

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Terziyan, Vagan; Kaikova, Olena; Golovianko, Mariia; Vitko, Oleksandra

Title: Can ChatGPT Challenge the Scientific Impact of Published Research, Particularly in the Context of Industry 4.0 and Smart Manufacturing?

Year: 2024

Version: Published version

Copyright: © 2024 the Authors

Rights: CC BY-NC-ND 4.0

Rights url: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Please cite the original version:

Terziyan, V., Kaikova, O., Golovianko, M., & Vitko, O. (2024). Can ChatGPT Challenge the Scientific Impact of Published Research, Particularly in the Context of Industry 4.0 and Smart Manufacturing?. *Procedia Computer Science*, 232, 2540-2550.
<https://doi.org/10.1016/j.procs.2024.02.072>



5th International Conference on Industry 4.0 and Smart Manufacturing

Can ChatGPT Challenge the Scientific Impact of Published Research, Particularly in the Context of Industry 4.0 and Smart Manufacturing?

Vagan Terziyan ^{a,*}, Olena Kaikova ^a, Mariia Golovianko ^b, Oleksandra Vitko ^b

^a Faculty of Information Technology, University of Jyväskylä, 40014, Jyväskylä, Finland

^b Department of Artificial Intelligence, Kharkiv National University of Radio Electronics, 61166, Kharkiv, Ukraine

Abstract

The released ChatGPT as a powerful language model is capable of assisting with a wide range of tasks, including answering questions, summarizing, paraphrasing, proofreading, classifying, and integrating texts. In this study, we tested ChatGPT capability to assist researchers in evaluating the academic articles' contribution. We suggest a dialogue schema in which ChatGPT is asked to answer research questions from the target article and then to compare its own answers with the answers from the article. Finally, ChatGPT is asked to integrate both solutions coherently. We experimented with Proceedings of ISM-2022 Conference on Industry 4.0 and Smart Manufacturing, utilizing explicit research questions. The chat context enabled assessing studied articles' contributions to Industry 4.0, uncovering advancements beyond the state-of-the-art. However, ChatGPT demonstrates limitations in content understanding and contribution evaluation. We conclude that while it collaborates with humans on academic tasks, human guidance remains essential, while ChatGPT's assistance efficiently complements traditional academic processes.

© 2024 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer-review under responsibility of the scientific committee of the 5th International Conference on Industry 4.0 and Smart Manufacturing

Keywords: Artificial Intelligence; ChatGPT; Industry 4.0; Smart Manufacturing; academic impact

1. Introduction

A powerful artificial intelligence (AI) driven language model and a chat bot when appropriate, ChatGPT [1] is currently one of the major newsmakers worldwide. It is based on the family of Generative Pre-trained Transformers

* Corresponding author. *E-mail address:* vagan.terziyan@jyu.fi

(GPT), such as GPT-2 [2] and GPT-3 [3], which are large-scale language models developed by OpenAI. GPT-2 was released in 2019 and has 1.5 billion parameters, while GPT-3, released in 2020, has up to 175 billion parameters, making it one of the largest language models to date. Both GPT-2 and GPT-3 are designed to generate coherent and natural language text in response to a given prompt, and are capable of performing a wide range of language tasks such as language translation, summarization, and question-answering. ChatGPT, on the other hand, is a specific implementation of the GPT architecture that is optimized for conversational interactions. It is trained on a large volume of data from the internet and is capable of carrying on a natural-sounding conversation on a variety of topics. It can produce sophisticated and seemingly intelligent texts and, therefore, it causes both excitement and controversy [4].

Transformers [5] are able to do such a job because they are based on a deep learning architecture that is specifically designed for natural language processing (NLP) tasks. The key innovation of the transformer architecture is the use of self-attention mechanisms [6], which allow the model to attend to different parts of the input at different levels of granularity. This enables the model to capture complex patterns and relationships between words and phrases. Due to massive amounts of data used to train transformers, the resulting robust representations of language that can be fine-tuned for specific tasks and domains. The combination of powerful neural architectures, sophisticated training techniques, and massive datasets allowed transformers to achieve state-of-the-art performance on many NLP tasks, and can be considered as a significant achievement in AI.

According to [7], the main advantages of ChatGPT are generalization and creativity with constant self-correction and self-improvement while being critical to potentially unsafe user questions and its own responses to them. Therefore, ChatGPT can greatly change human lives in many areas in the coming years. Being positioned as a universal assistant, ChatGPT is expected to facilitate improvements in production effectiveness and efficiency and influence almost all industries, including science, education, and medicine. As argued in [8], ChatGPT enhances humans in their everyday efficiency and accuracy at workplace and will save essential costs due to optimized processes.

The high performance of ChatGPT immediately brings it into the group of AI-based applications capable of providing assistance in academic work (teaching, learning, and research) [9]. However, dealing with such academic disciplines, whose development assumes closer interaction with actual physical reality (such as, e.g., Industry 4.0 and Smart Manufacturing) rather than its textual descriptions, may be a challenging task for ChatGPT. On the other hand, there is a strong belief that (if used as a tool) it can provide essential assistance in academic studies and research [10].

In this paper, we study whether ChatGPT is capable of addressing the research questions and evaluating the impact of published research within challenging for ChatGPT domain – Industry 4.0 and Smart Manufacturing. The objective was to check if and how ChatGPT could be used as an assistant in performing research in the domain. For testing ChatGPT, we have chosen the set of articles from popular conference proceedings “4th International Conference on Industry 4.0 and Smart Manufacturing” [11] published at the beginning of 2023. We made this choice so that “awareness” of ChatGPT (trained on the content dated up to the end of 2021) does not include the proceedings [11]. Therefore, we aimed to create the maximal uncomfortable condition for ChatGPT work, i.e., new content and a challenging domain. Another reason of choosing ISM conference content for the study is due to strong academic and industrial experts’ community traditionally publishing there as well as availability of well-cited open-source proceedings from excellent academic publisher.

