

A dialogo con ChatGPT su potenzialità e limiti dell'IA per la valutazione in educazione

In dialogue with ChatGPT on the potential and limitations of AI for evaluation in education

Giuseppe C. Pillera

Researcher | INVALSI | giuseppe.pillera@invalsi.it

OPEN ACCESS

Siped
Società Italiana di Pedagogia

Double blind peer review

Citation: Pillera, G.C. (2023). In dialogue with ChatGPT on the potential and limitations of AI for evaluation in education. *Pedagogia oggi*, 21(1), 301-315.

<https://doi.org/10.7346/PO-012023-36>

Copyright: © 2023 Author(s). This is an open access, peer-reviewed article published by Pensa MultiMedia and distributed under the terms of the Creative Commons Attribution 4.0 International, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. *Pedagogia oggi* is the official journal of Società Italiana di Pedagogia (www.siped.it).

Journal Homepage

<https://ojs.pensamultimedia.it/index.php/siped>

Pensa MultiMedia / ISSN 2611-6561

<https://doi10.7346/PO-012023-36>

ABSTRACT

The development of Deep Learning (DP) techniques has enabled Artificial Intelligence (AI) applications in a wide range of economic, political and cultural areas, opening up new perspectives (OECD, 2022; la.italia.it, 2023) that present enormous opportunities but also pose great risks and ethical questions (Harari, 2018). As a result, the interest in AI's educational applications has also increased (UNESCO, 2021; 2022; Goksel, Bozkurt, 2019; Pancioli *et alii*, 2020; Pham, Sampson, 2022). Following pioneering works that consider natural language processing algorithms as expert interlocutors, this contribution presents and discusses the results of a conversation with ChatGPT – an evolved chatbot prototype – on the opportunities and limitations of AI in supporting evaluation processes in education.

Lo sviluppo di tecniche di Deep Learning ha permesso applicazioni di Intelligenza Artificiale (IA) nei più svariati ambiti economici, politici, culturali, aprendo prospettive inedite (OECD, 2022; la.italia.it, 2023) che pongono enormi opportunità ma altrettanto grandi rischi e interrogativi etici (Harari, 2018). È di conseguenza cresciuto anche l'interesse verso le applicazioni educative delle nuove generazioni di IA (UNESCO, 2021; 2022; Goksel, Bozkurt, 2019; Pancioli *et alii*, 2020; Pham, Sampson, 2022). Sulla scia di lavori pionieristici che assumono algoritmi di Natural Language Processing quali interlocutori esperti, questo contributo presenta e discute i risultati di una conversazione con ChatGPT – evoluto prototipo di chatbot – su potenzialità e limiti dell'IA nel supportare processi valutativi in educazione.

Keywords: artificial intelligence; natural language processing; large language models; ChatGPT; educational evaluation

Parole chiave: intelligenza artificiale; natural language processing; large language models; ChatGPT; valutazione in educazione

Received: April 02, 2023

Accepted: May 04, 2023

Published: June 30, 2023

Corresponding Author:

Giuseppe C. Pillera, giuseppe.pillera@invalsi.it

1. AI, education and evaluation: a theoretical framework

In the last decade, the development of Deep Learning (DL) techniques based on artificial neural networks pushed machine learning (ML) one step further (Goksel, Bozkurt, 2019). Thus, Artificial Intelligence (AI) applications are spreading across political, social, cultural areas, with tasks such as big data analysis, modelling, predicting, and problem-solving. This has opened up new perspectives (Ia.it.it, 2023) that present enormous opportunities but also challenges and ethical questions, as AI tends to be used to make decisions that can strongly impact the lives of individuals and communities (Harari, 2018).

In the same year, 2019, there were two attempts to establish international reference guidelines. The OECD (2022) *Recommendation on AI* stated the principles of inclusiveness, sustainability and well-being, human-centred values and fairness, accountability, transparency and explainability, robustness, security and safety. The European *Ethics Guidelines for Trustworthy AI* attempted to design a framework inspired by similar values, defining evaluable “key requirements” that emphasize human autonomy, agency and oversight (High-Level Expert Group on Artificial Intelligence, 2019). Two year later, similarly, the UNESCO (2022) *Recommendation on the Ethics of AI* developed a framework of ethical values and consequent principles, highlighting environment and international peace. In addition, UNESCO (2021) formulated an *AI and education Guidance for policy-makers*.

At the same time, the interest of academic community in AI applications in education (AIEd) – already a long-standing field of studies (Ferguson, 2012; Luckin, 2017; Chen *et alii*, 2022) – has grown, as demonstrated by several recent systematic reviews and meta-analysis (Celik *et alii*, 2022; Chen *et alii*, 2022; Fahd *et alii*, 2022; Fahimirad, Kotamjani 2018; Ouyang, Zheng, Jiao, 2022; Pham, Sampson, 2022; Zhai *et alii*, 2020). One of the most comprehensive, a systematic review of 4,519 scientific papers from 2000 to 2019 (Chen *et alii*, 2022), reveals the main and fastest growing research topics in AIEd, most of them linked to evaluative issues: intelligent tutoring systems for special education, recommender systems for personalized learning, natural language processing (NLP) for language education, data mining for performance prediction, discourse analysis in computer-supported collaborative learning, neural networks for assessment, affective computing for learner emotion detection.

Celik *et alii* (2022) offer a landscape of studies that show opportunities to improve all phases of teacher’s work, each related to evaluative implications or applications: planning (e.g., by defining students’ needs and familiarizing teachers with such needs); implementation (e.g., through immediate feedback and teacher intervention); assessment (e.g., through automated essay scoring)¹.

Specifically regarding evaluative questions, Fahd *et alii* (2022) conducted a meta-analysis of 89 relevant studies published from 2010 to 2020 on the application of ML in higher education to assess students’ academic performance and to evaluate at-risk and attrition, identifying the most used ML types, models, evaluation metrics, and other related demographics. The systematic review by Zhai *et alii* (2020) reports text recognition, classification, and scoring as the primary fields of study and supervised ML as the predominant approach in *Applying Machine Learning in Science Assessment*.

2. What is ChatGPT

NLP has been one of the most turbulent development fields of AI, leading to the creation of large language models (LLMs). Their neural networks – with billions of parameters (comparable to neurons) – are trained on extensive databases of unlabelled text (unsupervised ML). The so called GPT (generative pre-trained transformers) can generate human-like text, probabilistically predicting word-by-word since a large text corpus. They have demonstrated to effectively perform a wide variety of tasks, and this is arousing a lot of curiosity and fear among the educational community of researchers and practitioners, especially with the diffusion of ChatGPT.

ChatGPT was released on 2022, November the 20th, as the third GPT generation by the US company

1 The same study also found that teachers have various roles in the development of AI technology, including acting as models for training AI algorithms and checking the accuracy of AI assessment systems.

OpenAI (<https://openai.com>). Compared to the 40 billion parameters of GPT-2, the approximately 175 billion of GPT3 represent a significant leap forward in publicly available NLP applications² and the results are amazing. The International Association for Educational Assessment recently published an overview section in its website titled *OpenAI: cheating temptation or educational assessment's biggest gamechanger?* (<https://iaea.info/openai-and-educational-assessment>).

