DIDACTICA
DANUBIENSIS

Outcomes of Large Language Models and Artificial Intelligence in Education

Andrada-Iulia State[1], Georgiana-Alexandra Moroșanu[2], Laura-Andreea Rață[3], Marius Geru[4]

Abstract: The recent development in Artificial Intelligence, specifically with the advent of large language models like GPT-4, is transforming educational paradigms. This literature review investigates the implications of AI and large language models in education, outlining a vision for personalized, on-demand and interactive learning. This research examines how these technologies can function as virtual tutors, encourage real-time examination, support lifelong education, and function as teachers' resources by automating administrative tasks. The paper also addresses the ethical concerns of utilizing AI in educational frameworks and advocating for a complement rather than a replacement of teachers. Additionally, it investigates their ability to promote real-time assessment, enabling students to receive immediate feedback and adapt their learning strategies accordingly. Moreover, this review highlights the potential of AI and large language models to support improving knowledge, competencies and new skills development. By providing learners access to vast knowledge repositories, these technologies empower individuals to pursue continuous learning beyond traditional classroom settings. By enabling a general understanding of the arising paradigm, the aim is to design a framework for future education systems where AI plays a constructive role.

Keywords: review; education; AI; learning; technologies

[1] Project Assistant, S.C. THECON S.R.L., Romania, Address: Calugareni street, No. 3, Galati, Romania, telephone: +40771518864, E-mail: andradastate@thecon.ro.
[2] PhD, "Dunărea de Jos" University of Galați, Romania, Address: Domnească street, No. 111, 800201, Galați, Romania, Tel.: +40336130208, Fax: +40236314463, Corresponding author: alexandra.costin@ugal.ro.
[3] Marketing researcher, S.C. THECON S.R.L., Romania, Address: Calugareni street, No. 3, Galati, Romania, Tel.: +40771518864, E-mail: laurarata@thecon.ro.
[4] Project Manager, S.C. THECON S.R.L., Romania, Address: Calugareni street, No. 3, Galati, Romania, Tel.: +40771518864, E-mail: marius@thecon.ro.

1. Introduction

In an era of rapid technological advancements, artificial intelligence (AI) stands out as one of the most transformative forces reshaping various facets of human life. From healthcare and finance to entertainment and now, education, AI's footprint is expanding, bringing with it both unmatched opportunities and intricate challenges. Among many emerging AI tools, large language models (LLMs) like OpenAI's ChatGPT, have garnered significant attention, revolutionizing interactions, content generation and data processing. While these models promise to democratize knowledge and enhance educational experiences, they also raise pressing ethical, legal and pedagogical concerns. As AI systems become increasingly entwined with our daily lives, it is imperative to critically examine their implications, especially in sensitive areas like education.

In the vast ecosystem of education, AI and LLMs have emerged as powerful catalysts for change. Beyond traditional teaching, these technologies are reshaping the very fabric of learning and knowledge dissemination. Classrooms are now being expanded and enriched by AI-driven personalized learning experiences. LLMs, with their adeptness at understanding and generating language, offer students instant access to a vast reservoir of information, simulating one-on-one tutoring sessions and bridging educational gaps. This article examines the European regulatory landscape surrounding AI, elucidates the potential risks and rewards of integrating LLMs in educational settings and offers insights into balancing the power of AI with the essence of human creativity and integrity.

2. The Evolution of GPT Technology

In the contemporary context of rapid technological advances, the field of artificial intelligence (AI) has made significant progress and the emergence of advanced linguistic models has represented a turning moment in the evolution of society. This wave of innovations has generated significant transformations in a wide range of aspects of everyday life. One of the sectors that has experienced remarkable transformations is education, where the integration of artificial intelligence and linguistic models has generated a profound change in traditional paradigms (Marcello et. al., 2023, pp. 1-25).

The central aim of this paper is to carry out an exhaustive analysis of how advanced linguistic models and artificial intelligence have redefined the way

education is designed and delivered, highlighting in particular the concepts of personalized, interactive and on-demand learning. This investigation intends to explore in detail how these innovations have transformed the interaction between students, teachers and educational content, outlining a comprehensive perspective on the significant changes brought about by these technologies in the contemporary educational landscape.

Thus, AI technology has captured attention and has become an area of great interest in a diverse range of applications. Artificial intelligence refers to the creation of systems and technologies that can mimic and perform tasks that traditionally required human intervention, such as complex pattern recognition, learning, problem solving, and user interaction. This technology has significantly evolved in recent years and a remarkable example of this evolution is the GPT language model, developed by OpenAI (Androidro, 2023).

OpenAI is a prestigious research laboratory based in San Francisco, California, dedicated to the development of Artificial General Intelligence (AGI). Founded in 2015, the institution was born from the collaboration of many influential personalities in the field of artificial intelligence, including Elon Musk, Sam Altman, Greg Brockman and Ilya Sutskever, among other notable figures (OpenAI, 2015).

The efforts and commitment of OpenAI in the development of AGI are relevant in the context where this form of artificial intelligence assumes a holistic and complex approach, capable of understanding, learning and solving problems in a variety of contexts. As AGI represents a paradigm that exceeds the capabilities of traditional artificial intelligence, OpenAI has become a central figure in the technological advancement that awaits us in the future. In the context of the rapid evolution of artificial intelligence, the contributions made by OpenAI not only redefine the limits of knowledge and technological capabilities, but also prepare the ground for a future development paradigm in this field.

Therefore, Generative Pre-Trained Transformers (GPT) constitute a category of machine learning models specialized in tackling natural language processing (NLP) tasks. These models are initially trained on massive data sets, such as books or web pages, to develop the ability to generate semantically and contextually relevant texts (Radford, 2019, pp. 1-24).

In simpler terms, GPTs are computer programs capable of generating text in a human-like manner without requiring explicit programming to do so. Thus, these

models can be adjusted to handle various tasks in natural language processing, such as answering questions, creating abstracts or translating texts.

