

Responsible use of Artificial Intelligence systems in research work and support

Ethics, integrity, and compliance

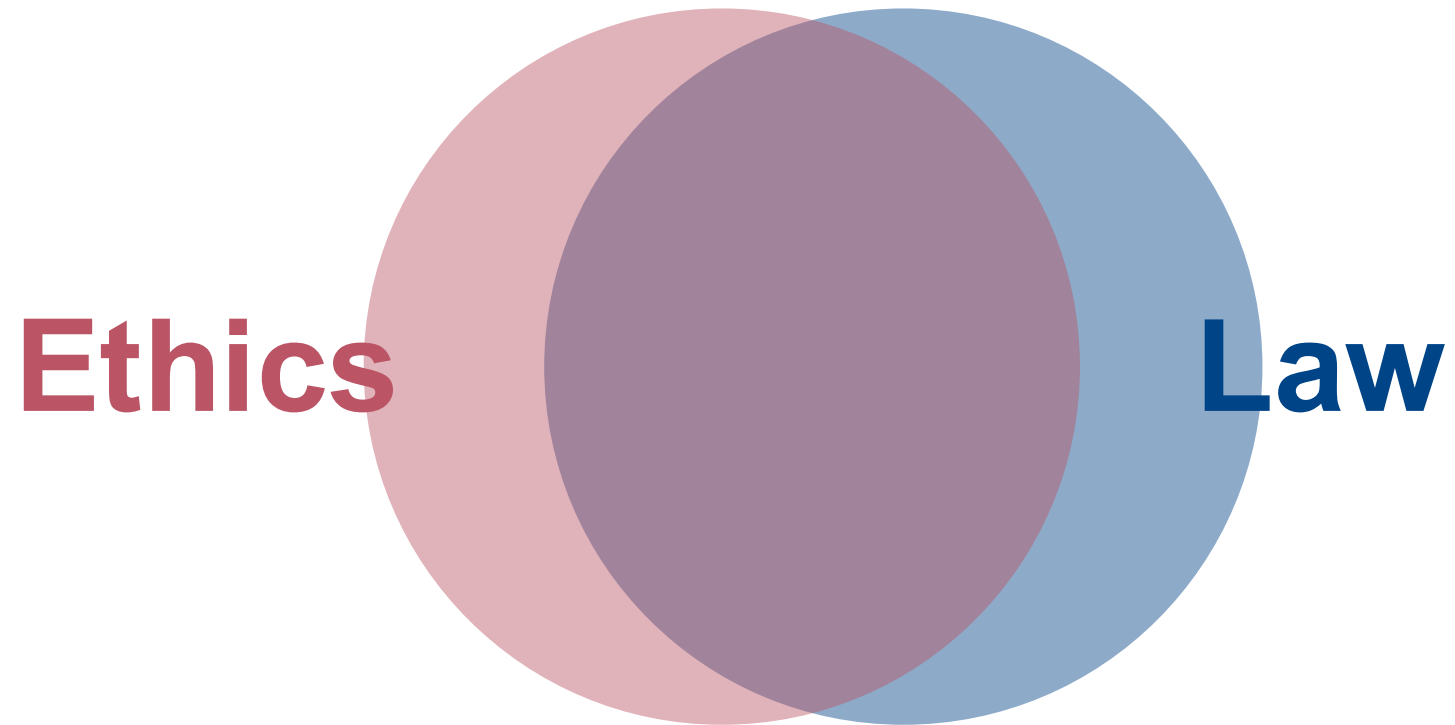
Dr. **Enrico Glerean**
Staff Scientist and Data Agent at **Aalto University**

About the speaker

Enrico Glerean, DSc.

- **Staff scientist and Data Agent at Aalto University** (background in neuroimaging), training and supporting researchers with handling personal data (anonymization, secure computing), medical images, clinical trials, research ethics and research integrity (AI and new technologies), statistics, open science.
- Other affiliations/COI: **CodeRefinery** (Nordic network to teach computational reproducibility), **Finnish Reproducibility Network** (National network to raise awareness in reproducibility), **Support Pool of Experts** at the **European Data Protection Board** (open training materials on personal data, AI, and cybersecurity), **Data Steward training program** for **Tampere University**