In this study, we have developed a particular dialogue schema to get needed academic service and content from ChatGPT: addressing the research questions from published articles, comparing and integrating its own answers with the actually published answers from the articles.

We made a provocative title for this paper just to challenge both the human researchers and the AI. See how ChatGPT has answered our question:

Question: “*Can ChatGPT Challenge the Scientific Impact of Published Research, Particularly in the Context of Industry 4.0 and Smart Manufacturing?*” →

Answer: “*As an AI language model, ChatGPT is capable of providing insights and opinions on scientific research, but it does not have the ability to challenge the scientific impact of published research in a definitive way. The evaluation of the scientific impact of research is often based on a number of factors, such as the quality of the research methodology, the novelty and significance of the research findings, the rigor of the peer review process, and the impact of the research on the field. These factors can be subjective and may vary depending on the perspective and expertise of the individual evaluating the research. While*

ChatGPT can analyze and interpret scientific research, it is important to note that its responses are generated based on the information and data it has been trained on. Therefore, its insights and opinions may be limited by the quality and scope of the training data. In addition, ChatGPT is not capable of making definitive judgments on the validity or reliability of research findings. It is always important to critically evaluate the scientific literature and consider multiple sources of evidence before drawing conclusions or making decisions based on research findings” (ChatGPT [1], personal communication, March 1, 2023).

The following text of the paper is organized as follows: in Section 2, we present our dialogue schema as a kind of Human-ChatGPT collaborative research method; in Section 3, we present results of experimental dialogues regarding chosen articles from [11]; in Section 4, we discuss discovered strengths and weaknesses of ChatGPT for academic assistance; and we conclude in Section 5.

2. Researcher-ChatGPT Dialogue Schema for Academic Impact Evaluation

In this section, we present one possible schema of a dialogue with ChatGPT as part of a collaborative (Human-AI) academic impact evaluation process. The object of evaluation could be academic articles, which include explicit research questions and which are concluded by explicit answers to these questions. Research questions are typical for *quantitative research*, which is a research method that involves the systematic collection, analysis, and interpretation of numerical data using statistical methods. In quantitative research, researchers typically start by specifying research questions or hypotheses that can be tested using numerical data, and then use various methods to collect and analyze data that can be quantified, such as experiments, or observational studies [12]. Although explicit research questions seem to be important in some fields, the study in [13] noticed that such questions were seldom explicitly mentioned (under 2%) in articles in engineering and physical, life, and medical sciences. For example, in the Proceedings of 4th International Conference on Industry 4.0 and Smart Manufacturing [11], which we used as a test set for this study, 4% of papers had explicit research questions-answers, particularly articles [14], [15], [16], and [17].

The structure of the suggested dialogue schema is shown in Fig. 1, and the appropriate process goes as follows:

1. The researcher collects articles on the desired topic (or from the target conference or research group), which include explicit research questions (RQs) and answers to them (RAs). Let us assume that answers to RQs made by ChatGPT will be named as “chat answers” (CAs);
2. The researcher takes current RQ_i , i.e., a corresponding text marked as “XXX” in Fig.1 (RQ_i has also an explicit answer RA_i somewhere in the paper marked as “YYY” text in Fig.1), starts on chatting with ChatGPT and enters RQ_i there. ChatGPT answers the question by providing some text (marked as “ZZZ” in Fig.1) as an output CA_i . The researcher documents “XXX”, “YYY”, and “ZZZ” into the final table.
3. The researcher enters the following query to ChatGPT: “Consider the following research question RQ_i : “XXX” and compare the two following answers to it (RA_i and CA_i) where RA_i : “YYY” and CA_i : “ZZZ”. What do you see in these two answers to be common (similar) and what is different regarding addressing the research question?”. The answer to this query from ChatGPT is documented into the final table.
4. The researcher enters the following query to ChatGPT: “Could you integrate both texts (RA_i and CA_i) into one consistent story addressing the research question RQ_i ?”. The researcher documents the corresponding outcome from ChatGPT (i.e., integrated “YYY” and “ZZZ” texts) into the final table.
5. The researcher takes into consideration another RQ and continues the same loop starting from step 2.

In this way, the researcher will collect detailed analysis as a set of tables (one for each RQ). One may notice that the outcome of such a dialogue could be considered as some kind of assessment of the impact published by the research articles from the collection. For such an assessment, ChatGPT provides useful background information and helps to compare specific outcomes of each published research study with the state-of-the-art which ChatGPT pretends to be “aware of”. Similar dialogue schema could be used by researchers to double-check their own articles before submitting them for publication elsewhere.

Examples and results of such structural dialogues with ChatGPT will be presented in the following section.

