AGIK

Institutional Guardrails for AI:

Balancing Data Protection, Local LLMs, and GenAI Decision Guide at UniBE

Dr.-Ing. Federico Grasso Toro

© 2026 by Dr.-Ing. Federico Grasso Toro is licensed under CC BY-NC 4.0

Content

1. **On Current UniBE Support**
2. **New GenAI Decision Guide**
3. **Case Studies**
4. **Recommendations**
5. **Takeaways and Annexes Discussion**

Added Bad Example: When two years of academic work vanished with a single click

© 2026 by Dr.-Ing. Federico Grasso Toro is licensed under CC BY-NC 4.0

Support for AI-Tools at UniBE

On Data Protection:

- *Shift from commercial cloud risks to secure local solutions (TextLab): TFS Courses and Students/Researchers reception*
- *Upcoming UniBE Cloud Strategy*

On Local LLMs: GPUstack solution

- Current Proof of Concept for limited self-hosted Models
- Already Success stories for Students and Researchers

© 2026 by Dr.-Ing. Federico Grasso Toro is licensed under CC BY-NC 4.0

Support for AI-Tools at UniBE

1. «updated» ***Guidelines for the deployment of AI tools and technologies and trustworthy AI research*** [in revision]
   - ***Old version:*** UniBE - Research on and with AI (2024)

2. Step-by-Step Decision Guide for AI Tools:

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

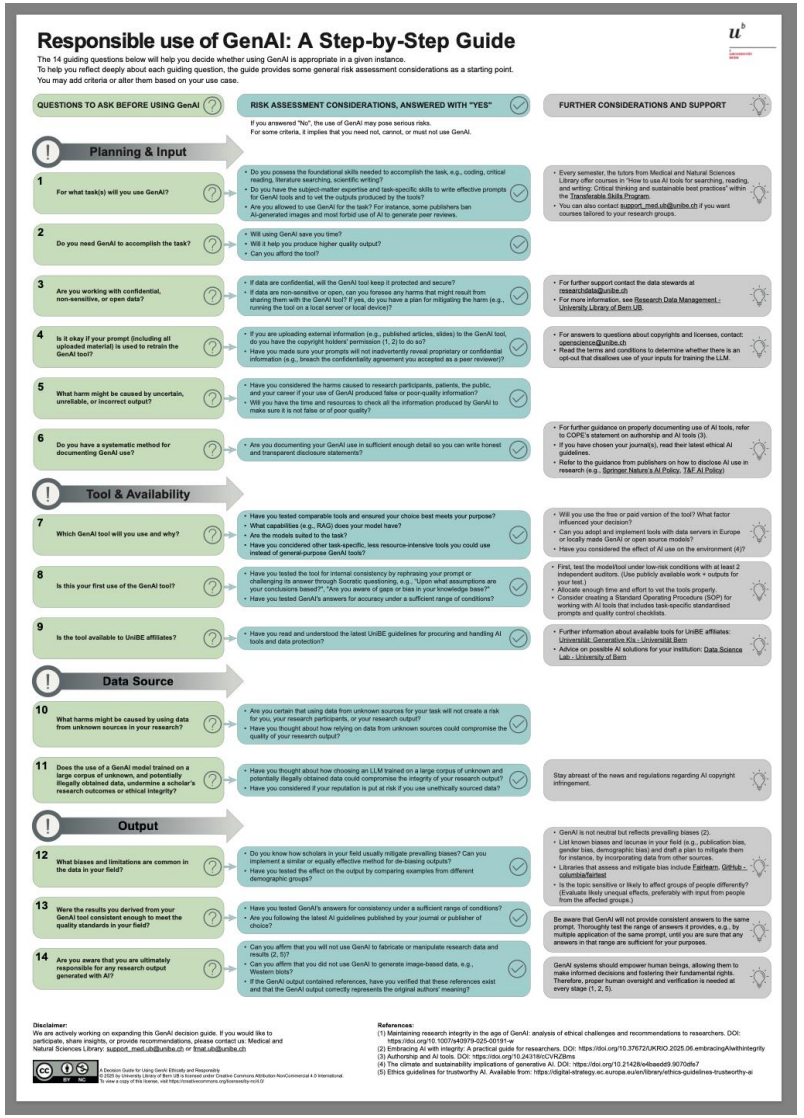
© 2026 by Dr.-Ing. Federico Grasso Toro is licensed under CC BY-NC 4.0

Appendix
New Guide

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

The 14 guiding questions **to help you decide whether using GenAI is appropriate**. (source)

1. Planning & Input -> *Knowledge Security*
2. Tool & Availability -> *Data Sovereignty*
3. Data Source -> *Bias Detection*
4. Output -> *Human Oversight*



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.

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 - University Library of Bern UB.
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). • 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, T&F AI Policy)
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?	• 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? • Have you considered the effect of AI use on the environment (4)?
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: Universität: 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 bias? • 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 fmat.ub@unibe.ch

CC BY NC
Decision Guide for Using GenAI Ethically and Responsibly
© 2025 by University Library of Bern UB is licensed under Creative Commons Attribution-NonCommercial 4.0 International.
To view a copy of this license, visit https://creativecommons.org/licenses/by-nc/4.0/

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/JKRIO.2025.06.embracingAIwithintegrity
(3) Authorship and AI tools. DOI: https://doi.org/10.24318/cCVKZ8ms
(4) The climate and sustainability implications of generative AI. DOI: https://doi.org/10.21428/e4baedd9.9070dfa7
(5) Ethics guidelines for trustworthy AI. Available from: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

To be revised (probably) twice per year.