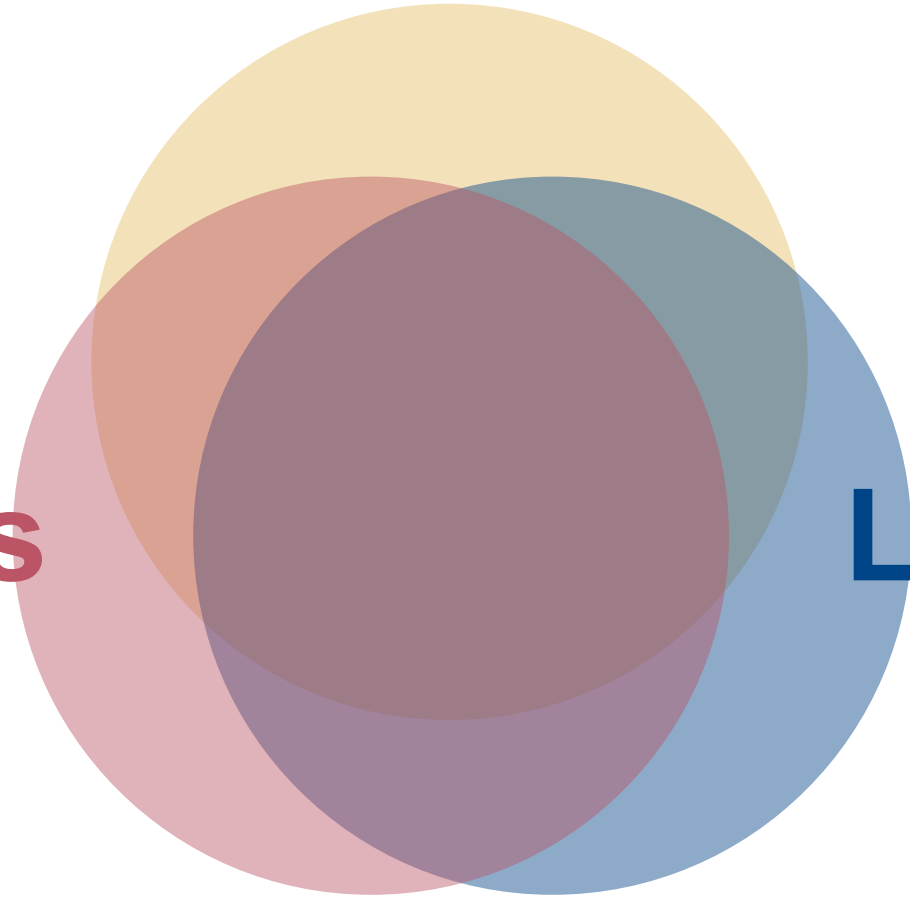
Ethics is not Law



Research

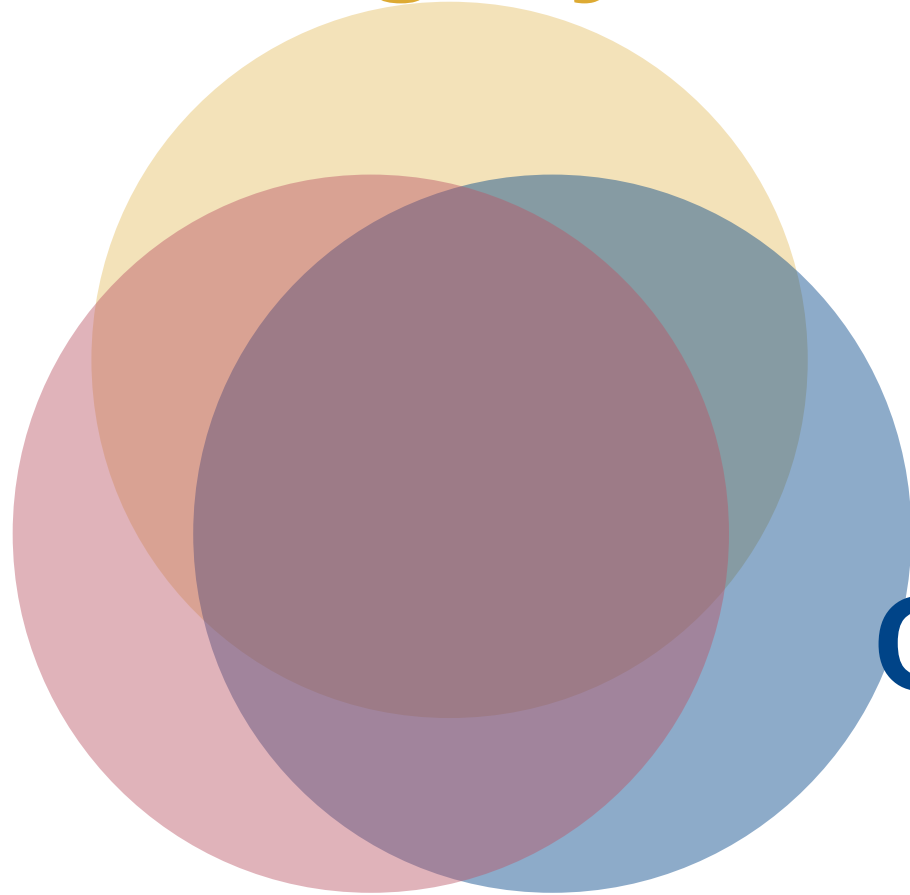
Ethics

Law



**Research
integrity**

**Research
Ethics**

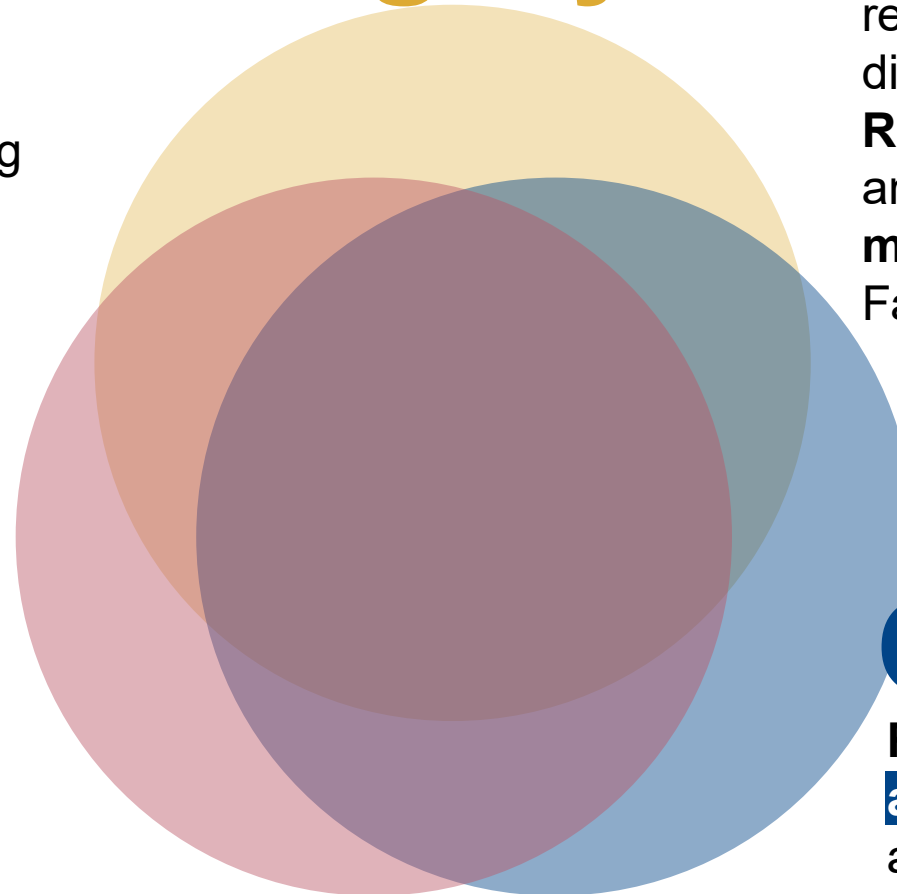


**Research
Compliance**

Research integrity

Research ethics is a set of **principles** that researchers must follow when conducting research: **Beneficence, Non-maleficence, Autonomy, Justice**, (*Beauchamp & Childress, 2019*)

