J.: Mathematical capabilities of chatgpt. arXiv preprint arXiv:2301.13867 (2023) [arXiv:2301.13867](https://arxiv.org/abs/2301.13867)
- [5] Pavlik, J.V.: Collaborating with chatgpt: Considering the implications of generative artificial intelligence for journalism and media education. Journalism & Mass Communication Educator **78**(1), 84–93 (2023) <https://doi.org/10.1177/10776958221149577>
- [6] Huh, S.: Are chatgpt's knowledge and interpretation ability comparable to those of medical students in korea for taking a parasitology examination?: A descriptive study. Journal of Educational Evaluation for Health Professions **20**(1) (2023)

<https://doi.org/10.3352/jeehp.2023>

- [7] Gilson, A., Safranek, C., Huang, T., Socrates, V., Chi, L., Taylor, R.A., Char-tash, D.: How well does chatgpt do when taking the medical licensing exams? the implications of large language models for medical education and knowledge assessment. medRxiv (2022). Preprint
- [8] Kung, T.-H., Cheatham, M., Medinilla, A., ChatGPT, Sillos, C., De Leon, L., Tseng, V.: Performance of chatgpt on usmle: Potential for ai-assisted medical education using large language models. medRxiv (2022). Preprint, 2022-12
- [9] Bommarito II, M., Katz, D.M.: Gpt takes the bar exam. arXiv preprint arXiv:2212.14402 (2022) [arXiv:2212.14402](https://arxiv.org/abs/2212.14402)
- [10] Gao, C.A., Howard, F.M., Markov, N.S., Dyer, E.C., Ramesh, S., Luo, Y., Pearson, A.T.: Comparing scientific abstracts generated by chatgpt to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers. bioRxiv (2022). Preprint
- [11] Terwiesch, C.: Would Chat GPT3 Get a Wharton MBA? A Prediction Based on Its Performance in the Operations Management Course. Wharton School, University of Pennsylvania. <https://mackinstitute.wharton.upenn.edu/wp-content/uploads/2023/01/Christian-Terwiesch-Chat-GTP.pdf>
- [12] Rudolph, J., Tan, S., Tan, S.: Chatgpt: Bullshit spewer or the end of traditional assessments in higher education? Journal of Applied Learning and Teaching **6**(1), 1–22 (2023)
- [13] Herft, A.: A Teacher’s Prompt Guide to ChatGPT: Aligned with What Works Best. <https://usergeneratededucation.files.wordpress.com/2023/01/a-teachers-prompt-guide-to-chatgpt-aligned-with-what-works-best.pdf>. Accessed on July 11, 2023 (2023)
- [14] U.S. Department of Education Office of Educational Technology: Artificial Intelligence and the Future of Teaching and Learning. <https://tech.ed.gov/ai-future-of-teaching-and-learning/>. Accessed on July 11, 2023 (2023)
- [15] Delic, H., Bećirović, S.: Socratic method as an approach to teaching. European Researcher **111**, 511–517 (2016) <https://doi.org/10.13187/er.2016.111.511>
- [16] The Economist: A New AI Language Model Generates Poetry and Prose. <https://www.economist.com/science-and-technology/2020/08/06/a-new-ai-language-model-generates-poetry-and-prose>
- [17] Times: Elon Musk, other AI researchers call for a ban on autonomous weapons. <https://time.com/6266679/musk-ai-open-letter/>