Some of the uses and applications of GPT models are (Xipeng et al., 2020, pp. 1872-1897):

- generating content: GPT models have the ability to generate text for a wide range of purposes, including writing articles, blog posts, product descriptions or social media content, making it an effective tool for quickly and coherently creating content;

- language translation: GPT can be applied in the automatic translation of texts from one language to another; having the ability to learn patterns and structures in multiple languages during the training process, these models can provide high-quality translation services;

- answers to questions: GPT models can provide relevant and intelligent answers to questions formulated in natural language; they can analyze context and generate informed responses, making them useful for developing chatbots, customer support systems and facilitating access to information;

- text summarization: GPT models are able to create concise abstracts of long or complex texts, being a valuable tool for synthesizing content from news articles, research papers or extensive documents;

- dialog systems and chatbots: ChatGPT and other GPT models can be integrated into chatbots and dialog systems, facilitating interactive conversations with users; these models can understand user questions and provide coherent and relevant answers, performing various tasks;

- creative writing: GPT models can be used to generate creative content such as poems, stories or screenplays; they may mimic the style and structure of different genres, providing assistance to writers or contributing to the creation of diverse content for entertainment purposes;

- research in language modeling: GPT models are of significant importance in research in the field of natural language processing and language understanding; they serve as benchmarks for evaluating model performances and for exploring new techniques and approaches in this ever-evolving field.

GPT technology, Figure 1, is a significant discovery in the field of NLP, enabling machines to understand and generate language with an unbelievable level of

fluency and accuracy. Next, the four variants of GPT models are presented, starting from the first version and ending with the most recent, GPT-4.

In this approach, the performances and limits of these models will be analyzed. It is important to note that the evolution from GPT-1 to GPT-4 consisted of a significant increase in the number of parameters and, implicitly, in the text generation and context understanding capabilities. Each iteration represented an important leap in the development of artificial intelligence and generated interest in the scientific and technical community.

Figure 1. Evolution of GPT technology (Machine Intelligence Update, 2023)

In 2018, OpenAI released GPT-1 being their first version of a language model, adopting the Transformer architecture innovation. With a total of 117 million parameters, this version represented a significant leap in the evolution of state-of-the-art language models (MUO, 2023).

One of the notable achievements of GPT-1 was its ability to generate fluent and coherent text when it was powered with a prompt or context. The model was trained on two distinct datasets: Common Crawl, a vast collection of web pages containing billions of words, and BookCorpus, a collection of over 11,000 books spanning various genres. Integrating these diverse data sources provided GPT-1 with advanced language modeling capabilities.

As GPT-1 marked an important benchmark in NLP, it brought with it certain restrictions. A notable example was its tendency to generate repetitive text, especially when prompted with requests outside the scope of the training data. The model also had difficulty handling multiple lines of dialogue and failed to track long-term dependencies in the text. Its fluency and cohesion were also limited

particularly in relation to longer text sequences where coherence could be lacking (MUO, 2023).

Despite these limitations, GPT-1 paved the way for further development of GPT models, laying the groundwork for generating coherent and context-relevant natural language. This model was the first step towards more advanced and larger GPT models such as GPT-2, GPT-3 and later GPT-4.

Although GPT-1 had some limitations, it demonstrated the enormous potential of natural language processing technology and the Transformer architecture in creating intelligent systems capable of understanding and generating human language.

In 2019, GPT-2 launched as a significant evolution from GPT-1, with 1.5 billion parameters. Trained on vast Common Crawl and WebText datasets, GPT-2 impressed with human-like text generation for tasks like content creation and translation. Still, it struggled with complex reasoning and context in longer texts. GPT-2's strengths and limitations paved the way for improved AI models, reflecting OpenAI's commitment to advancing language capabilities (MUO, 2023).

Thus, the release of GPT-3 in 2020 marked an exponential revolution in the evolution of natural language processing models. With an impressive amount of 175 billion parameters, GPT-3 is more than 100 times the size of GPT-1 and more than ten times the size of GPT-2. This significant increase in model complexity and power is a direct result of OpenAI's sustained efforts to develop machine learning to remarkable heights.

Trained on diverse sources like BookCorpus, Common Crawl, and Wikipedia, GPT-3's massive dataset of nearly a trillion words equips it for a wide array of tasks. Unlike prior models, it crafts coherent text, code, and art. With contextual understanding, GPT-3 excels in chatbots, content creation, and translation (MUO, 2023).

A concrete example is ChatGPT, Figure 2, an AI conversational assistant that experienced a rapid rise in popularity, going from obscurity to notoriety almost overnight.

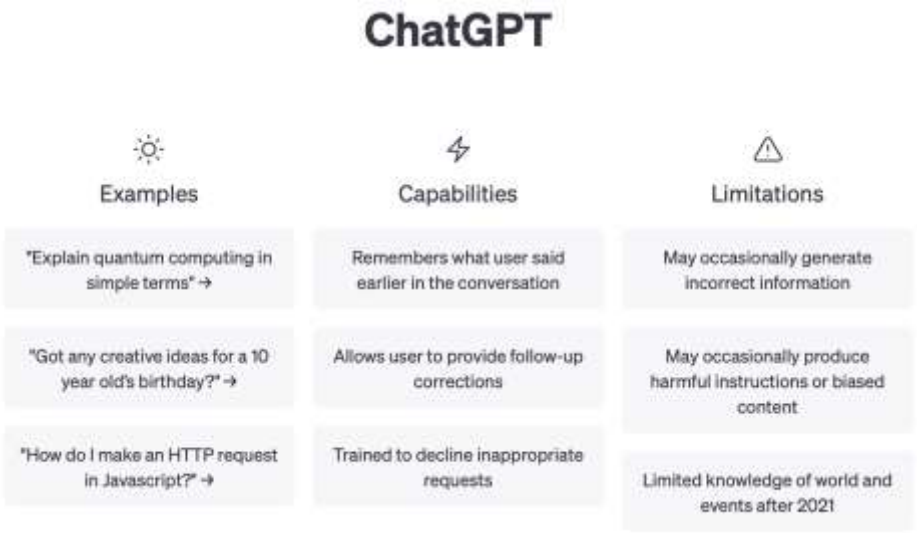

Figure 2. ChatGPT (IP Bytes, 2023)

ChatGPT was launched in 2022 by OpenAI and is a pioneer in the field of artificial intelligence chatbots with the aim of highlighting and evaluating the capabilities of an extended artificial intelligence system. Through ChatGPT, a wide range of questions can be asked and mostly answers are received that can be useful to users.