Research Ethics



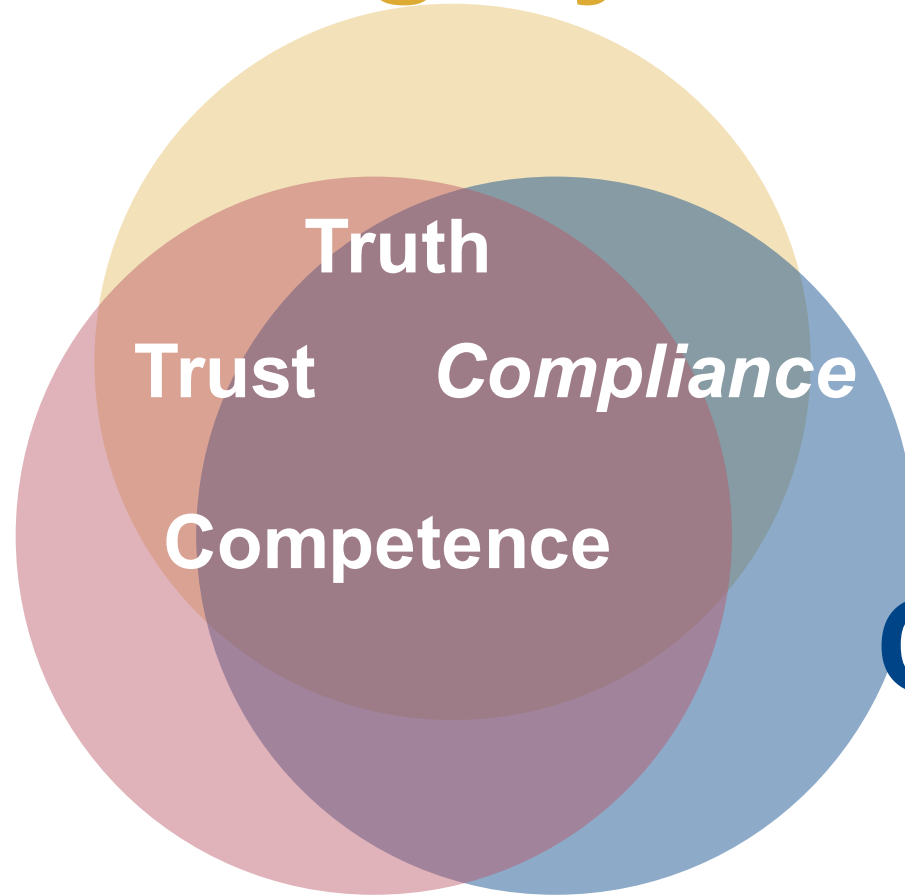
Research integrity primarily concerns the **practice**, **researchers' actions and behaviours** following the rules and regulations governing their disciplines: **Reliability, Honesty, Respect, Accountability** (ALLEA) and avoiding **Research misconduct** (Fabrication, Falsification, Plagiarism)

Research Compliance

Research compliance involves the **adherence** to regulations, directives, and national laws that apply to research and research outputs (General Data Protection Regulation, Dual Use and Export Control regulation, Sanctions regulation, Medical Device Regulation, Clinical Trial regulation, Animal Welfare directive, Finnish Act on the Secondary Use of Health and Social Data ...)

**Research
integrity**

**Research
Ethics**



**Research
Compliance**

A strong culture of open science and responsible research practices

Open science and responsible research practices

- Open science emerged as the ideal implementation of research integrity principles **Reliability, Honesty, Respect, Accountability** with strong foundations in **Research Ethics**, and adapting to a **more stringent legal landscape in the EU**
- **Budapest Open Access Initiative (2002)**
- **Creative Commons (2002)**
- **Transparency and Openness Promotion guidelines (2015, 2025)**
- **FAIR principles (2016, FAIR4RS)**

Open science has created strong opportunities for improving best practices, reducing falsification, fabrication, and plagiarism, and enabling reproducibility, replicability, data reuse, and European research infrastructures.

For years, we have taught researchers to use approved systems, secure/local storage, reproducible workflows, documentation, version control, and reference managers...

...and then came “AI”

A!

(Generative) AI

For simplicity we focus on **generative AI systems** that use **General Purpose AI models** such as **Large Language Models** that can be used in your research work.

Other uses of AI are not covered in this presentation (classifiers, recommendation systems, ML approaches to analyse data, etc...)

