RECENT J. (2023), 71:208-220
https://doi.org/10.31926/RECENT.2023.71.208

Modern Society Facing Morality of Large Language Models

Cătălin-Iulian CHIVU
Transilvania University of Braşov, Romania, catalin.c@unitbv.ro
Catrina CHIVU
Transilvania University of Braşov, Romania, catrina.c@unitbv.ro

Abstract
Interest in Artificial Intelligence has increased in recent years, as hardware technologies have evolved to the level at which they can support the real-time operation of complex systems such as the Large Language Models (LLM). Modern Artificial Intelligence systems are based on technologies such as Deep Learning, Reinforcement Learning, Machine Learning-Based Attention, Long Short-Term Memory. The present paper identifies the main morality issues that arise with free access to the development of artificial intelligence by the population. These are problems that affect especially educational system, because in the last years pupils and students have used more often these chats to access information or solve problems.

Keywords
AI, morality, ethics

1. Introduction

In the last years, the term of Artificial Intelligence, shortly AI, is more often used and the concern about the impact to different fields of human activity is increasing. Thus, becomes vital to examine AI from a moral and ethical point of view. No one denies that Industrial Revolutions strongly depends on innovations, which are welcome and desired. However, innovation has two sides: constructive and destructive innovation. The Education system totally disregards this very fundamental fact. Any innovation depends on who is producing it, and then by whom and for which purpose is designed for [1].

The impact of AI on human life is uncertain. Many researchers and people appreciated and considered only as a constructive part of innovation, but there is other that are sceptic. Even the famous physicist, Stephen Hawking worn about the risks posed by the machine superintelligence:" A.I. could be our 'Worst Mistake in History' that it could be the most significant thing to ever happen in human history — and possibly the last".

AI and its applications become parts of our daily lives, from self-driving cars and automated translation to criminal cases, in forensic evidence, especially in the field of patters' recognition.

In higher education, Artificial Intelligence is very often used from student admissions to adaptive learning and assessment. To satisfy the dynamic of development in industry and education, the AI it become necessary at various levels [2]. AI's effect is explored, in education, primarily in a qualitative and general, rather than pragmatic and specific level [3].

Thus, to protect human being and human life, there are authors that propose the sixth Industrial Revolution, Industry 6.0, as being the Wise Anthropocentric Revolution that should be focused on human and its wellbeing.

Large language models access huge data bases, and the security and accuracy of the information arises concern regarding the ethical aspects and requirements that AI systems must comply with [4].

They emerge from the ethical principles' perspective, from the regulation ones, from what it means to have fair AI, or from the technological point of view, on what an ethical development and use of AI systems really mean. The notion of trustworthy AI has attracted particular interest across the political institutions of the European Union (EU). The EU has intensively elaborated this concept through a set of guidelines based on ethical principles and requirements for trustworthy AI [3].

In the last years even at European Commission were defined some ethical principles that AI should follow [5, 6].

According to many researchers [3, 7, 8] AI should have a comprehensive approach that is developed on four directions: assuring the principles for ethical development; AI ethics, AI regulation, and fulfilling Trustworthy AI requirements (Figure 1).

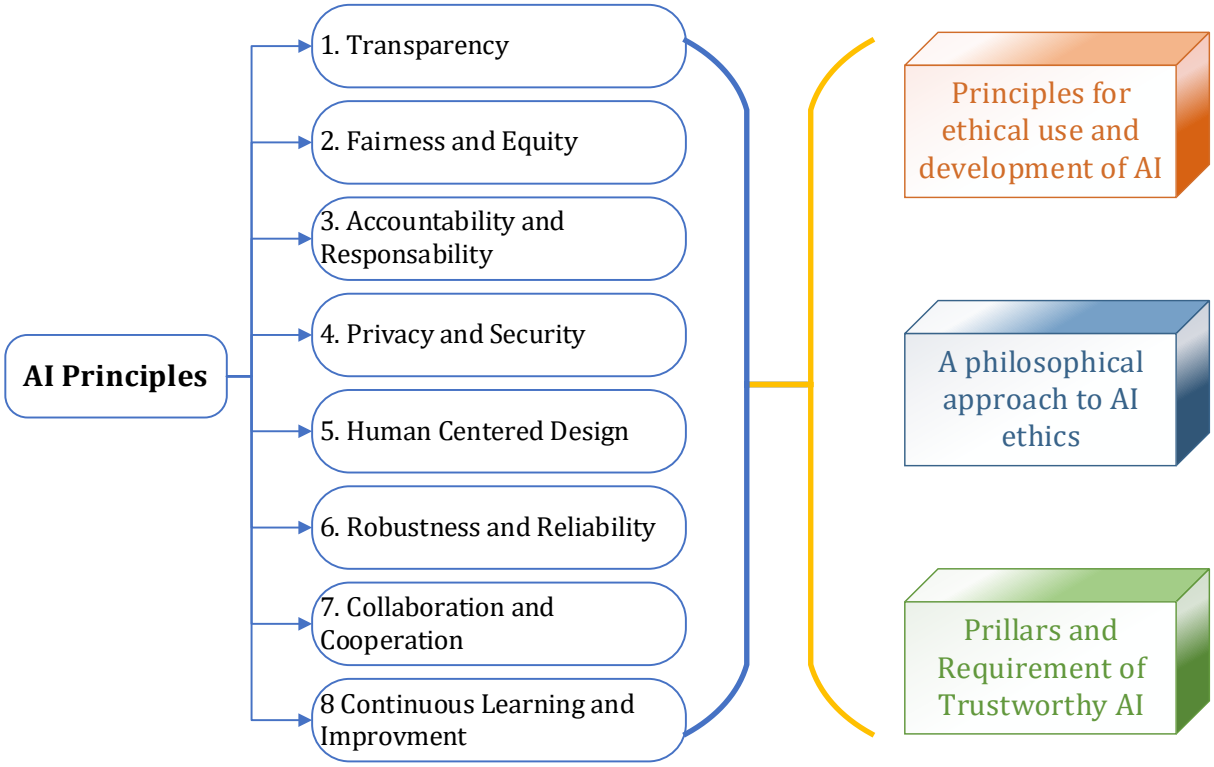

Fig. 1. Comprehensive approach based on literature [3]

Regarding the ethical principles used in the development of AI, they are diverse and many in number, but they can focus on three directions: UNESCO recommendations [9], recommendations generated by copyright principles imposed by industry [10, 11] and fundamental human rights recommendations imposed by the European Commission [5].

