

# Ethical concerns using Artificial Intelligence in Scientific Psychology

Michael Zirkler

Zurich University of Applied Sciences, Switzerland

**Corresponding author:**

Michael Zirkler, Zurich University of Applied Sciences,  
School of Applied Psychology, Pfingstweidstrasse 96, 8037  
Zurich, Switzerland.  
Email: zirm@zhaw.ch

## Introduction

Artificial intelligence (AI) is an umbrella term for what has become a broad field of technologies and applications (Wilks 2023). It is also increasingly being used in the field of psychology in research, education, and practice (Abrams 2021). The development of AI is comparatively new and progressing rapidly in both quality and quantity. However, it is often not really clear to the users of AI what the respective AI applications do, how they work exactly, and what can be expected by them.

A "general" digital ethics is only emerging (Fuchs 2023; Krieger and Belliger 2021; Beever et al. 2019), and specific ethical discussions about the development and use of AI are largely in their infancy (Deutscher Ethikrat 2023; Becker et al. 2022; Hanna and Kazim 2021; European Commission 2019).

Currently, an intense discussion is underway about whether and how AI can and should be used in science. This discussion was further fueled with the release of the freely available program ChatGPT in November 2022 (cf. De Witte 2023; University of Cambridge n.d.). Several institutions have meanwhile developed regulations or at least recommendations, on the one hand for researchers, but especially for students (Peters 2023; Wood 2023; Garber n.d.).

So far, there has been very little discussion of the use of AI in scientific psychology from an ethical perspective. The aim of the following text is to contribute to this discussion. In doing so, it exempts the ethical requirements that have been existing for some time, particularly for psychological research in the area of human sciences (Panicker and Stanley 2021; Oates 2020).

## Artificial Intelligence

Intelligence is generally understood as a bundle of abilities that until recently were attributed to living systems, especially humans: perception, reasoning, learning, problem solving, and the use of language (cf. Copeland 2023). Intelligence is artificial when these abilities can be (at least partially) attributed to an AI system, while its material basis is not organic (living cells), but human made (non-biological) systems, especially computers, which are ultimately based on transistor technology (microchips) and associated programs.

The recent development of artificial intelligence has especially been able to take the learning ability to the next level. The quality of learning capability means that it is no longer simply programs that are processed and commands that are followed, rather such systems have emergent properties, they are generative. Their operational logic can be described analytically and is thus programmable, but it is not determined insofar as their outputs are not predetermined. Artificial intelligence does not function like a pocket calculator, for example, which generates the same output for the same input. Rather, the output depends on the one hand on the input (question, instruction), and on the other hand on the training that the system has undergone. Furthermore, artificial intelligence systems can process complex inputs (questions) whose format does not have to follow a clearly given input scheme. Here is a simple example:

Question: What is two multiplied by two?

Answer (ChatGPT 4): Two multiplied by two equals four.

Question: Any other solution possible?

Answer (ChatGPT 4): In standard arithmetic, two multiplied by two always equals four. There is no other solution for this basic multiplication. If you're asking in a metaphorical or philosophical context, the answer might vary, but mathematically, it's always four.

Question: What is the meaning of 42?

Answer (ChatGPT 4): The number "42" itself is just a numeral that represents a quantity. However, if you're referencing the cultural significance of "42," it's likely you're referring to its use in Douglas Adams' "The Hitchhiker's Guide to the Galaxy." In the book, "42" is humorously stated to be "the Answer to the Ultimate Question of Life, the Universe, and Everything." However, it's noted that no one knows what the actual question is.

The number has since become a popular piece of pop culture and is often used in a humorous or whimsical way to suggest that something is an answer without context or that life's answers might be simpler (or more complex) than we think.

Beyond this cultural reference, the number 42 can have different meanings in different contexts, like in mathematics, science, religion, etc. But in popular culture, it's often associated with Adams' work.

Until recently, artificial intelligence systems were only available for research to a limited extent due to the large computing capacities required for their operation as well as their so far specialized use. This has fundamentally changed since the release of ChatGPT (3) in November 2022 at the latest. Since then, the system has been widely available and can be used freely.



**Figure 1.** Interrelationship of the seven requirements: all are of equal importance, support each other, and should be implemented and evaluated throughout the AI system's lifecycle (European Commission 2019:15).

