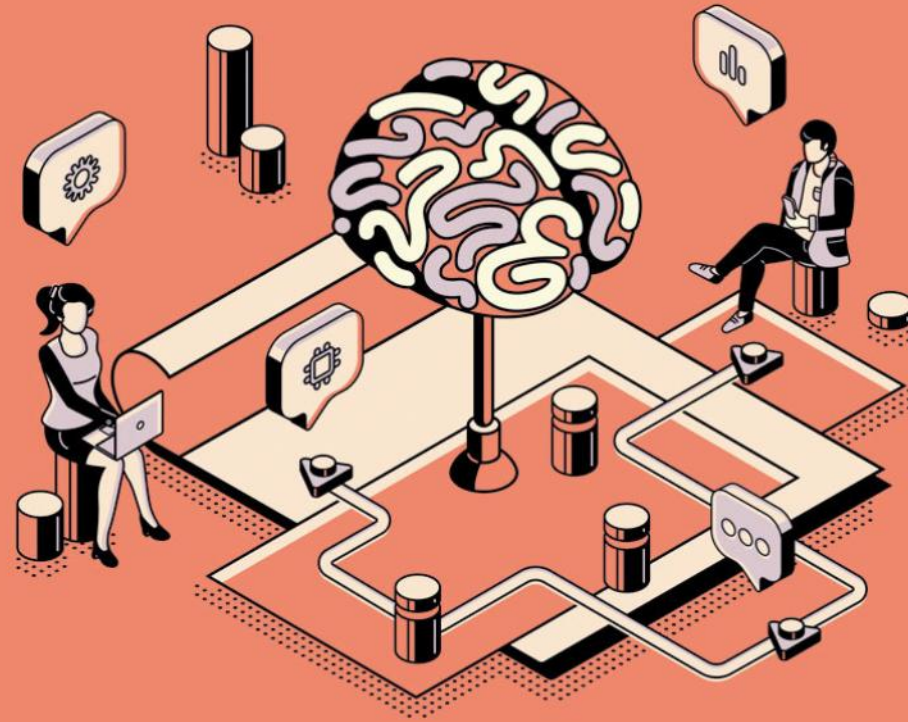




Automated Governance in Science? The Impact of Generative AI on Epistemic Authority and Responsibility

Linda Nierling, Angelina Sophie Dähms, Dana Mahr

10.-11.6.2026, Berlin

8th Weizenbaum Conference: Generative AI and Society. What is at stake?

Agenda

- 1. What is at stake? Challenges of Epistemic Authority and Responsibility in Science**
- 2. Empirical Study with Science Governance Actors**
 - Methodical Approach
 - First Results
- 3. Conclusions: On the Way to Automated Governance?**

Background: DiTraRe – Project

- Leibniz ScienceCampus "Digital Transformation of Research" is a joint project of FIZ Karlsruhe and the Karlsruhe Institute of Technology, funded by the Leibniz Association (2023-2027) (<https://www.ditrare.de/en>)
- Investigating the effects and potential of the increasing digitalization of scientific work in four interdisciplinary research clusters
- Includes the theoretical-conceptual level, methodological issues as well as consequences for science and society
- Based on interdisciplinary use cases: sports science, climate research and meteorology, biomedical engineering, chemistry
- Focus: Reflection & Resonance, Publication Cultures