Text → Text

Text → Code

Speech → Text

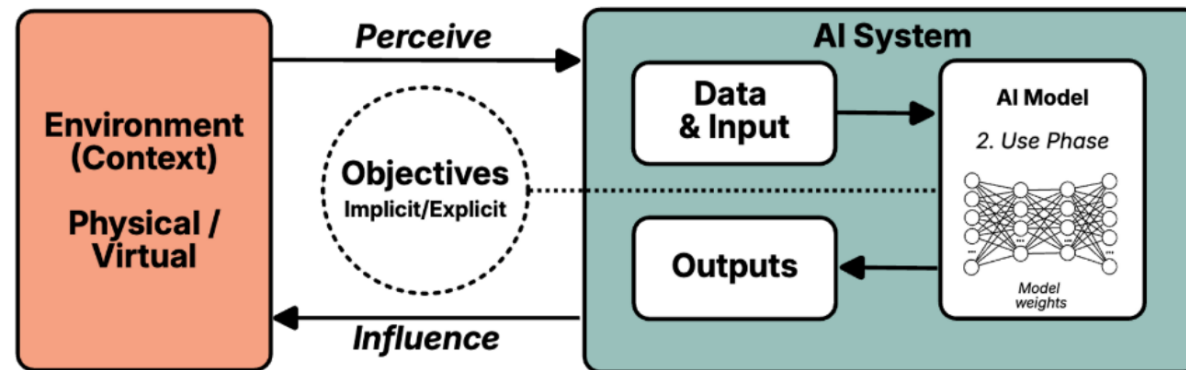
Text → Speech

Text → Image

Image → Text

A!

...

