On the use of AI-based tools like ChatGPT to support management research

Bastian Burger
HHL Leipzig Graduate School of Management, Leipzig, Germany
Dominik K. Kanbach
HHL Leipzig Graduate School of Management, Leipzig, Germany and
Woxsen School of Business, Hyderabad, India
Sascha Kraus
Faculty of Economics and Management, Free University of Bozen-Bolzano,
Bolzano, Italy and
Department of Business Management, University of Johannesburg,
Auckland Park, South Africa
Matthias Breier
Witten/Herdecke University, Witten, Germany, and
Vincenzo Corvello
Department of Engineering, University of Messina, Messina, Italy

Abstract
Purpose – The article discusses the current relevance of artificial intelligence (AI) in research and how AI improves various research methods. This article focuses on the practical case study of systematic literature reviews (SLRs) to provide a guideline for employing AI in the process.
Design/methodology/approach – Researchers no longer require technical skills to use AI in their research. The recent discussion about using Chat Generative Pre-trained Transformer (GPT), a chatbot by OpenAI, has reached the academic world and fueled heated debates about the future of academic research. Nevertheless, as the saying goes, AI will not replace our job; a human being using AI will. This editorial aims to provide an overview of the current state of using AI in research, highlighting recent trends and developments in the field.
Findings – The main result is guidelines for the use of AI in the scientific research process. The guidelines were developed for the literature review case but the authors believe the instructions provided can be adjusted to many fields of research, including but not limited to quantitative research, data qualification, research on unstructured data, qualitative data and even on many support functions and repetitive tasks.
Originality/value – AI already has the potential to make researchers’ work faster, more reliable and more convenient. The authors highlight the advantages and limitations of AI in the current time, which should be present in any research utilizing AI. Advantages include objectivity and repeatability in research processes that currently are subject to human error. The most substantial disadvantages lie in the architecture of current general-purpose models, which understanding is essential for using them in research. The authors will describe...
the most critical shortcomings without going into technical detail and suggest how to work with the shortcomings daily.

**Keywords** AI-Aided research, Management research, Systematic literature review, I10 general, O33 innovation

**Paper type** Research paper

---

**Introduction**

The use of artificial intelligence (AI) in research is becoming increasingly prevalent as more and more researchers recognize its potential as a valuable tool for data analysis and literature reviews, as demonstrated in recent studies, such as, for e.g. systematic literature reviews (SLR) (Burger et al., 2023). AI can methodically and productively support academic research. Despite being in its early stages, AI is already showing great promise and has the potential to revolutionize the way we conduct research, particularly for no-code research applications (Calo, 2017). As we delve further into the possibilities of AI, we consider the technology at a stage where no technical skills are required to use AI to enhance and improve research (Dwivedi et al., 2023).

The point of contention of the current discussion is the Generative Pre-trained Transformer (GPT) models by the private company OpenAI. ChatGPT, a GPT-3.5-based application, has gathered much attention lately as of early 2023. This editorial aims to discuss how researchers can utilize GPT-3 and similar models to improve their research, thus outlining a practical guide for using this technology in scientific research. Instead of discussing the pros and cons of AI, which have been covered extensively elsewhere (see, e.g. Bostrom and Yudkowsky, 2018), we will focus on how GPT-3 and similar models can be a tool to support research. Several studies across different research fields published recently even listed a GPT derivates as a co-author (e.g. Kung et al., 2022; Transformer and Zhavoronkov, 2022; Transformer et al., 2022).

We will not delve into the nature of AI, as it is a complex and loosely defined collection of a vast number of methods and concepts that, given sufficient computing power, can solve different problems of classical computing (see e.g. Moroney, 2020). Our goal is to show researchers how they can harness the power of High-Level Transformer- and similar models to improve their research and to provide a clear and actionable guide on how to do so. We believe that providing this guide can help researchers better understand AI’s potential and how to use it to improve their research. We will not discuss AI’s pros and cons on any moral dimension, which numerous outlets discussed in great detail.

**Status quo of AI in research and beyond**

**Current developments in AI**

Recent trends in AI outside of research have seen significant advancements in capabilities. The current hype around transformer models can potentially revolutionize areas such as entertainment, art and advertising. Additionally, AI is increasingly integrated into various industries, such as finance, healthcare and manufacturing, to automate repetitive tasks, optimize processes and improve decision-making.