Leibniz ScienceCampus
Digital Transformation
of Research



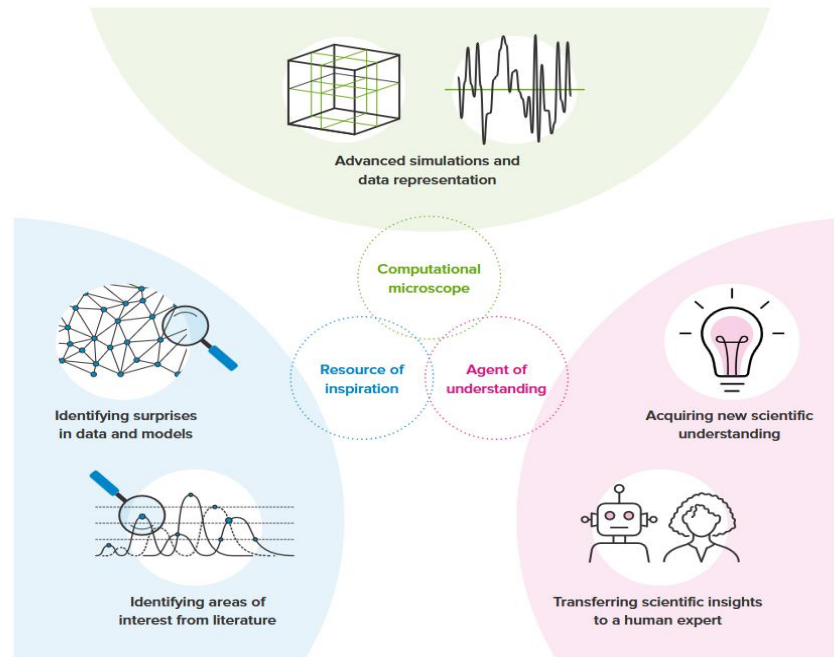
FIZ Karlsruhe

Leibniz Institute for Information Infrastructure



Karlsruhe Institute of Technology

1. Problem orientation



Epistemic risks



Authority?
Responsibility?

*“Just as plant monocultures are more vulnerable to pests and disease, **scientific monocultures** make our understanding of the world more vulnerable to error, bias and missed opportunities of innovation” (Messeri & Crockett 2024: 53).*

The Royal Society (2024): Science in the age of AI: How artificial intelligence is changing the nature and method of scientific research., p. 31

1. Generative AI in science – an overview

Digitalisation of research affects: How science is conducted but also how it is perceived?

- Transparency, data openness, and the potential for reproducibility enhance public trust.
- The reliance on black-box AI systems and large-scale datasets raises concerns about the reliability and accessibility of scientific findings.

(Generative) AI challenges workflows, infrastructures and outputs in science

- New chances for knowledge generation and processing
 - Increased efficiency and productivity in scientific working practices with various possible applications
- But:
 - Technical limits: Transparency and traceability: Detection, authentication, “hallucinations”
 - Social changes: AI-literacy of researchers & potential loss of competences
 - Legal issues: Data protection, limits of regulation
 - Ethical questions: Overreliance, spread of false information, biased knowledge, source misrepresentation

Challenges for research integrity on two levels

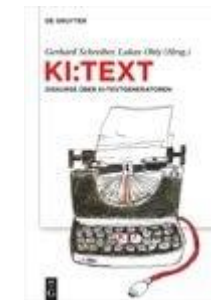
- Epistemic authority
- Responsibility



<https://www.ditrare.de/de/symposium-2025>



https://www.itas.kit.edu/veranstaltungen_2024_nta11.php



<https://doi.org/10.1515/9783111351490>



<https://doi.org/10.1007/978-3-658-48491-0>

1. Generative AI in science – challenging epistemic authority?

What is epistemic authority? A working definition

...refers to socially and institutionally recognized capacity to define, validate, assess, legitimate knowledge claims as credible, relevant, reliable. It is produced through:

- Disciplinary standards incl. methodological transparency
- Practices of critique and validation incl. peer review
- Publication cultures, reputation systems & research infrastructures

Generative AI introduces a new layer between knowledge production & scientific judgement

- Literature review, writing and argumentation, data interpretation
- Quality assessment, peer review, publication workflows

But: What is at stake?