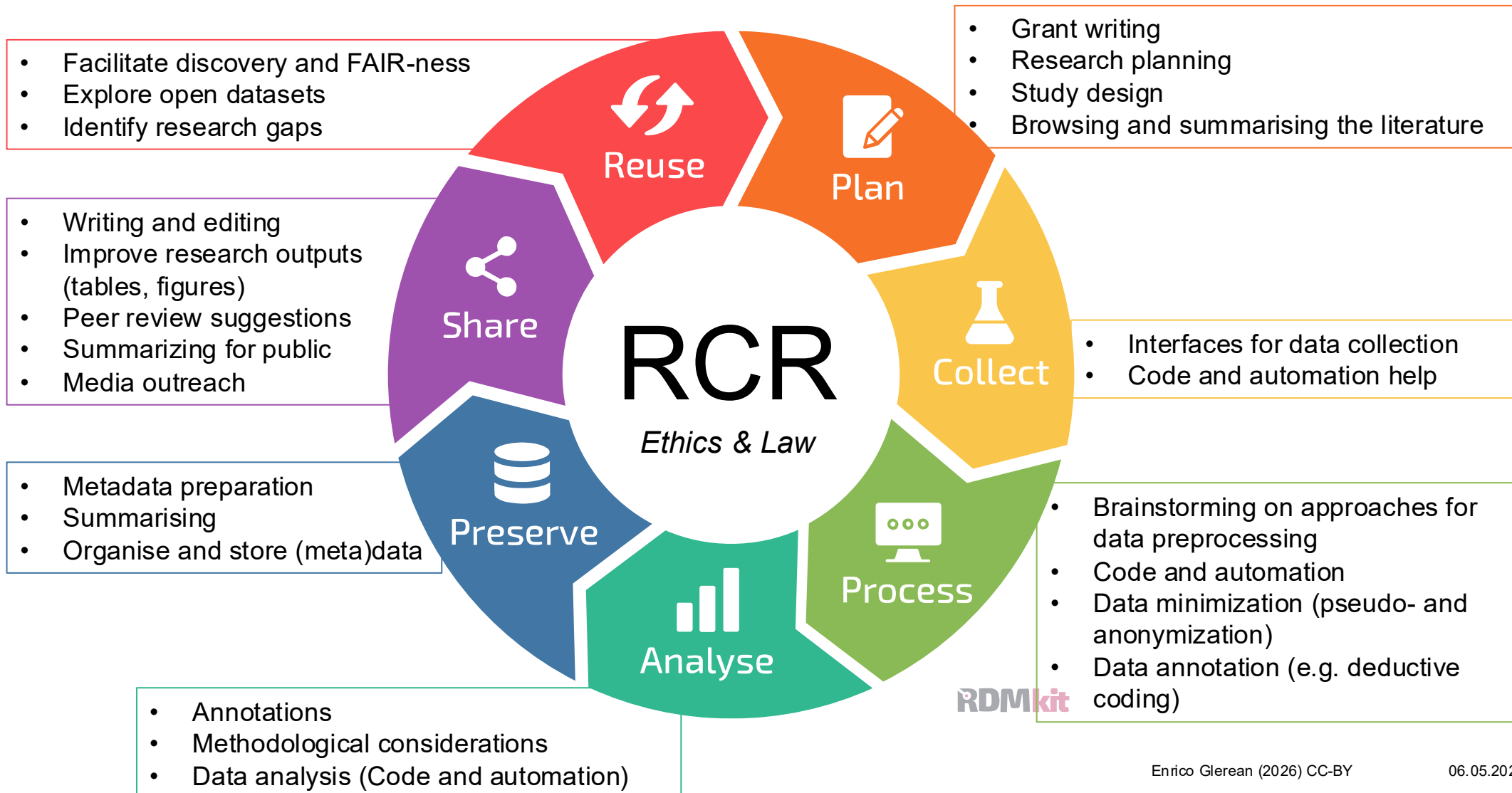


Use phase, post deployment

Figure from Glerean (2025) "Fundamentals of secure AI systems with personal data" European Data Protection Board

The Research Process aided by AI tools: Examples

with Responsible Conduct of Research at its core



AI!

Generative AI is the ideal misconduct machine

Generative AI: the ideal misconduct machine?

- **Generative AI systems** (like ChatGPT) are ideal fabrication, falsification, plagiarism machines: they are trained to always produce text and leave the quality control on the validity and reliability of the output to the final users.
- **Fabrication/Falsification:** GenAI tools can also perform data synthesis or data analysis, i.e. they can be used (**willingly or not**) to **fabricate data or falsify results** (*how many r's in strawberry?*) without the ability to verify and reproduce the analysis process.
- **Plagiarism: GenAI LLMs systems can *verbatim* reproduce text data in the training dataset** without citing the original source
- **Security and data protection risks:** commonly used generative AI system process your data remotely. You should protect the confidentiality and integrity of your data, your work, and the data of subjects, or other parties you work with. Use approved Aalto AI tools.

Misconduct is a core element of these systems

- *“(LLMs) may encode the correct answer, yet consistently generate an incorrect one” (Orgad et al. 2024)*
- *“...current models degrade documents during delegation: even frontier models (Gemini 3.1 Pro, Claude 4.6 Opus, GPT 5.4) corrupt an average of 25% of document content by the end of long workflows” (Laban et al. 2026)*
- *“... Harry Potter is so memorized that, using a seed prompt consisting of just the first line of chapter 1, we can deterministically generate the entire book near-verbatim.” (2025, Stanford University)*

Generative AI tools eroding truth, trust, and competence

- Hallucinated citations are polluting the scientific literature (Naddaf & Quill, Nature news, 2026)
- Same Prompt, Different Outcomes: Evaluating the Reproducibility of Data Analysis by LLMs (Cui & Alexander, 2026)
- The transparency dilemma: How AI disclosure erodes trust (Schilke & Reimann, 2025)
- The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers (Lee et al. 2025)

Can we do something before AI slop takes over all fields of research?

A framework for identifying risks of AI systems in research

Fundamental digital risks	AI-amplified risks	AI-native risks
<ul style="list-style-type: none">• Non-approved remote tools• Loss of confidentiality• Increased cybersecurity risks• Unethical / illegal solutions sold by vendors	<ul style="list-style-type: none">• Fabrication• Falsification• Plagiarism/IPR infringement• Deskillling• Environmental cost• Irreproducible research	<ul style="list-style-type: none">• Training data bias• No explainability• Erosion of autonomy• Impacts justice

Mitigating risks of AI systems in research

Fundamental digital risks

- Cybersecurity and data protection **trainings**
- **Locally hosted** tools
- Investments in **infrastructures** and **support experts**

AI-amplified risks

- **Trainings** on responsible use of AI
- **Better tools** for working with information sources (finding and citing)
- Resist and **just say no to AI for certain tasks**

AI-native risks

- **Abandon proprietary general purpose AI models**
- Adopt **automation** combined with **specialised models**
- Work on **interfaces to enable the work** rather than tools to replace the workers

What can data professionals do?

- **Learn and use AI tools as researchers do** (ChatGPT, Claude, Codex, Claude code, Opencode, AI agents...)
- **Raise awareness** on the research integrity risks that AI tools can amplify as well as other risks (security, deskilling, sustainability)
- **Encourage** the disclosure of AI already in the DMP
- **Develop new tools** such as integrating machine-readable DMPs with a controlled agentic component, or better tools for finding and managing citations
- **Support their organisations** when internal data are going to be used for developing new AI systems

Resist the hype, use AI tools responsibly (if you want to)

- AI literacy **training**
- **Safer versions of the tools you use** (e.g. local)
- **Opportunities to openly discuss** how people use AI (or avoid using AI)
- **Academic freedom** to let researchers explore (or ignore) such tools

A!