© 2026 by Dr.-Ing. Federico Grasso Toro is licensed under CC BY-NC 4.0

AI Literacy in Research:
Selecting, Using, and Reporting GenAI

Context & The Golden Rule

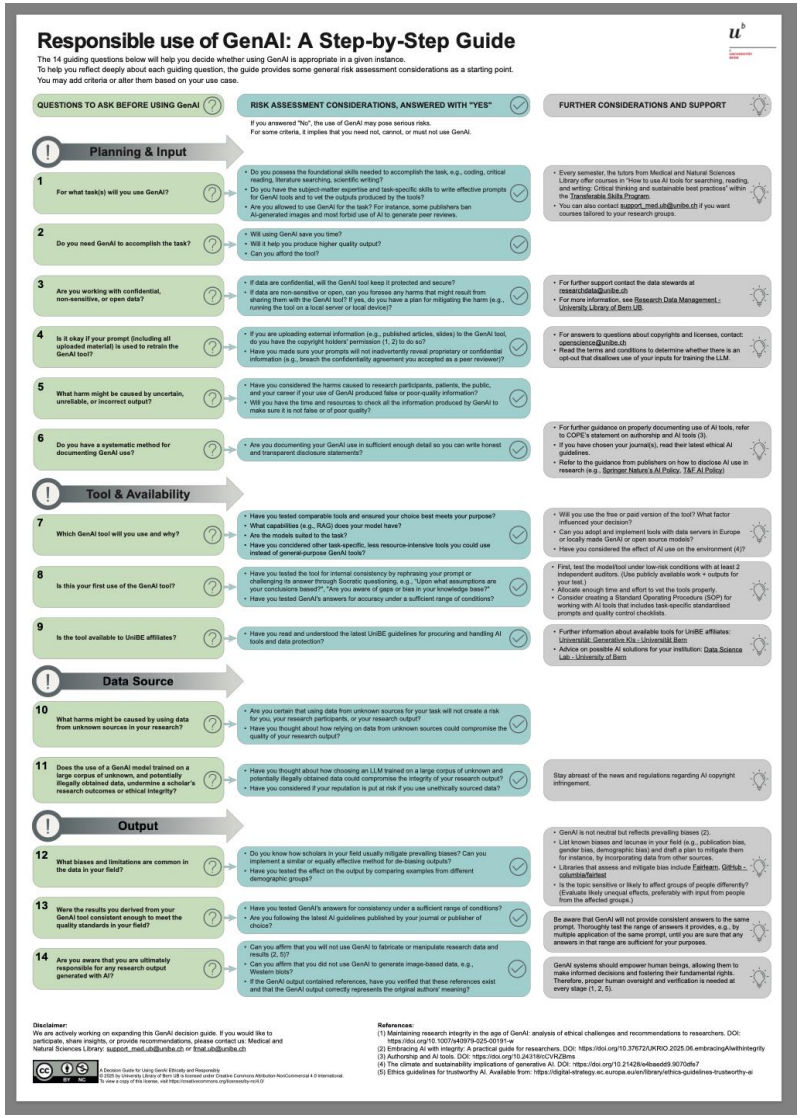
- **Objective:** To introduce the overarching "principles of responsible GenAI" use in praxis by Decision Guide.

The Golden Rule:

No output from GenAI can be considered a Research Result.

The Paradigm Shift:

GenAI systems should "empower" humans, i.e., proper human oversight and verification are always required at every stage.



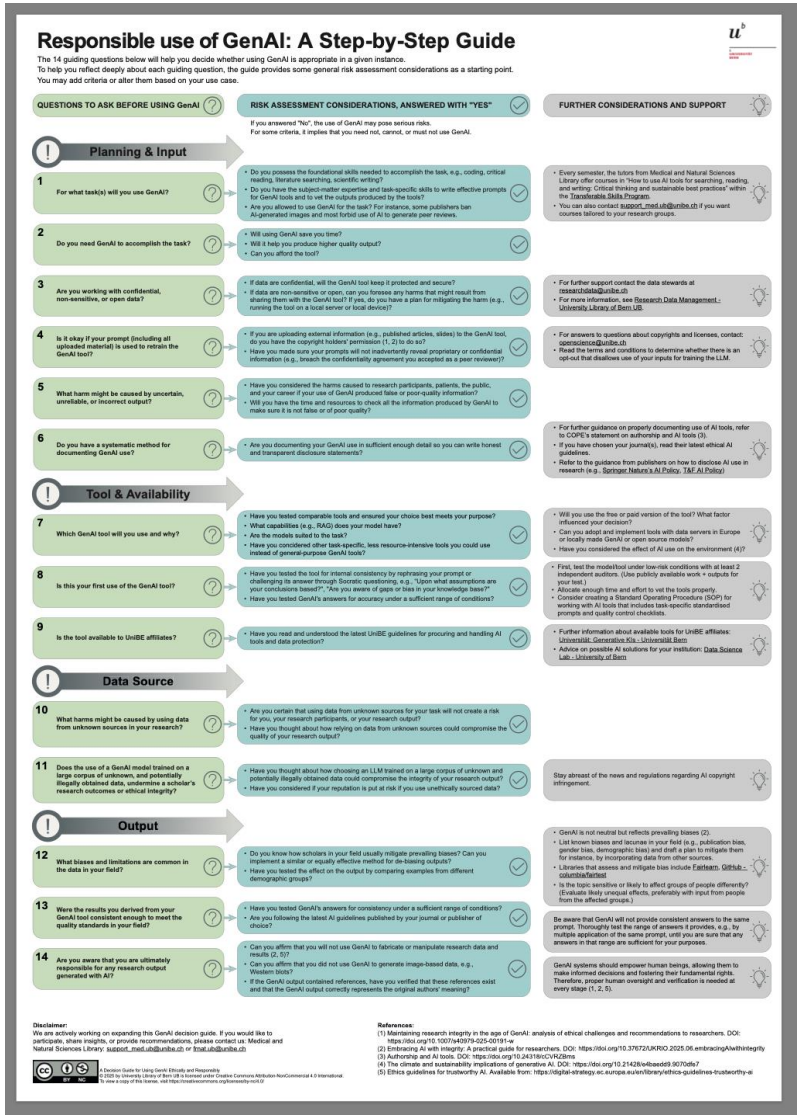
© 2026 by Dr.-Ing. Federico Grasso Toro is licensed under CC BY-NC 4.0

AI Literacy in Research:
Selecting, Using, and Reporting GenAI

Why a Step-By-Step Decision Guide?

- **Risk Assessment:** Answering "No" to specific risk assessment questions indicates that using GenAI may pose serious risks.
- **Ultimate Responsibility:** Researchers are ultimately responsible for any research output generated with AI.