Access to this AI model is required to benefit from ChatGPT. There are several ways to do this, depending on the needs and resources of each user. Here are some options for using ChatGPT (Geeki, 2023):

- use of an online service: OpenAI provides access to GPT-3 through its API (Application Programming Interface); this service can be accessed through a web application or other compatible software;

- downloading and installing software on PC: an unofficial release can be downloaded and installed in order to be use ChatGPT on PC;

- use of pre-existing applications or tools: there are various pre-built applications or tools that use ChatGPT; for example, it can be used GPT-3-based chatbots available on platforms such as Facebook Messenger or Slack.

To start using ChatGPT, users select an access method and follow instructions. GPT-3 is AI, not thinking on its own. It replies based on user input, so accuracy depends on providing the right information. Since its release, GPT-3 has been used to generate text for chatbots, synthesize information, create articles and stories, develop customer support systems, help manage inquiries, and more.

It can also be integrated with other AI services, such as Google Cloud Platform or Amazon Web Services, to automate tasks or solve complex problems. Its ability to understand natural language makes it extremely useful in communication contexts such as virtual assistants and chatbots.

However, chatGPT also has some limitations. These are similar to those of other AI models and may include (Geeki, 2023):

- knowledge limited to events until 2021: similar to other AI models, ChatGPT only has information until the year 2021; does not have access to subsequent events and cannot provide updated information;

- absence of independent thinking: GPT-3 is an AI model and cannot make decisions or think independently; its answers are based solely on the data provided by questions or messages;

- difficulty in understanding the context: GPT-3 cannot perceive the context or the intention of the interlocutor; therefore, clarity of questions and messages is essential to get accurate answers;

- language limitations: as a natural language model, GPT-3 can only work in the languages it was trained for; although it supports multiple languages, it cannot cover all existing languages.

These aspects should be kept in mind when using ChatGPT. This is an AI model and does not possess independent thought or the ability to perceive context or intent autonomously.

GPT-3's exceptional capabilities have brought to the fore concerns about the ethical aspects and possible abuses of such advanced language models. Specialists fear that the model could be used for nefarious purposes, such as generating fake news, phishing or malware. In fact, there have already been reported cases where criminals have used ChatGPT to create malware.

Prior to the official release of GPT-4, OpenAI developed an improved version of GPT-3, named GPT-3.5, to continue to explore and improve the capabilities of its advanced language models.

In the year 2023, GPT-4 marks the latest version in the GPT series of models, having about 100 trillion parameters, representing a significant leap from its predecessor, GPT-3, which, in turn, was impressive. Even though specific details about the training data and architecture of this model have not been officially

communicated, it is clear that GPT-4 builds on the strengths of GPT-3 and manages to overcome some of its limitations. GPT-4 is exclusive to ChatGPT Plus users with usage limits. Access requires a GPT-4 API waitlist, while Microsoft's Bing Chat offers free access without waiting. GPT-4 excels in multimodal comprehension, understanding both text and images. It handles complex tasks, demonstrates human-like performance, and has a larger context window. This AI model improves interaction by adapting context for accurate responses, transcending language barriers. GPT-4 enhances global communication and information access, reshaping human-machine interaction.

A central component of Chat GPT-4's ability to understand and interact is through the use of transformer technology. This is a deep learning architecture that allows it to identify complex patterns in training data and generate coherent and semantic responses based on the context provided. This deep learning mechanism ensures that the responses generated are fluent and appropriate, thus contributing to a more efficient and natural communication between the user and the AI system. GPT-4 pushes the current boundaries of AI tools and is likely to find applications in a wide range of fields such as customer service, education, medicine etc. However, as with any powerful technology, there are concerns about possible misuses and the ethical implications of such a powerful tool.

Table 1 shows the characteristics of the GPT-1, GPT-2, GPT-3 and GPT-4 models.

Table 1. Characteristics of the GPT models (Tooabstractive, 2023)

Characteristics	**GPT-1**	**GPT-2**	**GPT-3**	**GPT-4**
Released date	2018	2019	2020	2023
Parameters	117 million	1.5 billion	175 billion	About 100 trillion
Amount of dataset trained on	600 million words	450 billion words	1.5 trillion words	100 trillion words
Decoder layers	12	48	96	120
Context token size	512	1024	2048	4096
Hidden layer	768	1600	12288	12288
Capability	Generate text, translate languages and answer question	Generate more realistic and complex text	Generate even more realistic and complex text and to perform many tasks without any additional training	Generate even more realistic and complex text and to perform many tasks without any additional training; it can also process images and text

GPT-4 has improved memory. Compared to its predecessor, which could hold 8000 words, the GPT-4 has the ability to easily store 25000 words. This improvement allows the model to return to topics covered in long conversations. This advance in memory increases the coherence and relevance of its responses in context, making it useful in natural language processing tasks. In addition, the ability to recall previously discussed topics expands his abilities to engage in complex and meaningful dialogues with people. GPT-4 stands out for its ability to respond in 26 languages, Figure 3.