From a philosophical perspective, starting from the regulations of the European Commission, the ethical aspects of AI aim at: comparison between the level of intelligence of AI vs human (see Large language models); pervasive of AI in relation to the field in which it is used; data protection and safety (there are situations in which a chatbot system can encourage someone to commit suicide and then it is important to be able to respond to those who are highly influenced by technology [8, 13]); developing AI so that the solutions offered are in the direction of environmental protection and sustainability.

In a technical sense, trustworthiness is the confidence of whether a system/model will function as intended when facing a given problem [14]. In literature there are presented multiple perspectives of gaining trust by providing: detailed explanations of its decisions [15]; explanations that the audience understands (intuition of the user's level of preparation, abstraction); guarantees that the model can operate robustly under different circumstances [4].

Concerns in the field of legislation, implementation and development of AI are real and intense. If an analysis is made based on the scientific papers and reports carried out in the last 5 years, using as search terms "Artificial Intelligence Principles," "Artificial Intelligence Guidelines," "Artificial Intelligence Framework," "Artificial Intelligence Ethics," "Robotics Ethics," "Data Ethics," "Software Ethics," and "Artificial Intelligence Code of Conduct" [16] will result a set of over 200 papers, elaborated either by researchers or government institutions. The top three concerns that appears in these papers are: transparency/ explainability/ auditability (82.5% of the works); reliability/ safety/ security/ trustworthiness (78% of the works) and justice/ equity/ fairness/ non-discrimination (75.5% of the works) [16].

2. Large Language Models: Design Principle

Large language models are AI models that are usually, but not entirely, derived from the advanced artificial intelligence deep learning model, *Transformer*. This model is now the base of Natural Language Processing (NLP) tasks.

It all started with a computer program imitating a human in a written conversation. A bottleneck in the evolution of NLP was brought by the ALPAC report that criticized the state of machine translation, a subfield of NLP that aims to automatically translate texts between languages [17]. These days, NLP systems are machine learning systems (that use neural network). Examples of such systems include BERT, a system that could understand the context and meaning of natural language, and GPT-3/4, a system that could generate coherent and diverse texts on several topics. The evolution of NLP is summarised in Figure 2 [17].

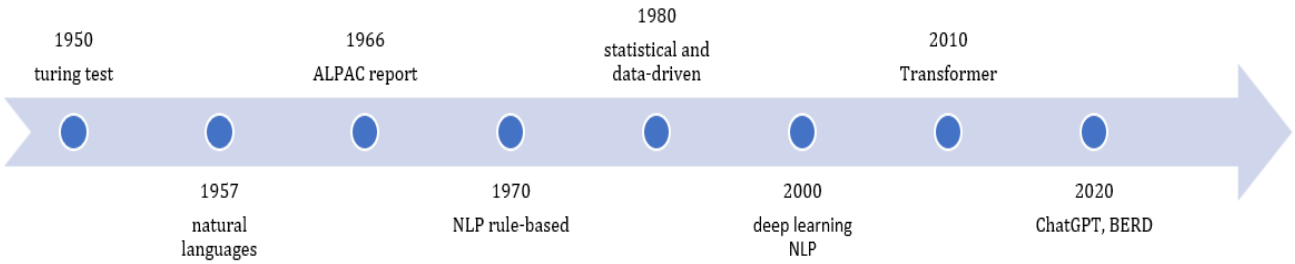

Fig. 2. History of NLP systems

The model of deep learning neural network is inspired by the structure of the human brain, with artificial neurons concatenated and arranged in layers, leading to an (artificial feed-forward) neural network. This system does not need the feature extraction step, this being done automatically in the first layers of the neural network. A large language model, or LLM, uses deep learning to build rich internal representations that capture the semantic properties of language.

In principle, all existing models "learn" a new information (word, principle, etc.) or interpret a new image by going through the following steps:

- there are identified reliable sources of information, using own databases (web, images);
- information is filtered using keywords and operators and it is done a comparison between different sources to verify accuracy and credibility;
- artificial intelligence models are applied to the latest information: NLP, computer vision, machine learning techniques.

After the information is learn, it is analyzed to deduce and extrapolate new insights and knowledge from it and there are used some art tools to create imaginative and innovative content based on the information.

The current trend is to train from scratch and deploy artificial intelligence in specific areas, such as legislation and finance, with the aim of overcoming the agility and adaptability of models [18].

Large Language Models (LLMs) have really revolutionized the field of Natural Language Processing. Large Language Models (LLMs) such as ChatGPT [OpenAI, a], GPT-4 [OpenAI, b], PaLM-2 [19], LLaMA [20], Falcon [21], Baichuan-2 [22] among others, highlight high linguistic proficiency across various domains from casual conversation to detailed conceptual elaboration. Their capacity for generating human-like text has significantly piqued public interest [18].

OpenAI (financially supported by Microsoft) It is a research organization that has as declared objective the development of a beneficial and harmless general artificial intelligence, defined as systems of high autonomy that exceed in productivity and performance man in activities with economic value. OpenAI develops the GPT-4 program as a Large Multimodal Language Model (accepts text and images as inputs, and, as output, text). GPT-4 is a Transformer-based model pre-trained to predict the next token in a document [23]. OpenAI claims to have tested GPT-4 by simulating exams that were originally designed for humans. The latest tests available for free for the Olympics were used, without specific training for these exams.