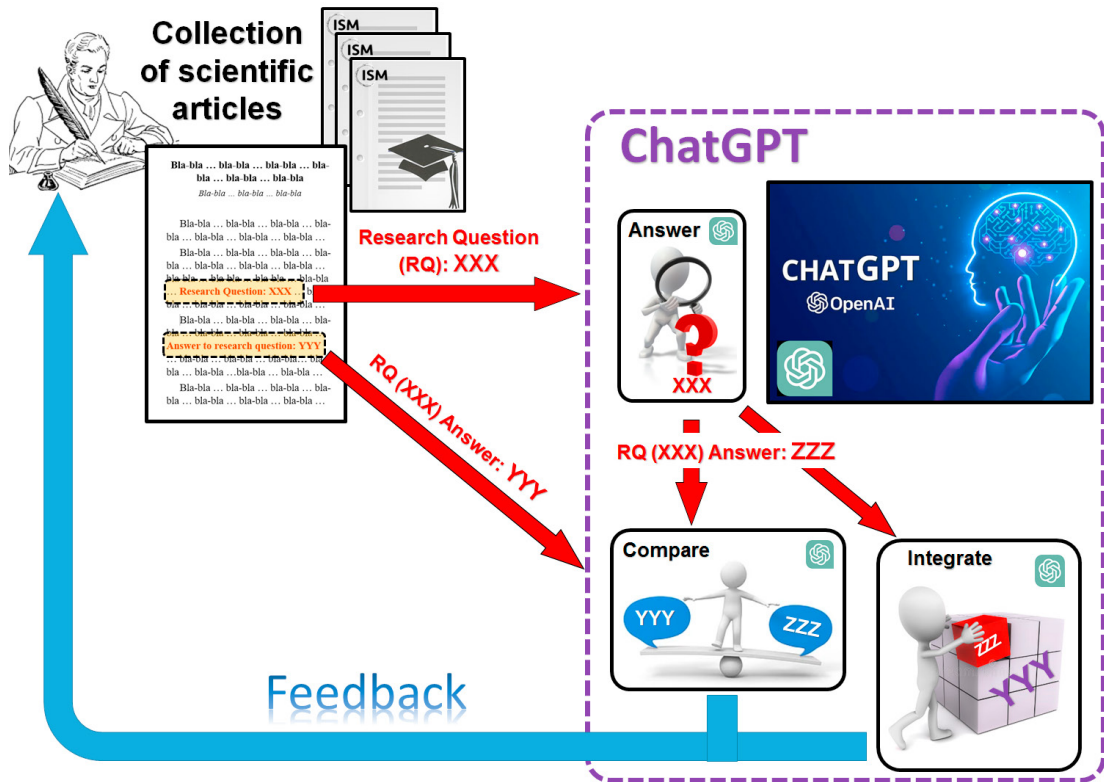


Fig. 1. The dialogue schema “Researcher – ChatGPT” is intended to evaluate collectively the impact of published research articles. Research questions from the articles are asked from ChatGPT and then the answers from chat are compared (and further integrated) with the original answers from the articles. Both comparison and integration operations are supported by ChatGPT.

3. Experimental dialogues with ChatGPT Regarding Academic Impact Evaluation

We tested our suggested dialogue schema with the articles [14], [15], [16], and [17] from the Proceedings of 4th International Conference on Industry 4.0 and Smart Manufacturing [11]. Results are collected within the following tables: Table 1 for [14]; Table 2 for [15]; Table 3 for [16], and Table 4 for [17].

The result shows that ChatGPT gives interesting and useful background information to compare the impact of published papers, which goes beyond state-of-the-art research.

For example, ChatGPT admits that articles [14] and [15] addressed the RQs more thoroughly than ChatGPT managed itself on the basis of known state-of-the-art data on the same topic. However, at the same time, ChatGPT suggested a few useful updates to the original formulation of RAs from [14] and [15] as a compromise between the original article and its own “opinion”.

Regarding the article [16], ChatGPT pointed out some interesting similarities and differences in the concept of “digital clone” used in the article and “cobot”, which ChatGPT recommends considering as a benchmark for the “digital clone” concept while addressing similar objectives. In the integrated text, ChatGPT suggests extending the collective intelligence term, which was considered in [16] as a collaborative effort provided by humans with their digital clones, taking into account also the “cobots”.

Regarding the article [17], ChatGPT in fact suggested extending a purely connectionist AI view on causality-aware neural networks (provided by the so-called causality-matrix as a component of neural network architectures) towards combining it with symbolic AI, particularly causal inference performed on top of causal graphical models.

In all cases, ChatGPT provides useful and inspiring information for both, evaluating the impact of existing studies and for planning further studies by using the suggested by ChatGPT extension to the existing ones.

Table 1. The dialogue with ChatGPT [1] (March, 2023) regarding article [14] from the ISM-2023 [11] collection