The recent hype in late 2022 and early 2023 is about transformer networks, a type of neural network architecture. These models train on a large dataset of text, such as books, articles and web pages, and they learn to predict the next word in a sentence based on the previous words (Vaswani et al., 2017). The model uses an attention mechanism that allows it to focus on specific parts of the input when making predictions, allowing it to understand the context of the text better. Once the model is trained, it can be fine-tuned for a specific task, such as language translation or text summarization, by training it on a smaller dataset of text specific to that task. The fine-tuning process allows the model to adapt to the specific characteristics of the task and to
learn patterns and structure of the text specific to that task. The model can then be used to generate text by providing it with a prompt or a seed text, which it uses to generate a coherent and fluent response. GPT-3 and similar models can also perform other language-based tasks such as question answering, text completion and sentiment analysis. Thus, the model is, of course, not “thinking” but simply predicting what to say next.

**Current developments in research**

While there has been significant progress in the development of AI technologies in recent years, the potential of AI in research is by far not yet fully realized and many opportunities remain unused. We see important and welcome trends in research, such as the use of open science practices, where researchers make their data, code and methods available to others. This procedure enables greater transparency, reproducibility and collaboration in research (Ramachandran et al., 2021). Additionally, we see a growing emphasis on using rigorous experimental design and statistical methods to validate research findings. This experimentation includes using randomized controlled trials, large sample sizes and powerful statistical methods to improve the accuracy and generalizability of research findings.

Furthermore, there is a growing trend in using meta-research, which involves studying the research itself. This meta-research includes the study of biases, reproducibility and research transparency. All of which aims to improve the overall quality of the scientific enterprise. All these developments improve the quality of research, but they also put on new challenges for the researchers. As the computer helps us work with larger data sets and protocol our procedures, we will need to eventually adapt to AI to create a new level of research.

However, many researchers are still unfamiliar with the current state-of-the-art AI techniques and lack the necessary skills to leverage them effectively in their research. Therefore, much more work must make AI a more valuable tool for researchers.

**Applying AI to the research process**

*Why to add AI to the research process*

There are many good reasons to consider the use of AI in research. The most obvious is that AI can take on tedious tasks and everyday chores. Adding AI to the research process has numerous contribution advantages as well. The most important of which is reducing human error. Unlike humans, AI will never skim data due to tiredness or be distracted, ensuring repeatable results if the input is provided thoroughly and correctly (Wu et al., 2018). Another advantage is that AI—as well as most other computational systems—can deliver repeatable results. With the proper parameters, the AI will always produce the same responses given the same history and input. This consistency is especially valuable as it eliminates the possibility of human variability in interpreting research. Furthermore, AI can offer a second set of eyes, providing an additional layer of precision in research protocol without requiring additional time, resources, or money. While AI may not reach the same accuracy as double-blind research conducted by multiple researchers, it offers a cost-effective alternative for organizations looking to improve their research process.

*Where to add AI*

At the current stage of development, AI can serve in various use cases. It can prove helpful in almost all cases of data analysis and classification. We will follow with a short description of the most apparent applications: (1) In data classification in interviews, AI can help identify patterns and themes in data classification in interviews that may not be immediately apparent to human researchers (Cui and Zhang, 2021). (2) In image analysis, computer vision tools can analyze images, such as microscopy images, to identify patterns and insights (Davenport and Kalakota, 2019). (3) In emotion analysis in chats and other unstructured data, AI can help identify sentiment, emotion and
other information that may not be obvious to human analysts (Gopalakrishnan et al., 2019). (4) In quantitative data analysis, AI can help identify patterns and trends in the data that may not be immediately apparent to human researchers. (5) In pattern recognition, AI can analyze large amounts of data, such as medical records, to identify patterns and insights that may not be immediately apparent to human researchers (Kong et al., 2020). (6) Finally, tools can automate repetitive tasks such as data entry and annotation, freeing up researchers’ time to focus on more critical tasks. This enumeration does not claim to be exhaustive but might give a feeling for the versatility of this new technology.