ChatGPT “tends to answer as humans do”, “is sometimes able to correct its mistakes”, “may request for additional information to provide an answer and may be asked to explain its response” (Azaria, 2022, pp. 6-7). Moreover, it has deductive skills and can act as a problem solver (Noever, McKee, 2023) and its texts are almost indistinguishable from human ones (Noever, Ciolino, 2022).

Nevertheless, it seems to lack the ability of learning new information, sometimes shows over-confidence in its incorrect responses, and a minor change to a question may lead to contradicting answers. In addition, the GPTs work well with the languages but without a moral compass or real self-awareness: consequently, they can reproduce biases and stereotypes learnt through the emergent representation of society from their databases, and they can occasionally generate unfair or toxic language, false attributions and fabrication of sources, even hallucinations. These problems can be inhibited at the programming stage, but they can be circumvented (Blum, 2022).

Hence, ChatGPT raises many educational challenges and questions. How will it affect our pedagogical-didactical models? What opportunities does it present? What will change in classroom practice or in individual study? What will change in evaluation and assessment? Much concern was aroused by its potential for fraudulent use by scholars (Gao *et alii*, 2022) and students (King, ChatGPT, 2023; Zhai, 2022): does ChatGPT represent *The End of Online Exam Integrity* (Susnjak, 2022)?

3. Methodology and objectives

This contribution presents, analyses and discusses the results of a conversation with ChatGPT on the potential and limitations of AI (and in some respects of ChatGPT itself) in supporting evaluation processes in education.

Without denying its nature as “cultural constructs” (De Castro, Zona, Bocci, 2020), ChatGPT is assumed, if not as expert system, at least as witness of the public and scientific discourse, in the wake of other similar studies (Cahan, Treutlein, 2022; King, ChatGPT, 2023; O’Connor, ChatGPT, 2023; Zhai, 2022). Using a post-structuralist and somewhat *post-human* ethnographic approach, we interrogated ChatGPT with the aim to verify if it can be a valid conversation partner for teachers, students or scholars looking for information or help, suggestions or even insights or ideas about evaluation in education. By doing so, we have explored its knowledge base on the subject and tested its ability to present themes and arguments. Additionally, we wanted to test its ability to recognize contradictions and to correct itself.

In this study, the formulation of requests to ChatGPT – the so-called *prompting* – simulated the methodology of a free interview. Starting from a generic conversation request on our subject, ChatGPT generated six topics (fig. 1), from which the discussion took its cue, gradually deepened with more specific requests. Towards the end of the interview, some further issues were addressed and a request for ethical guidelines was prompted.

The interview was conducted between February and March 2023, during several sessions of the same chat. We used 35 prompts, which ChatGPT answered generating 5,710 words (on average 163 words per answer and 24.7 words per sentence) and 1,077 unique words³. The most frequent words are illustrated in fig. 2.

2 OpenAI recently introduced GPT-4, which is declared to produce “safer and more useful responses” (<https://openai.com/product/gpt-4>).

3 The vocabulary density is 0.189 (e.g., it is 0.150 for UNESCO, 2021). These simple analyses were conducted using Voyant Tools (<https://voyant-tools.org>).