RQ: What skills and competences do students need to develop in order to face Industry 4.0 working challenges? [14]	
<i>Answer by the authors (TEXT_1):</i>	<i>Answer by ChatGPT (TEXT_2):</i>
<p>“Skills and Competences Expected in Students and Professionals Towards an Industry 4.0 Education: (A) Competences Skills Technical-technological: Media and coding skills, process understanding, state-of-the-art knowledge, and understanding of Information Technology security. Methodological: Analytical skills, autonomous learning, conflict- and problem solving, creativity, critical thinking, decision-making, and entrepreneurial thinking. Social-communicative: Ability to work in teams, cooperate and collaborate, cross-cultural understanding, digital and information literacies, share knowledge, as well as communication, intercultural, language, leadership, and networking skills. (B) Personal: Ability to work under pressure and adapt, be resilient, compliant, flexible, responsible, and tolerant, as well as keep a mindset focused on sustainability, keep motivated to learn (i.e., lifelong learning engagement), and produce knowledge” [14].</p>	<p>“The Fourth Industrial Revolution (Industry 4.0) is characterized by the integration of advanced technologies such as robotics, artificial intelligence, the Internet of Things (IoT), and big data analytics. As a result, students need to develop a range of skills and competences to adapt to the changing work environment. Some of these skills include critical thinking, problem solving, creativity, communication, collaboration, adaptability, and digital literacy. Additionally, they need to develop knowledge in fields such as data science, cybersecurity, cloud computing, and machine learning”.</p>
Comparison made by ChatGPT	
<p>Similarities: Both texts emphasize the need for students to develop a range of skills and competences to adapt to the changing work environment of Industry 4.0. Both texts mention the importance of critical thinking, problem-solving, creativity, communication, and digital literacy. Both texts highlight the importance of developing knowledge in specific fields such as data science, cybersecurity, and information technology security.</p> <p>Differences: TEXT_1 provides a more detailed list of competences and skills expected in students and professionals, categorizing them as technical-technological, methodological, social-communicative, and personal. TEXT_1 also emphasizes the importance of personal skills such as resilience, adaptability, responsibility, and lifelong learning engagement, which are not explicitly mentioned in TEXT_2. TEXT_2 emphasizes the importance of collaborative skills and competences, while TEXT_1 emphasizes the importance of networking, cross-cultural understanding, and leadership skills.</p> <p>Overall: Both texts share a common goal of preparing students for Industry 4.0 challenges, but <i>TEXT_1</i> [14] provides a more detailed and comprehensive list of skills and competences that students should develop.</p>	
Compromise between TEXT_1 and TEXT_2 suggested/generated by ChatGPT	
<p>The Fourth Industrial Revolution (Industry 4.0) requires students and professionals to develop new skills and competencies. These include technical and technological skills such as media and coding, state-of-the-art knowledge, and understanding of IT security, as well as methodological skills like analytical and problem-solving abilities, critical thinking, and entrepreneurship. Students must also possess social-communicative skills, including cross-cultural understanding, teamwork, digital and information literacy, and effective communication. Personal competencies are also essential, such as adaptability, resilience, responsibility, tolerance, sustainability, and a lifelong learning mindset. With advanced technologies like robotics, AI, IoT, and big data analytics, students must develop knowledge in areas such as cybersecurity, cloud computing, machine learning, and data science to adapt to the changing work environment.</p>	

Table 2. The dialogue with ChatGPT [1] (March, 2023) regarding article [15] from the ISM-2023 [11] collection

RQ: How advanced technologies improve the healthcare performance during COVID-19 pandemic? [15]	
<i>Answer by the authors (TEXT_1):</i>	<i>Answer by ChatGPT (TEXT_2):</i>
<p>“These technologies improved the healthcare performance during COVID-19 pandemic:</p> <p>(a) Color Coding (can be used to improve process efficacy and accuracy by using the computerized system in classification and categorization, which improves health service quality);</p> <p>(b) Robotics (can be used to assist medical practitioners in complex and precise operations, which improve the safety and quality of health service quality);</p> <p>(c) Big Data and Facial Recognition (using big data in machine learning and facial recognition will help medical decision making especially in diagnosing and tracking patients, which improves the performance and security of the health service quality at the minimum cost);</p> <p>(d) Artificial Intelligence (AI using ML and Data mining enables medical practitioners and researchers to infer disease characteristics which will help in diagnosing illness in the early stages);</p> <p>(e) Autonomous Vehicles (can be used to deliver food, supplies, drugs, and medical personnel while at the same time controlling the infection);</p> <p>(f) Social Media (collecting data and improving awareness about certain medical situations is much easier and more efficient using social media)” [15].</p>	<p>“Advanced technologies that have improved healthcare performance during the COVID-19 pandemic include:</p> <p>(a) Telemedicine: this technology enables doctors to consult with patients remotely, reducing the risk of viral transmission and enabling patients to receive care from their homes.</p> <p>(b) Artificial Intelligence (AI): AI has helped healthcare professionals to analyze large amounts of data quickly and accurately, enabling them to identify patterns and trends in the spread of the virus. AI has also been used to develop predictive models that help hospitals to anticipate and prepare for surges in demand for critical care.</p> <p>(c) Robotics: Robots can perform tasks such as cleaning and disinfecting hospital rooms, reducing the risk of viral transmission. They can also be used to deliver medication and supplies to patients, minimizing contact between healthcare professionals and patients.”</p>
Comparison made by ChatGPT	
<p>Similarities: Both texts describe how advanced technologies have improved healthcare performance during the COVID-19 pandemic. The texts have some common technologies mentioned, such as telemedicine and robotics, and also mention artificial intelligence.</p> <p>Differences: TEXT_1 provides a more extensive list of technologies, including color coding, facial recognition, autonomous vehicles, and social media, which are not mentioned in TEXT_2. TEXT_1 also provides more specific examples of how these technologies are being used to improve healthcare performance, such as using autonomous vehicles to deliver supplies and using facial recognition for medical decision making.</p> <p>Overall: Both texts provide solid information on how advanced technologies have improved healthcare performance during the COVID-19 pandemic. However, <i>TEXT_1 [15] is more detailed and comprehensive in its coverage of the topic, making it more effective in addressing the question of how advanced technologies have improved healthcare performance during the COVID-19 pandemic.</i></p>	
Compromise between TEXT_1 and TEXT_2 suggested/generated by ChatGPT	
<p>The COVID-19 pandemic has highlighted the crucial role of technology in improving healthcare performance. Telemedicine has enabled doctors to consult with patients remotely, reducing viral transmission risk and allowing patients to receive care from home. Robotics have been utilized to perform tasks such as cleaning and disinfecting hospital rooms, and delivering medication and supplies, minimizing contact between healthcare professionals and patients. Other technologies such as color coding, big data, facial recognition, AI, and autonomous vehicles have improved healthcare performance in various ways. Color coding and computerized systems have improved the accuracy and efficacy of processes, while big data and facial recognition have aided medical decision-making and tracking of patients. AI has enabled quick and accurate analysis of large amounts of data, and robots have assisted medical practitioners in complex and precise operations. Social media has also facilitated efficient collection of data and awareness-building of certain medical situations. The pandemic has highlighted the importance of utilizing advanced technologies to improve healthcare performance.</p>	

