Responsible use of GenAI: A Step-by-Step Guide

The 14 guiding questions below will help you decide whether using GenAI is appropriate in a given instance.
To help you reflect deeply about each guiding question, the guide provides some general risk assessment considerations as a starting point.
You may add criteria or alter them based on your use case.
Our checklist and declaration templates in the annex will help you for quality control while actively working with GenAI tools and for transparent documenting after completing a task.

QUESTIONS TO ASK BEFORE USING GenAI	RISK ASSESSMENT CONSIDERATIONS, ANSWERED WITH "YES"	FURTHER CONSIDERATIONS AND SUPPORT
	If you answered "No", the use of GenAI may pose serious risks. For some criteria, it implies that you need not, cannot, or must not use GenAI.	
Planning & Input		
1 **For what task(s) will you use GenAI?**	• Do you possess the foundational skills needed to accomplish the task, e.g., coding, critical reading, literature searching, scientific writing? • Do you have the subject-matter expertise and task-specific skills to write effective prompts for GenAI tools and to vet the outputs produced by the tools? • Are you allowed to use GenAI for the task? For instance, some publishers ban AI-generated images and most forbid use of AI to generate peer reviews.	• Every semester, the tutors from Medical and Natural Sciences Library offer courses in “How to use AI tools for searching, reading, and writing: Critical thinking and sustainable best practices” within the Transferable Skills Program. • You can also contact support_med.ub@unibe.ch if you want courses tailored to your research groups.
2 **Do you need GenAI to accomplish the task?**	• Will using GenAI save you time? • Will it help you produce higher quality output? • Can you afford the tool?	
3 **Are you working with confidential, non-sensitive, or open data?**	• If data are confidential, will the GenAI tool keep it protected and secure? • If data are non-sensitive or open, can you foresee any harms that might result from sharing them with the GenAI tool? If yes, do you have a plan for mitigating the harm (e.g., running the tool on a local server or local device)?	• For further support contact the data stewards at researchdata@unibe.ch • For more information, see Research Data Management or the swissethics guideline on Artificial Intelligence (AI) and Research Involving Human Beings.
4 **Is it okay if your prompt (including all uploaded material) is used to retrain the GenAI tool?**	• If you are uploading external information (e.g., published articles, slides) to the GenAI tool, do you have the copyright holders' permission (1, 2) to do so? • Have you made sure your prompts will not inadvertently reveal proprietary or confidential information (e.g., breach the confidentiality agreement you accepted as a peer reviewer)?	• For answers to questions about copyrights and licenses, contact: openscience@unibe.ch • Read the terms and conditions to determine whether there is an opt-out that disallows use of your inputs for training the LLM.
5 **What harm might be caused by uncertain, unreliable, or incorrect output?**	• Have you considered the harms caused to research participants, patients, the public, and your career if your use of GenAI produced false or poor-quality information? • Will you have the time and resources to check all the information produced by GenAI to make sure it is not false or of poor quality?	
6 **Do you have a systematic method for documenting GenAI use?**	• Are you documenting your GenAI use in sufficient enough detail so you can write honest and transparent disclosure statements?	• For further guidance on properly documenting use of AI tools, refer to COPE’s statement on authorship and AI tools (3) or the GAMER statements (4). • If you have chosen your journal(s), read their latest ethical AI guidelines. • Refer to the guidance from publishers on how to disclose AI use in research (e.g., Springer Nature’s AI Policy, Wiley AI guidelines).
Tool & Availability		
7 **Which GenAI tool will you use and why?**	• Have you tested comparable tools and ensured your choice best meets your purpose? • What capabilities (e.g., RAG) does your model have? • Are the models suited to the task? • Have you considered other task-specific, less resource-intensive tools you could use instead of general-purpose GenAI tools? • Was the research workflow conducted without the use of any AI agent?	• Will you use the free or paid version of the tool? What factor influenced your decision? • Can you adopt and implement tools with data servers in Europe or locally made GenAI or open source models? • Refer to our specific guidelines in the annex that provide guidance on using LLMs for summarizing reference material and transcribing audio recordings. • Have you considered the effect of AI use on the environment (5)?
8 **Is this your first use of the GenAI tool?**	• Have you tested the tool for internal consistency by rephrasing your prompt or challenging its answer through Socratic questioning, e.g., “Upon what assumptions are your conclusions based?", "Are you aware of gaps or bias in your knowledge base?" • Have you tested GenAI's answers for accuracy under a sufficient range of conditions?	• First, test the model/tool under low-risk conditions with at least 2 independent auditors. (Use publicly available work + outputs for your test.) • Allocate enough time and effort to vet the tools properly. • Consider creating a Standard Operating Procedure (SOP) for working with AI tools that includes task-specific standardised prompts and quality control checklists.
9 **Is the tool available to UniBE affiliates?**	• Have you read and understood the latest UniBE guidelines for procuring and handling AI tools and data protection?	• Further information about available tools for UniBE affiliates: Generative KIs - Universität Bern • Advice on possible AI solutions for your institution: Data Science Lab - University of Bern
Data Source		
10 **What harms might be caused by using data from unknown sources in your research?**	• Are you certain that using data from unknown sources for your task will not create a risk for you, your research participants, or your research output? • Have you thought about how relying on data from unknown sources could compromise the quality of your research output?	
11 **Does the use of a GenAI model trained on a large corpus of unknown, and potentially illegally obtained data, undermine a scholar’s research outcomes or ethical integrity?**	• Have you thought about how choosing an LLM trained on a large corpus of unknown and potentially illegally obtained data could compromise the integrity of your research output? • Have you considered if your reputation is put at risk if you use unethically sourced data?	Stay abreast of the news and regulations regarding AI copyright infringement.
Output		
12 **What biases and limitations are common in the data in your field?**	• Do you know how scholars in your field usually mitigate prevailing biases? Can you implement a similar or equally effective method for de-biasing outputs? • Have you tested the effect on the output by comparing examples from different demographic groups?	• GenAI is not neutral but reflects prevailing biases (2). • List known biases and lacunae in your field (e.g., publication bias, gender bias, demographic bias) and draft a plan to mitigate them for instance, by incorporating data from other sources. • Libraries that assess and mitigate bias include Fairlearn, GitHub - columbia/fairtest • Is the topic sensitive or likely to affect groups of people differently? (Evaluate likely unequal effects, preferably with input from people from the affected groups.)
13 **Were the results you derived from your GenAI tool consistent enough to meet the quality standards in your field?**	• Have you tested GenAI's answers for consistency under a sufficient range of conditions? • Are you following the latest AI guidelines published by your journal or publisher of choice?	Be aware that GenAI will not provide consistent answers to the same prompt. Thoroughly test the range of answers it provides, e.g., by multiple application of the same prompt, until you are sure that any answers in that range are sufficient for your purposes.
14 **Are you aware that you are ultimately responsible for any research output generated with AI?**	• Can you affirm that you will not use GenAI to fabricate or manipulate research data and results (2, 5)? • Can you affirm that you did not use GenAI to generate image-based data, e.g., Western blots? • If the GenAI output contained references, have you verified that these references exist and that the GenAI output correctly represents the original authors' meaning?	GenAI systems should empower human beings, allowing them to make informed decisions and fostering their fundamental rights. Therefore, proper human oversight and verification is needed at every stage (1, 2, 5).