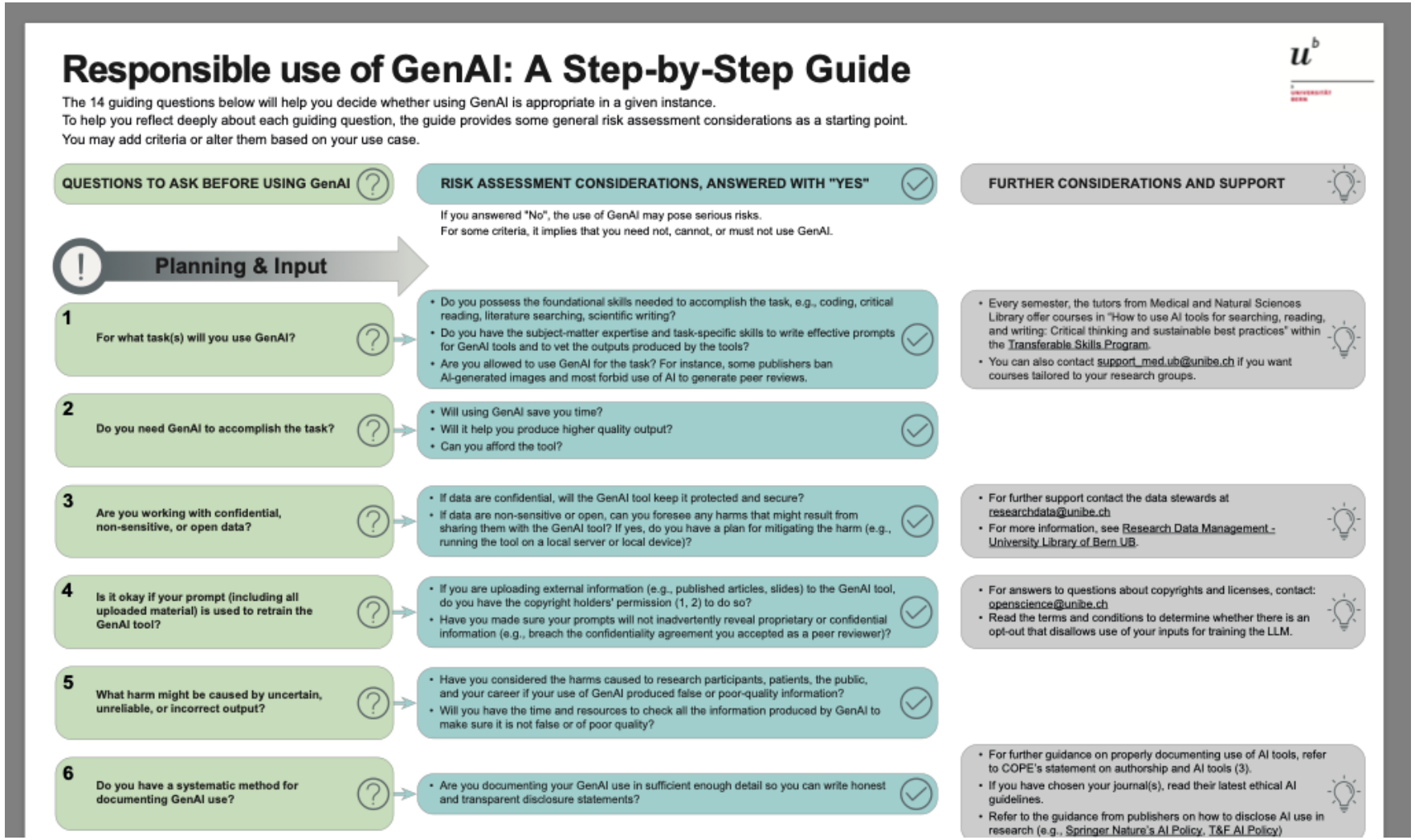
The Bottom Line: In some cases, researchers need not, cannot, or must not use GenAI tools.



© 2026 by Dr.-Ing. Federico Grasso Toro is licensed under CC BY-NC 4.0

AI Literacy in Research: Responsible use of GenAI

Selecting, Using, and Reporting GenAI



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

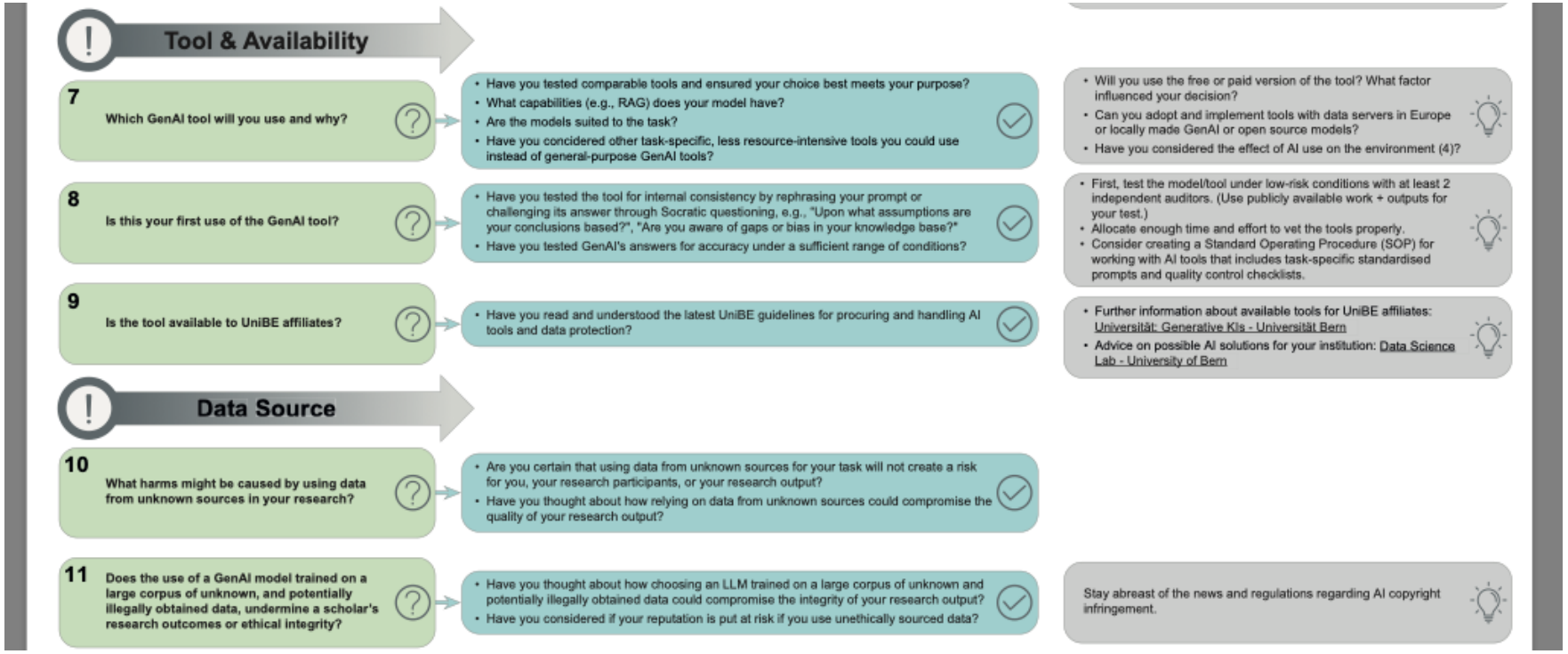
The 14 guiding questions below will help you decide whether using GenAI is appropriate in a given instance.
To help you reflect deeply about each guiding question, the guide provides some general risk assessment considerations as a starting point.
You may add criteria or alter them based on your use case.