- Reliability, transparency, traceability
- Bias, hallucinations, epistemic pollution
- Loss or transformation of expertise, scientific monocultures

“The figure of the scientist is thus threatened with displacement from the centre of knowledge creation. This is not simply a technical redistribution of tasks but a deeper reconfiguration of epistemic authority.” (Khutsishvili 2026, p. 51)

1. Generative AI in science – challenging responsibility?

What is responsibility in science? A working definition

...refers to the distributed obligation of individual and institutional actors to ensure the integrity, reliability, robustness of knowledge production. It is distributed across:

- Individual researchers, research organizations, research associations
- Funding bodies, publishing entities, infrastructure providers
- Regulatory and policy actors

New responsibilities emerge around maintaining human judgement by

- AI literacy, documentation of AI use
- Verification of outputs, bias and hallucination checks
- Defining adequate guidelines

“The novelty of AI-mediated inquiry creates a new demand within the sciences themselves: the need for epistemological orientation.”
(Khutsishvili 2026, p. 53)

But: Who is accountable when AI-mediated processes influence scientific decisions?

- The researcher? The institution?
- The publisher? The funder?
- The tool provider? The guideline-setting body?

2. Empirical study on science governance actors

Research questions:

- *How is GenAI use affecting epistemic authority and responsibility across actors, institutions, and infrastructures in the science system?*
 - How is scientific knowledge produced, validated and legitimised? (1)
 - How are steering processes organised? (2)

Methodical approach:

- Semi-structured expert interviews (April - June 2025) with international actors from science governance:
 - 1. Oversight & funding bodies: Ombudsmann für die Wissenschaft (1), DFG (2), EU (9)
 - 2. Publishing entities: Springer Nature (6), MIT Library (7), Frontiers/COPE (8)
 - 3. Research associations: HGF (3,5), Leibniz (4), FZ Jülich (10)
- Analysis: qualitative content analysis & code structure: Mayring (2022), use of MAXQDA

2. How is scientific knowledge produced, validated and legitimised?

Producing knowledge: Skills of researchers

- AI literacy
 - *There is a general need for more training on how to approach this **in the context of research ethics, academic integrity, and good research practice.** (Research association, translation)*
- New skills
 - ***because we need to develop new skills, given that science is very skills-driven.** I'm not sure whether it's already established enough for us to talk about careers, **or whether it's a transitional phase.** (Oversight and funding body, translation)*
- Prompting competences
 - ***prompt engineering** [...] is based on our expertise and [...] our kind of benchmarking of what actually works, to what extent, and **paired with that human expertise.** (Publisher)*
- Oversight knowledge
 - *I would say the role of humans is that if you're going to use AI, you have to **put your very strong critical thinking hat on** when you verify everything that the AI has generated, and to be able to make sure it's fully accurate, it's valid, and it's cites the right places. (Publisher)*

→ Open culture for learning (hybrid knowledge) and critical human judgement essential

2. How is scientific knowledge produced, validated and legitimised?

Producing knowledge: New models

- In-house models
 - *What I am increasingly seeing is that certain **weaknesses** in AI, particularly with regard to **confidentiality and data protection**, are being addressed through these in-house models. (Oversight and funding body, translation)*
- Tailored models
 - *We now need to start investing in these technologies ourselves. **Science-driven technologies and models that are also compatible with our values and practices.** (Research association, translation)*
- Enrichment with interdisciplinary competences
 - *What I see is that the more we use AI, the **more complexity we are adding.** [...] We need to include much more different profiles in our research team. [...] in the moment that you want to create your own model, or to train a model, or fine-tune a model, you need **to include people, research engineers, or data steelworks,** [...] That will have an influence on **how we fund science, how science is managed,** emerging is also this interdisciplinarity and hybrid profiles. (Oversight and funding body)*