One additional field of application we must consider is (systematic) literature reviews, in which AI is helping researchers quickly identify relevant papers and articles and analyze them.

How to add AI: applying AI to the SLR process

Based on our experience, the use of AI in the research process can be categorized into three main steps: (1) the AI preparation phase, (2) the use of AI in the research initiation phase and (3) the use of AI in the data analysis phase as outlined in the following:

(1) The AI preparation phase

For this discussion, we will introduce the concept of AI snapshots. The history of prompts matters for the response of a transformer model and can change the model’s output. Hence, we need to be able to follow the path from the bare model (vanilla version) to the current response. The snapshots should be documented for review. As most developers and providers of these technologies do not provide a seed code or something compatible to describe the AI, researchers must be careful in their documentation. Document the date of usage, the version of the AI, the exact parameters used and the exact prompts used. We require three AI snapshots for context-sensitive AI: (a) Briefed AI: knows the research question but does not have access to the data. (b) Informed AI: has accessed the data but did not synthesize it. (c) Synthesized AI: has conducted a data synthesis.

(2) Use of AI in the research initiation phase

We start our research by formulating the research question. As with all these steps, test them against the AI, so the AI understands it clearly and correctly. There must not be ambiguity in the research question. OpenAI uses the term “temperature” to control the randomness of the output. In a nutshell, a high value produces varying wording in a response, while a value of zero produces perfectly deterministic results. Even with a high temperature, the AI must produce repeatable and reliable reactions to it. From there, we build our research string as outlined in standard procedures for review studies (e.g. Kraus et al., 2020, 2021, 2022; Tranfield et al., 2003).

To build a research protocol for using AI in the research process, it is crucial to formulate the protocol thoroughly. The protocol will be the exact input strings for the AI and it is crucial to adjust the formulations until the AI returns the expected results on sample data. In this regard, the research protocol becomes the exact, interpretation-free program of the research. Responding to the objection that the actual conduct is unknown by the outset of the research, we urge that the foreseeable next steps must be programmable. Alternatively, research an in vitro sample and derive the protocol from there. As the researcher continues to refine the protocol, it is important to increase the non-determinism to ensure consistent results. This step ensures that the protocol is free of interpretation, at least for the model in use. Following these steps ensures a robust research protocol that will help get the most accurate and reliable results possible.

(3) Use of AI in the data analysis phase

The following steps are methodology specific: For the SLR, we collected candidate papers and screened the abstracts following the previously mentioned method papers. What human researchers cannot do easily, though, is screen all candidates in full text. The AI can, however.
Therefore, we use AI to full-text screen all papers. Feeding content is straightforward with hypertext markup language (HTML) formatted papers. GPT-3 is capable of interpreting HTML and cascading stylesheets (CSS). However, it is advantageous for readability, and beyond everything else cost, to strip all syntax. GPT-3 and ChatGPT both support Markdown, which is a preferable formation for input strings. Properly formatted portable document format (PDF) documents also can easily be extracted and transferred to Markdown syntax. Note that Markdown is not required. Both OpenAI products work properly on unformatted text (it will also strip rich text to unformatted text).

Keep in mind that OpenAI models cannot access the Internet at the time of this writing. The “new Bing” announcement by the Microsoft Corporation and Google’s BARD will change this fact. At the time of writing, the authors had only the opportunity to use the first one for a limited time. Hence, we cannot draw any conclusions yet. The AI must read the full text instead. Therefore, writing a script to download the papers and scrape their contents is advised. This scripting is possible with limited technical knowledge and a quick web search.

Also, remember that the maximum input per prompt for OpenAI is 2,048 Tokens. A Token converts in the ideal case to one word, in special cases to less. Thus, each paper might need several input prompts. The demands of all papers could amount to substantial financial investments, and some researchers might be motivated to switch to a less powerful model, which we strongly discourage. The strongest models can barely add practical value to our research, whereas the simpler models are not. OpenAI also offers a “Tokenizer” that lets everyone translate texts to tokens.