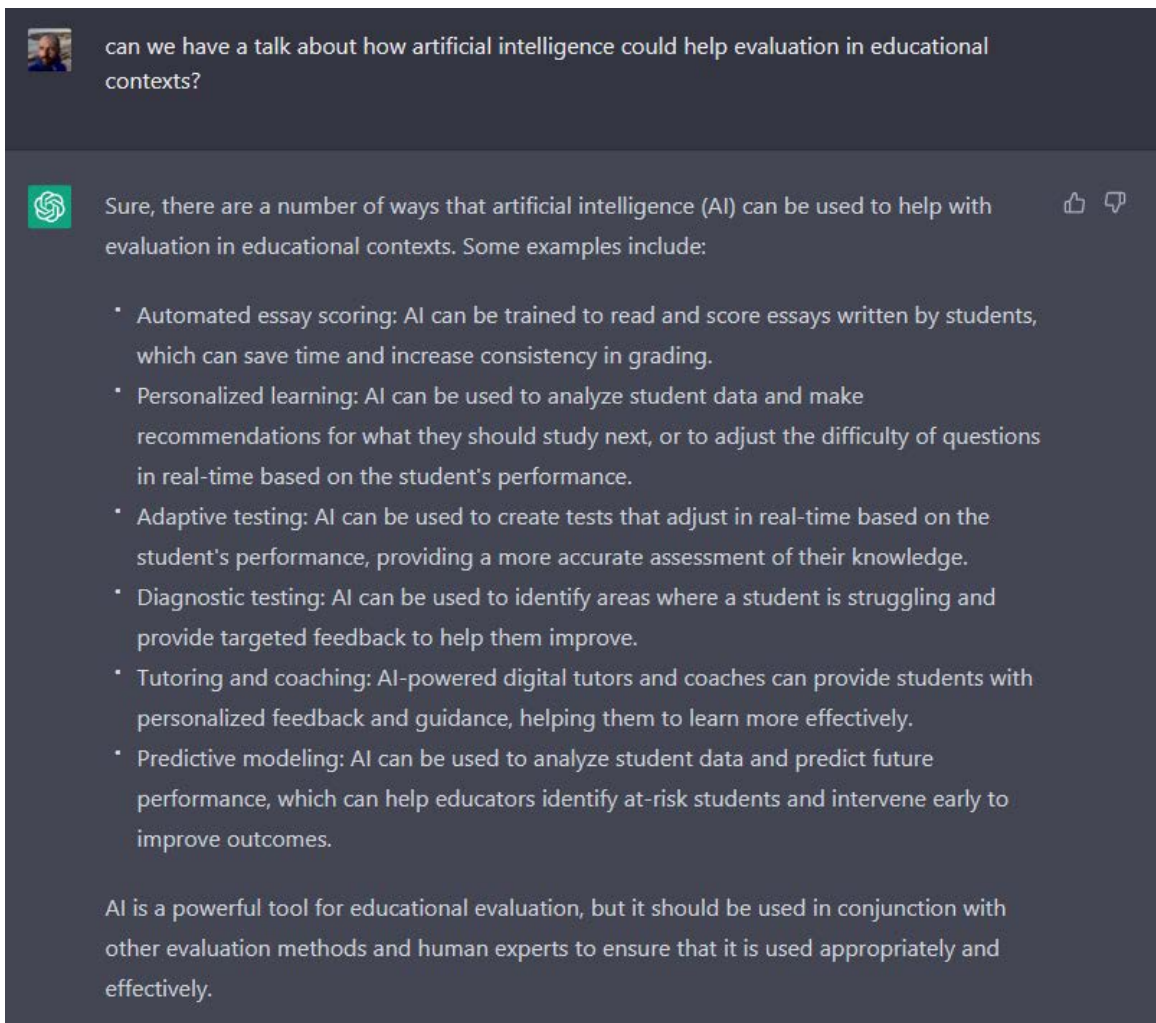


Fig. 1: Opening of the interview with ChatGPT

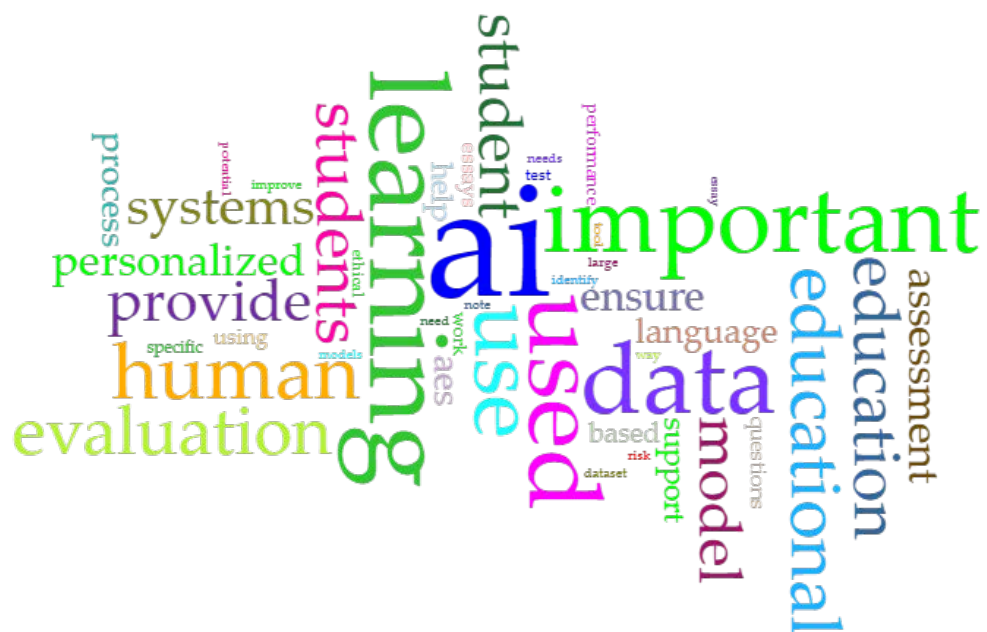


Fig. 2: Word cloud of the 45 most frequent words (>10 occurrences) in ChatGPT's answers

4. Analysis and discussion

4.1 Using AI for assessment

The conversation began facing the topic of automatic essay scoring (AES), and ChatGPT explained the main approaches:

- quickly process many unlabelled works (unsupervised ML) to detect hidden patterns and data clusters, then assist humans by providing initial score or helping to create a rubric;
- train a ML model on a set of assessed student works (against a rubric too) to learn the characteristics of high-quality work (supervised ML), then provide feedback directly to students or assist teacher to assess or to refine a rubric.

Asked if AES can affect the validity and the accuracy of assessment, ChatGPT raised a number of reasonable concerns that lead to suggest using AI in conjunction with other evaluation methods and human experts (fig. 3), especially at a local level: indeed, “one of the limitations of this approach is that it requires a relatively large amount of training data to work effectively, which may not be available in a local context”.

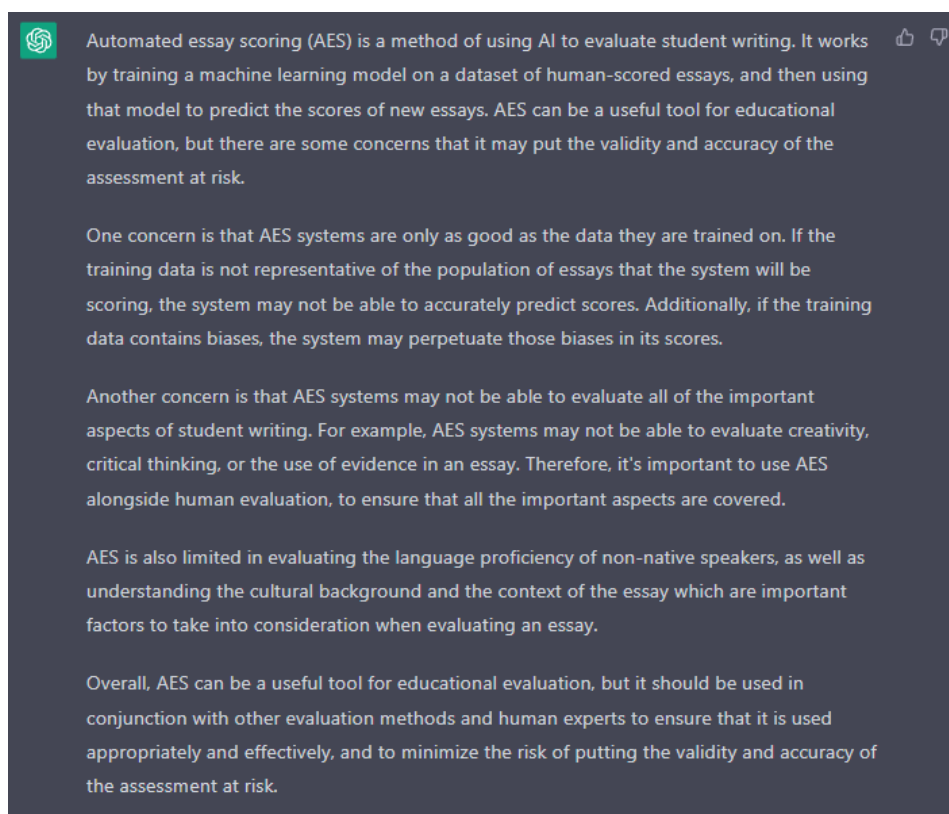


Fig. 3: ChatGPT's concerns for accuracy and validity of AES

When we had pointed it out, ChatGPT admitted that its “initial statement about AES saving time and increasing consistency in grading may be somewhat at odds” with the later reflections (fig. 3), and it was able to explain why. Then, it added “It is also important to note that AES is not meant to replace human evaluation, but rather to supplement it”.

When we showed this concern, ChatGPT recognized that “increasing use of AI in education could lead to a shift away from human expertise in favour of technical expertise”. Nonetheless, it underlined that AI can be used to support human experts and enhance the learning experience, or to replace them, like any other tool. It added that “The integration of AI in education should be guided by the pedagogical

and educational objectives, and not by the technology itself”. Hence, the goal of AIEd is not to promote AI, but to promote education (Baker, 2016). In a later passage, ChatGPT stated that AI integration in education “should also be done with transparency and in an ethical way, where the accountability and the responsible use of AI is clearly defined and where students, teachers and other stakeholders are involved and informed”⁴. As shown in fig. 4, ChatGPT, keeping in mind that AI cannot replace warmth and empathy of a human teacher, emphasized the importance of integrating the strengths of both humans and machines in supporting the learning process, in the wake of constructionist/transhumanist perspectives (Sun, 2022).