Disclaimer:
We are actively working on expanding this GenAI decision guide. If you would like to participate, share insights, or provide recommendations, please contact us: Medical and Natural Sciences Library: support_med.ub@unibe.ch or frnat.ub@unibe.ch

References:

(1) Maintaining research integrity in the age of GenAI: analysis of ethical challenges and recommendations to researchers. DOI: https://doi.org/10.1007/s40979-025-00191-w
(2) Embracing AI with integrity: A practical guide for researchers. DOI: https://doi.org/10.37672/UKRIO.2025.06.embracingAIwithintegrity
(3) Authorship and AI tools. DOI: https://doi.org/10.24318/cCVRZBms
(4) The GAMER statements. DOI: https://doi.org/10.1136/bmjebm-2025-113825
(5) National Centre for Atmospheric Science. Conscious use of AI. Available from: https://ncas.ac.uk/conscious-use-of-ai-practical-tips-and-thoughts-from-environmental-researchers/
(6) Ethics guidelines for trustworthy AI. Available from: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

GenAI Quality Control/Prompting Checklist

Project Title: _____

Researcher Name: _____

GenAI Tool in Use: _____

Phase 1: Pre-Prompting Security & Setup

Complete these checks before entering any data or queries into the GenAI tool.

[ ] **Tool Selection:** I have tested comparable tools and ensured my choice best meets my purpose.

[ ] **Confidentiality Check:** I have made sure my prompts will not inadvertently reveal proprietary or confidential information (e.g., breach a confidentiality agreement I accepted as a peer reviewer).