In the data extraction, first, generate the data. As usual, find a way to manage all data in a database, spreadsheet, or data dump. The minimum data for ranking by importance are the journal of publication, the number of citations and the publishing date. For classification, the minimum data are the outcome/finding, the methodology, the keywords, the abstract and the full text. Next, ask the AI for the same data and compare the data. This limitation is partly due to the architecture of transformer models, which we will examine later (Vaswani et al., 2017). The most exciting data are conflicting data. These can give a hint for a misunderstanding. While until GPT-3, Generative Artificial Intelligence (GAI) models are not strong enough to raise concerns. They do provide a sanity check.

In data synthesis, first, create a model for synthesis according to the literature (Kraus et al., 2020) and condense the data on that basis. The AI at snapshot two can help with two data checks. Check one: Add AI to the data sheet, let it do the same cluster and check for differences. Check two: Let the AI at snapshot one generically summarizes findings for each dimension (without feeding data) and check for differences. In this check, the AI will probably not create a complete picture but might hint at something missing from the review. In all cases, the resolution for most conflicts will probably be an error by AI, but it still should be documented.

The resulting review will add a layer of methodology. Even without the component of AI, it will reduce ambiguity and increase documentation. The researcher still must be in the driver’s seat and decide about the correctness of the analysis.

**Current limitations and drawbacks**

**Progress in AI application**

The field of AI continues to evolve and improve. One of the most notable developments has been the recent release of the GPT-3 model, which has received significant attention due to its advanced capabilities. While GPT-3 has many strengths, it is not without its limitations. One of the most notable challenges is its lack of advanced context sensitivity, which can sometimes make it less effective in certain situations.

However, progress has in recent months – and weeks – with the release of ChatGPT, new Bing, Bard and others, a newer and more advanced AI model. Unlike GPT-3, ChatGPT is
context-sensitive, which allows it to integrate more deeply with the system in use. This advanced capability has been praised by many experts as a major step forward in the field of AI and promises to lead to many exciting and innovative applications shortly.

Limitations

The AI landscape is rapidly changing, and there are several key points that researchers must keep in mind when working with AI models. One of the first things to be aware of is the cut-off date for most models, including GPT-3 and ChatGPT, beyond which they have no training data. This cut-off date means that if the field of research has changed substantially after that point, AI may not provide insightful data. Additionally, it is essential to note that all training data is open-access only and closed-access papers are usually not part of the AI training. Exceptions can be made for papers accessible through alternative measures, but publishing houses that use paywalls to protect articles are hindering progress in yet another way. Predatory journals often publish Open Access and might enter the generative AI training set and create a bias, despite the developers’ best efforts to discount misinformation sources.

It is also worth mentioning that, even when using OpenAI models, the AIs may sometimes fabricate untruthful responses. This “hallucinating” is particularly true when asking for a literature reference; GPT-3 and ChatGPT can come up with compelling elaborate titles and authors that are, however, wholly made up.

To ensure repeatability, documentation of the exact wording of the AI prompts, and all input parameters of the AI are crucial. Peer reviewers are also encouraged to reproduce the results of the AI quickly and adjust the prompts slightly, as the most popular AI models are owned by private companies and can be adjusted and changed at will without snapshots of previous versions. Open-source implementations like Stable Diffusion for image generation are available but not as popular and not in all fields of AI.

After all, it is important to remember that the researcher is always fully responsible for the results they get from AI models. The researcher must always keep these points in mind and exercise caution when working with AI, as the field is constantly evolving and new challenges are always emerging.

In that line of argument, AI will not and cannot provide a reliable explanation. In general, the researcher needs to understand the model in use. For example, this includes reading the base papers of each architecture to understand the capabilities and limitations. We understand that this is a bone-dry task, even by research standards. However, with the speed of development in AI, there is no alternative to it. The most common types of neural networks are:

1. Feedforward Neural Networks (MLP – not to be confused with NLP): A type of artificial neural network where the data flows in only one direction, from input to output, without looping back.

2. Convolutional Neural Networks (CNN) (LeCun et al., 1998): A type of neural network used in image recognition and processing that applies a convolution operation to the input data to extract features for classification.

3. Recurrent Neural Networks (RNN) (Le et al., 2015): A type of artificial neural network that uses recurrent connections to allow the network to process sequences of input data, such as time series or natural language.