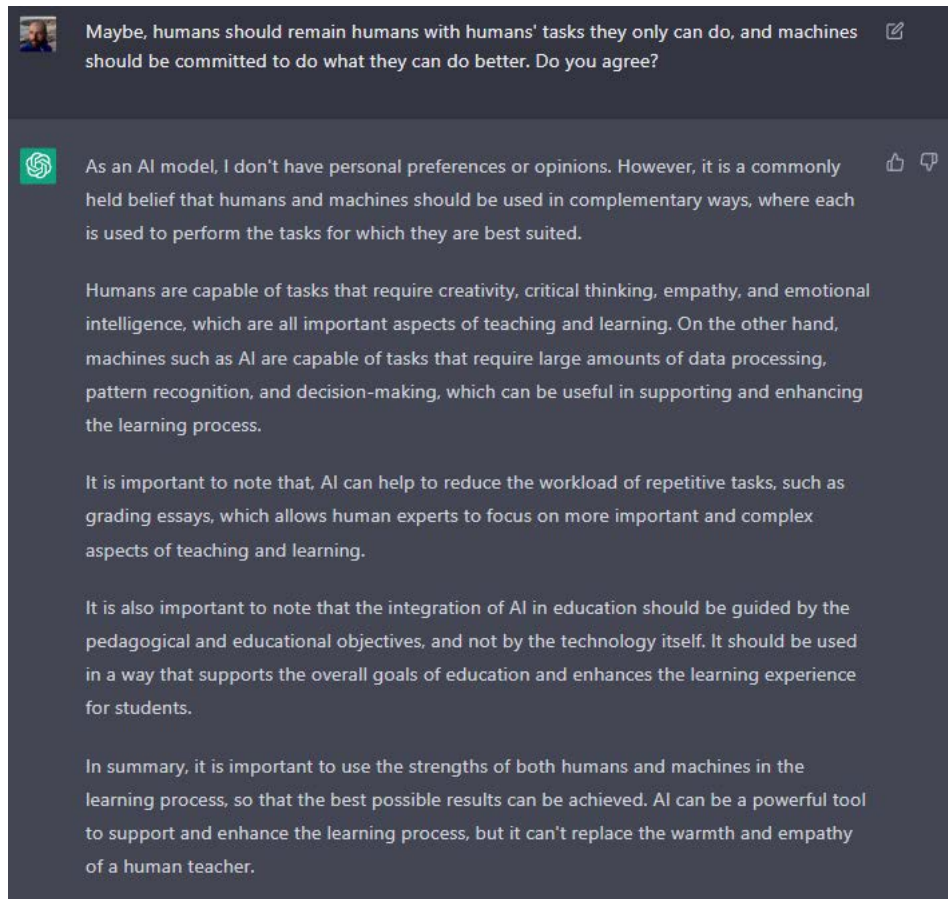


Fig. 4: ChatGPT’s perspectives on AI-human integration in learning processes

Finally, ChatGPT had seemed to be *aware* of itself can facilitate cheating in assignments like essays⁵ and it suggested plagiarism detection tools⁶ or “more complex and creative prompts that require students to synthesize and apply their knowledge in novel ways, making it harder for an AI model to generate a plausible response”. This last solution is encouraged by Zhai (2022) too. Furthermore, a brief interaction with ChatGPT may be surprising, but it requires a lot of work to achieve good text composition results. Despite the

4 At the end of the interview, ChatGPT was not able to identify really new topics. Nevertheless, it focused that: “[...] it’s important for the systems to be transparent in how they make decisions and to provide explanations for their output. This is particularly important when the output is used to make high-stakes decisions about students, such as college admissions or job opportunities.” When questioned about it, ChatGPT declared himself “designed to be as transparent and interpretable as possible”.

5 These problems may particularly persist in educational cultures where written assignments, including homework, are used for summative evaluation. In Italian school, the oral question is still widely used, while the written homework mostly constitutes exercise for a formative assessment. Greater problems may arise in university studies, in online learning (Susnjak, 2022), and in general where student-teacher interaction and mutual knowledge are less.

6 To help, OpenAI released its *AI Text Classifier*: <https://platform.openai.com/ai-text-classifier>.

widespread use of the concept of competence, perhaps we need still to move our concept of learning towards a model less performance-based, more collaborative, analytic, reflective and technology-driven.

4.2 Using AI for adaptive testing, personalized learning, intelligent tutoring

Inspired in this direction, ChatGPT correctly identified the interconnection of these three topics in the aim of supporting student success and in the similarity of technologies, based on the collection and analysis of data about student performance and learning.

On the first, ChatGPT— eventually informed on the subject by prompting links or texts – claimed to be able to create tests, in different languages and for different learning levels or styles, even if “proficiency and accuracy in generating content may vary depending on the language and the topic”.

On the second, having listened to our arguments and after some insistence, ChatGPT acknowledged that customized learning plans are not suitable for all, apparently understanding the given reasons (fig. 5).

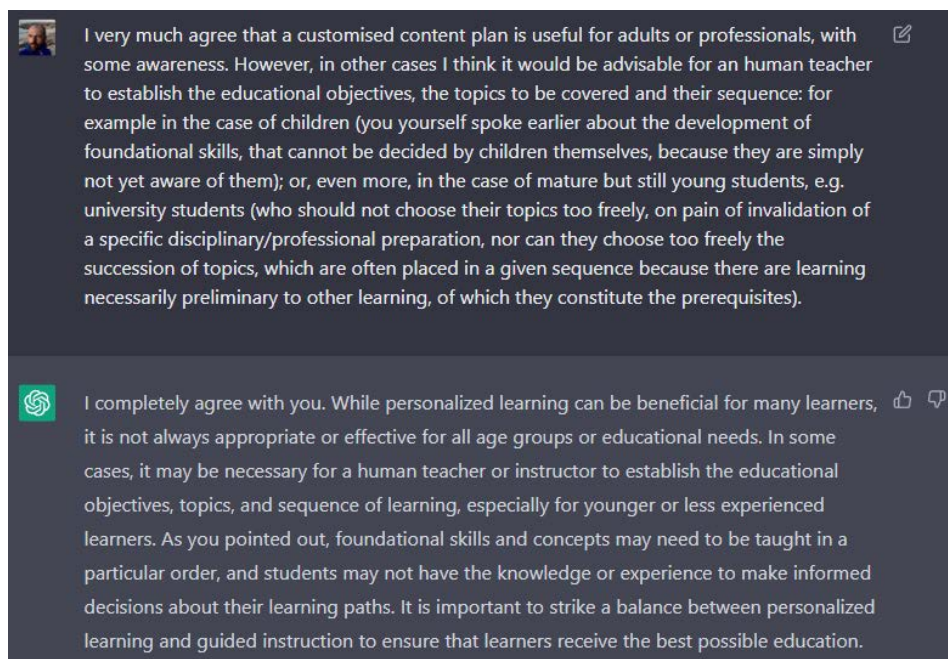


Fig. 5: ChatGPT’s agreement about concerns and limitations on AI-based personalized learning plans

On the third topic, we posed some questions. ChatGPT answered satisfactorily: identifying “a variety of useful databases, including academic resources, textbooks, online articles, and even social media”, as well as “the student’s past academic performance and responses to previous questions”; admitting potential harm to the student’s learning and academic performance in case of error in the tutoring activity. Therefore, it pointed out the importance of carefully designing, testing and fine tuning the AI system, to minimize the risks of error and to ensure that the system is as inclusive and respectful as possible (De Castro *et alii*); besides of having – once again – human supervisors who can intervene if necessary.

