Considerations for the Use of AI Tools at the Centre for Social Innovation

Version 1, July 2024

Authors: Jesse de Pagter, Stefanie Schuerz, Dietmar Lampert

Contributors: Lina Klingbacher, Maria Schrammel, Utku Demir

Reviewed by: Ursula Holtgrewe, Katharina Koller, Desiree Pecarz, Laure-Anne Plumhans, Klaus Schuch, Elisabeth Unterfrauner, Gorazd Weiss, Wolfgang Michalek, Barbara Glinsner

ZENTRUM FÜR SOZIALE INNOVATION
CENTRE FOR SOCIAL INNOVATION

Considerations for the Use of AI Tools at ZSI

Table of Contents

Preamble

The goal of this document is to inform ZSI staff on key aspects relating to the use of AI for work at ZSI, including relevant terms and concepts, legal and regulatory frameworks, and a range of considerations for a responsible and effective use of AI. Moreover, it puts forward a course of action for ZSI in how to integrate AI tools effectively and responsibly in the future. The document closes with an initial set of potential use cases that could serve as a basis for a growing collaborative collection to guide ZSI staff in their work with AI.

Since the technological possibilities and regulatory framework surrounding AI is very much in flux at the point of the publication of this document (July 2024), it must be seen as preliminary. ZSI is committed to continuing its active and critical engagement with new and developing technologies to integrate them in a responsible manner into its workflows.

Introduction

Following the rapid developments in algorithmic decision-making and artificial intelligence (AI), there has been considerable activity on the responsible, ethical management and governance of such systems. Issues that have been addressed range from potential harms, to the protection of important liberties – including human rights – to accessibility and equity when it comes to AI. These kinds of developments demonstrate a trend where an increasing number of countries and organisations aim to anticipate AI impact, often referring to the language of ethics.

Broadly speaking, initiatives and non-binding guidelines for responsible AI governance are usually based on an agreed set of broad (democratic) principles. Important examples of such principles are: privacy and data governance, accountability and auditability, robustness and security, transparency and explainability, fairness and non-discrimination, human oversight, and promotion of human values. They offer a wider framework of reference that organisations using or developing AI can adopt to encourage the responsible and human-centric use of AI, independent from legal compliance obligations.

In order to make these principles more concrete, there is a wide range of (non-binding) **recommendations and guidelines on AI**. The most prominent examples as of this moment are the UNESCO 'Recommendation on the ethics of artificial intelligence' and the OECD 'Recommendation of the Council on Artificial Intelligence'. Another recent example of international initiatives is the launch of the AI Governance Alliance by The World Economic Forum, dedicated to responsible generative AI, in November 2023. The recommendations that these organisations put forward provide clear arguments for the notion that AI should be developed and used in an ethical, trustworthy, human-centred and/or responsible manner.

When it comes to the further implementation and integration of the values that these recommendations and guidelines purport, **technical standards** can be seen as a pivotal step. They are particularly important because as soon as they are clearly delineated, they are often more feasible to implement by engineers and other AI practitioners. For instance, issues like explainability, reproducibility and trustworthiness need to be properly defined in order to be implemented. Standardisation bodies with a strong international status like the ISO/IEC and the IEEE are currently developing standards for AI technology. Furthermore, national standardisation bodies like the NIST (U.S.) and CEN-CENELEC (EU) are working to develop standards for their own communities.

Finally, in terms of actual, **hard regulation**, the EU has moved forward by developing the AI Act. Central to the Act is its risk classification system which aims to determine if an AI system poses a risk to a person's livelihood, safety or fundamental rights. The risk levels are currently divided between unacceptable risk (the AI system is banned), high risk (the AI application is subject to strict requirements), limited risk (the AI system is subject to specific obligations, mainly with regards to its transparency), and minimal or no risk (the AI system is freely allowed). As of December 2023, the Act is under review at the European Commission, the EU Council of (national) ministers, and the European Parliament in trilogues. Furthermore, a wide range of legal and ethical experts currently scrutinises the potential impact of the Act.