Table 3. The dialogue with ChatGPT [1] (March, 2023) regarding article [16] from the ISM-2023 [11] collection

RQ: Is it possible to integrate the two concepts (the automation-centric Industry 4.0 and the human-centric Industry 5.0) and their corresponding technological artifacts into one efficient and sustainable hybrid? [16]	
<i>Answer by the authors (TEXT_1):</i>	<i>Answer by ChatGPT (TEXT_2):</i>
The paper raises the question of whether it's possible to create a sustainable hybrid integrating Industry 4.0 and Industry 5.0, and argues that it is possible through the use of digital artifacts such as cognitive clones. These clones, created using adversarial learning, serve as a bridge between automation-centric Industry 4.0 and human-centric Industry 5.0 by allowing donors to remain involved in decision-making processes while also making the process more efficient. The paper also argues that collective intelligence, including both humans and their autonomous digital clones, can serve as an enabler for a collective-intelligence-centric and resilient Industry 4.0 - Industry 5.0 hybrid [16].	“The integration of Industry 4.0 and Industry 5.0 is possible and can create a more holistic approach to industry. Industry 4.0 emphasizes automation and digitalization, while Industry 5.0 values human workers and their skills. To create a sustainable hybrid, it is essential to balance these two concepts by leveraging technologies that enhance human-machine collaboration and valuing human workers. Cobots can work alongside human workers to allow them to focus on tasks requiring human skills, while upskilling and promoting a culture of innovation can create a positive work environment. The integration can lead to a more efficient and sustainable manufacturing process that reduces waste and optimizes production”.
Comparison made by ChatGPT	
Similarities: Both TEXT_1 and TEXT_2 discuss the integration of Industry 4.0 and Industry 5.0 and their corresponding technological artifacts into a more efficient and sustainable hybrid. They both emphasize the importance of balancing automation with human workers and their skills (via “clones” in TEXT_1 and “cobots” in TEXT_2).	
Differences: However, TEXT_1 focuses more on the use of digital artifacts such as cognitive clones as a bridge between Industry 4.0 and Industry 5.0, as well as the role of collective intelligence in creating a resilient hybrid. TEXT_2 emphasizes the importance of technologies that enhance human-machine collaboration and upskilling human workers to create a positive work environment.	
Overall: In terms of which text better addresses the research question, it depends on the specific focus of the research. Taking into account that the study is more focused on the use of digital artifacts and collective intelligence, TEXT_1 is more relevant. If the research would be focused on balancing automation with human workers and creating a positive work environment, then TEXT_2 would be more relevant. Ultimately, both texts provide valuable insights into the integration of Industry 4.0 and Industry 5.0 and how it can lead to a more sustainable and efficient manufacturing process. One potential benefit of using digital cognitive clones (TEXT_1 [3]) is that corresponding processes are expected to be more efficient. On the other hand, there are also potential drawbacks to the use of digital cognitive clones. For example, there may be concerns about privacy and data security, as well as the ethical implications of creating such clones. Additionally, there may be concerns about the potential for digital clones to be used to replace human workers, rather than working alongside them. <i>The use of digital cognitive clones is a complex and controversial idea that requires careful consideration of the potential benefits and drawbacks, as well as the ethical and societal implications.</i>	
Compromise between TEXT_1 and TEXT_2 suggested/generated by ChatGPT	
The possibility of integrating Industry 4.0 and Industry 5.0 to create a sustainable hybrid is the main focus here. While Industry 4.0 emphasizes automation and Industry 5.0 values human workers and their skills, we argue that it is possible to create a bridge between the two through the use of digital artifacts such as cognitive clones, which serve as a means of involving donors in decision-making processes while also making the process more efficient. The integration of these two concepts can lead to a more efficient and sustainable manufacturing process that reduces waste and optimizes production. To achieve this, it is essential to balance these two concepts by leveraging technologies that enhance human-machine collaboration and valuing human workers. Cobots can work alongside human workers to allow them to focus on tasks requiring human skills, while upskilling and promoting a culture of innovation can create a positive work environment. In summary, the collective intelligence of both humans and their autonomous digital clones can serve as an enabler for a collective-intelligence-centric and resilient Industry 4.0 - Industry 5.0 hybrid that values both automation and human skills.	