→ Targeted innovations to overcome current (technical) shortcomings

2. How is scientific knowledge produced, validated and legitimised?

Validation & legitimacy

- Peer review
 - *We think of high potential for using AI for summarization, **to assist peer review**, not to replace peer review[...] Things like assisting editors in **identifying peer reviewers** based on analysis on literature, **providing reviewers with summaries, abstracting, and assisting in verifying citations, protocols**, et cetera, throughout the whole process as a companion to peer review. (Publisher)*
- Human authority
 - ***That you make a generated text your own. At the very least by reading it in its entirety.** That, as a researcher, you agree with every sentence and you cannot simply assume that it is correct, because when using these systems, you cannot assume that it is already correct. It has to be checked, so to speak. (Oversight and funding body, translation)*
- Technical detection
 - *Even the OpenAI company that created a tool to try to detect if text was created by ChatGPT cannot detect if a text was created by ChatGPT. The accuracy that I've seen the best, I think, is at 60 or 70 percent. **It's basically you flip a coin, and you decide if it's been generated by AI or not.** (Publisher)*
- Return to personal networks
 - *Ultimately, in my view, **it all comes down to going back to your personal network.** So that you trust people because you know them. Because you're part of a research community where everyone knows one another. Then things get back on track. Which lab is someone from? Does the lab have a good reputation, or has the person come out of nowhere?(Research association, translation)*

→ Limitations of hybrid assurance practices

2. How are steering processes organised?

Challenges for scientific steering

- Variety of (epistemic) risks
 - These include epistemic risks relating to the **dissemination of scientific expertise** or the reliance on such expertise, as well as **other forms of risk**, such as the spread of misinformation, bias, plagiarism, or the misrepresentation of sources. (Research Association, translation)
- Power Structures
 - These are tools that can be useful for science and knowledge production. **Who controls them and who covers them?** Having **concentrated economic interests**, having most of the work and done in **closed systems and on closed data**, it gives a **lot of advantages for potential monopolies** which is bad for who get the returns on science and knowledge, but also bad for incentives to develop basic research. (Publisher)
- Transparency
 - Take prompt discussion [...] Quite seriously, it doesn't benefit anymore when it comes to transparency. **Does it help for someone to read my ever-more-refined prompt?** Does that really serve the cause of transparency? (Oversight and funding body, translation)
- Accountability
 - If you say, **no generative AI for your figures or no generative AI for your text**. **You need a way for people to be accountable for it**, and you need a way for people to be able to detect if they are doing it or not. (Publisher)

→ Limits of traceability (data use and persecution)

2. How are steering processes organised?

Guidelines as main steering instrument

- Fragmentation
 - There could be a **big fragmentation**. In the first scoping that we did before even launching the idea, we saw that there were already 20 or 30 different guidelines from universities in Europe in the United States. (Oversight and funding body)
- Distributed Authority
 - We have stakeholders that represent different families of stakeholders, like universities, publication businesses, research organizations, and so on. **We need a goal there in order to co-create these guidelines.** In fact, we would have been very happy **if any other stakeholder or country had taken the lead role**, and we could have supported them. (Oversight and funding body)
- Limits of guidance
 - A guideline [...] can recommend how the use has been handled. But it can actually only **influence actual scientific work to a very limited extent**. (Oversight and funding body, translation)
- Pacing problem
 - Part of the challenge of this is that the **technology moves so quickly in this space, and it leaves a gap on one side, the legal side, but also what the guidance can actually produce.** (Publisher)

→ Guidelines are in place but lack both, liability and granularity

3. First conclusions

- **GenAI is on the agenda of decision-makers in the science system**
 - **Increased efficacy is embraced**
 - AI literacy, hybrid skills, tailored models
 - **Differentiation of roles**
 - **Oversight & Funding bodies:** Development of interdisciplinary frameworks for process efficiency and control
 - **Publishing entities:** Dealing with challenges in academic writing, peer-review and quality assurance practices
 - **Research associations:** Research sovereignty, while guiding researchers by practices and tools for use (e.g. own models)
- **Scientific practices are reshaped**
 - New patterns of knowledge production emerge
 - Hybrid competences & ultimate human judgement
 - But
 - Systemic and practical limits of validation
 - Increase of human verification (verification tax)
 - Diffusion of responsibilities between different actors