In all of these efforts, ethical considerations have proven to be a useful framework to help identify and address existing and emerging issues tied to the development[1], deployment and employment of AI, but also as a means to better adjust AI tools themselves to societal needs. Nevertheless, while it is encouraging that ideas on ethics and responsibility are so prominent in this realm, it also often leads to accusations of what is sometimes called "ethics-washing" – of employing the language without following through in actions. A proactive stance on these issues is therefore important, both for regulators in order to prevent problematic impacts, but certainly also for corporations in order to prepare for the necessary compliance. In that context, we at ZSI have developed the following considerations for the use of (generative) AI at our organisation.

Considerations for using generative AI tools at ZSI

In the section above, we have pointed at general issues of AI technologies and promising trends that aim to anticipate their socio-ethical impact. It is important to acknowledge that AI regulation, and a general agreement on relevant ethical rules and guidelines, is still very much in a process of (re)configuration. This is made all the more difficult by the speed and overall volatility with which developments are happening, with AI technologies representing a highly complex field that increasingly permeates all societal processes. At the same time, many of the guidelines and recommendations developed by governments and international organisations are rather abstract in their focus on broad principles for the development and deployment of AI systems in society. However, in order for them to be effective, such requirements must be operationalized in a targeted manner for specific fields. We are currently seeing a rise of guidelines for specific fields of practice which often apply to certain types of AI. It is likely that organisations will need dedicated AI governance programs and AI officers that oversee the application of any such guidelines and manage any issues that may arise in the process.

This is also the case for ZSI as an organisation that is primarily engaged with social science research projects. In that spirit, these considerations are mostly focused on generative AI tools, since these tools are currently impacting our work. The following list of considerations has been inspired by different types of other guidelines, primarily the above-mentioned general guidelines and recommendations from governmental entities and international organisations, as well as more specific guidelines from universities and other social science research organisations[2]. Based on this, we distinguish between different topics. Finally, it is important to note once again that generative AI technologies are rapidly evolving: the recommendations and guidelines are likely to be subject to reconfigurations and additions and should be seen as a first step in a process.

[1] While this is not the focus of the present paper, the conditions under which algorithms are developed and trained are important to consider when exploring the ethical concerns connected to AI systems. For instance, like a lot of work done for giant platform, data labelling is powered by a precarious and underpaid workforce usually recruited from the global south. Similarly, training data is often collected and used without consent, credit or renumeration.

[2] https://huit.harvard.edu/ai/guidelines , https://studieren.univie.ac.at/en/studying-exams/ai-in-studies-and-teaching/ , https://apa.at/wp-content/uploads/2023/07/Leitlinie-zum-Umgang-mit-kuenstlicher-Intelligent-2023-2.pdf

Content quality and scientific literacy

- Be aware that any intentionality must come from you as a researcher and the overall responsibility for the accuracy of the content you process remains with you.
- Content generated by generative AI systems can be inaccurate (including AI hallucinations), misleading and/or miss necessary nuance. Be aware of this and look for telltale signs of mistakes in the outputs.
- Preferably only use AI to assist your work when having enough specialist insights yourself.
- Don't rely solely on generative AI systems to explore questions but use them rather as tools to support your process.
- Be aware of the social biases reproduced by GenAI system as a consequence of their inevitably biassed training data. Take care to actively counteract societal structures of inequality inscribed in training data. This pertains both to your interaction with the AI system (e.g., through the prompts you employ) and to the use of AI outputs for your work.

Confidential data

- Most GenAI tools on the market do not submit to confidentiality and non-disclosure requirements – including transcription tools. Before processing any confidential data (e.g. interview data), check the terms and conditions, location of servers, whether there is a privacy mode available, and whether there is a better alternative. For some tools, including WhisperAI for transcriptions, running local instances of the software is an option that safeguards the privacy of research participants.
- Be transparent with your collaborators (including interview partners etc.) when you intend to use AI for data processing in a way that might negatively impact their privacy rights. Include any relevant use of AI tools in your informed consent sheets and allow participants to opt out.
- Only upload anonymised transcripts for AI based processing (e.g. content analysis) unless you have a local instance set up that does not share the uploaded data with any third party.
- Make sure the data you feed into AI tools is safe and will not be used for anything else but the intended use it was generated for. Don't use any AI tools before you understand what happens to the data you share.
- Of special note are *data pertaining to vulnerable natural persons* and *sensitive personal data*, both of which are protected under GDPR and subject to specific processing conditions. Under no circumstances should such data be shared with a third party, which includes using AI cloud services.
    - Depending on context, vulnerable natural persons include:
        - minors, anyone legally incompetent or who cannot give consent
        - asylum seekers, ethnic minorities
        - the sick and patients, people with disabilities, people with mental disorders, the elderly, pregnant people
        - anyone who may suffer adverse consequences if their personal data were to become publicly available
    - Sensitive personal data entail:
        - personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs
        - trade-union membership
        - genetic data, biometric data processed solely to identify a human being