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 - University Library of Bern UB.
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). • 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, T&F AI Policy)

© 2026 by Dr.-Ing. Federico Grasso Toro is licensed under CC BY-NC 4.0

AI Literacy in Research: Responsible use of GenAI

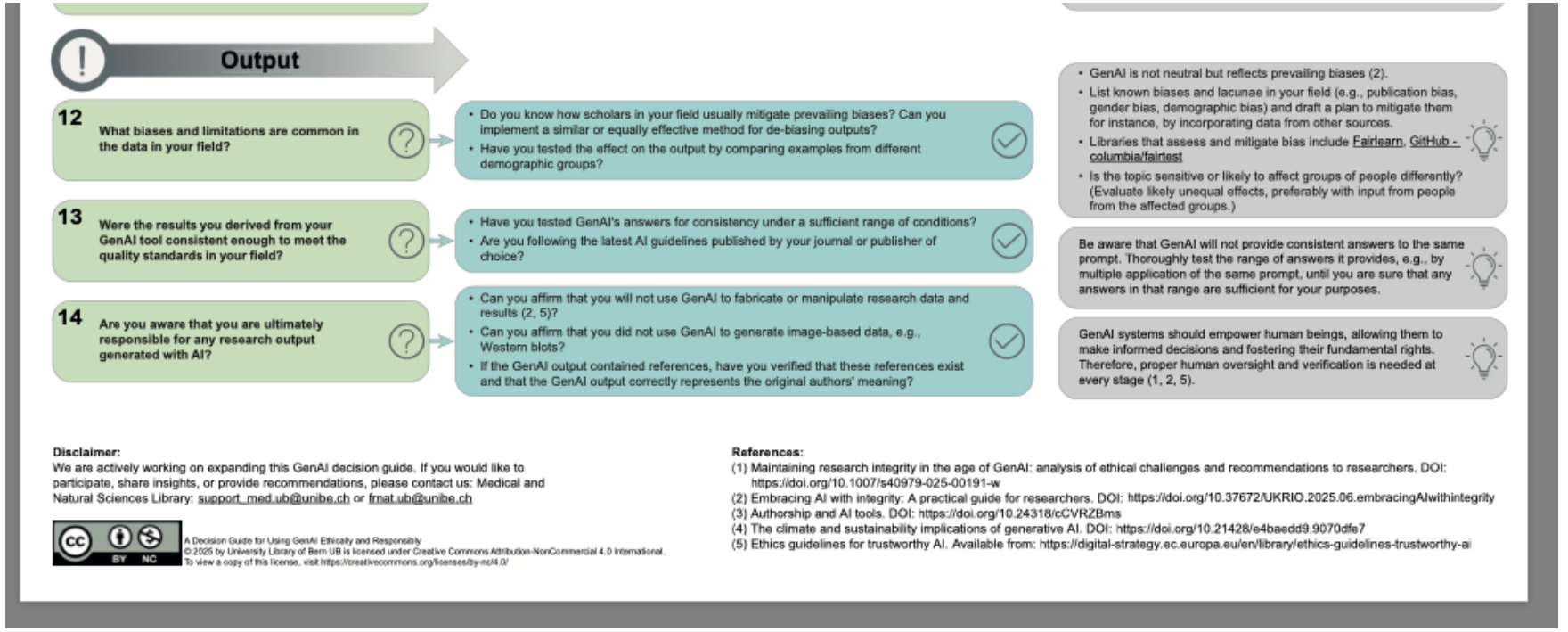
Selecting, Using, and Reporting GenAI



© 2026 by Dr.-Ing. Federico Grasso Toro is licensed under CC BY-NC 4.0

AI Literacy in Research: Responsible use of GenAI

Selecting, Using, and Reporting GenAI



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

A Decision Guide for Using GenAI Ethically and Responsibly
© 2025 by University Library of Bern UB is licensed under Creative Commons Attribution-NonCommercial 4.0 International.
To view a copy of this license, visit https://creativecommons.org/licenses/by-nc/4.0/

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 climate and sustainability implications of generative AI. DOI: https://doi.org/10.21428/e4baedd9.9070dfe7
(5) Ethics guidelines for trustworthy AI. Available from: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

© 2026 by Dr.-Ing. Federico Grasso Toro is licensed under CC BY-NC 4.0

Case Study A – AI for Audio Transcription

- **Scenario:** Using GenAI tools for qualitative interview transcription.
- **Data Security:** If interview data is confidential, researcher must ensure the GenAI tool keeps the raw data protected and secure.
- **Mitigating Risk:** Consider using AI tools with data servers in Europe, as well as locally made/hosted GenAI tools, and/or open-source models.
- **Terms and Conditions:** Always read the terms of external cloud-based AI tools to find if opt-outs that are relevant can be defined, i.e., prevent inputs from training private LLM.

© 2026 by Dr.-Ing. Federico Grasso Toro is licensed under CC BY-NC 4.0

Case Study B – AI for Literature Summaries

- **Scenario:** Using GenAI tools to summarize published articles.
- **Foundational Skills:** Researcher must possess the skills (i.e., critical reading), to vet the quality of any AI output.
- **Copyright Compliance:** Ensure copyright holders' permission before uploading/using published articles to an AI tool.
- **Evaluating Output:** GenAI is not neutral and reflects prevailing biases.
- **Reference Verification:** Researchers must verify that any generated references exist and correctly represent the original authors' meaning.

© 2026 by Dr.-Ing. Federico Grasso Toro is licensed under CC BY-NC 4.0

Case Study C – AI for Data Analysis & Coding

- **Scenario:** Using GenAI tools for data processing tasks, such as coding or formatting.
- **Data Integrity:** Researchers must explicitly affirm that GenAI will not be used to fabricate and/or manipulate research data and results.
- **Image Generation Restrictions:** You must not use GenAI to generate image-based research data blindly, such as research figures.
- **The Standalone Rule:** No output from GenAI can be presented as a standalone Research Result; all conclusions must be the direct result of human oversight.