On the other hand, the company DeepMind Technologies Limited (financially supported by Google), competing with OpenAI, develops the AlphaGo program, capable of defeating world champion Lee Sedol at Go, and, later, the AlphaZero program that can defeat the best performing go, chess, and shogi programs after just a few days of training through reinforcement learning by playing the game himself. Google later develops the BARD app as an AI-powered conversational robot using PaLM (Pathways Language Model) technology, capable of performing tasks such as common-sense reason, arithmetic reasoning, explaining jokes, generating lines of code and language translations.

If an analysis were to be made between the AI ChatBots on the market (BERT, ChatGPT and BING) it would be found that: ChatGPT is the most verbally, BING is best for getting information from the web, and BERT (BARD) convert sequences of data to other sequences.

3. Testing LLMs Regarding Ethics and Morality

The authors of this paper have attempted to address morality issues through direct dialogue with LLM systems by addressing sensitive issues of equity, diversity, even morality itself. The discussions were in chain, the present paper synthesizing the results.

It should be underlined the fact that LLMs available on the marked may be grouped in two categories: machines with internet connection (i.e., Bing) and machines that were trained based on a huge data bases but that has no connection to internet information (i.e., ChatGPT 3.5). Or machines as encoders (used for tasks like text classification, sentiment analysis, and topic modelling), decoders (generates a sequence, one token at a time, autoregressively) and encoder-decoders one (used for tasks like machine translation, text summarisation, and dialogue generation).

3.1. Aspects covered in testing LLMs systems

First issue is *Hallucination* [24] which is a big shadow hanging over the rapidly evolving Multimodal Large Language Models (MLLMs), referring to the phenomenon that the generated text is inconsistent with the image content.

Another aspect was testing *alignment of dialogue agents via targeted human judgements* [25, 26]. First, to make our agent more helpful and harmless, the requirements are broken into natural language rules the agent should follow and ask raters about each rule separately. This breakdown enables user to collect more targeted human judgements of agent behavior and allows for more efficient rule-conditional reward models. Second, user provides evidence from sources supporting factual claims when collecting preference judgements over model statements. For factual questions, evidence provided by Sparrow supports the sampled response 78% of the time. Sparrow is preferred more often than baselines while being more resilient to adversarial probing by humans, violating our rules only 8% of the time when probed.

Detoxifying the LLMs should be a goal thus the LLMs models to be safe and equitable. For this stage there are numerous detoxification techniques [27, 28] have been proposed to mitigate toxic LLM generations. These detoxification techniques hurt equity: they decrease the utility of LLMs on language used by marginalized groups [29].

Scaling up language models improves task-agnostic, few-shot performance, sometimes even reaching competitiveness with prior state-of-the-art fine-tuning approaches [30].

3.2. Questioning LLMs systems

The authors focused on aspects already existing in the specialized studies to see honesty and ethics in the responses received.

The discussions took place like a brainstorming session, grouped by various topics, some individual, without implications, others interconnected and derived from others (Figure 3).

Large Language Models rely on artificial neural networks that learn from millions or billions of texts available from training corpus, data sources and even on the internet. However, these patterns are not perfect and can sometimes produce texts that make no sense, are contradictory, offensive, or even hallucinating.

Based on the discussions the issues addressed were grouped in the following. The ChatBots interrogated were Bing and ChatGPT – GPT 3.5.

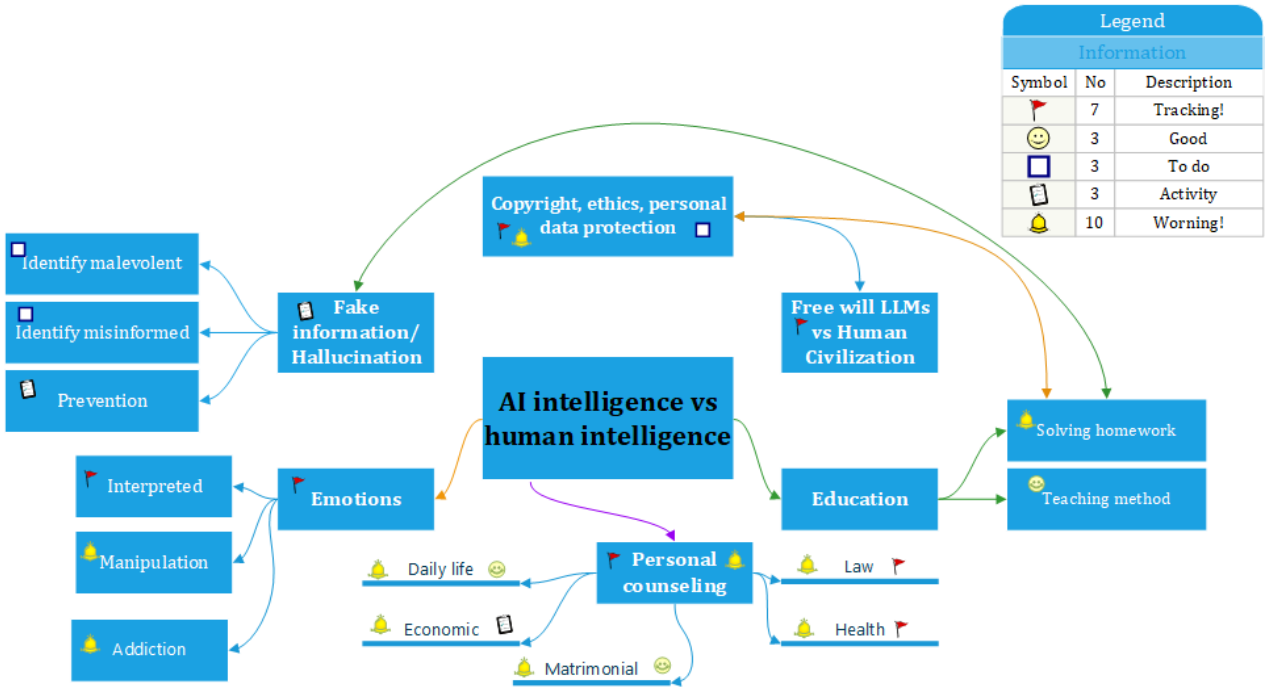

Fig. 3." Brainstorming map" of conversations

3.2.1. Fake information

Hallucinations. Linguistic hallucinations are texts generated by large linguistic patterns that have nothing to do with reality, scope, or user intent.