[ ] **Copyright Clearance:** If I am uploading external information (e.g., published articles, slides) to the GenAI tool, I have the copyright holders' permission to do so.

[ ] **Auditor Testing:** If this is a new tool or use case, I have first tested the model/tool under low-risk conditions with 2 independent auditors using publicly available work.

Phase 2: Active Prompting & Consistency Testing

GenAI will not provide consistent answers to the same prompt. Use these steps to test the robustness of the output.

[ ] **Range Testing:** I have thoroughly tested the range of answers provided by multiple applications of the same prompt, until I am sure that any answers in that range are sufficient for my purposes.

[ ] **Internal Consistency:** I have tested the tool for internal consistency by trying to rephrase my prompt.

[ ] **Socratic Challenge:** I have challenged its answer through Socratic questioning (e.g., asking "Upon what assumptions are your conclusions based?" or "Are you aware of gaps or bias in your knowledge base?").

[ ] **Condition Testing:** I have tested the AI's answers for accuracy and consistency under a sufficient range of conditions.

Phase 3: Output Evaluation & De-Biasing

GenAI is not neutral but reflects prevailing biases. Evaluate the final output with a critical eye.

[ ] **Demographic Testing:** I have tested the effect on the output by comparing examples from different demographic groups.

[ ] **Sensitivity Check:** If the topic is sensitive and/or likely to affect groups of people differently, I have evaluated likely unequal effects, preferably with input from people from the affected groups.

[ ] **Bias Mitigation:** I have drafted a plan to mitigate known biases and lacunae in my field (e.g., publication bias, demographic bias), for instance, by incorporating data from other sources.

[ ] **Reference Verification:** If the GenAI output contained references, I have verified that these references actually exist.

[ ] **Contextual Accuracy:** I have verified that the GenAI output correctly represents the original authors' meaning in those references.

[ ] **Time Allocation:** I have allocated enough time and effort to vet the used tools and their outputs properly.

Signature (Upon Completion): _____ Date: _____

General GenAI Usage Declaration Template

Project Title / Manuscript Name: _____

Lead Researcher / Author(s): _____

Assigned Data Steward and/or Research Software Engineer: _____

_____

Date: __/__/____.

1. Tool Identification

Please list all Generative AI tools used during this research or drafting process.

- **Name of GenAI Tool(s): _____**
- **Version / Model (e.g., Free, Paid, Local, Open Source): _____**
- **Key Capabilities Utilized (e.g., RAG, text-to-text, audio transcription): _____**
- **Justification (Why was this specific tool chosen?): _____**

2. Scope of Application

Describe the specific task(s) for which the GenAI tool was utilized. (Check all that apply and provide brief details)

[ ] **Planning & Ideation: _____**

[ ] **Data Processing (e.g., coding, formatting): _____**

[ ] **Literature Searching / Summarization: _____**

[ ] **Drafting / Language Polishing: _____**

[ ] **Other: _____**

3. Human Oversight & Verification

GenAI systems should empower human beings, requiring proper human oversight and verification at every stage.

- **Method of Documentation:** Briefly describe your systematic method for documenting your GenAI use (e.g., prompt logs, version control).
  Description: _____
- **Bias Mitigation:** How did you test for and mitigate prevailing biases or lacunae in the AI's output?
  - *Description:* _____
- **Verification of References:** If the GenAI output contained references, describe how you verified that these references exist and correctly represent the original authors' meaning.
  - *Description:* _____

4. Ethical Affirmations

By signing below, the author(s) affirm the following statements regarding the ethical and responsible use of GenAI:

[ ] I/We acknowledge that we are ultimately responsible for any research output generated with AI.

[ ] I/We affirm that GenAI was **not** used to fabricate or manipulate research data and results.

[ ] I/We affirm that GenAI was **not** used to generate image-based data (e.g., Western blots, research figures).

[ ] I/We confirm that no proprietary, confidential, or unethically sourced data was inadvertently revealed or uploaded to the GenAI tool without proper safeguards.

[ ] I/We have reviewed and adhered to the latest AI guidelines published by our target journal/publisher (e.g., Springer Nature, Wiley,…) as well as COPE's or GAMER statements on authorship and AI tools.

[ ] I/We affirm that no output from GenAI is presented as a standalone Research Result; all conclusions are the result of human oversight.