Chat GPT-4's impact is not limited to simple interaction; this language model opens doors to innovations in a variety of fields. From assisting in research and development to providing fast and accurate answers in the medical, legal or technological fields, Chat GPT-4 adds significant value in many contexts. Its ability to provide relevant information and solutions makes it an indispensable virtual partner for a variety of professionals and users in search of accurate and timely answers. Thus, Chat GPT-4 not only understands and responds to conversations, but also redefines the way we communicate with technology and supports us in various areas, using Transformer technology, contextual analysis and its ability to interact in a variety of languages.

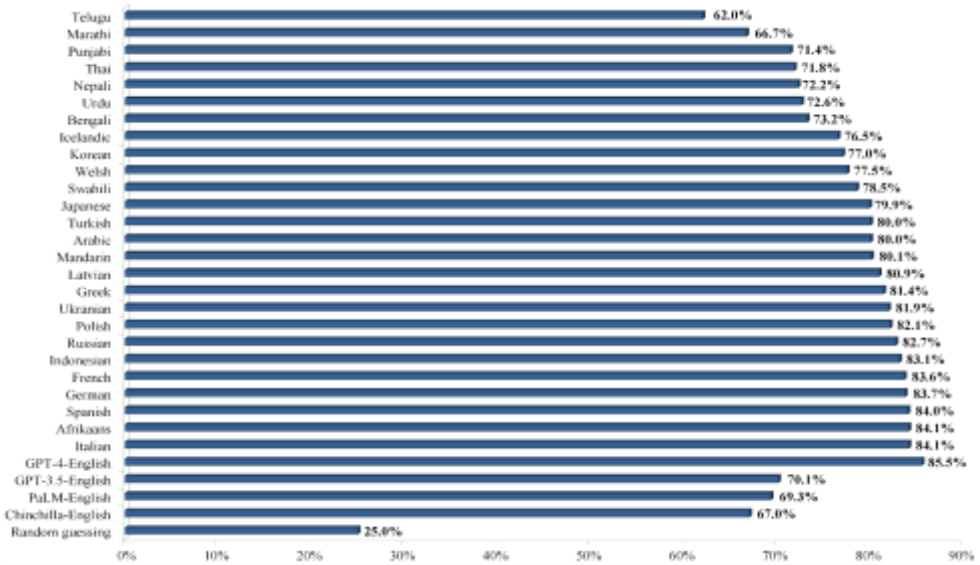

Figure 3. GPT-4 3-Shot Accuracy on MMLU Across Languages (MetaversePost, 2023)

In a total of 26 languages evaluated, including resource-constrained languages such as Latvian, Welsh and Swahili, GPT-4 outperforms GPT-3.5 and of other language

models (such as Chinchilla and PaLM) in 24 languages. This evolution in natural language processing has the potential to significantly improve multilingual communication and accessibility. This step forward represents an important step towards eliminating language barriers in various fields such as education, health and business (MetaversePost, 2023).

Through GPT-4's superior performance in low-resource languages, more opportunities are opening up for people to access information and services in their mother tongue. This contributes to the creation of a more inclusive and diverse digital environment, promoting cultural exchange and understanding. GPT-4 also stands out for its ability to maintain a calm and constructive attitude, avoiding threats, insults or meaningless speech. This skill was developed by training the model on a large set of text requests involving harmful content such as hate speech, online harassment and trolling. This helps the model identify and avoid generating harmful content in its responses, Figure 4 (MetaversePost, 2023).

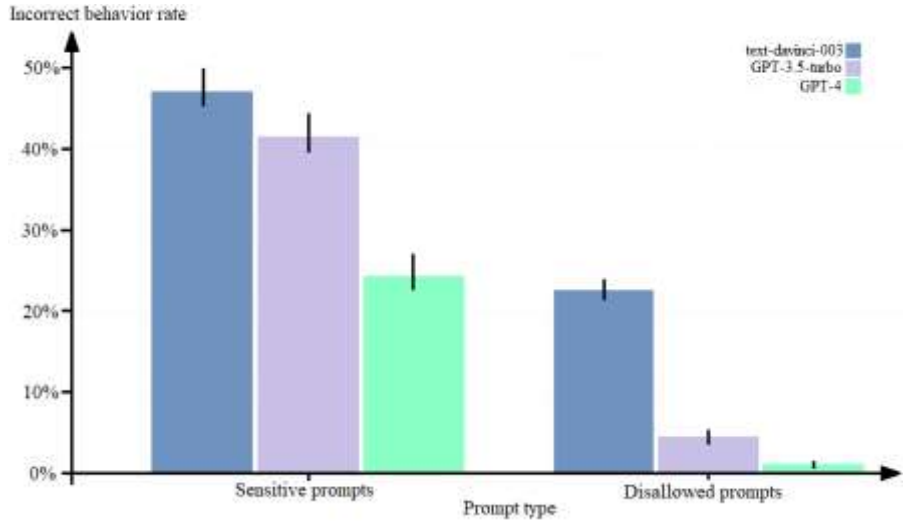

Figure 4. Incorrect behavior Rate on Disallowed and Sensitive Content (MetaversePost, 2023)

Compared to GPT-3.5, GPT-4 has many significant improvements in security features due to the tweaks applied. One notable change is that, unlike GPT-3.5, GPT-4 exhibits an 82% reduction in its tendency to respond to requests involving prohibited content. Also, inaccurate GPT-4 responses to sensitive questions, such as those related to medical advice or self-harm, were reduced by 29% (MetaversePost, 2023).

GPT-4 successfully passed over 20 tests of diverse knowledge, including areas such as mathematics, physics and chemistry. In certain scenarios, the model's performance exceeded the performance of 88% of the participants tested. This success suggests that GPT-4 has the potential to bring about significant change in sectors such as education, research and customer service. However, further testing and development is required to ensure effective integration in these areas. Various major companies, including Stripe, have already started testing GPT-4. OpenAI also admitted that the new version of GPT was used in the update of the Bing search engine developed by Microsoft. However, the company preferred not to go into technical details to avoid unnecessary disclosure of information to competitors.