© 2026 by Dr.-Ing. Federico Grasso Toro is licensed under CC BY-NC 4.0

Case Study D – AI for Vibe-Coding (for RSEs)

- **Scenario:** Using GenAI coding assistants (e.g., GitHub Copilot, Cursor, ChatGPT) for "vibe-coding"—the practice of generating, refactoring, or documenting research software architectures and scripts via high-level natural language prompts.
- **Foundational Skills:** RSEs must have the foundational coding skills necessary to deeply understand and vet the syntax, logic, and security of the outputs produced by GenAI tools.
- **Confidentiality & IP:** Researchers must ensure their prompts do not inadvertently reveal proprietary algorithms, unreleased data, or confidential information (e.g., breaching a confidentiality agreement). Always read the terms and conditions to determine if an opt-out is necessary to prevent your code from training public LLMs.

© 2026 by Dr.-Ing. Federico Grasso Toro is licensed under CC BY-NC 4.0

Case Study D – AI for Vibe-Coding (for RSEs)

- **Citations Verification:** GenAI is prone to hallucinating non-existent software packages or libraries. RSEs must verify that any generated dependencies or references actually exist and are secure.
- **The Standalone Rule – expanded into Vibe-Coding:**
  - No output from GenAI can be considered a Research Result.
  - All AI-generated code must never be deployed blindly;
  - Proper human oversight (e.g., thorough unit testing), and code verification are "required at every stage".

© 2026 by Dr.-Ing. Federico Grasso Toro is licensed under CC BY-NC 4.0

The "During" Phase – Quality Control (Annex A)

- **Building a Workflow:** Research groups should create a Standard Operating Procedure (SOP) for working with GenAI Tools.
- **The Quality Control Checklist:** Using it while actively working with a GenAI Tools.
- **Testing:** GenAI will not fully provide consistent answers to the same prompt.
- **Auditing:** Test new models under low-risk conditions with two independent auditors using publicly available work.
- **Slowing Down:** The checklist allows groups to slow down, verify, and thoroughly test the range of answers provided.

© 2026 by Dr.-Ing. Federico Grasso Toro is licensed under CC BY-NC 4.0

The "After" Phase – Transparency & Reporting (Annex B)

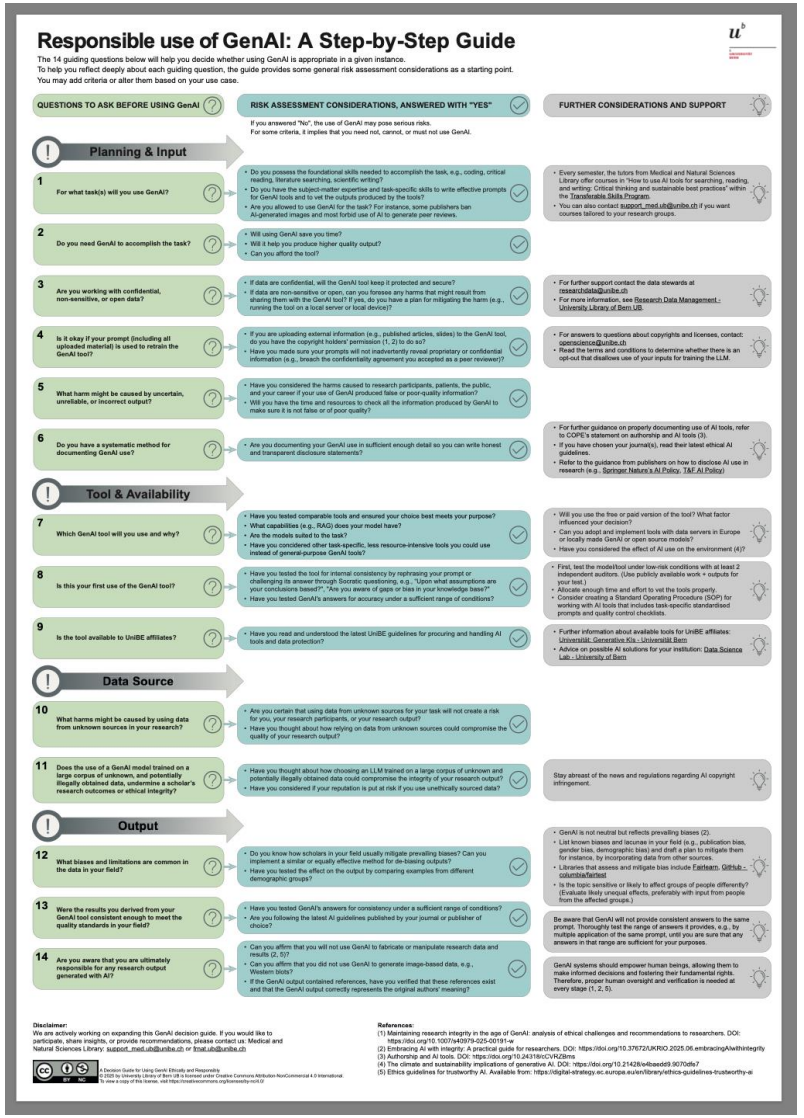
- **The Usage Declaration Template:** A systematic method to document GenAI use after completing a task as part of a workflow (i.e., Transparency Report).
- **Honest Disclosure:** Documenting use in sufficient detail is required to write honest and transparent disclosure statements.
- **Publishing Requirements:** Refering to publisher guidelines (e.g., Springer Nature) and/or COPE's statement on authorship for GenAI tools (MDPI's personal example).
- **Ethical Affirmations:** Researchers and general users must formally affirm they did not use GenAI blindly to fabricate and/or manipulate results.

© 2026 by Dr.-Ing. Federico Grasso Toro is licensed under CC BY-NC 4.0

AI Literacy in Research:
Selecting, Using, and Reporting GenAI

Takeaways

- **The Golden Rule Re-stated:** GenAI is just a tool for processing language into outputs; ONLY human oversight implies a Research Result.
- **Ethical Integrity:** The decision guide and checklists are not red tape; they are frameworks to protect your work (i.e., reputation and ethical integrity).



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.