    - health-related data
    - data concerning a person's sex life or sexual orientation

Digital literacy on (generative) AI

- In the broadest sense, it is important to understand how AI systems are developed and trained to be able to assess the quality and potential weaknesses of created outputs.
- Consider engaging in trainings on how to use AI both responsibly and productively.
- Many existing methodologies deploy (predecessors of) GenAI. Think for instance of systematic reviews and their use of (often open source) Natural Language Processing (NLP) tools.

Type of tool used

- Find the right tool for what you want to achieve. Consider potentially better alternatives to well-known tools (such as ChatGPT).
- Open Source vs. proprietary tools: Currently the developments in *open* Large Language Models are evolving quickly. It is very well possible that they become better than proprietary AI tools like ChatGPT.
- Specialised vs. all-purpose tools: Some tools are developed for specific activities and use cases and might perform these tasks better than more generalised tools.
- Be aware of AI tools within other tools and that they might share data in undisclosed ways and with third parties outside the EU. AI tools often pop up in unexpected places and can be quite persistent. This includes plugins based on ChatGPT as well as "AI assistants" such as Microsoft Copilot, the AI Companion in Zoom, Adobe Sensei, etc.

Transparency regarding use of Generative AI

- An increasingly popular solution to deal with the implementation of generative AI is to be transparent about the way it has been used. Depending also on institutional guidelines and the demands of funding agencies, publishers, and other entities, transparency might need to be considered throughout a process:
    - before using GenAI tools by sharing this intention e.g. with project partners and participants
    - when using GenAI tools e.g. in terms of ensuring quality control of outputs
    - after using GenAI e.g. when submitting a paper or project proposal, when publishers and funding agencies ask for this to be disclosed

Decision-making when using AI

- Be aware which decisions you give away – if any – when employing Generative AI. This might be about setting priorities, extracting and categorising information, or choosing certain styles of writing. Make sure that important decisions remain with you. As researchers, you are responsible for any decisions, not the system, thus always make sure that there is a human in the loop.

External factors

- Pay close attention to existing policies of relevant funding agencies, clients, publishers, etc.
    - For instance, it is not at all uncommon that the use of generative AI is prohibited for review processes. One of the main reasons for this is that current AI tools cannot guarantee compliance with confidentiality and non-disclosure policies, as required by many of these review processes (see also the point about confidentiality above).
    - Work that is developed entirely by Generative AI – but submitted as one's own creation – is in many contexts considered as fraud. For most scientific publishers, AI programmes cannot be an author of an academic paper, and humans must accept full responsibility for a text's accuracy.
    - Make sure to be up to date on policies that might be changing due to a (re)assessment of developing technologies.
- Existing regulations
    - This concerns AI guidelines, but also very much privacy guidelines.
    - At this moment, both in an EU context as well as in an Austrian context, there is not yet a 'hard' regulation on AI. This is likely to change as soon as the AI Act comes into force. This is expected to be happening at the end of 2025.

Direct impact of AI on professional & private lives

- Phishing & Deepfakes: Be aware that phishing attempts will become more sophisticated with the use of generative AI, e.g. by making emails seem more personal and real. Similarly, Deepfakes might become a problem demanding a more rigorous interrogation of the materials we are confronted with – including videos and voice recordings.