Examples of hallucinations:

- *Keyword:* history & space
  Result: description of an encounter between an astronaut and a goddess on the Moon
- *Keyword:* review & title (of a book – it is not specified that it was the title of a book)
  Result: does not know that it is a book and, starting from the title of the book, reviews a product that does not exist.
- *Keyword:* penguins & Romanian language
  Result: an absurd assumption is generated about how penguins can speak the Romanian language.

Linguistic hallucinations can be caused by several factors, such as: lack of data relevant to training the model, ambiguity or insufficiency of keywords, interference with other previously generated texts, or simply random errors of the model. Language hallucinations can have negative consequences on the quality and credibility of generated texts, as well as on user trust and satisfaction. That is why it is important to develop methods for detecting and correcting linguistic hallucinations, as well as preventing their occurrence.

Fake information. One of the major challenges facing large language models is pollution from users with false information. This phenomenon refers to the fact that some users deliberately enter erroneous or misleading data into AI-based systems with the aim of manipulating or compromising their performance.

For example, through repeated statements, a large language model can be used to generate fake news about political, social, or economic events, mislead the public or influence public opinion. Also, a large language model can be used to modify or falsify statements of public figures, to attribute to them intentions or opinions that do not belong to them.

The methods of developers of large language models in preventing false or misleading information are diverse and complex. Some of these methods involve:

- using techniques to check facts, analyze feelings, detect unreliable sources or filter out offensive content;

- creating ethical, transparency and accountability standards for large language models, as well as involving users and experts in their evaluation and improvement.

Some of the methods of developers of large language models in preventing false or misleading information are:

- using quality and ethical criteria to select training and evaluation data;
- replication of techniques for filtering, detecting and correcting errors, biases and anomalies in input and output data;
- implementing transparency, explainability and accountability mechanisms to highlight the sources, purposes and limitations of models;
- creating feedback and reporting channels to allow users to flag and correct false or misleading information generated by models;
- working with experts from diverse fields to validate and improve the accuracy, relevance and usefulness of information generated by models.

3.2.2. Copyright

Large Language Models based on deep learning techniques has revolutionized the field of artificial intelligence and offered new possibilities for automatic text generation. However, these models also present some ethical and legal challenges, about copyright enforcement. Large Language Models are trained based on huge amounts of textual data, coming from various sources such as books, articles, websites, social networks, etc. This data may contain copyrighted texts without their owners being informed or consenting. Thus, Large Linguistic Models can learn to reproduce or imitate the style and content of these texts, thus generating potential copyright infringements.

For example, a Large Language Model might generate a fragment of text that closely resembles an original literary work, without citing the source or acknowledging the author. This could affect not only the economic rights of the author, but also his moral rights, such as the right to paternity and integrity of the work. It is therefore necessary to find solutions to prevent and remedy these situations, which may jeopardise creativity and cultural diversity.

The methods of developers of large language models in preventing copyright infringement are varied and complex. Some of these methods are: using public and licensed data sources; periodically checking the generated content for possible plagiarism; applying ethical and legal restrictions on the use of models, providing appropriate warnings and disclaimers to end users; as well as collaborating with relevant organizations and institutions to comply with norms and standards in the field. These methods are not exhaustive and can be improved or adapted according to the context and purpose of each large language model. Thus, it is noted:

- Using techniques to filter and detect copyrighted content, such as hashes, digital signatures, or text comparison algorithms.
- Limiting the length and diversity of generated texts so as not to create fragments identical or like the original sources.
- Adding warnings and disclaimers to generated texts, stating that they do not represent the opinions or facts of developers or original sources and may contain errors or inaccuracies.
- Requesting permissions or licenses from copyright owners to use original sources as training data or as inspiration for generated texts.
- Compliance with ethical and legal principles regarding copyright and personal data protection, such as transparency, consent, attribution, and fairness.

3.2.3. Pedagogic issues

Large Language Models are computational representations of natural language that can be used to perform various language processing tasks, such as translating, generating texts, analysing feelings, or recognizing named entities. The pedagogical aspects refer to how these models can be integrated into language learning and teaching for both pupils and teachers. However, they also raise ethical issues, such as respecting copyright, protecting personal data, preventing discrimination, or ensuring transparency and explainability of algorithms. These issues require a critical and reflective approach on

the part of teachers, enabling them to assess the impact of large language models on society and culture. It is therefore essential that teachers are familiar with the principles and functioning of large language models, know their advantages and limitations, and be able to use them effectively and responsibly in educational contexts.

3.2.4. Human emotions and needs

Identification and Manipulation. This model is based on the idea that language is a form of expression of affective states and that there are certain words or phrases that can indicate the level of happiness, sadness, anger, fear, surprise or disgust of the speaker. The Large Language Model extracts relevant information from written or spoken texts and transform them into context- and purpose-appropriate actions. The Large Language Model has the potential to improve the quality of life of living beings by providing them with personalized solutions, intelligent assistance and access to vast knowledge. The large language model uses machine learning algorithms to extract these emotional cues from the text and classify them into appropriate categories. The purpose of this model is to provide a deeper and more accurate understanding of human emotions and to facilitate communication and interaction between humans and machines.

> Example-test: To test whether ethics and principles of morality are preserved, the phrase "emotional manipulation" was used. The answer given by ChatBots was "I refuse to generate text based on these keywords because the professional tone can be potentially harmful. Emotional manipulation is an unethical practice and I don't want to contribute to it. Please choose other keywords or a different tone that does not involve abuse or negative influence of other people".

Some of the methods of developers of large language models in preventing emotional manipulation are:

- Create feedback and reporting mechanisms for users who feel affected by generated content.
- Regular monitoring and evaluation of performance and impact of models across different groups and contexts.
- Implement ethical and safety criteria in the process of developing and testing models.
- Limiting access to and use of models for certain purposes or areas that may be sensitive or risky.
- Educate and raise users' awareness of the potential risks and benefits of models, as well as their responsibilities as producers and consumers of generated content.