- [18] Hinton, G.E., Sejnowski, T.J.: Learning and relearning in boltzmann machines. In: Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 1, pp. 282–317. MIT Press, ??? (1986)
- [19] Review, T.: Geoffrey Hinton tells us why he’s now scared of the tech he helped build. <https://www.technologyreview.com/2023/05/02/1072528/geoffrey-hinton-google-why-scared-ai/>. Accessed on July 11, 2023 (2023)
- [20] Bard by Google. <https://bard.google.com/>
- [21] Washington Post: Google critic Timnit Gebru’s exit exposes rifts in its ethical AI unit. <https://www.washingtonpost.com/technology/2020/12/23/google-timnit-gebru-ai-ethics/>
- [22] Fortune: ChatGPT banned from workplace by Apple, Goldman over risk to privacy. <https://fortune.com/2023/05/19/chatgpt-banned-workplace-apple-goldman-risk-privacy/>. Accessed on July 11, 2023 (2023)
- [23] Connected World: ChatGPT: The Most Disruptive Tech of the Century? <https://connectedworld.com/chatgpt-the-most-disruptive-tech-of-the-century/?amp=1>
- [24] The Telegraph: Forget deepfakes, AI-generated text is the future. <https://www.telegraph.co.uk/technology/2020/08/26/forget-deepfakes-ai-generated-text-should-worried/>
- [25] Chomsky, N., Roberts, I., Watumull, J.: Noam chomsky: The false promise of chatgpt. The New York Times (2023)
- [26] Chollet, F.: Tweet by Francois Chollet. <https://twitter.com/fchollet/status/1624817201906868225?lang=en>
- [27] The Telegraph: ChatGPT creator Sam Altman admits world ‘in danger’. <https://www.telegraph.co.uk/business/2023/05/16/chatgpt-creator-sam-altman-admits-world-in-danger/>
- [28] Susnjak, T.: Chatgpt: The end of online exam integrity? (2022) [arXiv:2212.09292](https://arxiv.org/abs/2212.09292) [cs.CL]
- [29] The Verge: Education regulators ban ChatGPT for student safety and accuracy concerns. <https://www.theverge.com/2023/1/5/23540263/chatgpt-education-fears-banned-new-york-city-safety-accuracy>
- [30] The National News: Dubai university set to reinstate pen and paper exams after students cheat with ChatGPT. <https://www.thenationalnews.com/uae/education/2023/04/28/dubai-university-set-to-reinstate-pen-and-paper-exams-after-students-cheat-with-chatgpt/>

- [31] Baltic News: Can artificial intelligence get higher education? <https://bnn-news.com/can-artificial-intelligence-get-higher-education-245552>
- [32] University of Bath: Generative AI Case Studies. <https://teachinghub.bath.ac.uk/generative-ai-case-studies/>
- [33] Engadget: Mark Zuckerberg says generative AI is coming to every single one of our products. <https://www.engadget.com/mark-zuckerberg-says-generative-ai-is-coming-to-every-single-one-of-our-products-204741820.html>. Accessed on July 11, 2023 (2023)
- [34] Financial Times: AI machine learning gets mixed report card from academics. <https://www.ft.com/content/a8cdf3a-a445-476c-b4a7-367468ac1398>
- [35] Liu, Y., Wang, Y., Zhang, Y., Zhang, Y., Li, Y.: The ai revolution in education: Will ai replace or assist teachers in higher education? arXiv preprint arXiv:2305.01185 (2023)
- [36] Carneiro, R., Draxler, A.: Education for the 21st century: Lessons and challenges. *European Journal of Education* **43**(2), 149–160 (2008) <https://doi.org/10.1111/j.1465-3435.2008.00348.x>
- [37] Organisation for Economic Co-operation and Development: Schooling for Tomorrow—Personalising Education. OECD Publishing, ??? (2015). <https://www.oecd.org/education/schooling-for-tomorrow-9789264245914-en.htm>
- [38] Stanford HAI: AI Definitions. <https://hai.stanford.edu/sites/default/files/2020-09/AI-Definitions-HAI.pdf>
- [39] Dhavare, U., Kulkarni, U.: Natural language processing using artificial intelligence. *International Journal of Emerging Trends Technology in Computer Science (IJETTCS)* **4**, 203–205 (2015)
- [40] IBM: What is Machine Learning? IBM (n.d.). <https://www.ibm.com/topics/machine-learning>
- [41] EurekAlert!: Can we trust scientific discoveries made using machine learning? <https://www.eurekalert.org/news-releases/611930>. Accessed on July 11, 2023 (2019)
- [42] Magazine, Q.: Common Sense Comes Closer to Computers. *Quanta Magazine* (2020). <https://www.quantamagazine.org/common-sense-comes-to-computers-20200430/>
- [43] Gunning, D.: Machine Common Sense Concept Paper (n.d.)
- [44] Central, P.: Apophenia: What It Is and More. *Psych Central* (2021). <https://psychcentral.com/health/apophenia-overview>