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 - University Library of Bern UB.
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). • 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, T&F AI Policy)
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?	• 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? • Have you considered the effect of AI use on the environment (4)?
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: Universität 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 fmat.ub@unibe.ch

CC BY NC
A Decision Guide for Using GenAI Ethically and Responsibly
© 2026 by University Library of Bern UB is licensed under Creative Commons Attribution-NonCommercial 4.0 International.
To view a copy of this license, visit https://creativecommons.org/licenses/by-nc/4.0/

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/JUKRO.2025.06.embracingAIwithintegrity
(3) Authorship and AI tools. DOI: https://doi.org/10.24318/cCVJK2Bms
(4) The climate and sustainability implications of generative AI. DOI: https://doi.org/10.21428/e4baedd9.9070dfa7
(5) Ethics guidelines for trustworthy AI. Available from: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Dr.-Ing. Federico Grasso Toro

Data Steward for Data Science and IT

Federico.grasso@unibe.ch

© 2026 by Dr.-Ing. Federico Grasso Toro is licensed under CC BY-NC 4.0

AI Literacy in Research: **ANNEX A**

Selecting, Using, and Reporting GenAI

Annex A: 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:** _____

© 2026 by Dr.-Ing. Federico Grasso Toro is licensed under CC BY-NC 4.0

AI Literacy in Research: **ANNEX B**

Selecting, Using, and Reporting GenAI

Annex B: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, T&F) as well as COPE's statement 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:** _____

© 2026 by Dr.-Ing. Federico Grasso Toro is licensed under CC BY-NC 4.0

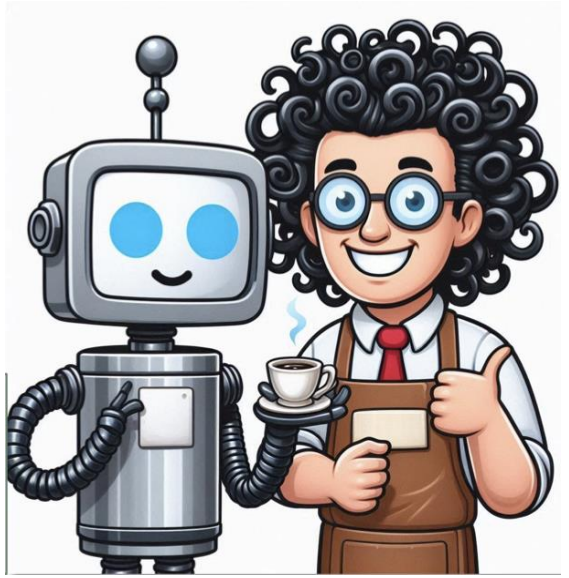
Contact

Dr.-Ing. Federico Grasso Toro

Data Steward for Data Science and IT

Federico.grasso@unibe.ch

Or by support.dsl@unibe.ch



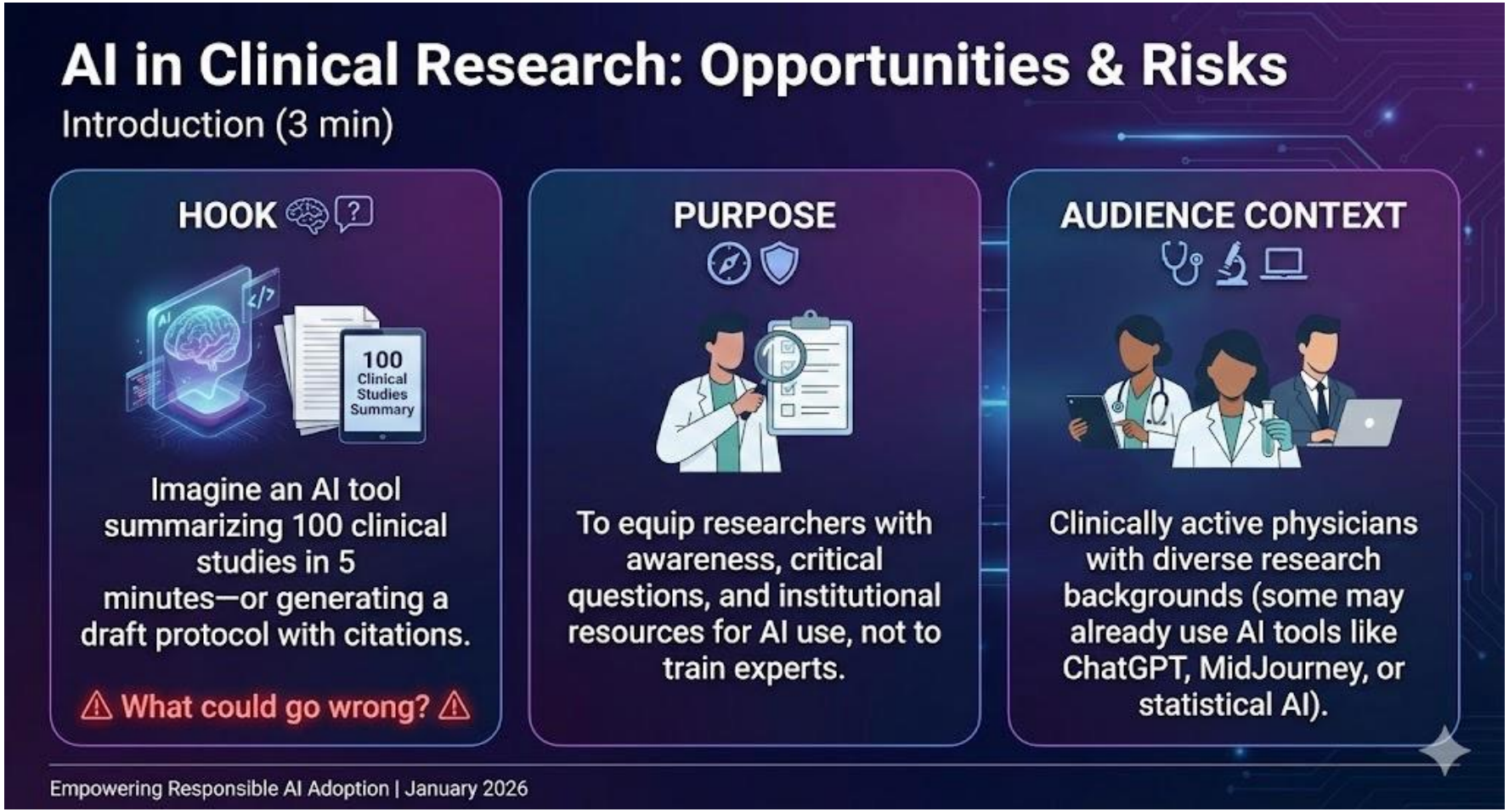
TL;DR:

Summary: Guardrails aren't meant to stop AI use; they are meant to enable it safely. By combining secure local infrastructure (TextLab) with critical thinking frameworks (the 14 Questions), UniBE safeguards both their workers and the principles of Open Science.