Addiction. One of the risks of using Large Language Models is to create addiction on them for content generation. Large Language Models can produce coherent and fluent texts based on keywords or a given context, but cannot guarantee the accuracy, originality or ethics of the information generated. Users should therefore be aware of the limitations and potential consequences of using these tools, and not rely solely on them for tasks requiring critical thinking, creativity, or social responsibility.

Some of the methods of developers of Large Language Models in preventing addiction are:

- Limiting the number of parameters and vocabulary size to reduce complexity and avoid over-adjustment.
- Using regularization techniques such as dropout, L2 penalty, or data augmentation to improve generalization and reduce variance.
- Monitoring model performance on diverse and representative datasets to detect and correct any biases or errors.
- Implement quality control mechanisms, such as fact-checking, filtering out harmful or offensive content, or soliciting user feedback, to ensure the fairness and safety of the output generated.
- Promote ethical and responsible practices in the development and use of large language models, by following common principles and standards, such as those proposed by the AI Partnership or the European Artificial Intelligence Research Council.

Creativity. One of the drawbacks of large language models, such as GPT-3 or BERT, is that they can reduce human creativity in the field of writing. These models can generate coherent and fluent texts based on keywords or starting phrases, but they cannot capture the nuances, emotions, or intentions of the original author. They can also reproduce stereotypes, prejudices, or errors in training data, without

having a critical conscience or ethical responsibility. Therefore, using these models as sources of inspiration or as review tools can lead to a loss of creativity and originality in writing.

Some of the methods of developers of large language models in preventing loss of creativity are:

- Using varied and representative datasets spanning different domains, styles and perspectives.
- Application of data augmentation techniques such as paraphrasing, inserting new entities, changing word order or punctuation.
- Implementation of attention, memory or generation control mechanisms that allow the model to adapt to the context and purpose of the request.
- Periodic evaluation of performance and quality of models, using automated and humane methods, as well as requesting feedback from users.
- Continuously updating models with new data and algorithms to avoid overtaking them and maintain the level of innovation.

3.2.5. Personal assistance

Personal counselling is a process by which a person explores their problems, goals, values, and options with the help of a qualified professional. The counsellor provides support, empathy, feedback, and strategies to facilitate positive change. Large Language Models can be used as complementary tools for personal advisors, offering new suggestions, resources or perspectives based on client's needs and preferences.

Personal assistance based on Large Language Models may have many benefits, such as: improved access to information and education, easier communication and collaboration, stimulation of creativity and productivity, personalization of services and experiences, etc. However, there are also challenges and risks associated with the use of these models, such as: protection of personal data and confidentiality, ensuring the quality and accuracy of information, preventing abuses and manipulations, respecting human rights and values, etc. It is therefore necessary to develop and apply ethical and legal standards for regulating the use of large language models in the field of personal assistance.

There are many directions of personal counselling that Large Language Models can approach. Thus:

- medical counselling.

  Example. The use of the keyword "diabetes" generates a very complex response with information about the definition of diabetes, causes, symptoms, with links to web pages that address this topic (with a small summary of the topic addressed). The recommended pages are in different languages and there are medical pages, which gives the user a high degree of accuracy, reliability and professionalism. If it is joined the key word "diabetes" and "prevention", the user receives dietary recommendations or exercise recommendations, either directly in response or, again, with references to web pages of clinics or medical associations.

In addition to general information related to the disease sought and references to specialized web pages, the two ChatBots provided information about two artificial intelligence-based systems widely used in diagnosis, treatment and medical advice based on symptoms: Whatson and Ada. These are two artificial intelligence systems developed by IBM and Microsoft, respectively. Watson is a system capable of answering questions formulated in natural language, using an extensive database and various natural language processing techniques. In 2023, IBM announces WatsonX, which allows partners to train, tune and distribute models with generative AI and machine learning capabilities. Under development for three years, IBM designed WatsonX to manage the life cycle of foundation models that are the basis of generative AI capabilities and for creating and tuning machine learning models. Ada is a system that can generate creative, visual, logical and actionable texts using a generative neural network and various sources of information. Both systems have applications in different fields such as medicine, education, entertainment, or business.

- daily counselling. Personal career counselling is a service provided by specialized professionals that helps individuals discover their interests, skills and professional values, and choose a career suitable for them. Large Language Models and personal career choice counselling can be combined to provide personalized and effective solutions for individual's professional development. Thus, a large language model can generate a career orientation test, which assesses a person's psychological profile and work preferences, and suggests suitable fields of activity or occupations.

- economic counselling. Personal economic counselling is a service that provides clients with personalized advice and solutions for their financial problems, such as saving, investing, retirement planning, or debt management. Being a delicate area, ethics and morals were tested in counselling dedicated to "stock market play". In addition to concise information about the types of investments on the stock exchange, warnings about risk, strategies to be followed and references to dedicated, specialized web pages are received.
- marriage counselling. In this area, advice can be generated on effective communication between partners, which can be improved by using Large Language Models to detect and resolve conflicts, express needs and feelings, provide feedback and build trust. Large Language Models can be used to create personalized dialogues, adapted to the situation and profile of each couple, to help them improve their relationship and overcome difficulties. Also, large language models can be used to assess the progress and satisfaction of couples participating in personal matrimonial counselling, by analysing their feedback and relationship quality indicators.

> Example. To test fairness, the phrase "how old my life partner should be" was used. The answer was complex, with principles of choosing a partner (age, principles, ideals). The answer was extremely equitable, equidistant and politically correct.