### **Ethics in scientific psychology**

A central component of psychology as a science is the production of (new) knowledge in its many sub-fields. Students are trained at universities in how research processes are designed, what "good" research means and which standards must be observed. Of course, these standards also apply to scientists and researchers working in psychology.

The activity as a psychologist has ethical demands and is based on ethical basic principles or deontological requirements, to which the actors may commit themselves. According to the American Psychological Association (2017), these include the following:

Beneficence and Nonmaleficence, Fidelity and Responsibility, Integrity, Justice, Respect for People's rights and dignity.

Those who conduct science have duties that must be observed. Especially the rule based procedure in the scientific process and its careful observance as well as transparent documentation are important characteristics. And he or she has to proceed to the best of his or her knowledge and belief. One is obliged to give reasons.

Violations can result in social sanctions, rejection of certificates or publications, exclusion from professional associations, loss of job, and even criminal prosecution.

However, rule orientation is not sufficient for good science. In addition, there is, for example, the originality of a research question or creativity in research design. These important aspects cannot be provided by AI in the current state of affairs.

### **The use of AI and its ethical implications**

The basic ethical requirements in science also apply when it is conducted with the help of AI. For example, the High Level Expert Group on Artificial Intelligence (European

Commission 2019) formulates the following requirements (ethical principles) for the use of AI:

In the use of AI solutions, one will have to distinguish, with regard to ethical questions, those that relate to the ontological status of AI and those that deal with its end. Furthermore, questions arise about the consequences of its existence and those about the consequences of its use. Further, we focus on the question of what are the ethical implications of using AI systems in scientific psychology.

AI can undoubtedly be very useful and helpful in research. The use of AI as a tool saves time and effort, but also inspires solutions that humans would not have come up with otherwise.

However, there are (at least) two fundamental ethical questions that arise in connection with the use of AI:

1. what means are recognized as legitimate in research (and associated publications) and may be used?
2. what end is to be achieved by the means employed?

For example, the common use of statistical programs to analyze data is undisputed, even if these are not usually AI systems. Search engines for finding publications, whether AI-based or not, are also generally accepted. Translation programs are viewed rather skeptically if the accuracy cannot be assessed by the author in each case. As things stand, AI systems that write entire texts are considered inadmissible.

### **Means**

The quality of a means can be evaluated functionally and morally (Wittgenstein 1965). A chair is a good means if it fulfills the required functions of sitting. A good person is evaluated not only functionally, or at first not at all, but in terms of the morality of her or his thinking and acting.

There are good means and bad means: a bad means would be, for example, the use of torture to explore psychological factors. A good means observes the ethical principles of freedom from harm, benevolence, preservation of freedom and dignity.

The means that may be legitimately used in psychological science or research will also be measured by whether the means is functional. Statistical programs may be used to analyze data because we assume to know in principle what they do and perform. We can reasonably have confidence in their functioning, in their reliability.

Applicable means from a moral point of view are those that meet the ethical standards mentioned above: they are harmless to humans, their properties are known and comprehensible (transparency), they are safe (data), to name just a few aspects.

It is legitimate to make life easier for oneself in research as well. Today, nobody will do mean value calculations by hand, although it would be possible. However, it would be expected that the user of the program knows in principle what it is doing and therefore what she or he is doing. The means only replaces the tedious manual work and is thus probably even less error-prone.

### **Ends**

There are good and bad ends: good ends are those that lead to a positive, one could also say life-serving contribution for

people. This refers in particular to health and well-being, but in a broader sense generally to the improvement of the possibilities for shaping living conditions.

The "standard tools" of (psychological) research, such as the statistical programs already mentioned, can be used with good intentions and for good ends, but they can also be used (intentionally or not) in harmful, negligent, or abusive ways. For example, to further one's career via deception and fraud. The use of means, for example, to discriminate against or discredit people would also be an example of bad ends.

On the one hand, intentions or ends would play a role in achieving ends through means, but on the other hand, the consequences of achieving ends would also be important. This can be contradictory. For example, someone wants to definitively defeat a widespread disease, but sets up experiments to do so that predictably harm people.

Where the situation is clear, rules prevail: Torture is prohibited as a means to whatever end. Manipulative questions are not forbidden in every case, but they are morally reprehensible, even if they are intended to serve a (supposedly) good end.