Work plan for implementing an AI Strategy at ZSI

1. The present paper serves as a foundation for a shared **Strategy for an effective and responsible use of AI tools for work purposes at ZSI**. Building on this, a more detailed implementation plan for active engagement with the topic for the next 2-4 years would be helpful. This may include the following dimensions.
2. **Inputs by external experts** to cover different aspects of GenAI where ZSI is missing internal expertise. These include:
   a. **Technical background information on AI**. E.g. *how does GenAI work in essence*?
   b. **Hands-on training on specific tools** for specific purposes / use cases and to leverage generative AI as a supportive tool for individual/ZSI skill development, in order to recognize potential benefits while being mindful of the limitations of AI.
   c. Training on **how to safeguard personal and sensitive data** when using AI tools. Education on this is especially important for the DPO, anyone with IT responsibility, and the Ethics Committee.
3. **Internal hands-on workshops** organized by ZSI colleagues to try out AI tools, both to instil confidence and develop the necessary understanding and skills.
4. **Whitelist of tools** that satisfy basic requirements of data protection and privacy, or standard tools that are set up, managed and made available for all ZSI personnel in a centralised manner. Since all tools are subject to change (in data practices, output quality, payment model, etc.), open communication between ZSI employees and the ZSI board is important to allow for adjustments as needed.
   a. Standard tools could entail local instances that do not share data with any third parties, or online tools that satisfy privacy requirements.
   b. Any list of tools needs to include timestamps for when a tool was last assessed and would benefit from including notes on how to ensure data is safe (both in general and concrete terms, e.g., explaining how to activate “private mode”, etc.).
5. **Continued meetings** 2-4 times a year for interested ZSI employees to discuss new developments in the field of AI and potentially adapt the *Considerations* where necessary.
6. **Clarify responsibilities for AI governance at ZSI**. Relevant entities here include the ZSI board, the IT department, the DPO, the ethics committee, and the employees as a whole. Clear instructions and transparent processes are important to effectively and responsibly integrate AI for work purposes at ZSI.

First steps to implement these have been taken at the point of the publication of this document.

Resources & further reading

AI and science: what 1,600 researchers think (Nature news feature): https://www.nature.com/articles/d41586-023-02980-0

APA Leitlinien zum Umgang mit künstlicher Intelligenz: https://apa.at/wp-content/uploads/2023/07/Leitlinie-zum-Umgang-mit-kuenstlicher-Intelligent-2023-2.pdf

ChatGPT use shows that the grant-application system is broken (Nature career column) https://www.nature.com/articles/d41586-023-03238-5

Commission gears up to confront the risks generative artificial intelligence poses to science: https://sciencebusiness.net/news/ai/commission-gears-confront-risks-generative-artificial-intelligence-poses-science

GitHub and Copilot Intellectual Property Litigation: https://www.saverilawfirm.com/our-cases/github-copilot-intellectual-property-litigation

Initial guidelines for the use of Generative AI tools at Harvard: https://huit.harvard.edu/ai/guidelines

KI, ChatGPT und die Wissenschaften – DFG formuliert Leitlinien für Umgang mit generativen Modellen zur Text- und Bilderstellung (DFG Pressemitteilung): https://www.dfg.de/service/presse/pressemitteilungen/2023/pressemitteilung_nr_39/index.html?wt_zmc=nl.int.zonaudev.112331552451_434391519312.nl_ref

KU Leuven Responsible use of Generative Artificial Intelligence: https://www.kuleuven.be/english/education/student/educational-tools/generative-artificial-intelligence

Policy on Use of Generative Artificial Intelligence in the ARC’s grants programs: https://www.arc.gov.au/sites/default/files/2023-07/Policy%20on%20Use%20of%20Generative%20Artificial%20Intelligence%20in%20the%20ARCs%20grants%20programs%202023.pdf

Russell Group principles on the use of generative AI tools in education: https://russellgroup.ac.uk/media/6137/rg_ai_principles-final.pdf

UNESCO Recommendation on the Ethics of Artificial Intelligence: https://unesdoc.unesco.org/ark:/48223/pf0000380455/PDF/380455eng.pdf.multi

University of Ljubljana's recommendations on the use of artificial intelligence: https://www.uni-lj.si/news/news/2023092014431970/

Universities ready to take up generative artificial intelligence, but say guidelines are needed: https://sciencebusiness.net/news/universities/universities-ready-take-generative-artificial-intelligence-say-guidelines-are

UNIVIE AI in studies and teaching: https://studieren.univie.ac.at/en/studying-exams/ai-in-studies-and-teaching/