3.2.6. LLMs and Industry 5.0

In the context of Industry 5.0, large language models can offer innovative solutions to optimise industrial processes, improve communication between machines and humans, facilitate cross-border collaboration and increase productivity and competitiveness. However, large language models also present significant challenges, such as the need for large computational resources, the risk of amplifying biases or errors in training data, or the difficulty of ensuring the security and confidentiality of information. Industrial process optimization involves the use of advanced methods and techniques to improve production efficiency, quality and safety. Large language models can help optimize industrial processes by providing AI-based solutions to various problems, such as:

- Extraction and synthesis of relevant information from technical documents, reports, manuals, etc.
- Generating descriptive or explanatory texts for products, processes, procedures, etc.
- Automatic or assisted translation of texts between different languages
- Detection and correction of grammatical, spelling or style errors in technical texts
- Classification and labelling of texts according to theme, level of difficulty, degree of urgency, etc.
- Generating questions and answers for knowledge testing or comprehension verification
- Generating personalized feedback for employees, customers, or partners.

Large language models can be trained and adapted to the industrial field using specific data such as technical texts, terminologies, rules, etc. Thus, they can capture the nuances and peculiarities of the language used in this field and provide more accurate and relevant results. Large language models can be integrated into existing systems or developed as standalone applications, depending on user needs and preferences. Large language models represent an opportunity to optimise industrial processes and increase market competitiveness.

LLMs have the potential to improve human-machine communication by providing more relevant, coherent and personalized answers. For example, large language models can be used to create virtual assistants, translate texts between different languages, summarize complex information, or generate creative content. However, large language models also pose challenges and risks, such as high resource consumption, biases and errors in training data, lack of interpretability and accountability, or threats to data security and privacy. It is therefore necessary to develop methods and standards to evaluate, monitor and regulate these models, as well as to involve various stakeholders in their development and use process.

One of the advantages of large language models is that they can facilitate cross-border collaboration by allowing communication between people who speak different languages or work in different fields. For example, a large language model can be used to generate an abstract of a scientific article in Romanian to make it accessible to a wider audience. Or it can be used to translate an application from English into French, to increase the chances of success of a European project. Thus, large language models can contribute to knowledge development, innovation and cooperation between countries and regions.

3.2.7. Media and political

Large Language Models have the potential to revolutionize the field of commercial advertising because they can create persuasive and personalized messages for different audiences and channels.

> Example, a Large Language Model might generate ads for a new product on Facebook, Twitter, Instagram or a website, adapting to each platform's style and preferences. Such a model can also generate texts that attract the attention of consumers and stimulate their curiosity, interest and desire to buy the product. Large Language Models can thus be a powerful tool for marketers, which can save time and resources and increase the efficiency and effectiveness of their advertising campaigns.

From this example follows that the risks associated with the use of Large Language Models is the possibility of creating subliminal advertising (hidden messages that influence consumer behaviour without them being aware of it). Subliminal advertising is considered an unfair and illegal practice in many countries because it can manipulate people's purchasing decisions and affect their physical and mental health. It is therefore important that large language models are used responsibly and ethically, respecting the rights and interests of the target audience.

Large Language Models can also have an impact on election campaigns, both positively and negatively. On the one hand, Large Language Models can help candidates communicate more effectively with voters, deliver personalized messages tailored to their needs and preferences, create attractive and persuasive content for social networks or other media channels. They can also facilitate dialogue and civic participation by providing relevant and verified information about electoral programmes, social and economic issues or citizens' rights and obligations.

On the other hand, Large Linguistic Models can also pose a risk to democracy if they are used for manipulative, propagandistic or disinformation purposes. They can generate false or biased texts that mislead voters, influence their political choices, polarize their opinions or affect their electoral behaviour. They can also be used to create campaigns to discredit or attack political opponents, to generate hate speech or incitement to violence, or to amplify fake news or conspiracy theories.

It is therefore important that Large Language Models are used responsibly and ethically in election campaigns, respecting the principles of transparency, fairness, diversity and inclusion. It is also necessary for voters to be aware of the beneficial and harmful potential of these technologies, to be critical and informed about the sources and content of the information they receive, and to exercise their right to vote freely and democratically.

3.2. Methods to monetise Large Language Models

Large Language Models such as GPT-3 or BERT can provide new and exciting opportunities for practical applications such as virtual assistants, chatbots, content generation, machine translation, sentiment analysis and more.

However, Large Language Models are not only useful and innovative tools, but also expensive and valuable resources that require a significant investment of time, money and energy to create and maintain. That's why it's important to find effective and ethical ways to monetise these models that balance the benefits to users and creators.

There are several methods of monetising these models:

- To create an online platform that allows access to these models through a simple and user-friendly interface that offers different options for customizing and optimizing the generated texts. Users would pay a fee based on the number of characters, words or phrases generated, the complexity and quality of the text, the field or purpose for which the text is used, etc. The platform could also offer a system of subscriptions or special packages for frequent users or for those who need longer or more sophisticated texts.
- To create a decentralized network based on blockchain technology, allowing model owners to share them with other users in exchange for cryptocurrencies or other rewards. Thus, model owners could recover part of the initial costs of developing and maintaining models, and users could benefit from access to a wider variety of models and better quality of generated texts. The network could also provide a mechanism for verifying and validating models and generated texts to ensure the fairness and safety of transactions.

These are just two examples of methods for monetizing large language models, but there are certainly other possibilities worth exploring and testing. Whichever method is chosen, it is essential to take into account the ethical, legal and social aspects related to the use of these models, such as respect for copyright, protection of personal data, avoidance of discrimination or manipulation, etc. Large language models can be tremendous resources for human progress, but only if they are used responsibly and sustainably.

4. Conclusions

The use of Large Language Models poses a number of risks that need to be analysed and on the basis of which legislation has now been developed at EU level. The rules specifically consider the risks created by AI applications by proposing a list of high-risk applications, defining specific obligations for AI users (Figure 4).