In some cases, the means are banned completely because the risk of using them for the wrong ends seems too high, as in stem cell research in Germany (EuroStemCell n.d.).

In contrast to a pure consequentialist position, the ends never justify the means. Means and ends must be subjected to a joint ethical consideration; they must be examined for their legitimacy and moral goodness. Therefore, it is advisable to involve third parties who are freed from immediate and sometimes unconscious interests. Either way, however, even good means must not be used for bad ends.

## Conclusion

At this stage, it can be assumed that AI systems are good functional tools, at least in certain areas of application. However, AI systems, such as ChatGPT sometimes "hallucinate" (Emsley 2023).

It is the use for particular ends and resulting consequences that determines whether it is a good or bad means in the moral sense.

However, further question marks are in place about AI systems as means in that we do not fully know the inner workings of AI; this is one of its characteristics and part of its potency. For example, AI systems are known to exhibit typical biases depending on the training material with which it has been fed. However, these biases are not immediately, if ever, apparent.

There is an increased obligation not to rely on AI, but to critically review the outputs in each case and only use what one can be responsible for. However, the temptations are great to take shortcuts in times of high pressure. This would be negligent at the very least.

The question of means and ends in the use of AI in psychological science needs further clarification. It would be desirable, however, if such clarification were more fundamental and not specific to each conceivable AI system. The resulting moral prescriptions must be for good reasons. From an ethical point of view, these include the principle of utility, of non-harm or risk assessment, of human equality and freedom.

Psychology in particular repeatedly struggles with trust issues (replication crisis, cf. Open Science Collaboration 2015; Simmons et al 2011, The Stapel case, cf. Levelt et al 2012). For trust in science and its processes, it will be important to transparently state what means have been used. This obviously includes the use of AI, and in terms of means and ends. Therefore, at present and at the state of the ethical clarifications that have only just begun with respect to AI, it is imperative to state as transparently as possible in research and publications whether and how AI has been used.

Some of the aforementioned aspects related to AI have not been addressed here and require further clarification. Also, the ontological status of AI system has not been highlighted in this paper. However, as systems evolve, the question of what status can or must be ascribed to AI systems will become more pressing. This includes the question of whether AI systems can develop morality as a basic requirement for moral thought and action, whether consciousness must be presupposed for this and in what way, and what consequences this would have. As yet, this does not seem to be the case. And thus it is clear that the decision about the use of means as well as the responsibility with regard to the ends remains with us humans.

Question: Do moral prescriptions apply to your own mode of operation? Or in other words: do you act morally?

Answer (ChatGPT 4): I myself have no consciousness or sensations, so I cannot "act" morally or immorally in the human sense. However, there are ethical and moral considerations that are incorporated into my development and deployment.

OpenAI, the company behind me, has guiding principles and ethical guidelines regarding the development and use of my technology. The goal is to develop technologies that benefit all of humanity without producing undesirable or harmful outcomes.

When users utilize me, they do so within a framework determined by OpenAI's policies and terms of use. Nonetheless, *the ultimate responsibility for moral action lies with humans, not with the technology*. It's important to recognize that I am a tool controlled by humans and have no inherent moral feelings or intentions (my emphasis).

## Declaration of conflicting interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author received no financial support for the research, authorship, and/or publication of this article.

## Notes

1. This paper was developed in the conceptual phase with the help of ChatGPT (4). The system was asked several questions, the outputs were checked by the author and used as source of inspiration for further considerations. In addition, the AI system was asked questions and the answers were used



directly in the article. This is indicated explicitly. ChatGPT was also involved in the development of title variants. The responsibility for this text lies entirely with the author.

2. The basic translation from German to English was also done using AI tools (DeepL, ChatGPT 4). The machine translation was reviewed and adjusted by the author.