- [45] Pasquinelli, M.: How a machine learns and fails—a grammar of error for artificial intelligence (2019)
- [46] Aryal, K., Gupta, M., Abdelsalam, M.: Analysis of label-flip poisoning attack on machine learning based malware detector. arXiv preprint arXiv:2301.01044 (2023) [arXiv:2301.01044](https://arxiv.org/abs/2301.01044)
- [47] Jebreel, N.M., Domingo-Ferrer, J., Sánchez, D., Blanco-Justicia, A.: Defending against the label-flipping attack in federated learning. arXiv preprint arXiv:2207.01982 (2022) [arXiv:2207.01982](https://arxiv.org/abs/2207.01982)
- [48] Qiu, S., Liu, Q., Zhou, S., Huang, W.: Adversarial attack and defense technologies in natural language processing: A survey. *Neurocomputing* **492**, 278–307 (2022) <https://doi.org/10.1016/j.neucom.2022.04.020>
- [49] Matthias, A.: The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology* **6**(3), 175–183 (2004) <https://doi.org/10.1007/s10676-004-3104-7>
- [50] Shen, Z., Liu, J., He, Y., Zhang, X., Xu, R., Yu, H., Cui, P.: Towards out-of-distribution generalization: A survey. arXiv preprint arXiv:2108.13624 (2021) [arXiv:2108.13624](https://arxiv.org/abs/2108.13624)
- [51] Lab, D.D.: What is Model Drift in Machine Learning? Dominodatalab. Retrieved June 20, 2023 (n.d.). <https://www.dominodatalab.com/data-science-dictionary/model-drift>
- [52] Mardziel, P.: Drift in Machine Learning. Why is it hard and what to do about it? *Towards Data Science* (2021). <https://towardsdatascience.com/drift-in-machine-learning-e49df46803a>
- [53] Explorium.ai: Understand and Handling Data Drift and Concept Drift. Explorium.ai (2020). <https://www.explorium.ai/blog/understanding-and-handling-data-and-concept-drift/>
- [54] Madaan, A., Tandon, N., Gupta, P., Hallinan, S., Gao, L., Wiegrefe, S., Alon, U., Dziri, N., Prabhunoye, S., Yang, Y., Gupta, S., Majumder, B.P., Hermann, K., Welleck, S., Yazdanbakhsh, A., Clark, P.: SELF-REFINE: Iterative Refinement with Self-Feedback (n.d.)
- [55] Kim, G., Baldi, P., McAleer, S.: Language models can solve computer tasks. arXiv preprint arXiv:2303.17491 (2023) [arXiv:2303.17491](https://arxiv.org/abs/2303.17491)
- [56] IFLScience: The Eliza Effect: How A Chatbot Convinced People It Was Real Way Back In The 1960s. IFLScience. <https://www.iflscience.com/the-eliza-effect-how-a-chatbot-convinced-people-it-was-real-way-back-in-the-1960s-64155>