Call to Action: Encourage UniBE employees to activate their Campus Accounts on UBELIX to try TextLab as a better alternative (self-hosted solution) and to bookmark the 14-Questions Guide for before using GenAI.

AI aided Summary – (1/3)

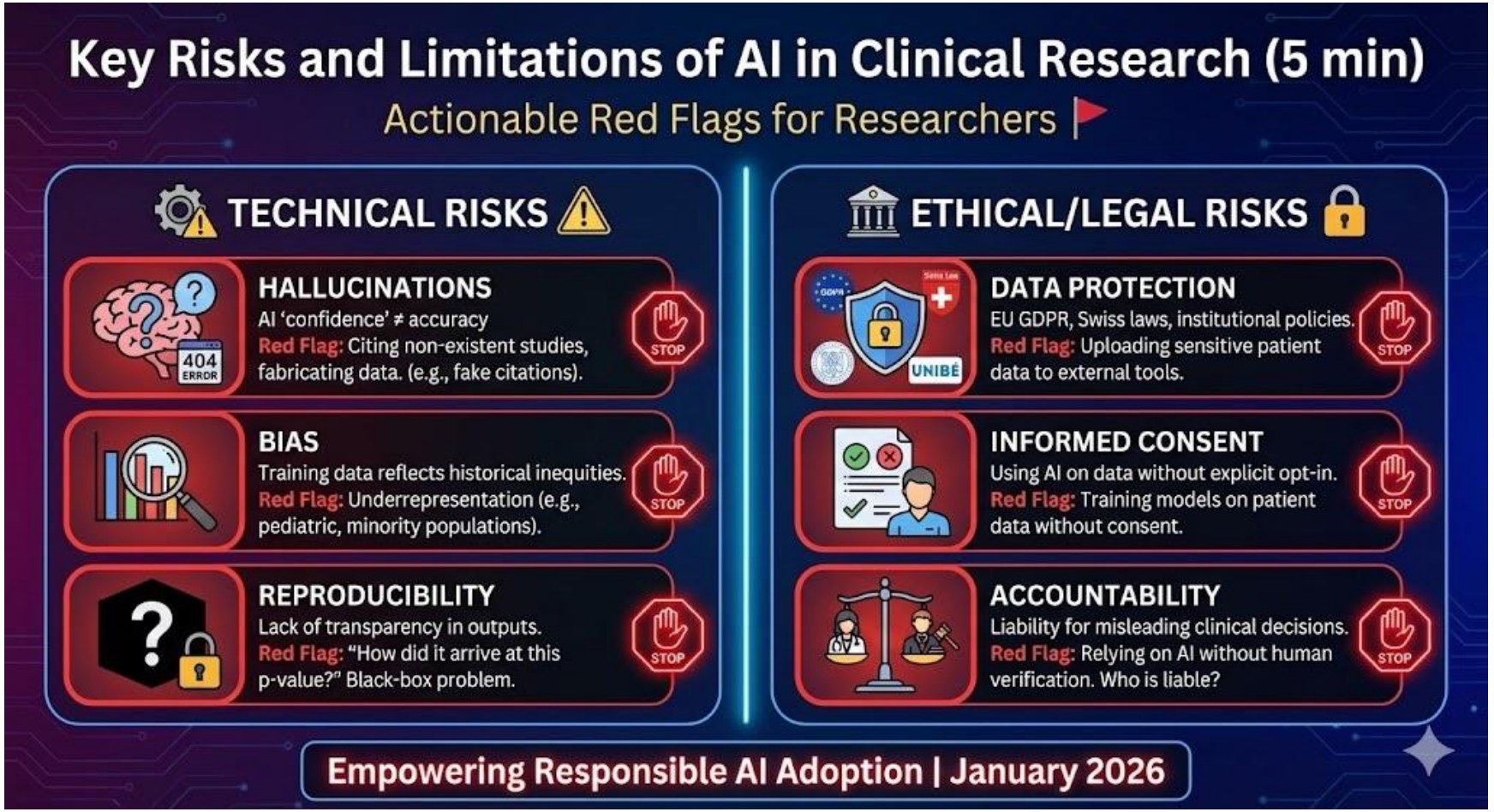
Introduction and Overview



© 2026 by Dr.-Ing. Federico Grasso Toro is licensed under CC BY-NC 4.0

AI aided Summary – (2/3)