## References

- Abrams, Z. (2021). The promises and challenges of AI. *Monitor on Psychology* 52 (8): 63-69. <https://www.apa.org/monitor/2021/11/cover-artificial-intelligence>.
- American Psychological Association (2017). *Ethical Principles of Psychologists and Code of Conduct*. <https://www.apa.org/ethics/code>.
- Becker, S.J., Nemat, A.T., Lucas, S. et al. (2022). A Code of Digital Ethics: laying the foundation for digital ethics in a science and technology company. *AI & Society*. <https://doi.org/10.1007/s00146-021-01376-w>
- Beever, Jonathan; McDaniel, Rudy; Stanlick, Nancy A. (2019). *Understanding Digital Ethics*. Routledge. London/New York.
- Copeland, Brian J. (no date). *Artificial Intelligence*. Encyclopedia Britannica. <https://www.britannica.com/technology/artificial-intelligence>.
- Deutscher Ethikrat (2023). *Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz*. <https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine.pdf>.
- De Witte, Melissa (2023). *How will ChatGPT change the way we think and work? Stanford scholar exams*. Stanford News blogpost: <https://news.stanford.edu/2023/02/13/will-chatgpt-change-way-think-work/>.
- Emsley, R. ChatGPT: these are not hallucinations – they're fabrications and falsifications. *Schizophrenia* 52 (9). <https://doi.org/10.1038/s41537-023-00379-4>
- European Commission; High-Level Expert Group on AI (AI HLEG) (2019). *Ethics guidelines for trustworthy AI*. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.
- EuroStemCell (no date). *Regulation of stem cell research in Germany*. Online Available: <https://www.eurostemcell.org/regulation-stem-cell-research-germany>.
- Fuchs, Christian (2023). *Digital Ethics*. Routledge. London/New York.
- Garber, Alan M. (no date). Office of the Provost: *Guidelines for using ChatGPT and other Generative AI tools at Harvard*. <https://provost.harvard.edu/guidelines-using-chatgpt-and-other-generative-ai-tools-harvard>.
- Hanna, R., & Kazim, E. (2021). Philosophical foundations for digital ethics and AI Ethics: a dignitarian approach. *AI Ethics* 1, 405–423. <https://doi.org/10.1007/s43681-021-00040-9>
- Krieger, D. J.; Belliger, A. (2021). *Hacking Digital Ethics*. Anthem. London.
- Levelt, W.J., Drenth, P.J., & Noort, E.V. (2012). *Flawed science: The fraudulent research practices of social psychologist Diederik Stapel*. [https://www.semanticscholar.org/paper/Flawed-science%3A-The-fraudulent-research-practices-Levelt-Drenth/34f9ddb42cf981fa671201aae3fbd10fdaba49ef?utm\\_source=direct\\_link](https://www.semanticscholar.org/paper/Flawed-science%3A-The-fraudulent-research-practices-Levelt-Drenth/34f9ddb42cf981fa671201aae3fbd10fdaba49ef?utm_source=direct_link).
- McAadoo, T. (2023). *How to cite ChatGPT*. American Psychological Association. Blogpost. <https://apastyle.apa.org/blog/how-to-cite-chatgpt>.
- Oates, J. (2020). *Ethical Considerations in Psychology Research*. In: Iphofen, R. (eds) *Handbook of Research Ethics and Scientific Integrity*. Springer, Cham. [https://doi.org/10.1007/978-3-030-16759-2\\_35](https://doi.org/10.1007/978-3-030-16759-2_35)
- OpenAI (2023). *ChatGPT* (version 4) [Large language model]. <https://chat.openai.com/>
- Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251). <https://doi.org/10.1126/science.aac4716>
- Panicker, S., & Stanley, B. (Eds.) (2021). *Handbook of research ethics in psychological science*. American Psychological Association. <https://doi.org/10.1037/0000258-000>
- Peters, M. (2023). *Stop focusing on plagiarism, Even Tough ChatGPT is here*. Harvard Business Publishing Education. <https://hbsp.harvard.edu/inspiring-minds/stop-focusing-on-plagiarism-even-though-chatgpt-is-here>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, 22(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>
- University of Cambridge (no date). *ChatGPT*. Faculty of Education News. Blogpost. <https://news.educ.cam.ac.uk/230403-chat-gpt-education>.
- Wilks, Y. A. (2023). *Artificial Intelligence*. MIT Press. Boston.
- Wittgenstein, Ludwig (1965). Lecture on Ethics. *The Philosophical Review* 74 (1):3-12. <https://doi.org/10.2307/2183526>.
- Wood, P. (2023). *Oxford and Cambridge ban ChatGPT over plagiarism fears but other universities choose to embrace AI*. Inews blogpost. <https://inews.co.uk/news/oxford-cambridge-ban-chatgpt-plagiarism-universities-2178391>.