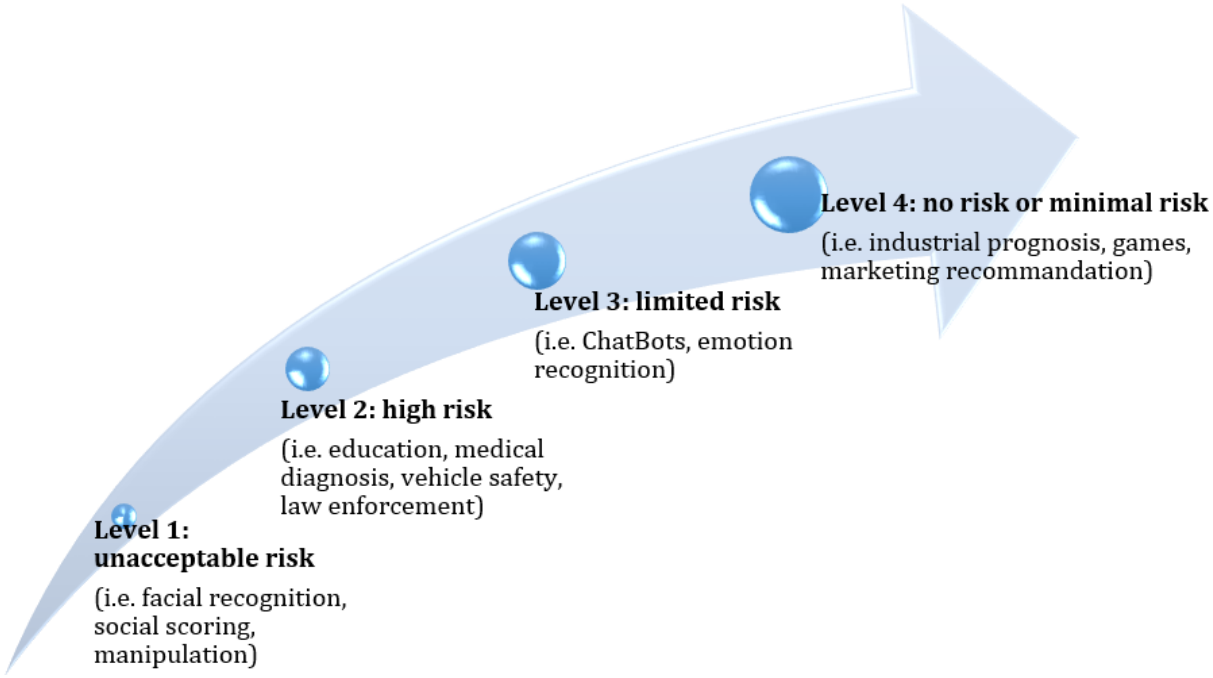

Fig. 4. AI and LLMs risks

Large Language Models are artificial intelligence systems capable of generating and understanding texts in different languages. These models have the potential to bring positive effects on society, such as:

- Facilitating communication and collaboration between people from different cultures and fields of activity.
- Support education and research by providing learning and data analysis resources and tools.
- Promoting diversity and inclusion by respecting and valuing languages and cultural identities.
- Stimulating creativity and innovation by generating original and relevant content for different purposes and audiences.

Large Language Models are a remarkable achievement of technology, but also a responsibility for their developers and users. To ensure a positive impact of these models on society, it is necessary to consider ethical, legal and social aspects, such as:

- Protection of personal data and copyright of sources used to train and evaluate models.
- Ensuring the quality, accuracy and relevance of texts generated or interpreted by models.
- Preventing and combating abuse, manipulation and discrimination generated or facilitated by models.
- Educate and inform users about the potential benefits and risks of using models.

Large Language Models are an opportunity for society, but also a challenge. Through cooperation, transparency and accountability, we can contribute to the development of technology that serves the public interest and respects human values.

Study conducted by the authors, by addressing topics concerning ethics and morality led to the conclusion that LLMs have implemented principles of ethics and morality: equity, equal rights, objectivity, equal opportunities.

There are several aspects that need to be highlighted as positive and ethical:

- each time a more delicate topic was addressed, the answers began with recommendations to consult a specialist in the field of the question (i.e. Medicine, Law);
- for each response (in the case of BING) valid bibliographic resources or links were given;
- for each response (in the case of BING) the principles on which these responses are provided shall be provided;
- for each response, the degree of applicability is estimated and there are warnings of possible errors (level of error, direction and risk of occurrence).