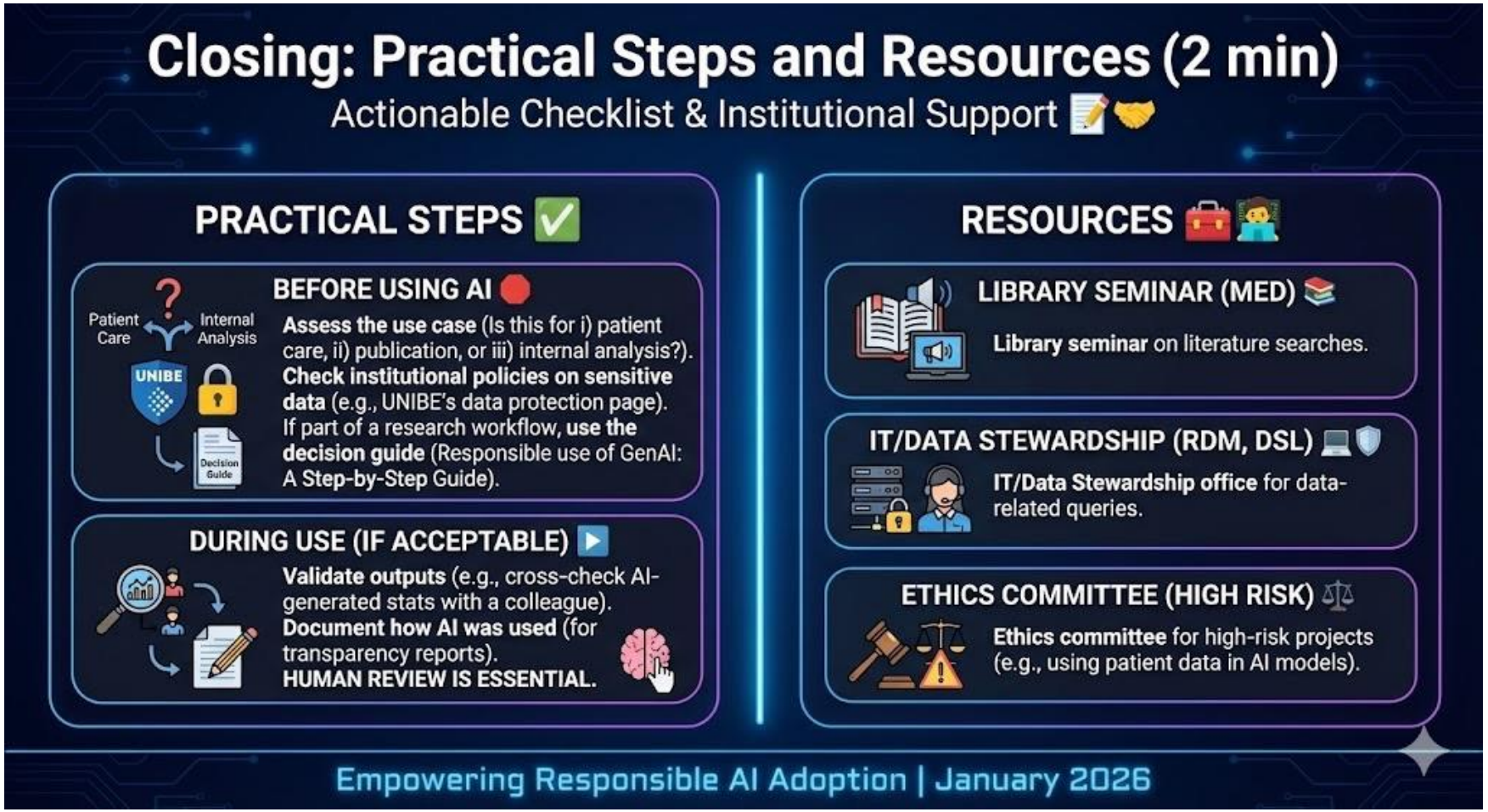
Current Key Aspects



© 2026 by Dr.-Ing. Federico Grasso Toro is licensed under CC BY-NC 4.0

AI aided Summary – (3/3)

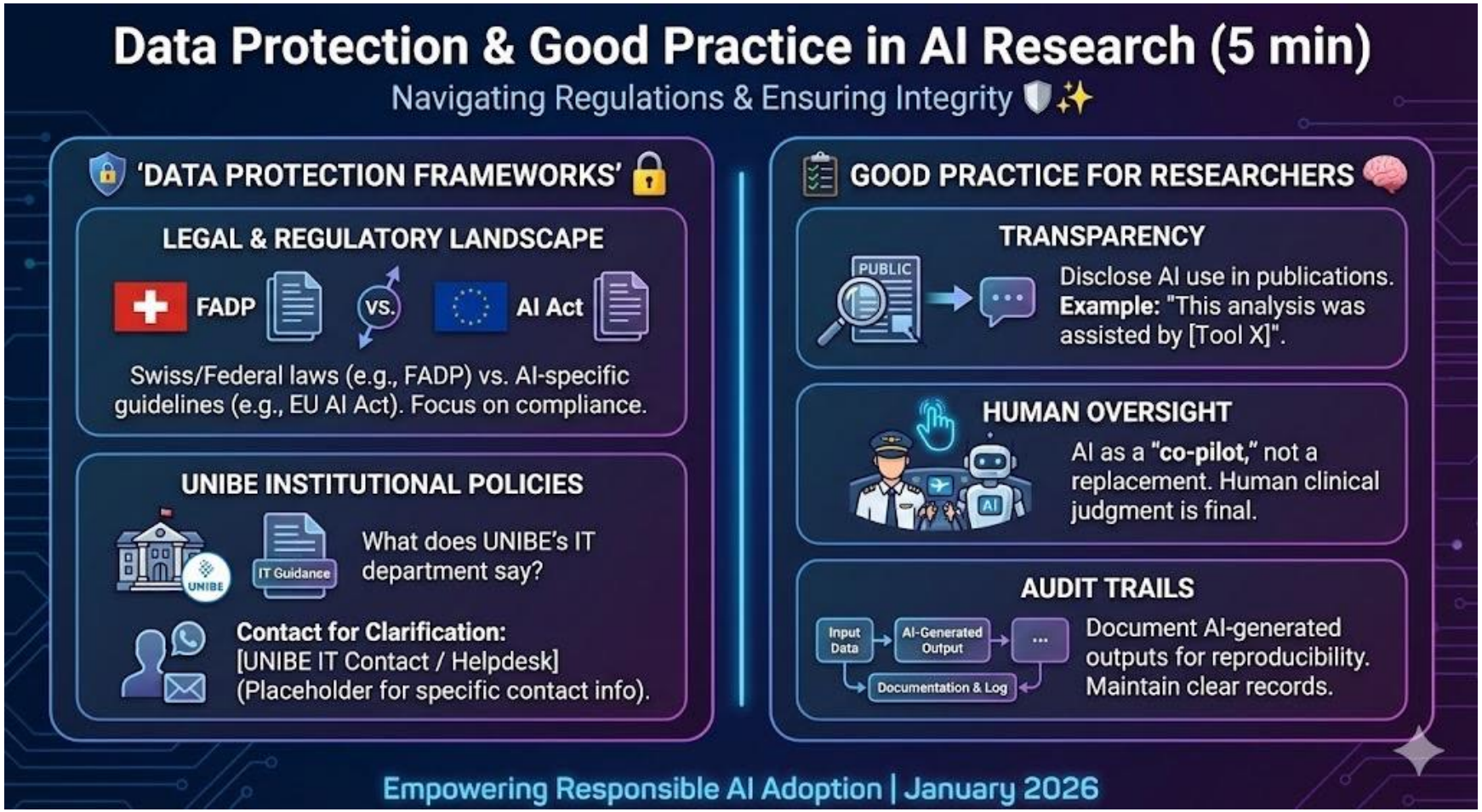
Institutional Safeguards and Future Outlook



© 2026 by Dr.-Ing. Federico Grasso Toro is licensed under CC BY-NC 4.0

AI aided Summary – (4/3)

Institutional Safeguards and Future Outlook



© 2026 by Dr.-Ing. Federico Grasso Toro is licensed under CC BY-NC 4.0