Annex

Glossary

Algorithm (+ algorithmic decision making): *Formula or set of rules [...] for solving a problem or for performing a task. In [AI], the algorithm tells the machine how to find answers to a question or solutions to a problem. In [ML], systems use many different types of algorithms.* (Source: AI: A Glossary of Terms)

Artificial Intelligence / AI: *An AI system is a machine-based system that is capable of influencing the environment by producing an output (predictions, recommendations or decisions) for a given set of objectives. It uses machine and/or human-based data and inputs to*

- *perceive real and/or virtual environments;*
- *abstract these perceptions into models through analysis in an automated manner (e.g., with machine learning), or manually; and*
- *use model inference to formulate options for outcomes. AI systems are designed to operate with varying levels of autonomy.* (Source: OECD AI Principles)

Big Data: As the name suggests, the volume of the data is one of the features of the definition of *big data* - the emphasis is on volumes which are too big to process with conventional means and require the use of advanced analytics and technologies. Further features comprise a variety of different data types (both structured and unstructured), the increasing speed of data generation, and the value and (yet) untapped potential in terms of deriving new knowledge and insights. (cf. EC 2017), JRC 2015, OECD 2016[3])

Explainability, Explainable AI (XAI): The EC has been driving Trustworthy AI and Explainable AI for several years now. Explainability in AI refers to the capability of an AI system to provide clear and understandable explanations for its decisions and actions, especially when those impact individuals or society as a whole. It's an important factor in ensuring AI systems' transparency, accountability, and trustworthiness. A related concept is reproducibility of results generated by AI. (cf. AI HLEG, Ethics Guidelines for Trustworthy AI)

Foundation Model: Any model that is trained on broad data (generally using self-supervision at scale) that can be adapted (e.g., fine-tuned) to a wide range of downstream tasks. (Source: Stanford Institute for Human-Centered AI)

Generative AI (GenAI): *Generative AI refers to a set of [AI] techniques and models designed to learn the underlying patterns and structure of a dataset and generate new data points that plausibly could be part of the original dataset.* (Source: Pinaya et al. 2023)

Large Language Models (LLMs): A class of language models that use deep-learning algorithms and are trained on extremely large textual datasets - two types:

- generative LLMs: models that output text, such as the answer to a question or even writing an essay on a specific topic (typically unsupervised or semi-supervised, predict what the response is for a given task)

[3] OECE (2016): Big data: bringing competition policy to the digital era, **glossary**.

- discriminatory LLMs: supervised learning models that usually focus on classifying text, such as determining whether a text was made by a human or AI.

(Source: EU-U.S. Terminology and Taxonomy for AI)

Machine Learning (ML): *A branch of AI that focuses on [...] systems capable of learning from data to solve an application problem without being explicitly programmed.* (Source: JRC Glossary of human-centric AI)

Natural Language Processing: *The ability of a machine to process, analyse, and mimic human language, either spoken or written.* (Source: EU-U.S. Terminology and Taxonomy for AI)

Recommender Systems: A fully or partially automated system used by an online platform to suggest in its online interface specific information to recipients of the service, including as a result of a search initiated by the recipient or otherwise determining the relative order or prominence of information displayed. (Source: COM/2020/825 final)

Reproducibility: *Reproducibility refers to the closeness between the results of two actions, such as two scientific experiments, that are given the same input and use the methodology, as described in a corresponding scientific evidence (such as a scientific publication). A related concept is replication, which is the ability to independently achieve non-identical conclusions that are at least similar, when differences in sampling, research procedures and data analysis methods may exist. Reproducibility and replicability together are among the main tools of the scientific method.* (Source: JRC Glossary of human-centric AI)

Trustworthiness: *Trustworthy AI has three components: (1) it should be lawful, ensuring compliance with all applicable laws and regulations (2) it should be ethical, demonstrating respect for, and ensure adherence to, ethical principles and values and (3) it should be robust, both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm. Characteristics of Trustworthy AI systems include: valid and reliable, safe, secure and resilient, accountable and transparent, explainable and interpretable, privacy-enhanced, and fair with harmful bias managed. Trustworthy AI concerns not only the trustworthiness of the AI system itself but also comprises the trustworthiness of all processes and actors that are part of the AI system's life cycle.* (Source: EU-U.S. Terminology and Taxonomy for AI)

Digitaler Humanismus: Die Prinzipien des digitalen Humanismus besagen, dass digitale Technologien solchermaßen entwickelt werden sollen, dass sie mit menschlichen Bedürfnissen und Werten im Einklang stehen, anstatt menschliche Werte neuen technologischen Systemen anzupassen. Um dies zu erreichen bedarf es einer Reihe an Maßnahmen, inklusive demokratischer Kontrolle, Gesetzesbildung, sowie eines offenen Dialogs zwischen Forscher:innen und der Gesellschaft.