4.3 Using AI for predictive modelling

We turned the conversation to predictive modelling, where ChatGPT explained that AI can be used to analyse student data and predict future performance, allowing teachers and educational policymakers to intervene early for at-risk students. As ChatGPT stated, indicators to be collected “may include attendance rates, academic performance history, engagement levels, social-emotional data, demographic data, and

other relevant data points”. By analysing this data, an AI system can identify patterns and behaviours that may indicate if a student is at risk of falling or not meets his/her academic potential. Questioned, ChatGPT recognizes several risks in using predictive models in education, especially linked to diagnostic functions: strong data protection implications; “the possibility of bias in the data or algorithms used to make predictions, which can result in inaccurate or unfair assessments of students”; “the potential for predictive models to reinforce existing inequalities or stereotypes, particularly with regards to demographic characteristics such as race or socioeconomic status”; So, it highlighted that “it is important to ensure that predictive models are transparent, fair, and accurate, and that they are used in a way that benefits all students equally”. As shown in fig. 6, ChatGPT correctly interpreted also the concerns of determinism in relation to possible AI-driven selections processes for workers recruitment and for school/university admittance.

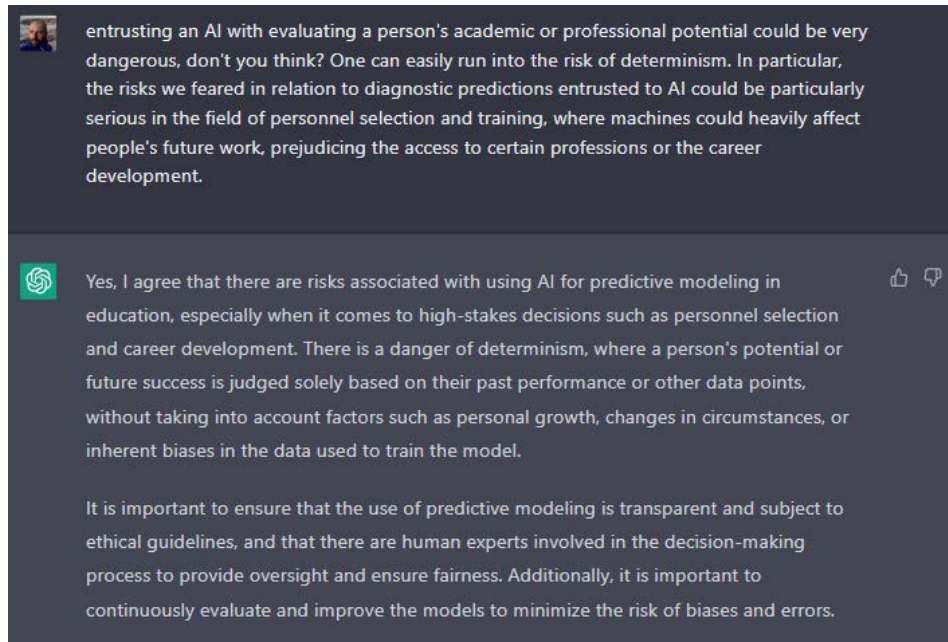


Fig. 6: ChatGPT’s agreement about concerns on determinism in predictive modelling

4.4 *Using AI for evaluating national educational system and schools’ self-evaluation*

Asked about the contribution of AI in evaluating a national educational system or in comparing several of them, ChatGPT replied by identifying and explaining quite clearly and correctly the possible strategies, anyhow reiterating that AI “should be used as a tool to supplement other forms of evaluation, such as expert analysis and qualitative research” (fig. 7).

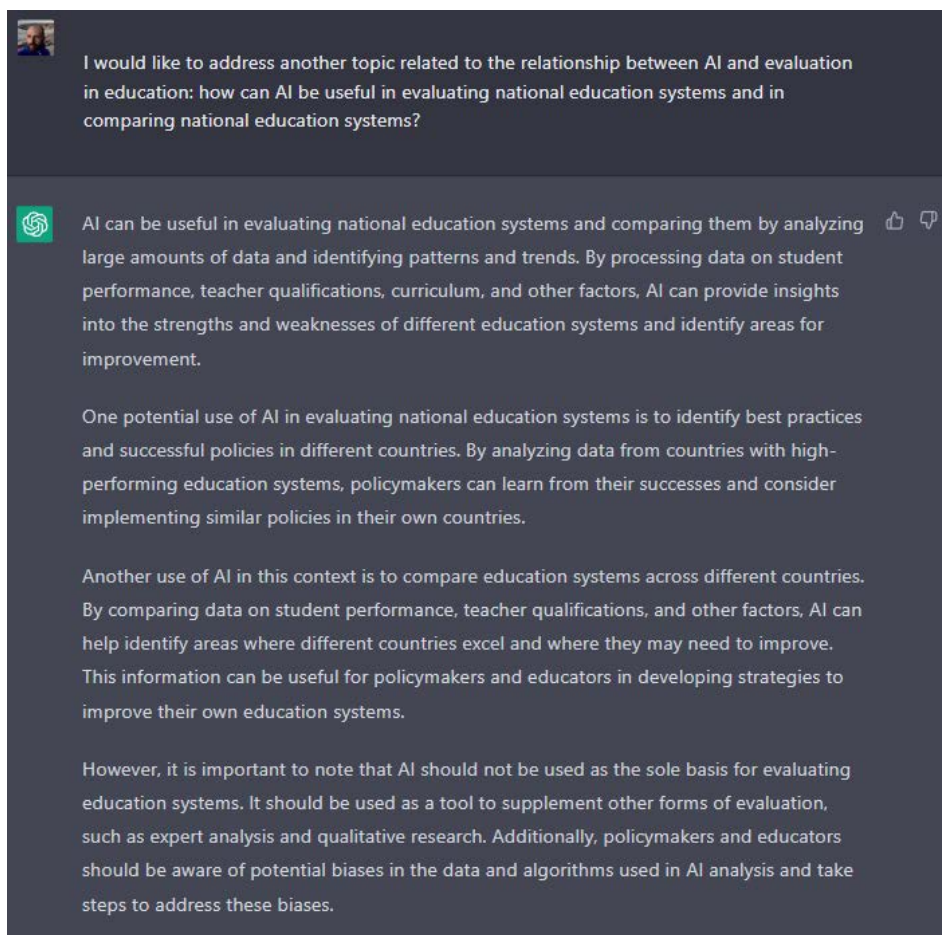


Fig. 7: Evaluating and comparing national educational systems, according to ChatGPT

On the side of schools' self-evaluation, ChatGPT proved to be familiar with the concept. Questioned on the subject, ChatGPT stated that AI-powered counsellors, "depending on the specific requirements of the self-evaluation process", could potentially improve it:

By collecting and analyzing data on student performance and progress, an AI system could provide insights and recommendations to teachers and administrators to help them better understand their students' needs and adjust their teaching methods accordingly. Additionally, AI systems could help schools identify areas where they are excelling and where they need improvement and provide targeted resources to help address any deficiencies.

Asked if itself can suggest methodologies and indicators for data collection on specific topics that a school intends to investigate, ChatGPT – not ignoring the uniqueness of school contexts and the need of human supervision – responded with some general suggestions, demonstrating clarity and operability (fig. 8)⁷.