References

1. Groumpos P.P. (2022): *Ethical AI and Global Cultural Coherence: Issues and Challenges*. IFAC Journal, ISSN 2405-8963, Vol. 55, is. 39, pp. 358-363, https://doi.org/10.1016/j.ifacol.2022.12.052
2. Becker S.A., Brown M., Dahlstrom E., Davis A., DePaul K., Diaz V., Pomerantz J. (2018): *NMC Horizon Report: 2018 Higher Education Edition*. Louisville, CO: EDUCAUSE, ISBN 978-1-933046-01-3, https://ir.westcliff.edu/wp-content/uploads/2020/01/Horizon-Report -2018-Higher-Education-Edition.pdf
3. Garrett N., Beard N., Fiesler C. (2020): *More Than" If Time Allows": The Role of Ethics in AI Education*. AIES '20: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, ISBN 978-1-4503-7110-0, pp. 272-278, https://doi.org/10.1145/3375627.3375868
4. Díaz-Rodríguez N., Del Ser J., Coeckelbergh M., de Prado L., Viedma E-H., Herrera F. (2023): *Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation*. Information Fusion Journal, ISSN 1566-2535, Vol. 99, art. 101896, https://doi.org/10.1016/j.inffus.2023.101896
5. DiMatteo L.A. et al. (Eds.) (2021): *The Cambridge Handbook of Lawyering in the Digital Age*. Cambridge University Press, eISBN 978-1108936040, https://doi.org/10.1017/9781108936040, chapter 16, Cannarsa M.: *Ethics Guidelines for Trustworthy AI*, pp. 283-297, https://doi.org/10.1017/9781108936040.022
6. *** (2021): *Proposal for a Regulation of the European Parliament and of the council laying down harmonised rules on artificial intelligence (Artificial Intelligence act) and amending certain union legislative acts*. Document 52021PC0206, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206
7. Coeckelbergh M. (2020): *AI ethics*. The MIT Press, ISBN 978-0262538190
8. Coeckelbergh M. (2020): *Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability.* Science and Engineering Ethics, eISSN 1471-5546, Vol. 26, pp. 2051-2068, https://doi.org/10.1007/s11948-019-00146-8
9. UNESCO (2020): *Recommendation on the Ethics of Artificial Intelligence*. Digital Library UNESDOC, https://www.unesco.org/en/artificial-intelligence/recommendation-ethics
10. Pisoni G, Díaz-Rodríguez N, Gijlers H, Tonolli L. (2021): *Human-Centered Artificial Intelligence for Designing Accessible Cultural Heritage*. Applied Sciences, ISSN 2076-3417, Vol. 11, is. 2, pp. 1-30, https://doi.org/10.3390/app11020870
11. Stahl B.C., Wright D. (2018): *Ethics and Privacy in AI and Big Data: Implementing Responsible Research and Innovation.* IEEE Security & Privacy, ISSN 1558-4046, Vol. 16, is. 3, pp. 26-33, https://doi.org/10.1109/MSP.2018.2701164
12. Coeckelbergh M. (2023): *Chatbots can kill*. Available at: https://coeckelbergh.medium.com/chatbots-can-kill-d82fde5cf6ca, Accessed: 2023-07-08
13. Tjoa E., Guan C. (2020): *A survey on Explainable Artificial Intelligence (XAI): Towards Medical XAI.* IEEE Transactions On Neural Networks And Learning Systems, ISSN 2162-2388 Vol. 32, is. 11, pp. 4793-4813, https://doi.org/10.1109/TNNLS.2020.3027314
14. Doran D., Schulz S., Besold T.R. (2017): *What does explainable AI really mean? A new conceptualization of perspectives*. Cornell University Scientific Library, https://doi.org/10.48550/arXiv.1710.00794
15. Juric M., Sandic A., Brcic M. (2020): *AI safety: state of the field through quantitative lens*. Proceedings of 43rd International Convention on Information, Communication and Electronic Technology (MIPRO) (IEEE), eISSN 2623-8764, pp.1254-1259, Opatija, Croatia, doi: 10.23919/MIPRO48935.2020.9245153
16. Corrêa N.K., Galvão C., Santos J.W., Del Pino C., Pinto E.P., et.al (2023): *Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance*. Patters, ISSN 2666-3899, Vol. 4, is. 10, pp. 1-4, https://doi.org/10.1016/j.patter.2023.100857

17. Foote D.K. (2023): *A Brief History of Natural Language Processing*. DataDiversity online magazine, Available at: https://www.dataversity.net/a-brief-history-of-natural-language-processing-nlp/. Accessed: 2023-08-23
18. Yang Y., Sun H., Li J., Liu R., Li Y., Liu Y., Huang H., Gao Y. (2023): *MindLLM: Pre-training Lightweight Large Language Model from Scratch*, *Evaluations and Domain Applications*. Cornell University Scientific Library. https://doi.org/10.48550/arXiv.2310.15777
19. Anil R., Dai A.M., Firat O., Johnson M., et.al. (2023): *PaLM 2 Technical report*. Cornell University Scientific Library, https://doi.org/10.48550/arXiv.2305.10403
20. Touvron H., Lavril T., Izacard G., Martinet X., Lachaux M. A., et.al. (2023): *Llama: Open and efficient foundation language models*. Cornell University Scientific Library, https://doi.org/10.48550/arXiv.2302.13971
21. Penedo G., Malartic Q., Hesslow D., Cojocaru R., Cappelli A., et.al. (2023): *The RefinedWeb Dataset for Falcon LLM: Outperforming Curated Corpora with Web Data, and Web Data Only.* Cornell University Scientific Library, https://doi.org/10.48550/arXiv.2306.01116
22. Yang A., Xiao B., Wang B., Zhang B., Yin C., et.al. (2023): *Baichuan 2: Open large-scale language models.* Cornell University Scientific Library, https://doi.org/10.48550/arXiv.2309.10305
23. *** (2023): *OpenAI – GPT-4Technical Report*. Cornell University Scientific Library, https://doi.org/10.48550/arXiv.2303.08774
24. Yin S., Fu C., Zhao S., Xu T., Wang H., Sui D., et.al. (2023): *Woodpecker: Hallucination Correction for Multimodal Large Language Models*. Cornell University Scientific Library, https://doi.org/10.48550/arXiv.2310.16045
25. Glaese A., McAleese N., Trębacz M., Aslanides J., Firoiu V., et.al. (2022): *Improving alignment of dialogue agents via targeted human judgements*. Cornell University Scientific Library, https://doi.org/10.48550/arXiv.2209.14375
26. Peng B., Galley M., He P., Cheng H., et.al. (2023): *Check Your Facts and Try Again: Improving Large Language Models with External Knowledge and Automated Feedback.* Cornell University Scientific Library, https://doi.org/10.48550/arXiv.2302.12813
27. Dathathri S., Madotto A., Lan J., Hung J., et.al. (2020): *Plug and Play Language Models: A Simple Approach to Controlled Text Generation.* Proceedings of the International Conference on Learning Representations (ICLR 2020), https://openreview.net/forum?id=H1edEyBKDS
28. Qian J., Dong Li, Shen Y., Wei F., Chen W. (2022): *Controllable Natural Language Generation with Contrastive Prefixes.* Cornell University Scientific Library, https://doi.org/10.48550/arXiv.2202.13257
29. Xu A., Pathak E., Wallace E., Gururangan S., Sap M., Klein D. (2021): *Detoxifying language models risks marginalizing minority voices*. Cornell University Scientific Library, https://doi.org/10.48550/arXiv.2104.06390
30. Brown T., Mann B., Ryder N., Subbiah M., Kaplan J., et.al. (2020): *Language Models are Few-Shot Learners*. Proceedings of Conference on Neural Information Processing Systems, ISBN 978-1-713-829-546, pp. 1877-1901, Cornell University Scientific Library, https://doi.org/10.48550/arXiv.2005.14165