Table 4. The dialogue with ChatGPT [1] (March, 2023) regarding article [17] from the ISM-2023 [11] collection

<p>RQ: How to discover and evaluate possible causal relationships within individual images and how to use such causality-awareness for better training convolutional neural networks (CNNs) for image classification? [17]</p>	
<p><i>Answer by the authors (TEXT_1):</i></p>	<p><i>Answer by ChatGPT (TEXT_2):</i></p>
<p>“We suggest an additional component (causality matrix as a group of special neurons) to update the traditional CNN architecture towards causality-awareness within the image classification process. The resulting causality-aware CNN (CA-CNN) architecture is expected to distinguish between the classes of images, taking into account the causal relationships between the features from the images. The features itself are taken after the last pooling layer of the traditional CNN architecture and they are considered as the objects representing the scene shown in the image. The assumption is made that, in addition to the mutual appearance of each pair of such objects in the image, there might be a hidden causal relationship within each couple. Causality matrix computing is done to set-up (initialize) the hypothesis on potential causalities hidden within the input image, and the CA-CNN training process is a way to actually prove some of the hypothesis. We suggested two computation schemas for the causality matrix, with either more light or more heavy computations. The choice of the analytics to compute the initial values for the causality matrix is based on heuristics so that the estimates for conditional probabilities between the pairs of features could be used as a measure for potential causal relationship between the features” [17].</p>	<p>“Some techniques can be used to infer causal relationships from images, such as analyzing the spatial relationships between objects, recognizing patterns or structures in the image, and understanding the context in which the image was captured. To evaluate possible causal relationships within individual images, one can use causal inference techniques. These methods aim to identify causal relationships between variables by controlling for confounding variables that may affect the outcome of interest. One approach is to use a causal graphical model, which represents the causal relationships between variables as a directed acyclic graph. It helps to identify which variables are potential causes or effects and to determine the strength and direction of the causal relationships. Using causality-awareness for better training CNNs for image classification can help to improve their performance and interpretability. One approach is to incorporate causal information into the training process, either by using causal inference techniques to select relevant features or by designing CNN architectures that explicitly model causal relationships. For example, one can penalize the CNN's parameters that violate the causal relationships learned from the data, leading to a more interpretable and accurate model.”</p>
<p>Comparison made by ChatGPT</p>	
<p>Similarities: Both answers highlight the importance of using causal inference techniques to identify relevant causal relationships between features in images.</p> <p>Differences: However, <i>TEXT_1</i> proposes an additional component, a causality matrix, to update the traditional CNN architecture, whereas <i>TEXT_2</i> does not suggest any specific modification to the CNN architecture. <i>TEXT_1</i> also focuses on computing the causality matrix to initialize the hypothesis on potential causalities within the input image and then using the CA-CNN training process to prove some of these hypotheses. In contrast, <i>TEXT_2</i> discusses the use of causal inference techniques, such as causal graphical models, to represent causal relationships between variables and using this information to select relevant features or penalize the CNN's parameters that violate the causal relationships.</p> <p>Overall: Both answers offer useful insights into how to discover and evaluate causal relationships within individual images and how to use causality-awareness for better training convolutional neural networks for image classification, but they differ in the specific approach suggested for achieving this goal.</p>	
<p>Compromise between TEXT_1 and TEXT_2 suggested/generated by ChatGPT</p>	
<p>Discovering and evaluating causal relationships within individual images can improve the performance and interpretability of CNNs for image classification. There are several techniques available to infer causal relationships from images, such as analyzing the spatial relationships between objects and recognizing patterns or structures in the image. Causal inference techniques can also be used to identify causal relationships between variables by controlling for confounding variables. One approach is to use a causal graphical model, which represents the causal relationships between variables as a directed acyclic graph. Incorporating causality-awareness into the training process of CNNs can help to improve their performance and interpretability. This can be achieved by using causal inference techniques to select relevant features or by designing CNN architectures that explicitly model causal relationships. Alternatively, one can introduce an additional component, such as a causality matrix, to update the traditional CNN architecture. The choice of the approach depends on the specific problem and available data. Overall, incorporating causality-awareness into CNNs can lead to more reliable and useful image classification models.</p>	

4. Strengths and Weaknesses of ChatGPT

Experimental dialogues with ChatGPT inspired us because of the high level of the AI tool performance and because of obtaining some useful information as a result. We documented into Table 5 our general observations in the form of pros and cons related to ChatGPT performance as a research assistant. These observations partially confirm common knowledge and expectations regarding capabilities of ChatGPT, however add more critical (pessimistic) points.

One positive thing to admit (among others) is that ChatGPT is not yet at the level to replace humans in their academic work. However, as a research assistant, ChatGPT appeared to be a very capable tool.

Table 5. The pros and cons of using ChatGPT as an assistance in academic research

General use of ChatGPT	
Pros	Cons
<i>Efficiency:</i> ChatGPT can quickly generate responses to a wide range of prompts without the need for human input. Efficiency will depend on the specific use case and the nature of the task.	<i>Lack of accuracy:</i> ChatGPT may generate inaccurate or misleading responses due to its reliance on statistical patterns in its training data.
<i>Versatility:</i> ChatGPT can be used to generate text on a wide range of topics and in different languages.	<i>Bias:</i> ChatGPT may reflect biases present in its training data, potentially leading to discriminatory or problematic responses.
<i>Accessibility:</i> ChatGPT can be accessed easily online without the need for specialized equipment or software.	<i>Lack of critical thinking:</i> ChatGPT does not have the ability to evaluate the accuracy or quality of the information it generates.
<i>Creative inspiration:</i> ChatGPT can be used as a tool to generate creative ideas and perspectives that may not have been considered before.	<i>Limited understanding of context:</i> ChatGPT may generate responses that do not fully take into account the context of the conversation or prompt.
Academic use of ChatGPT	
Pros	Cons
Assistance in formulating and refining research questions	Limited ability to fully understand the context and nuance of academic research
Assistance in identifying and analyzing relevant literature	May rely too heavily on algorithmic assistance, leading to a lack of critical thinking and independent analysis
Assistance in generating new hypotheses and ideas	May not be able to fully understand the ethical implications of certain research questions or methods
Potential to save time in literature review and research outline	May not be able to identify biases in the training data used to create the model
Ability to generate new text for manuscripts or proposals	May not be able to provide guidance on experimental design or statistical analysis