- [57] LeCun, Y.: A path towards autonomous machine intelligence version 0.9.2, 2022-06-27. Technical report, Facebook AI Research (2022)
- [58] Turing, A.M.: Computing machinery and intelligence. *Mind* **59**(236), 433–460 (1950) <https://doi.org/10.1093/mind/LIX.236.433>
- [59] TechTarget: What is the Turing Test? — Definition from TechTarget. TechTarget. <https://www.techtarget.com/searchenterpriseai/definition/Turing-test>
- [60] Beller, M.: Heisenberg’s indeterminacy principle and life. *Journal of the History of Ideas* **43**(1), 147–160 (1982)
- [61] Think, B.: What Einstein and Bohr’s debate over quantum entanglement taught us about reality. <https://bigthink.com/hard-science/einstein-bohr-quantum-entanglemen/> (2022)
- [62] Popper, K.R.: *The Open Universe: An Argument for Indeterminism*. Hutchinson, ??? (1982)
- [63] Zalta, E.N.: Relativism. *Stanford Encyclopedia of Philosophy* (2015). <https://plato.stanford.edu/entries/relativism/>
- [64] Jung, H., Lee, J., Park, C.: Deriving design principles for educational chatbots from empirical studies on human-chatbot interaction. *Journal of Digital Contents Society* **21**, 487–493 (2020) <https://doi.org/10.9728/dcs.2020.21.3.487>
- [65] Weizenbaum, J.: Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM* **9**(1), 36–45 (1966)
- [66] Goda, Y., Yamada, M., Matsukawa, H., Hata, K., Yasunami, S.: Conversation with a chatbot before an online efl group discussion and the effects on critical thinking. *Journal of Information Systems Education* **13**, 1–7 (2014) <https://doi.org/10.12937/EJSISE.13.1>
- [67] Nagata, R., Hashiguchi, T., Sadoun, D.: Is the simplest chatbot effective in english writing learning assistance? In: Nguyen, L.M., Tojo, S., Phan, X.H., Hasida, K. (eds.) *16th International Conference of the Pacific Association for Computational Linguistics (PACLING)*. *Communications in Computer and Information Science*, vol. 1215, pp. 245–246 (2020). <https://doi.org/10.1007/978-981-15-6168-9>
- [68] A.L.I.C.E. <https://www.chatbots.org/chatbot/a.l.i.c.e/>
- [69] Abbasi, S., Kazi, H.: Measuring effectiveness of learning chatbot systems on student’s learning outcome and memory retention. *Asian Journal of Applied Sciences* **3**, 57 (2014) <https://doi.org/10.15590/AJASE/2014/V3I7/53576>
- [70] Abbasi, S., Kazi, H., Hussaini, N.N.: Effect of chatbot systems on student’s

- learning outcomes. *Sylwan* **163**(10) (2019)
- [71] Bii, P.K., Too, J.K., Mukwa, C.W.: Teacher attitude towards use of chatbots in routine teaching. *Universal Journal of Educational Research* **6**(7), 1586–1597 (2018) <https://doi.org/10.13189/ujer.2018.060719>
- [72] Howie. <https://github.com/linzhp/Howie>
- [73] Bii, P., Too, J., Langat, R.: An investigation of student’s attitude towards the use of chatbot technology in instruction: The case of knowie in a selected high school. *Education Research* **4**, 710–716 (2013) <https://doi.org/10.14303/er.2013.231>
- [74] Na-Young, K.: A study on the use of artificial intelligence chatbots for improving english grammar skills. *Journal of Digital Convergence* **17**, 37–46 (2019) <https://doi.org/10.14400/JDC.2019.17.8.037>
- [75] Winkler, R., Soellner, M.: Unleashing the potential of chatbots in education: A state-of-the-art analysis. In: *Academy of Management Annual Meeting Proceedings*, vol. 2018, p. 15903 (2018). <https://doi.org/10.5465/AMBPP.2018.15903>
- [76] Quizlet QChat. <https://quizlet.com/labs/qchat>
- [77] Pérez-Marín, D.: A review of the practical applications of pedagogic conversational agents to be used in school and university classrooms. *Digital* **1**(1), 18–33 (2021) <https://doi.org/10.3390/digital1010002>
- [78] Hobert, S.: How are you, chatbot? evaluating chatbots in educational settings - results of a literature review. In: Pinkwart, N., Konert, J. (eds.) 17. Fachtagung Bildungstechnologien, DELFI 2019 - 17th Conference on Education Technologies, DELFI 2019, pp. 259–270 (2019). https://doi.org/10.18420/delfi2019_289
- [79] Hobert, S.: Small talk conversations and the long-term use of chatbots in educational settings - experiences from a field study. In: Folstad, A., Araujo, T., Papadopoulos, S., Law, E., Granmo, O.-C., Luger, E., Brandtzaeg, P. (eds.) 3rd International Workshop on Chatbot Research and Design, CONVERSATIONS 2019. *Lecture Notes in Computer Science*, vol. 11970, pp. 260–272 (2020). https://doi.org/10.1007/978-3-030-39540-7_18
- [80] Hobert, S., Wolff, R.: Say hello to your new automated tutor - a structured literature review on pedagogical conversational agents. In: Pipek, V., Ludwig, T. (eds.) 14th International Conference on Wirtschaftsinformatik (2019)
- [81] Carayannopoulos, S.: Using chatbots to aid transition. *International Journal of Information and Learning Technology* **35**, 118–129 (2018) <https://doi.org/10.1108/IJILT-10-2017-0097>