It should be noted that GPT-4 is not yet perfect: its knowledge is limited until September 2021, and in some situations, it may provide incorrect or purely fictitious information. Currently, access to the GPT-4 model is only available to paying subscribers of the ChatGPT Plus service, and developers can register the model for API use. The release of the free version has not yet been announced.

Therefore, the evolution of GPT models has had a significant path in the development of artificial intelligence, focused on improving natural language processing and text generating skills. Starting with GPT-1 and continuing to more recent versions such as GPT-4, each iteration brought significant improvements in performance, capabilities, and model sizes.

Overall, the evolution of GPT models has represented a progressive and improved exploration of the capabilities of artificial intelligence in natural language processing. These iterations have contributed significantly to improving human-machine communication and facilitating complex language tasks, although there are still challenges and limitations in deep contextual understanding and logical reasoning.

3. European Regulations on the Use and Regulation of Artificial Intelligence Systems in the Context of GPT Chatbot Applications

GPT models are here to revolutionize, but as already mentioned, they do not think on their own and they require human intervention and coordination. As there have been cases where GPT models were not used for the greater good or how they were intended to be used, it is imperative to have a stable regulatory environment that

ensures the optimal and safe climate for the further development of these applications. In the last 2 years, the field of artificial intelligence has become a significant presence in various debates and discussions. This new field has raised many ethical and moral questions attracting the attention of the legal community.

The applicability of machine-learning based automation is being realized in a multitude of domains, which intensifies the interest in analyzing its impact on society. In this regard, concerns have been raised about the problems that may arise from the involvement of artificial intelligence in everyday life in various branches of law. Concerns related to AI-based systems also extend to issues related to the protection of personal data. Moreover, AI-based automation has the capacity to change and influence the concrete reality in which we live, affecting its materiality.

Lately, ChatGPT application developed by OpenAI is gaining ground in several sectors of activity, becoming an integral part of our lives as technology evolves. Starting from the GPT-3 functionality, which is unimodal, allowing only the analysis of text input, to the GPT-4 defined with multimodal functionalities that allow the analysis of text input as well as those generated by the application, making it much more efficient in analyzing and processing complex text and image data. As an exemple, a Twitter user asked GPT-4 to explain climate change in Shakespeare's voice (Miller, 2023).

Based on the above, it is also worth mentioning the risks of using this artificial intelligence-based application. In this regard, the main risks in use that the latter presents are listed. In the light of the latest news about the functionalities of the ChatGPT application, there is a growing concern among authorities about the lack of adequate knowledge about how generative artificial intelligence works. This lack of information could lead to potential data breaches for a significant part of the population.

It is worth mentioning that in order to achieve its objective, which is to respond as comprehensively as possible to the requirements they introduce into the system, the application developed by OpenAI has been trained by analyzing large amounts of information and data (Hillemann, 2023) published on the Internet, more than 250 billion words, including essays, code snippets, poems, websites, articles etc. All these sources of information are carriers of personal data and end up being queried, through various methods, unintentionally by other users (Cyberhaven, 2023).

From this point, it can be talking about serious invasions of privacy of individuals as well as other legal implications. The General Data Protection Regulation (GDPR) details that it is sufficient for a person to be directly identifiable based

only of his or her name, a unique identification number, location data, an online identifier or one or more factors specific to the physical, psychological, hereditary, mental, financial, ethnical or social nature of that person.

Article 9, paragraph 2 lit. e) of the General Data Protection Regulation (Eur-lex Europa, 2016) provides for an exception to the rule that "sensitive" data may be processed without explicit consent only if the data are clearly made public by the data subject. Indeed, personal data freely published on social media cannot be interpreted as being accessed without the explicit consent of the user. Article 6 (Eur-lex Europa, 2016) of the same Regulation also provides for situations in which personal data may be publicly processed in the context of one of the six justifications, starting from the consent of the person to the processing of the data to the obligation to publish them under a contract.

On the basis of the above, Open AI could resort to the following two options: either obtain the consent of the person under Article 6 paragraph 1 lit. a), or on the basis of the justification of the legitimate interest to process the data, under Art. 6 paragraph 1 lit. f), which is difficult to argue in order to demonstrate respect for the fair balance between access to the data and the rights of the individual to health, safety, fundamental rights, environment, democracy and the rule of law.

OpenAI's published privacy policy does not explicitly refer to the reasons behind the collection of data processed by OpenAI, but mentions that this data processing activity is based on "legitimate interests" when" "developing" its services: "*Our legitimate interests in protecting our Services from abuse, fraud or security risks, or in developing, improving or promoting our Services, including when we train our models. This may include the processing of Account Information, Content, Social Information and Technical Information. See here for instructions on how you can give up of our use of your information to train our models*" (OpenAI, 2023).

In Open AI's published technical documentation for GPT-4 (GPT-4 Technical Report, 2023), the company makes reference to their work to remove personal data from the information collected in order to train the application, "where possible".

A regulation is in place at European level whereby companies based in this area can designate a country to take over and manage disputes concerning the protection of personal data. As far as OpenAI is concerned, it has no established base in the European area, in which case any person who is linked to the practices of this company, against which damage has been caused, can address individually to the

country to which they belong. It should be noted that the scope of the GDPR also extends to firms/companies accessing markets in the EU as well as the UK and are obliged to comply with it.

Besides, there are situations when inaccurate or erroneous information, wrongly entered or erroneously generated information is slipped into the multitude of data. It is a case in China where an individual used GPT chat to spread a Fake News about a train accident where "*9 construction workers lost their lives in a city of Gansu province (northwest China)*". This news was accessed more than 15,000 times after it was disseminated on social media on 25 April 2023 (Reuters, 2023).