It also worth mentioning that ChatGPT not only capable of providing textual explanation regarding comparisons of answers to research questions but also able to provide numeric similarity estimation if asked. For example, ChatGPT has evaluated similarity of two answers (TEXT_1 and TEXT2) to RQ formulated in Table 1 to be 0.455. The appropriate computation applies simple Jaccard Similarity utilizing the following attributes for each answer: total unique words and words related to the RQ. A more thorough evaluation and explanation could be required from ChatGPT, which is based on the semantic analysis using pre-trained word vectors (Word2Vec). In the example, the semantic similarity scores indicate that both ANSWER_1 and ANSWER_2 (Table 1) exhibit a reasonable degree of similarity to the RQ. ANSWER_1 has a slightly higher cosine similarity score (0.685) with the RQ, indicating that it shares a slightly stronger semantic alignment with the content and concepts present in the RQ. On the other hand, ANSWER_2 also demonstrates a meaningful degree of semantic similarity with a cosine similarity score of 0.660, showing that it captures relevant aspects of the RQ's content. Final semantic similarity (cosine similarity) between ANSWER_1 and ANSWER_2 from Table 1 is computed to be 0.749.

Michael Wooldridge, the famous contributor to AI in general and to the autonomous agents and multi-agent systems domain in particular, provides his bold opinion [18] about the capabilities of GPT-3 as a foundation model for ChatGPT. He admits that such foundation models have demonstrated impressive performance in some tasks. However, he believes that “they are not the end of the road in AI” because they are “limited with respect to what they have learned and what they can do”. He provided an example by discussing the concept of “wetness”. Is it possible to a system to understand such a concept (i.e., in the context of “rain is wet”) without having actually experienced it in the real world? Michael believes that such systems will be limited without such experiences, in addition to purely symbolic concept representations.

We visualized our opinion on the source of the ChatGPT limitations in Fig. 2, where it is shown that certain important experiences related to cognitive capabilities (i.e., real world observation through sensors, [emotional] processing and understanding these observations, etc.) before verbalizing these experiences by texts, are missing in GPTs and, therefore, in ChatGPT.

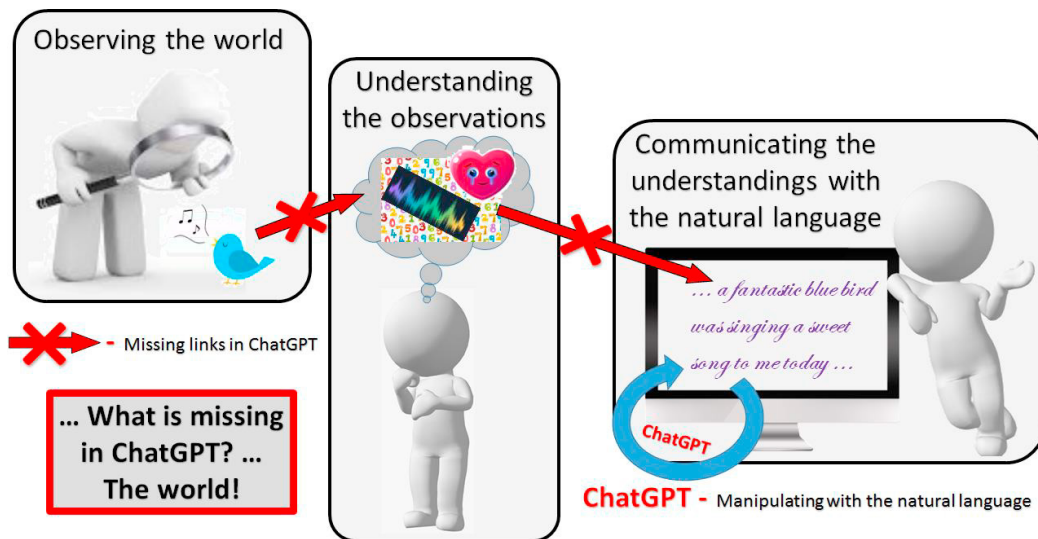


Fig. 2. A visual explanation of the limitations related to ChatGPT.

5. Conclusions

The currently released ChatGPT [1] is meant to assist humans with a wide range of tasks related to smart manipulation with natural language content. In this study, we tested the capabilities of ChatGPT in assisting researchers to evaluate the contribution of published academic research particularly in the domain of Industry 4.0 and Smart Manufacturing. We suggested the dialogue schema in which ChatGPT is asked to answer the research questions from a target article and then to compare and integrate its own answers with the explicit answers from the article. Based on the chat schema, we gained useful context to study the contribution of particular articles.

After experimenting with ChatGPT, we can admit that the use of it can offer significant benefits to researchers working in the domain of Industry 4.0 and Smart Manufacturing, helping them to collaborate more effectively and conduct efficient and impactful research.

While ChatGPT can be useful for academic studies, it has certain limitations that must be taken into account, e.g.: limited understanding of context; lack of domain-specific knowledge; bias in training data; the need for human input and oversight, etc. Therefore, human creativity and intuition cannot be fully replicated by AI.

It is worth mentioning that many academic journals start introducing the issue of ChatGPT in the guidelines for the authors. For instance, the popular journal “Agriculture and Food Economics” clearly states the following: “Large Language Models (LLMs), such as ChatGPT, do not currently satisfy our authorship criteria. Notably, an attribution of authorship carries with it accountability for the work, which cannot be effectively applied to LLMs” (cited from: <https://agrifoodecon.springeropen.com/submission-guidelines/preparing-your-manuscript/reviews>; accessed 13.08.2023).