- [82] Davies, J.N., Verovko, M., Verovko, O., Solomakha, I.: Personalization of e-learning process using ai-powered chatbot integration. In: Shkarlet, S., Morozov, A., Palagin, A. (eds.) Selected Papers of 15th International Scientific-practical Conference, MODS 2020: Advances in Intelligent Systems and Computing. Advances in Intelligent Systems and Computing, vol. 1265, pp. 209–216 (2020). https://doi.org/10.1007/978-3-030-58124-4_20
- [83] Huang, W., Hew, K.F., Gonda, D.E.: Designing and evaluating three chatbot-enhanced activities for a flipped graduate course. *International Journal of Mechanical Engineering and Robotics Research*, 813–818 (2019) <https://doi.org/10.18178/ijmerr.8.5.813-818>
- [84] Matsuura, S., Ishimura, R.: Chatbot and dialogue demonstration with a humanoid robot in the lecture class. In: Antona, M., Stephanidis, C. (eds.) 11th International Conference on Universal Access in Human-Computer Interaction, UAHCI 2017. Lecture Notes in Computer Science, vol. 10279, pp. 233–246 (2017). <https://doi.org/10.1007/978-3-319-58700-4>
- [85] Kerly, A., Hall, P., Bull, S.: Bringing chatbots into education: Towards natural language negotiation of open learner models. *Knowledge-Based Systems* **20**, 177–185 (2007) <https://doi.org/10.1016/j.knosys.2006.11.014>
- [86] Chang, M.-Y., Hwang, J.-P.: Developing chatbot with deep learning techniques for negotiation course. In: 2019 8th International Congress on Advanced Applied Informatics, IIAI-AAI 2019, pp. 1047–1048 (2019). <https://doi.org/10.1109/IIAI-AAI.2019.00220> . IEEE
- [87] Almahri, F.A.J., Bell, D., Merhi, M.: Understanding student acceptance and use of chatbots in the united kingdom universities: A structural equation modelling approach. In: 2020 6th IEEE International Conference on Information Management, ICIM 2020, pp. 284–288 (2020). <https://doi.org/10.1109/ICIM49319.2020.244712> . IEEE
- [88] Fryer, L., Nakao, K., Thompson, A.: Chatbot learning partners: Connecting learning experiences, interest and competence. *Computers in Human Behavior* **93**, 279–289 (2019) <https://doi.org/10.1016/j.chb.2018.12.023>
- [89] Mikolov, T., Karafiát, M., Burget, L., Cernocký, J., Khudanpur, S.: Recurrent neural network based language model. Eleventh Annual Conference of the International Speech Communication Association (2010)
- [90] Bengio, Y., Ducharme, R., Vincent, P., Jauvin, C.: A neural probabilistic language model. *Journal of Machine Learning Research* **3**, 1137–1155 (2003)
- [91] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, , Polosukhin, I.: Attention is all you need. *Advances in Neural Information Processing Systems* **30**, 5998–6008 (2017)

- [92] Bommasani, R., Hudson, D.A., Altman, E.A.R., Arora, S.: On the Opportunities and Risks of Foundation Models (2023)
- [93] Bowman, S.R.: Eight things to know about large language models. arXiv preprint arXiv:2304.00612 (2023) [arXiv:2304.00612](https://arxiv.org/abs/2304.00612)
- [94] Adadi, A., Berrada, M.: Peeking inside the black-box: A survey on explainable artificial intelligence (xai). *IEEE Access* **6**, 52138–52160 (2018) <https://doi.org/10.1109/ACCESS.2018.2870052>
- [95] Schiller, N.: Finding a socratic method for information literacy instruction. *College & Undergraduate Libraries* **15**(1-2), 39–56 (2008)
- [96] Moodle. <https://moodle.org/>. Accessed July 16, 2023 (n.d.)