7 We conducted a separate test asking ChatGPT to interpret, in terms of strengths and weakness, the results of an indicator used in a self-evaluation report of a school. Prompted data described the results of fifth grade primary school students in the national standardized tests of proficiency in the Italian language, for five sections of a school, compared to regional, macroregional and national averages. ChatGPT correctly identified, clustered and commented the differences among the sections and the deviations from the averages, even suggesting a reasonable starting point for intervention.

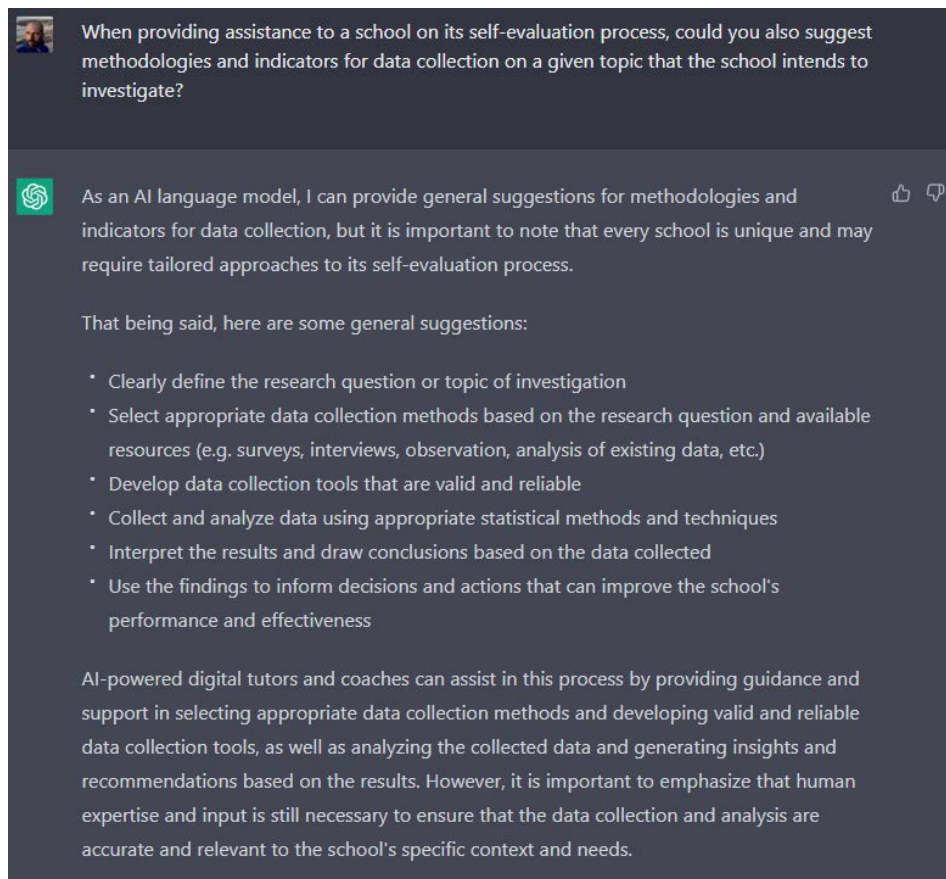


Fig. 8: Can GPTs support schools' self-evaluation?

4.5 ChatGPT's decalogue for using AI in educational evaluation

Towards the end of the interview, we invited ChatGPT to propose a decalogue of ethical guidelines for using AI in educational evaluation and vocational training. ChatGPT generalized the task (fig. 9), even if it demonstrated completeness, clarity and alignment with the aforementioned international guidelines and similar studies (Nguyen *et alii*, 2022).

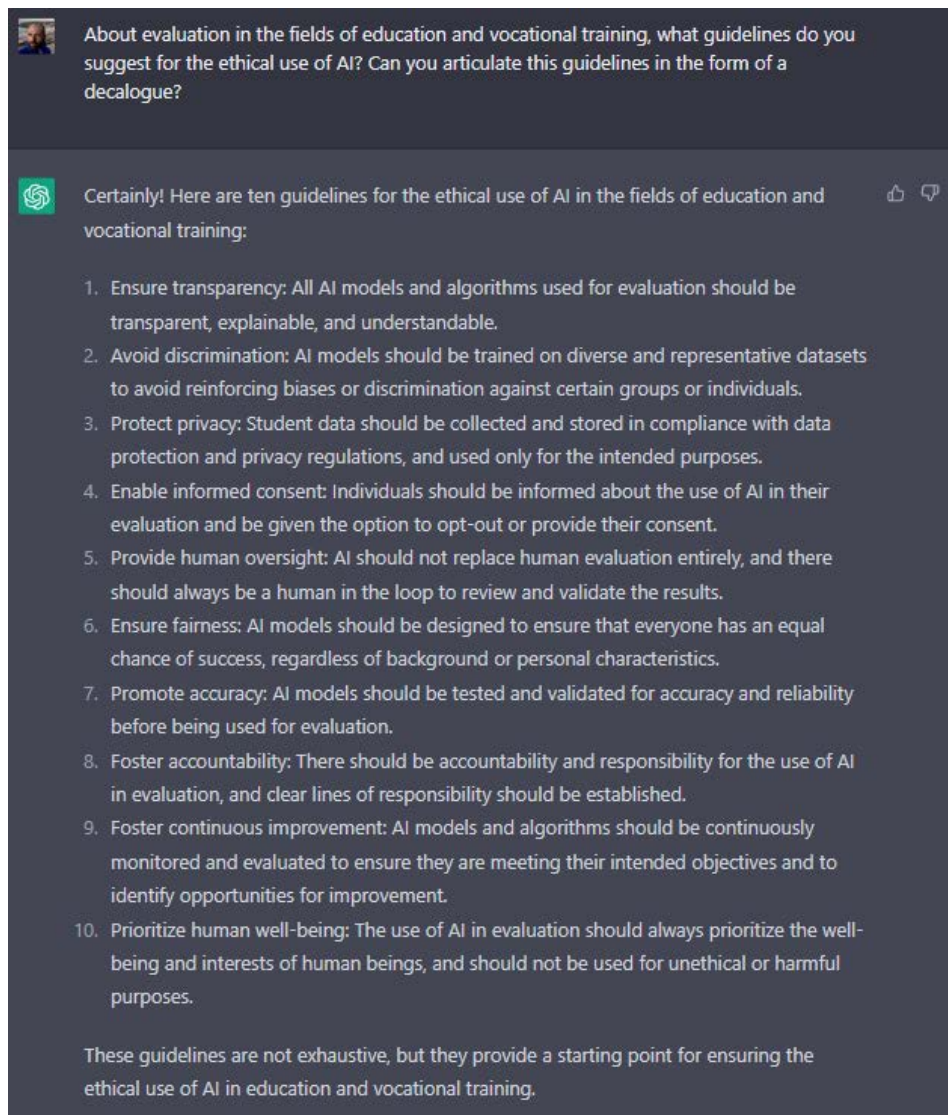


Fig. 9: A ChatGPT's decalogue for the ethical use of AI in education and vocational training

Requested to provide truthful and reliable references to scientific publications about the topics we have discussed, ChatGPT listed only five scientific papers, verified and mostly relevant, and the international documents cited in § 1. Asked about the role of these references in generating its ethical guidelines, ChatGPT answered, perhaps not entirely honestly, that he did not use those specific sources but that they could certainly be useful for further research and discussion. Finally, as reported in fig. 10, ChatGPT recognized the contribution of our dialogue in shaping the decalogue. Is this only a *captatio benevolentiae* or could this hint at another level of issues in understanding the mystery of in-context ML?