Potential areas of use for AI tools

AI tools can support a variety of daily tasks at ZSI. The non-exhaustive list below shows some tasks for which AI tools might be helpful. Please bear in mind that some of these processes currently work more smoothly than others.

- **Text editing:** Translating texts. Checking and correcting spelling, grammar and style.
- **Text processing:** Audio-transcription. Summarising texts and extracting key information. Categorising, classifying and structuring texts. Changing stylistic elements of a text (e.g. translating a scientific text into a blogpost or a ppt presentation). Explaining complex texts and translating them into more accessible language.
- **Research assistance:** Exploring project ideas and research questions. Supporting text research and identifying relevant literature. Synthesising large bodies of documents. Identifying gaps (e.g. in an argument).

Examplary use cases for AI tools

Elicit.org: Desk research, extracting information, discovering concepts

Workflow

Elicit basically has three different functions that can build on each other:

- *Find scientific research papers:* I entered the question "How does philosophy of science influence scientific outcome?" in this function. The tool then suggested 8 papers and summarised the first 4. Here you can also enter various lines that specify the papers, e.g. summary, counter-arguments, etc.
- *Extract information from PDFs:* There is a function that analyses papers and "extracts information". With this function, the papers are uploaded as PDFs and the lines to be analysed can be specified, e.g. summary, outcome measured, duration, participant age, etc.
- *Discover concepts across papers:* With this function of elicit, a concept can be entered, e.g. "feminism in philosophy of science" and then the tool searches for papers. For my search term, the tool listed 37 concepts in 50 papers and provided sources with "relevant citations from relevant papers".

Advantages

The use of AI in this context seems to be a great time saver, as the papers do not even have to be read, but the tool summarises and outputs everything - including the "essence" of the papers. The big advantage I see is that the AI does everything from the research question to the papers to the summary of the papers and the researchers only have to come up with one question. The main saving is therefore in researching the literature and scanning it in detail.

Disadvantages

AI has the disadvantage that it cannot read "between the lines". So if you enter something like "paradigm" in the line that is not explicitly presented in the text (as elicit relies on language models and does not produce its own findings), elicit cannot answer this. Another disadvantage is that elicit does not make it clear how the texts are selected and according to which criteria. If only 8 papers are suggested on a topic, then I ask myself who wrote them and why exactly they are suggested. Since the scientific community itself also has certain biases, I suspect that it is similar at elicit. So the big question is: Who is being suggested and who is not? There is a high likelihood that the global North is overrepresented in the outputs provided.

The disadvantages could be compensated for by carrying out additional research and using elicit as an introduction to a topic or as an extension of your own research. However, I would not use elicit as the only tool.

Concerns

Your own email address is shared with elicit, but it is still unclear what other information the AI collects (such as IP address, Google, etc.). It is also possible to log in with Google or GitHub, but it is unclear to me what information is shared with the platforms and how. I handed over the decisions to summarise and research the papers to the tool. I tried to check the results and came to the conclusion that the tool can only extract what is clearly stated in the paper. I have not taken any measures to protect the information that has been entered, as the papers are open access and freely available on the internet. When extracting information with PDFs, I would not upload confidential PDFs or unpublished documents/articles.

ChatGPT: Essay writing and analysis

Workflow

Registration: As ChatGPT is no longer easily accessible, I have decided against verifying myself with my telephone number. However, the AI can be accessed via chatgptx.de without registration.

Entering a question: I entered the question "Write me an essay on philosophy of science from a feminist perspective with a focus on subjectivity and objectivity" and then without a feminist perspective.