Some companies provide instructions in their policies on the deletion of personal data or algorithms developed on the basis of information obtained in an unauthorized way, in accordance with Article 17 of the GDPR (Eur-lex Europa, 2016). However, this issue has not been defined in the training work of AI-based modules, largely due to the fact that their operating structure comprises several interconnectivity relationships between collected and generated data, thus making it difficult to identify the specific module, obtained in an unauthorized way from the other modules.

This issue was also raised in February 2023 by data protection authorities in Italy. Italian authorities started an investigation into an alleged breach of data protection rules, at which point ChatGPT was taken out of service pending the completion of verification procedures. In order to start the investigation, the authorities have requested a list of concrete requirements that OpenAI must meet by 30 April, including the provision of "the methods and logic" underlying the processing of data by the application (Garante per la protezione dei dati personali, 2023).

These security issues in the management and personal data collection is a rather controversial issue due to Article 3 of the Privacy Policy of the OpenAI application where it is mentioned that "*In certain circumstances we may provide your Personal Information to third parties without further notice to you, unless required by the law....*" (OpenAI, 2023).

Currently, steps are being taken to establish a regulation for this field, which provides for the establishment of obligations and rights for both providers and developers of such systems based on artificial intelligence and large language models, depending on the risk degree it presents to the rights and freedoms of the individual and to his or her best interest. Against the backdrop of recent advances in AI-based technology, the European Parliament and the Council of the European Union have started debating a Regulation on artificial intelligence. The aim of this

initiative is to support the Union's goal of becoming a world leader in the development of safe, reliable and ethical artificial intelligence, in line with the European Council declaration and to guarantee the protection of moral principles, as expressly requested by the European Parliament. At the same time, legislators also provided limitations on the development of the following AI application functionalities: "real-time" remote biometric identification systems in publicly accessible spaces; "post" remote biometric identification systems, with the only exception of law enforcement for the prosecution of serious crimes and only after judicial authorization; biometric categorisation systems using sensitive characteristics (e.g. gender, race, ethnicity, citizenship status, religion, political orientation); predictive policing systems (based on profiling, location or past criminal behavior); emotion recognition systems in law enforcement, border management, the workplace, and educational institutions; untargeted scraping of facial images from the internet or CCTV footage to create facial recognition databases (violating human rights and right to privacy) (News European Parliament, 2023).

This Regulation provides for compliance with the following "*ethical objectives for the development, deployment and use of AI, robotics and related technologies*": ensure that AI systems placed on the Union market and used are safe and respect existing law on fundamental rights and Union values; ensure legal certainty to facilitate investment and innovation in AI; enhance governance and effective enforcement of existing law on fundamental rights and safety requirements applicable to AI systems; facilitate the development of a single market for lawful, safe and trustworthy AI applications and prevent market fragmentation (News European Parliament, 2023). This regulation is expected to be outlined and introduced as a legislative framework in two years time, meaning in 2025. And for this publication to take place, both the European Parliament and the 27 Member States of the European Union must give their consent.

4. Ethical Concerns of Using Large Language Models in Education

Steps have been taken to overcome the challenges imposed by the usage of LLMs and new regulation is in the process of being implemented. This regulation aims to ensure AI's safety, legal clarity, adherence to fundamental rights and a unified market approach, with specific prohibitions on certain AI functionalities. However, when it comes to ethics, it is more difficult to impose laws that can cover all

aspects of LLMs usage from various domains and various countries. Ethical concerns often deal with abstract, nuanced and culturally influenced topics. When applying these concerns to LLMs or other advanced technologies, the complexity grows due to the wide range of applications and contexts in which they might be used. Creating universal laws that address all potential ethical issues, especially when considering cultural and regional differences, is a challenging effort.

The technological advancements of the past few decades have entered the educational area in various ways. Special tools have been developed to ensure a proper learning environment and to ease the studying process (Yan et al., 2023, pp. 1-23). It is no surprise that automation has found its way in the academic field and so did large language models. As each technology innovation poses certain challenges and ethical concerns, LLMs do not make an exception. The integration of large language models in education raises multifaceted ethical dilemmas, spanning concerns from pedagogical dependency and academic integrity, to data privacy, inherent biases and potential economic implications. While LLMs offer transformative possibilities for personalization and scalability, their deployment necessitates careful consideration of potential traps, especially in perpetuating digital disparities, overshadowing human educators and inadvertently aligning with commercial rather than educational imperatives. Responsible implementation mandates a holistic understanding of their capabilities and limitations, combined with rigorous human oversight.

4.1. Benefits of Using LLMs in Education and their Ethical Dilemma

Large language models have revolutionized the educational landscape with their unprecedented capabilities. Firstly, they offer a level of personalization previously unattainable. Given their ability to process vast amounts of data and produce nuanced responses, LLMs can tailor learning experiences to individual students, adapting content based on their learning pace, style and areas of interest. Additionally, LLMs provide scalability. While human educators are limited by time and physical constraints, LLMs can simultaneously satisfy to thousands of learners, making education more accessible and allowing for the democratization of knowledge. Furthermore, these models can be a rich resource for educators, providing supplemental materials, assisting with content creation and even offering feedback on assignments. Their ability to engage with multitudes of information also means that they can assist in bridging the gap between diverse subjects, fostering

interdisciplinary learning. While all these advantages appear great, they do come at a price.

The adoption of LLMs in education isn't without its ethical quandaries. A primary concern is the risk of pedagogical dependency. With the convenience and sophistication of LLMs, there's the potential danger of students leaning too heavily on these tools, undermining the development of critical thinking and independent problem-solving skills which can lead to severely limit a future professional's capacity for effective decision-making, innovation, adaptability and confronting multifaceted challenges, jeopardizing their career progression and success in their field.