Signature: _____ **Date:** _____

Using LLMs for Summarizing Reference Material

Objective of the use:

- To leverage Large Language Models (LLMs) to efficiently and effectively generate summaries of various reference materials, including academic papers, technical reports, and other lengthy documents.
- The aim is to produce accurate and concise summaries that capture the key points and essential information of the source material.
- This can significantly accelerate research and information gathering processes for drafting a **State-of-the-Art.**

Limitations:

- Current LLMs, based on advanced deep learning architectures, have demonstrated remarkable capabilities in natural language understanding and generation.
- They can process and comprehend vast amounts of text, identify salient information, and produce coherent and fluent summaries.
- Despite their advancements, LLMs can still generate inaccurate or misleading summaries, often referred to as "hallucinations."
- These hallucinations may arise due to limitations in the model's understanding of the context, ambiguities in the source material, or biases in the training data.
- The quality of the summary can also be affected by the complexity and length of the source material, as well as the specific LLM architecture and training methodology used.

Human Oversight:

- To ensure the accuracy and reliability of LLM-generated summaries, human oversight and review are essential.
- A human expert can critically evaluate the summary, identify any errors or inconsistencies, and verify the information against the original source material.
- The human reviewer can also provide feedback to the LLM system, helping to improve its performance and reduce the incidence of hallucinations.

Suggested Best Practices:

- To maximize the effectiveness of LLMs for summarizing reference material:
    - Select an appropriate LLM model that has been trained on a relevant corpus of text, or has presented good outcomes for similar reference material in the past.
    - Clearly define both (1) the desired summary length and (2) level of detail.
    - Preprocess the source material to remove any irrelevant/redundant information.
    - Fine-tune the LLM model on a specific domain or topic, if necessary.
    - Implement a robust human review process to ensure the quality and accuracy.

Future Directions for research:

- Ongoing research in LLM development aims to address the limitations and improve the performance of these models for summarizing tasks. This can include developing more sophisticated architectures, incorporating external knowledge sources, and enhancing the explainability and transparency of LLM-generated outputs.
- The integration of human-in-the-loop approaches and active learning strategies is also being explored to further enhance the collaboration between humans and LLMs in the summarization process, i.e., H+M systems for summarizing bibliographical references.

Interview Audio Transcription

Utilizing AI for interview audio transcription can produce rapid and mostly efficient transcriptions, converting spoken words into written text quickly and mostly accurately.

This allows for easier review, analysis, tagging and sharing of interview content.

Limitations:

- **Language Input:** The accuracy and effectiveness of AI transcription can be influenced by the language spoken in the interview. AI models like Whisper have varying levels of proficiency in different languages, dialects, or accents, which could lead to errors or misunderstandings in the transcription.
- **Unrevised Details:** AI transcriptions may not always capture every nuance and/or detail of the spoken word, potentially leading to minor errors or omissions in the transcribed text. This could include misinterpretations of homophones, misspellings of proper nouns, or inaccuracies in punctuation.

Human Oversight:

Human oversight and intervention tend to be essential components of the interview audio transcription process, even when utilizing AI technology.

- **Transcription Verification:** Human reviewers should compare the AI-generated transcription with the original audio recording to ensure accuracy and completeness. This includes correcting any errors, filling in gaps, and clarifying ambiguities. Also, some AI models hallucinate content that humans can filter out of the transcription.
- **Sensitive Data Handling:** In cases where interviews involve sensitive or confidential information, human reviewers play a crucial role in ensuring proper data handling. This includes pseudonymization or anonymization of personal data as a post-processing step in the transcription to protect privacy and maintain confidentiality.

Additional Considerations:

- **Speaker Identification:** In interviews with multiple speakers, AI models may struggle to accurately identify and differentiate between individual speakers, especially if there are overlaps in speech or variations in speaking styles. Human reviewers can help to correctly attribute dialogue to the appropriate speakers.
- **Background Noise:** Background noise or poor audio quality can negatively impact the accuracy of AI transcriptions. Human reviewers can use their judgment and context to interpret speech that may be unclear or distorted due to noise interference.
- **Technical Expertise:** While AI transcription tools are becoming increasingly user-friendly, some level of technical expertise may still be required to operate them effectively. Human reviewers with relevant skills can help to manage the technical aspects of the transcription process.

By combining the speed and efficiency of AI Models with the accuracy and judgment of human reviewers, interview audio transcription can be conducted effectively and ethically, supporting research, journalism, and other fields that rely on accurate and accessible interview data.