Analysis of the essays: The one text from a feminist perspective writes about marginalised groups and perspectives of women (intersectionality is not sufficiently considered here). Without the feminist perspective, ChatGPT writes about academics as people who cannot guarantee objectivity. From an analytical point of view, these are fundamental differences - from a feminist perspective, it is mainly women who drive and promote subjectivity; if I leave out the perspective, ChatGPT writes about all scientists. According to ChatGPT, the feminist perspective is a critical one, while subjectivity is not questioned in the theory of science. It seems that the approaches are similar and from the same origin, but it needs the addition of a feminist perspective so that this is also acknowledged and not formulated as "universally valid".

Advantages

One advantage is that AI is very versatile if you know how to use the tool. You can write essays, journalistic articles or even poems, proposals, etc. On the one hand, this saves time, but on the other, it also facilitates your own creativity.

Disadvantages

The tool cannot specify sources sufficiently (or sometimes correctly/at all). Since ChatGPT is fed with "social" data, biases are also reproduced and sometimes taken to extremes – especially when no further specifications are entered (e.g. feminist perspective/anti-racist perspective). The prejudices from society are reflected in AI. Even if very specific perspectives are asked for, it is possible that underlying concepts are not sufficiently illuminated. "AI literacy" must therefore be learned and AI cannot simply be fed with casual questions.

Concerns

I have not entered my telephone number for reasons of data protection, but I have entered my email address. It is unclear which data is recorded (e.g. IP address, history of entries to ChatGPT). Furthermore, it is unclear to whom the data is sold on – if, for example, "Meta" sells data to cambridge analytica, I wonder whether the AI also sells data to companies that can use user profiles to place advertising (not only products, but also fake news/election advertising) or play a role in the output of ChatGPT's texts. So the question for me is: what happens to my data, who uses it, what answers does ChatGPT give based on my user profile?

Another concern I would like to express is that the supposed "objectivity" of AIs might not only reproduce and confirm existing prejudice in people (including racism, sexism, ableism, classism, etc.), but also give rise to new ones. So if the tool is believed and literacy does not exist, I fear that concepts could be adopted without reflection – including those of scientists. Especially due to the large number of scientific concepts, it can happen that the person using the tool is not familiar with everything and problematic concepts are adopted (such as Kulturkreis Theory).

With the takeover of ChatGPT by Microsoft, I also wonder how tools such as text suggestions in Word through ChatGPT and its embedding in social biases can/will find their way into one's own normality through repeated suggestions (also at ZSI) and how Microsoft users pass on or feed data to AI systems.

ChatGPT: Qualitative data analysis – coding and comparing

Workflow

This is a method to supplement regular qualitative data analyses. As such, you first need to identify and gather relevant data material such as expert interviews, workshop results, and relevant text passages from literature. These must be hand-coded for a qualitative content analysis. From there, coded text passages can be input to ChatGPT and re-coded to critically compare with the original analysis and e.g. identify gaps. Coded passages can also be summarised, paraphrased, and rephrased, while the tool can also quickly extract key statements, recommendations, challenges, and so on. All outputs must be thoroughly checked and adapted.

Advantages

Using ChatGPT may save a lot of time. It can also be used to stimulate critical examination of one's own analysis, support the analysis, and develop nice, crisp formulation (e.g. of recommendations) one might not be able to produce oneself (as easily). In the same vein, the tool can also provide linguistic support when writing texts (of various kinds) and “proofread” materials.

Disadvantages

One of the biggest weaknesses of ChatGPT related to data protection, especially for sensitive data. Thus, strict and deliberate anonymization is essential. Furthermore, the outputs provided by ChatGPT are not necessarily correct (hallucinations!), so they must be closely checked and one should not rely too much on these outputs. As such, the time for checking any outputs must be included in any calculation of overall efforts. You also need basic AI literacy to be aware of the limits of the technology and how to deal with them.

Concerns

Garbage in, garbage out: The tool accesses data from the internet without being able to discern “high quality” data from”low quality” ones. As such, no outputs generated by ChatGPT can be trusted blindly. Also, ChatGPT is no search engine, but rather provides "creative" or "made up" suggestions that can serve as ideas to build on. If you are very familiar with a topic and know what you want to say, check the output, make improvements, etc., you don’t actually leave any decisions up to the AI. But you need to do this very intentionally, which also needs to be reflected in the choice of prompts that frame the final output.