Regarding the research impact evaluation, in spite of the useful assistance one can get from ChatGPT for that, we can agree with the ChatGPT self-evaluation mentioned above, i.e., admitting one's own limitations on making definitive judgments on the validity or reliability of research discoveries. It is always important to critically evaluate and consider multiple sources of evidence before making decisions and to do this in combination with human expertise to ensure reliable outcomes. Therefore, summarizing the study, we may say the following. Our answer is “yes, to a certain extent” if asked “Is ChatGPT capable of addressing research questions?”. However, if asked “Is ChatGPT capable of evaluating the impact of published research?” our answer will be “not yet”.

In light of our findings, several promising avenues for future research emerge. Firstly, refining the interaction dynamics between researchers and ChatGPT holds the potential for elevating its collaborative potential. Additionally, exploring hybrid models that combine ChatGPT's capabilities with specialized domain knowledge could offer refined insights in specific fields. Lastly, the potential integration of ChatGPT within a larger academic ecosystem (including peer-review processes, collaborative writing and dissemination) could redefine scholarly communication, forging a synergistic partnership that revolutionizes knowledge dissemination in the digital age.

References

- [1] OpenAI (2021). “ChatGPT”. <https://chat.openai.com/chat>
- [2] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., and Sutskever, I. (2019). “Language models are unsupervised multitask learners”. *OpenAI Blog*, **1(8)**: 9. https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf
- [3] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwall, S., ..., and Amodei, D. (2020). “Language models are few-shot learners”. *Advances in Neural Information Processing Systems*, **33**: 1877-1901. <https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfc4967418bfb8ac142f64a-Paper.pdf>
- [4] van Dis, E. A., Bollen, J., Zuidema, W., van Rooij, R., and Bockting, C. L. (2023). “ChatGPT: five priorities for research”. *Nature*, **614(7947)**: 224-226. <https://www.nature.com/articles/d41586-023-00288-7>
- [5] Lin, T., Wang, Y., Liu, X., and Qiu, X. (2022). “A survey of transformers”. *AI Open*, **3**: 111-132. <https://doi.org/10.1016/j.aiopen.2022.10.001>
- [6] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). “Attention is all you need”. *Advances in Neural Information Processing Systems*, **30**: 5998-6008. <https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>
- [7] Zhou, J., Ke, P., Qiu, X., Huang, M., and Zhang, J. (2023). “ChatGPT: potential, prospects, and limitations”. *Frontiers of Information Technology & Electronic Engineering*, 1-6. <https://doi.org/10.1631/FITEE.2300089>
- [8] Deng, J., and Lin, Y. (2022). “The benefits and challenges of ChatGPT: An overview”. *Frontiers in Computing and Intelligent Systems*, **2(2)**: 81-83. <https://doi.org/10.54097/fcis.v2i2.4465>
- [9] Opara, E., Adalikwu, M., and Tolorunleke, C. (2023). “ChatGPT for teaching, learning and research: Prospects and challenges”. *Global Academic Journal of Humanities and Social Sciences*, **5(2)**: 33-40. <https://doi.org/10.36348/gajhss.2023.v05i02.001>
- [10] Rudolph, J., Tan, S., and Tan, S. (2023). “ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?”. *Journal of Applied Learning and Teaching*, **6(1)**: 1-22. <https://doi.org/10.37074/jalt.2023.6.1.9>
- [11] Longo, F., Affenzeller, M., Padovano, A., and Shen, W. (eds.) (2023). “4th International Conference on Industry 4.0 and Smart Manufacturing”. *Procedia Computer Science*, **217**: 1-1954. <https://www.sciencedirect.com/journal/procedia-computer-science/vol/217/suppl/C>
- [12] Dunn, P. K. (2021). “*Scientific research and methodology: An introduction to quantitative research in science and health*”. RStudio, PBC. <https://bookdown.org/pkaldunn/Book/>
- [13] Thelwall, M., and Mas-Bleda, A. (2020). “How common are explicit research questions in journal articles?”. *Quantitative Science Studies*, **1(2)**: 730-748. https://doi.org/10.1162/qss_a_00041
- [14] Treviño-Elizondo, B. L., and García-Reyes, H. (2023). “What does Industry 4.0 mean to industrial engineering education?”. *Procedia Computer Science*, **217**: 876-885. <https://doi.org/10.1016/j.procs.2022.12.284>
- [15] Ali, O., AlAhmad, A., and Kahtan, H. (2023). “A review of advanced technologies available to improve the healthcare performance during COVID-19 pandemic”. *Procedia Computer Science*, **217**: 205-216. <https://doi.org/10.1016/j.procs.2022.12.216>
- [16] Golovianko, M., Terziyan, V., Branytskyi, V., and Malyk, D. (2023). “Industry 4.0 vs. Industry 5.0: Co-existence, transition, or a hybrid”. *Procedia Computer Science*, **217**: 102-113. <https://doi.org/10.1016/j.procs.2022.12.206>
- [17] Terziyan, V., and Vitko, O. (2023). “Causality-aware convolutional neural networks for advanced image classification and generation”. *Procedia Computer Science*, **217**: 495-506. <https://doi.org/10.1016/j.procs.2022.12.245>
- [18] Wooldridge, M. (2022). “What is missing from contemporary AI? The world”. *Intelligent Computing*, **2022**: 9847630. <https://spj.science.org/doi/10.34133/2022/9847630>