3. First conclusions

- **Guidelines are the main steering instrument & are in place for many contexts (also for TA), but**
 - Limited reach; contradicting and few practical guidance
 - Traceability and liable tracing remains open
 - No single entity in charge, but distributed guidance for steering
- **Legitimation under pressure**
 - Standardisation of knowledge practices: Monocultures for the sake of traceability?
 - Hybrid measures & human oversight: Who provides final authorisation?
 - Increase of human verification: Where it is situated, executed, which resources are used?
 - Return to traditional assurance practices: Backlash to "old-boy networks"?
- **Shift of knowledge negotiation processes to technical infrastructures: Automated governance as future of science? What is needed?**
 - Continuous adaption & balance between epistemic authority & responsibility by different actors
 - Development and organisation of new negotiation arenas, incl. infrastructures
 - Risk of overburdening individual scientists
 - Need for development of binding (legal) anchors

3. On the way to automated governance? A hypothesis

Automated governance describes the partial delegation of evaluative, classificatory, decision-supporting or steering processes to algorithmic or AI-based infrastructures.

This may include AI-supported or automated processes in identification of relevant knowledge, quality assurance, e.g. manuscript screening, peer review, plagiarism and AI detection, research evaluation, e.g. ranking and metric systems, funding decisions.

Automated Governance is not "Full automation", rather governance is increasingly embedded in technical infrastructures, platform architectures, interfaces, automated workflows, standards and templates, recommender systems, scoring systems, detection tools.

3. On the way to automated governance?

- **Future questions:**
 - **Scientific judgement may remain formally human, but its conditions increasingly shaped by technical systems...**
 - How to deal with distributed responsibility?
 - How to organize the negotiation between different actors, incl. AI?
 - **Automated governance in science does not replace human judgement, but changes how it is exercised, delegated and justified...**
 - How to keep reflexive, contestable and responsible forms of scientific judgement?
 - How not to overburden individual scientists?
 - How to organize continuous adaption & balance between epistemic authority & responsibility by different actors?

Literature

Digital Transformation of Research. 2025. DiTraRe Symposium 2025. Karlsruhe. <https://www.ditrare.de/de/symposium-2025>, visited 11.06.2026.

Digital Transformation of Research. 2026. Leibniz-WissenschaftsCampus - Digital Transformation of Research (DiTraRe). <https://www.ditrare.de/de/symposium-2025>, visited 11.06.2026.

Institut für Technikfolgenabschätzung und Systemanalyse. 2024. 11. Konferenz des Netzwerks Technikfolgenabschätzung. Berlin. https://www.itas.kit.edu/veranstaltungen_2024_nta11.php, visited 11.06.2026.

Khutsishvili, Kristina. 2026. AI and the scientist. On the fracture of epistemic authority. In Understanding Science with Large Language Models? Potentials for the History, Philosophy, and Sociology of Science, Eds. Arno Simons, Adrian Wüthrich, Michael Zichert, Gerd Graßhoff, pp. 51-55. transcript: Bielefeld. <https://doi.org/10.14361/9783839447529>

Mayring, Philipp. 2022. Qualitative Inhaltsanalyse. Grundlagen und Techniken. Weinheim: Julius Beltz. ISBN 978-3-407-25898-4

Messeri, Lisa, and M. J. Crockett. 2024. Artificial intelligence and illusions of understanding in scientific research. In Nature 627: 49-58. <https://doi.org/10.1038/s41586-024-07146-0>

Ornetzeder, Michael, Mahshid Sotoudeh, and Walter Preissl (Eds.). 2025. Methoden für die Technikfolgenabschätzung. Bewährte Praxis, aktuelle Herausforderungen, neue Chancen. Wiesbaden: Springer VS. <https://doi.org/10.1007/978-3-658-48491-0>

Schreiber, Gerhard, and Lukas Ohly (Eds.). 2024. KI:Text. Diskurse über KI-Textgeneratoren. Berlin: De Gruyter. <https://doi.org/10.1515/9783111351490>

The Royal Society. 2024. Science in the age of AI. How artificial intelligence is changing the nature and method of scientific research. https://royalsociety.org/-/media/policy/projects/science-in-the-age-of-ai/science-in-the-age-of-ai_report_web.pdf

Thank you for your attention!

Contact for questions & comments:

linda.nierling@kit.edu

angelina.daehms@kit.edu

dana.mahr@kit.edu