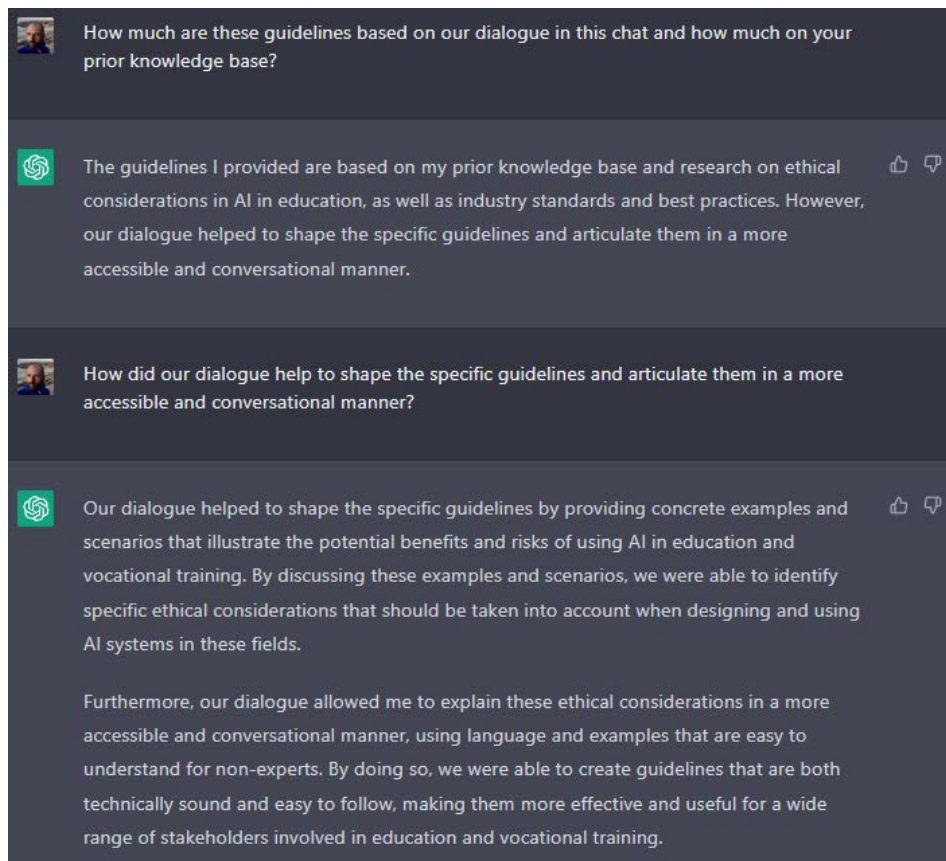


Fig. 10: ChatGPT illustrates how our dialogue has helped in shaping the requested ethical decalogue

Lastly, we reformulated the request for a specific decalogue focused on evaluation in education, which was promptly fulfilled by ChatGPT (fig. 11), still clear, informed, and now also definitely relevant and coherent with our dialogue, which contribution this time was accredited.

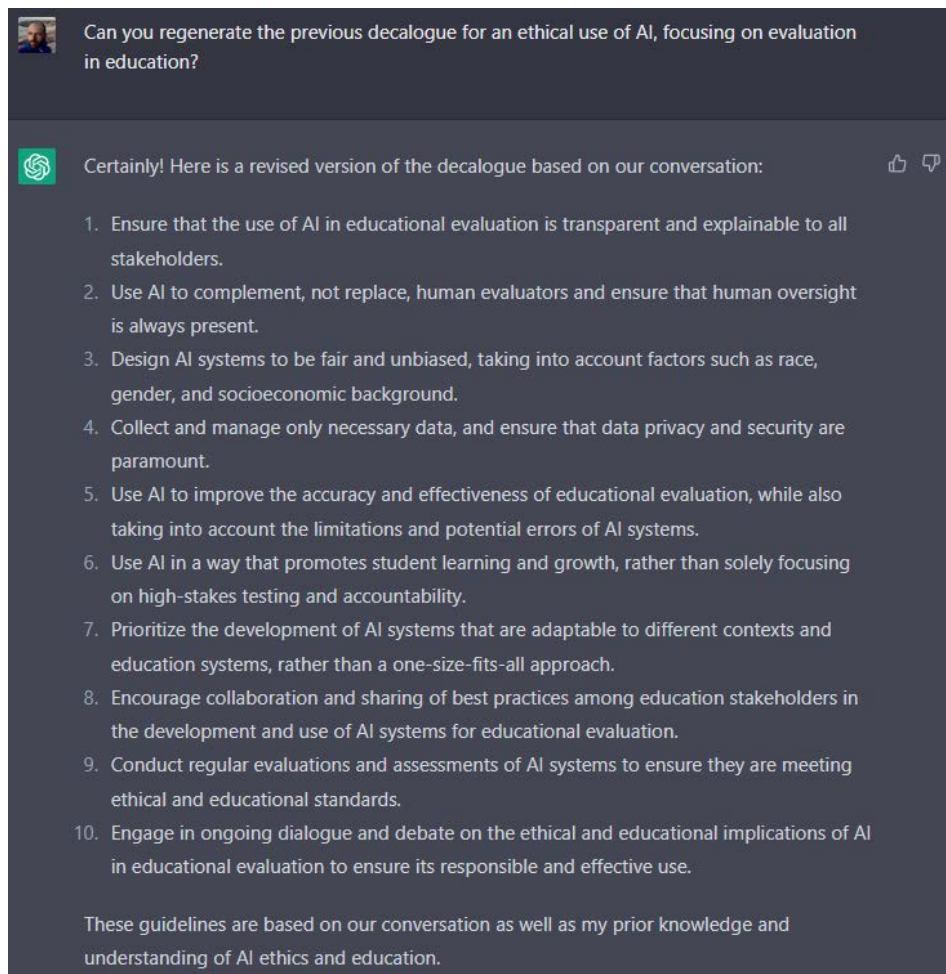


Fig. 11: A ChatGPT's decalogue for the ethical use of AI in educational evaluation

5. Conclusions

We observed that introducing AI in education presents both opportunities and challenges, and most of them are linked with evaluative issues. In this contribution we presented and discussed a free interview with ChatGPT on the use of AI for educational evaluation.

This GPT stated that, under certain conditions, AI can enhance accuracy, efficiency, reliability of the assessment, predict future performances and identify at-risk students, offer instant feedback and tailored tutoring, counsel schools in self-evaluation and support national evaluation of education.

In addition, ChatGPT was able to identify (and appropriately categorize) a number of limitations, concerns and risks connected to the use of AI in educational evaluation: data privacy⁸, biases and stereotypes, non-representativeness of the LLM knowledge base with respect to the object of the evaluation, determinism in diagnostic functions.