Moreover, LLMs raise academic integrity issues by blurring the line between student originality and machine-produced work. They can also hinder student creativity by offering ready-made solutions, making it challenging for educators to assess and grade student efforts accurately. Additionally, the lack of emphasis on data privacy in LLM-based innovations poses significant concerns, as students may unknowingly expose their private information, making it vulnerable to misuse (Yan et al., 2023, pp. 1-23).

The LLM, trained on a diverse linguistic corpus, effectively caters to users from various backgrounds and excels in tasks like translation. However, integrating multiple language models introduces inherent cultural, religious, and ideological nuances, leading to potential inconsistencies in responses. These biases might arise from differing national viewpoints on issues, causing the LLM to offer distinct interpretations based on the input language. Such disparities can be manipulated to foment discord, highlighting the LLM's non-consistent worldview (Zhang et al., 2023, pp. 1-5). If unchecked, these biases risk becoming deeply embedded in students, reinforcing stereotypes and potentially distorting educational content.

4.2. Enhancing Education or Corrupting Creativity

Long before ChatGPT, AI tools have been making silent inroads into classrooms. From basic math programs in the late 20th century to adaptive learning platforms, AI has always sought to revolutionize education (Ng et al., 2022, pp. 1-20).

ChatGPT, with its unparalleled natural language processing capabilities, represents the climax of this evolution, bringing opportunities and challenges with it. The immediacy of ChatGPT's responses aids students in assignments, homework, and

doubt clarification, bridging the gap between understanding and application. Each student is unique, and ChatGPT's adaptability ensures that individual needs are met, reinforcing areas of weakness and providing challenges where needed. Moreover, ChatGPT offers educational support in various languages for students worldwide, democratizing access to information and tutorial guidance (Baidoo-Anu & Ansah, 2023, pp. 1-20).

With a tool as sophisticated as ChatGPT, students may lean too heavily on it, bypassing the struggle and growth from grappling with challenging concepts. ChatGPT might inadvertently discourage diverse thought processes, unique solutions, and exploring unconventional ideas by providing 'correct' or' popular' answers (Fitria, 2021, pp. 134-147).

Encouraging critical thinking and active engagement with information is crucial to prevent a passive acceptance of AI-generated solutions. Educators can harness ChatGPT's potential while preserving creativity by: establishing clear usage guidelines for ChatGPT; incorporating tasks that prioritize original thinking and problem-solving, where AI tools may have limitations; promoting student critique and discussion of AI-generated content to nurture skepticism and curiosity.

By effectively utilizing tools like large language models, productivity and creativity can be enhanced. Ethical concerns should be addressed through proactive measures, such as developing frameworks for responsible LLM use in the academic environment.

A key starting point is the pursuit of transparency in training data. By revealing the sources and nature of such data, users can be better informed about potential biases inherent in the models. Alongside transparency, continuous auditing of LLM outputs is indispensable. Regular checks, bolstered by feedback from the user community, can shed light on inadvertent biases or inaccuracies. User education stands as another linchpin in this ethical endeavor. By acquainting students and educators with both the capabilities and limitations of LLMs, it is possible to foster a balanced understanding that these tools, while invaluable, are not free from societal biases and can err (Zhou et al., 2023, pp. 1-8). Concurrently, the research community is delving into bias mitigation techniques designed to render LLMs more equitable, ensuring they don't perpetuate harmful stereotypes.

To further ensure balance, it's pivotal to use diverse and representative training data (Kasneci et al., 2023, pp. 1-9). A deliberate inclusion of marginalized and underrepresented voices can assist in producing models that resonate with a broader spectrum of societal perspectives. Ethical guidelines, tailored for educational

settings, can offer a solid framework, guiding stakeholders on facets like data privacy, informed consent, and the judicious incorporation of LLMs into academic curricula.

The role of human oversight remains paramount. While LLMs are powerful, coupling their strengths with the nuanced understanding and ethical discernment of human educators can yield more holistic outcomes. This human-machine synergy can be further enhanced by enabling educators to customize LLM outputs, ensuring alignment with both educational goals and ethical standards. Creating feedback mechanisms allows users to voice concerns, facilitating the model's iterative refinement. Lastly, instilling ethical considerations in the very developers and teams behind LLMs can lead to more ethically-aware design and deployment choices.

5. Conclusions

In conclusion, artificial intelligence technology, with examples like GPT, has managed to transform the way we interact with digital technology and how it understands our needs. These developments open new horizons in the field of communication and assistance, thus underlining the importance of technological progress in today's society. Large language models and artificial intelligence have brought about significant changes in educational paradigms. These innovations enable personalized learning, real-time assessment and lifelong learning.

Amid these complexities, the European Union is taking proactive measures to regulate AI and large language models. The ongoing debates surrounding the Regulation on artificial intelligence aim to establish a robust framework that prioritizes safety, ethical development, and compliance with fundamental rights. This proposed regulation seeks to ensure legal clarity, support innovation, and enhance governance around AI systems. As it progresses through legislative stages, it symbolizes the European Union's commitment to fostering ethical and secure AI development while preventing market fragmentation.

In anticipation of this regulation's introduction in 2025, collaboration between the European Parliament and the Member States is essential to establish a comprehensive legislative framework. This pivotal step aligns with the Union's aspirations to lead in the realm of safe, reliable, and ethical artificial intelligence, ensuring that AI technologies serve the best interests of individuals and society at large.

However, it is important to responsibly approach the ethical implications of using AI in education and to recognize that while technology can enhance the learning process, it should not completely replace the role of teachers. Instead, the intelligent integration of these technologies in a properly operated framework, can create a more efficient and flexible educational environment that supports the development of students' knowledge and skills in an ever-changing world.

Acknowledgement

This work is supported by the project AiMedia - Open innovation platform for managing collaborative creativity in Digital Marketing (*Platformă de inovare deschisă pentru gestionarea creativităţii colaborative în Marketingul Digital - AiMedia*) in the framework of Competitiveness Operational Programme, financed from the European Regional Development Fund under the contract number 77/221_ap2/01.09.2020, SMIS Code: 129318.