**KEEP
CALM
AND
CARRY
ON**

References

- ALLEA. (2023). The European Code of Conduct for Research Integrity (Revised edition 2023). ALLEA. <https://doi.org/10.26356/ECOC>
- Beauchamp, T. L., & Childress, J. F. (2019). Principles of biomedical ethics (8th ed.). Oxford University Press.
- Barker, M., Chue Hong, N. P., Katz, D. S., Lamprecht, A.-L., Martinez-Ortiz, C., Psomopoulos, F. E., Harrow, J., Castro, L. J., Gruenpeter, M., Martinez, P. A., Honeyman, T., et al. (2022). Introducing the FAIR Principles for research software. *Scientific Data*, 9, Article 622. <https://doi.org/10.1038/s41597-022-01710-x>
- Budapest Open Access Initiative. (2002). Budapest Open Access Initiative. <https://www.budapestopenaccessinitiative.org/read/>
- Center for Open Science. (2025). Transparency and Openness Promotion (TOP) Guidelines. <https://www.cos.io/initiatives/top-guidelines>
- Cooper, A. F., Gokaslan, A., Ahmed, A., Cyphert, A. B., De Sa, C., Lemley, M. A., Ho, D. E., & Liang, P. (2025). Extracting memorized pieces of (copyrighted) books from open-weight language models. arXiv. <https://doi.org/10.48550/arXiv.2505.12546>
- Cui, J., & Alexanor, R. (2026). Same prompt, different outcomes: Evaluating the reproducibility of data analysis by LLMs. arXiv. <https://doi.org/10.48550/arXiv.2602.14349>
- European Data Protection Board. (2025). Fundamentals of Secure AI Systems with Personal Data (E. Glerean, Author). https://www.edpb.europa.eu/our-work-tools/our-documents/support-pool-experts-projects/fundamentals-secure-ai-systems-personal_en
- Glerean, E., & Silva, P. E. S. (2023). Artificial Intelligence and Research Work [Slides]. Zenodo. <https://doi.org/10.5281/zenodo.8413164>
- Keiski, R., Hämäläinen, K., Karhunen, M., Löfström, E., Näreaho, S., Varantola, K., Spoof, S. K., Tarkiainen, T., Kaila, E., & Aittasalo, M. (Eds.). (2023). The Finnish code of conduct for research integrity and procedures for handling alleged violations of research integrity in Finland 2023. Finnish National Board on Research Integrity TENK. https://tenk.fi/sites/default/files/2023-05/RI_Guidelines_2023.pdf
- Laban, P., Schnabel, T., & Neville, J. (2026). LLMs corrupt your documents when you delegate. arXiv. <https://doi.org/10.48550/arXiv.2604.15597>
- Lee, H.-P. H., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (Article 1121, pp. 1-22). ACM. <https://doi.org/10.1145/3706598.3713778>
- Muthanna, A., Chaaban, Y., & Qadhi, S. (2024). A model of the interrelationship between research ethics and research integrity. *International Journal of Qualitative Studies on Health and Well-being*, 19(1), Article 2295151. <https://doi.org/10.1080/17482631.2023.2295151>
- Naddaf, M., & Quill, E. (2026). Hallucinated citations are polluting the scientific literature. What can be done? *Nature*, 652(8108), 26-29. <https://doi.org/10.1038/d41586-026-00969-z>
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., ... Yarkoni, T. (2015). Promoting an open research culture. *Science*, 348(6242), 1422-1425. <https://doi.org/10.1126/science.aab2374>
- Orgad, H., Toker, M., Gekhman, Z., Reichart, R., Szpektor, I., Kotek, H., & Belinkov, Y. (2024). LLMs know more than they show: On the intrinsic representation of LLM hallucinations. arXiv. <https://doi.org/10.48550/arXiv.2410.02707>
- Research Data Management Kit. (n.d.). Data life cycle. ELIXIR. https://rdmkit.elixir-europe.org/data_life_cycle
- Schilke, O., & Reimann, M. (2025). The transparency dilemma: How AI disclosure erodes trust. *Organizational Behavior and Human Decision Processes*, 188, Article 104405. <https://doi.org/10.1016/j.obhdp.2025.104405>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, Article 160018. <https://doi.org/10.1038/sdata.2016.18>