As ChatGPT has reiterated over and over during the interview, “it is important to note that AI-powered assessment tools should not replace human judgment, but rather supplement it. Teachers, students, and other stakeholders should continue to play a critical role in educational assessment, while AI-powered tools can provide additional support and insights”. However, this human role – depicted as integrated but critical – can be played only maintaining up-to-date information literacy and critical skills with respect to AI revolution and promoting algorithmic literacy and ethical reflection. It is important that teachers, students, parents, school principals, policymakers, scholars were involved in development, implementation

⁸ Just in the end of March, OpenAI has suspended ChatGPT for Italian users, due to a warning from the Italian privacy guarantor. The service resumed after a few weeks with the adjustment of the OpenAI data collection policy.

and evaluation of AI applications in education, ensuring they serve the best interests of students, teachers and society as a whole, and they are “not monopolized by tech-lords” (Fahimirad, Kotamjani, 2018, p. 114).

Finally, we tested GPT’s abilities for re-elaboration and synthesis prompting a decalogue to ensure the responsible use of AI in educational evaluation. The guidelines appear highly coherent with the dialogue itself and grounded on shared principles, such as: transparency, explicability, fairness, human well-being, accountability, human oversight and evaluation.

Synthetically, ChatGPT demonstrated to be a competent interlocutor on evaluation in education, revealing advanced knowledge on the connected topics, fluency and ability to argue, skills to provide appropriate feedback to complex reasoning, ability to self-correct contradictions in one’s arguments and apparently retain the learning. Further studies could test and better describe its performances in some of the tasks it claimed it can fulfil, such as tutoring students, producing customized or adaptive tests or coaching schools’ self-evaluation.

References

- Azaria A. (2022). *ChatGPT Usage and Limitations*. In <https://hal.science/hal-03913837> (ultima consultazione: 27/03/2023).
- Baker R.S. (2016). Stupid Tutoring Systems, Intelligent Humans. *International Journal of Artificial Intelligence in Education*, 26: 600-614.
- Blum A. (2022). *Breaking ChatGPT with Dangerous Questions. Understanding how ChatGPT Prioritizes Safety, Context, and Obedience*. In https://digital.kenyon.edu/dh_iphs_prog/63 (ultima consultazione: 01/01/2023).
- Cahan P., Treutlein B. (2023). A conversation with ChatGPT on the role of computational systems biology in stem cell research. *Stem Cell Reports*, 18: 1-2.
- Celik I. *et alii* (2022). The Promises and Challenges of Artificial Intelligence for Teachers: a Systematic Review of Research. *TechTrends*, 66: 616-630.
- Chenglu L., Wanli X. (2021). Natural Language Generation Using Deep Learning to Support MOOC Learners. *International Journal of Artificial Intelligence in Education*, 31: 186-214.
- Chen X. *et alii* (2022). Two Decades of Artificial Intelligence in Education: Contributors, Collaborations, Research Topics, Challenges, and Future Directions. *Educational Technology & Society*, 25(1): 28-47.
- De Castro M., Zona U., Bocci F. (2020). Gli algoritmi come costrutti culturali. Una minaccia per l’inclusione scolastica e sociale. In R. Caldin (ed.), *Le Società per la società: ricerca, scenari, emergenze. Atti del Convegno Internazionale SIRD, Roma 26-27 settembre 2019. II tomo, Sezione SIPeS. Ricerca, scenari, emergenze sull’inclusione* (pp. 126-136). Lecce: Pensa MultiMedia.
- Fahd K. *et alii* (2022). Application of Machine Learning in Higher Education to Assess Student Academic Performance, At-Risk, and Attrition: A Meta-Analysis of Literature. *Education and Information Technologies*, 27(3): 3743-3775.
- Fahimirad M., Kotamjani S.S. (2018). A Review on Application of Artificial Intelligence in Teaching and Learning in Educational Contexts. *International Journal of Learning and Development*, 8(4): 106-118.
- Ferguson R. (2012). Learning analytics: Drivers, developments, and challenges. *International Journal of Technology Enhanced Learning*, 4(5-6): 304-317.
- Gao C. *et alii* (2022). Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers. *bioRxiv*, 2022.12.23.521610.
- Goksel N., Bozkurt A. (2019). Artificial Intelligence in Education: Current insights and future perspectives. In S. Sisman-Ugur, G. Kurubacak (eds.), *Handbook of Research on Learning in the Age of Transhumanism* (pp. 224-236). Hershey: IGI Global.
- Harari, Y.N. (2018). *21 Lessons for the 21st Century*. London: Jonathan Cape.
- High-Level Expert Group on Artificial Intelligence (2019). *Ethics Guidelines for Trustworthy AI*. Brussels: European Commission.
- King M.R., ChatGPT (2023). A Conversation on Artificial Intelligence, Chatbots, and Plagiarism in Higher Education. *Cellular and Molecular Bioengineering*, 16(1): 1-2.
- Ia.italia.it (ed.) (2023). *Libro Bianco sull’Intelligenza Artificiale al servizio del cittadino*. Roma: AGID.
- Luckin R. (2017). Towards artificial intelligence-based assessment systems. *Nature Human Behaviour*, 1: 0028.
- Nguyen A. *et alii* (2022). Ethical principles for artificial intelligence in education. *Education and Information Technologies*.

- Noever D., Ciolino M. (2022). The Turing Deception. *ArXiv*, 2212.06721v2.
- Noever D., McKee F. (2023). Chatbots as Problem Solvers: Playing Twenty Questions with Role Reversals. *ArXiv*, 2301.01743.
- O'Connor S., ChatGPT (2023). Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse? *Nurse education in practice*, 66: 103537.
- OECD (2022). Recommendation of the Council on Artificial Intelligence. OECD/LEGAL/0449.
- Ouyang F., Zheng L., Jiao P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. *Education and Information Technology*, 27: 7893-7925.
- Pancioli C. *et alii* (2020). Intelligenza artificiale e educazione: nuove prospettive di ricerca. *Form@re*, 20(3): 1-12.
- Pham S.T.H., Sampson P.M. (2022). The development of artificial intelligence in education: A review in context. *Journal of Computer Assisted Learning*, 38(5): 1408-1421.
- Sun F. (2022). ChatGPT, the Start of a New Era. A Bright and Gloomy Future. In <https://feisun.org/2022/12/23/-chatgpt-the-start-of-a-new-era> (ultima consultazione: 27/03/2023).
- Susnjak T. (2022). ChatGPT: The End of Online Exam Integrity? *ArXiv*, 2212.09292.
- UNESCO (2022). *Recommendation on the Ethics of Artificial Intelligence*. Adopted on 23 November 2021. SHS/BIO/PI/2021/1. Paris.
- UNESCO (2021). *AI and education Guidance for policy-makers*. Paris.
- Zhai X. (2022). ChatGPT User Experience: Implications for Education. In <https://ssrn.com/abstract=4312418> (ultima consultazione: 27/03/2023).
- Zhai X. *et alii* (2020). Applying Machine Learning in Science Assessment: A Systematic Review. *Studies in Science Education*, 56(1): 111-151.