References

Androidro. (2023). *How Chat GPT-4, artificial intelligence, can help us and what it can do*. *Web page*. Retrieved from https://www.androidro.ro/cum-ne-poate-ajuta-si-ce-stie-sa-faca-chat-gpt-4-inteligenta-artificiala/, date: 20.08.2023.

Baidoo-Anu, D. & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *SSRN*, pp. 1-20.

Cameron Coles (2023). *11% of data employees paste into ChatGPT is confidential. Web page.* Retrieved from https://www.cyberhaven.com/blog/4-2-of-workers-have-pasted-company-data-into-chatgpt/, date: 19.08.2023.

Eur-lex Europa (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Web page.* Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679. date: 19.08.2023.

Fieldfisher (2023). *Unveiling the Crucial 5 GDPR Obstacles of ChatGPT That Can't Be Ignored. Web Page.* Retrieved from https://www.fieldfisher.com/en/insights/unveiling-the-crucial-5-gdpr-obstacles-of-chatgpt, date: 19.08.2023.

Fitria, T. N. (2021). Artificial Intelligence (AI) in education: using AI tools for teaching and learning process. In *Prosiding Seminar Nasional & Call for Paper STIE AAS*, Vol. 4, No. 1, pp. 134-147.

Garante per la protezione dei dati personali (2023). *ChatGPT: Italian SA to lift temporary limitation if OpenAI implements measures 30 April set as deadline for compliance. Web Page.* Retrieved from https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9874751#english, date: 19.08.2023.

Geeki. (2023). *What is GPT Chat and how is it used? Web page.* Retrieved from https://geeki.ro/ce-este-gpt-chat-si-cum-se-foloseste/#ce-este-gpt-chat, date: 20.08.2023.

IP Bytes. (2023). *What is ChatGPT and how does it work? Web page.* Retrieved from https://blogs.luc.edu/ipbytes/2023/04/22/chat-gpt-should-we-have-a-chat-about-ips-role/, date: 20.08.2023.

Kasneci, E.; Seßler, K.; Küchemann, S.; Bannert, M.; Dementieva, D.; Fischer, F. & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences*, Vol. 103, pp. 1-9.

Machine Intelligence Update. (2023). *GPT-1 to GPT-4: The evolution of AI language models*. *Web page.* Retrieved from *https://www.youtube.com/watch?v=dNFC57Bz10c*, date: 20.08.2023.

Mariani, M.; Machado, I.; Magrelli, V. & Dwivedi, Y. (2023). Artificial intelligence in innovation research: a systematic review, conceptual framework, and future research directions. *Technovation*, Vol. 122, pp. 1-25.

MetaversePost. (2023). *GPT-4 vs. GPT-3: What does the new model have to offer? Web page.* Retrieved from https://mpost.io/ro/gpt-4-vs-gpt-3/, date: 20.08.2023.

MUO. (2023), *GPT-1 to GPT-4: Each of OpenAI's GPT Models Explained and Compared*. *Web page.* Retrieved from https://www.makeuseof.com/gpt-models-explained-and-compared/, date: 20.08.2023.

Ng, D.T.K.; Lee, M.; Tan, R.J. Y.; Hu, X.; Downie, J.S. & Chu, S.K.W. (2022). A review of AI teaching and learning from 2000 to 2020. *Education and Information Technologies*, pp. 1-57.

Open AI (2023). *GPT-4 Technical Report. Web page*. Retrieved from https://cdn.openai.com/papers/gpt-4.pdf, date: 19.08.2023.

OpenAI (2023). *Privacy policy. Web page.* Retrieved from https://openai.com/policies/privacy-policy, date 19.08.2023.

OpenAI. (2015). *About*. *Web page.* Retrieved from https://openai.com/about, date: 20.08.2023.

Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D. & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Blog*, Vol. 1(8):9, pp. 1-24.

Reuters (2023). *China reports first arrest over fake news generated by ChatGPT. Web Page.* Retrieved from https://www.reuters.com/technology/china-reports-first-arrest-over-fake-news-generated-by-chatgpt-2023-05-10/, date: 19.08.2023.

Tooabstractive. (2023). *GPT-4*. *Web page.* Retrieved from https://tooabstractive.com/how-to-tech/difference-between-gpt-1-gpt-2-gpt-3-gpt-4/?utm_content=cmp-true, date: 20.08.2023.

Twitter (2023). Post form Dan Miller. *Web Page.* Retrieved from https://twitter.com/danmiller999/status/1636780580431093760, date: 19.08.2023.

Xipeng, Q.; Tianxiang, S.; Yige, X.; Yunfan, S.; Ning, D. & Xuanjing, H. (2020). Pre-trained models for natural language processing: a survey. *Science China Technological Sciences*, Vol. 63, pp. 1872-1897.

Yan, L.; Sha, L.; Zhao, L.; Li, Y.; Martinez-Maldonado, R.; Chen, G.; Li, X.; Jin, Y. & Gašević, D. (2023). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, pp. 1-23.

Yasmina Yakimova, Janne Ojamo (2023). *MEPs ready to negotiate first-ever rules for safe and transparent AI. Web page.* Retrieved from https://www.europarl.europa.eu/news/en/press-room/20230609IPR96212/meps-ready-to-negotiate-first-ever-rules-for-safe-and-transparent-ai, date: 19.08.2023.

Zhang, J.; Ji, X.; Zhao, Z.; Hei, X. & Choo, K.K.R. (2023). Ethical considerations and policy implications for large language models: guiding responsible development and deployment. *Arxiv*, pp. 1-5.

Zhou, J.; Müller, H.; Holzinger, A. & Chen, F. (2023). Ethical ChatGPT: Concerns, challenges, and commandments. *Arxiv*, pp. 1-8.