4. Transformer Networks (Vaswani et al., 2017): A type of neural network architecture that uses self-attention mechanisms to process data sequences, such as in natural language processing tasks.
Generative Adversarial Networks (GAN) (Goodfellow et al., 2020): A type of deep learning algorithm where two neural networks are trained together, with one trying to generate data that is indistinguishable from actual data and the other trying to distinguish between real and generated data.

Autoencoder (Hinton and Salakhutdinov, 2006): A type of neural network architecture used for unsupervised learning where the network trains to reconstruct its input data while learning to reduce the dimensionality and extract the essential features.

Of course, neural networks are not the only methods in AI. However, to stick with the example of neural networks, the deeper the levels of the network become, the harder it is to interpret the decision-making vectors of each engine. Particularly complicated is this understanding, as the decision-making of neural networks is not semantic but statistical. What results is a lack of interpretability. In summary, AI will not provide a theory/model for any research and non-domain-specific tools cannot be made to do so.

**Conclusion**

In conclusion, using AI in (management) research can significantly improve the objectivity and accuracy of the results when applied correctly. By adding an automated component to both the research initiation phase as well as the data analysis phase, we can reduce the potential for human error and achieve better reproducibility. While the current state of AI – generalized AI models in particular – is not where a researcher can trust the results blindly, they often hint in the right direction and uncover oversights. The ability to process data at a deeper level, as well as the ability to achieve faster results, makes AI a valuable tool in the field of research. The theoretical implications of this research suggest that using neural networks in data analysis can bring us one step closer to research results that are free of human error and can lead to a higher level of objectivity in the field. We must also highlight that the use of AI is not limited to any methodology or a certain level of technical savvy required by the researcher. While it is true that a theoretical core understanding of AI in general and the AI employed in specific is necessary and basic skills in Phyton, including frameworks like Pandas, PyTorch, or TensorFlow are beneficial. The latter can be easily obtained and institutions should provide courses for their researchers.

**Practical implications**

To get started with AI, researchers first must familiarize themselves with the topic. OpenAI provides a “Playground” feature that allows users to play around with the AI models without Application Programming Interface (API) access. Getting a feeling for how AI works and its limitations and capabilities is the crucial first step in any AI aspiration. As with any other tool, we need to understand its features first. With non-linear models such as AIs, exact control of the parameters is crucial (compare the “butterfly effect” in chaos theory). For the same reason, prompts must be reliable and exact.

To be safe, we should always use a “fresh” process for each research task. Once we ask an AI process about the weather, future responses might be tainted. In ChatGPT, for example, this indicates using several chats side by side. One to try out and fine-tune prompts and one that strictly advances the research (and other chats as the researcher pleats).

Most importantly, AI will not provide reasoning for any research. It is always the researchers’ responsibility to provide causality and reason for all findings. In our experience so far, AI is also unfitted to suggesting causality, as the models are always, and by definition, trained on existing data and hence are unable to draw novel conclusions.
Future opportunities
The usage of AI in research has just begun. In the future, there will be plenty of applications that can support the research community with AI. Fed with suitable models and frameworks, AI could for example create continuous literature reviews that are always up to date and always sum up the current state of the research. Another practical application would be a list of every current call for paper to increase and simplify transparency of academic discussions. It can also help simplify research language so that scientific insights are accessible to a broader public as the linguistic barrier drops.

We want to call aspiring entrepreneurs to create research-specific AI. While generalized AI is an excellent way for many use cases, research demands strongly differ from the broad public. For example, research questions are not interested in the most common or most convenient answer. Also, research has training requirements on all research literature and current databases, which some organizations are opposed to sharing. A research AI must also be highly sensitive to new data, setting that data into a research context and knowing and applying methodologies. Finally, a research AI must be able to differentiate between its own “creative” writing (as observed before, GPT-3 tends to “invent” research papers) and recitation of sources.

Finally, as with other technologies, it is probable that AI will not entirely replace human work but will integrate with it, generating new and more complex forms of human-machine interaction. Critically analyzing these evolved forms of interaction and striving to quickly understand them, rather than being simplistically enthusiastic or hostile towards new technologies, is the right attitude to advance any area of activity, including scientific research.

References


**Corresponding author**

Vincenzo Corvello can be contacted at: vincenzo.corvello@unime.it

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com