A Practical Ethics of Thinking: Human-in-the-Loop Responsibility and Authorship in Academic Work with AI

Marcel K. Will[1]

[1] Cooperative State University Baden-Württemberg
marcel.will@dhbw.de

Abstract. This paper examines how generative AI [genAI] changes not only academic writing but the processes of thinking and knowledge creation. GenAI can serve as a productive space of resonance by challenging assumptions and introducing new perspectives, while its linguistic coherence can also create an Illusion of Evidence, in which plausibility is mistaken for justified knowledge. The paper argues that conventional Human-in-the-Loop approaches insufficiently capture researchers' responsibility once AI becomes involved in the formation and evaluation of ideas. It therefore develops a Practical Ethics of Thinking and proposes Human-in-the-Loop-Thinking as a framework for preserving intellectual autonomy, epistemic vigilance, intellectual self-assertion, transparency, and non-delegable responsibility. Its central claim is that while AI can support the creation of knowledge, knowledge itself, human judgement, and the execution of responsibility are not delegable to a machine.

Keywords: Practical Ethics of Thinking; Human-in-the-Loop-Thinking; Generative AI; Epistemic Responsibility; Intellectual Autonomy; Illusion of Evidence.

1 Introduction

More than a decade ago, Brynjolfsson and McAfee described digital technologies as increasingly capable of extending not only human physical but also cognitive capacities. They expected humans and intelligent machines to work together to create new knowledge, innovation, and solutions (Brynjolfsson and McAfee 2015, p. 120). Generative AI [genAI] has made this prospect considerably more tangible. That raises the question of what such collaboration means for human thinking and judgement which is more fundamental than asking how AI can make academic work more efficient.

GenAI is increasingly becoming part of academic work. Large language models [LLMs] can assist researchers in searching and structuring information, developing arguments, testing counterarguments, and revising texts. Thomas Metcalf describes some of the problems arising from the submission of fully AI-generated papers, particularly in the humanities. He argues that "LLMs will overproduce safe, technical, and insular work, crowding out radical innovation in humanities fields" but will reduce the variety of thoughts expressed and ideas pursued (Metcalf 2026, p. 2/23). Of course, it stands to reason that such a development does not have to remain limited to humanities.

Current AI ethics already extends far beyond the question of how machines themselves should behave. Floridi, for example, develops five principles - beneficence, non-maleficence, autonomy, justice, and explicability - that connect the development and use of AI to human responsibility and decision-making (Floridi 2023, pp. 58–64). Similarly, Mühlhoff understands AI ethics as a broader social and political practice concerned with the conditions under which humans develop, deploy, and live with AI systems (Mühlhoff 2020, p. 2). Both approaches

therefore move beyond mere machine ethics and explicitly include human agency and responsibility.

However, these approaches primarily deal with the ethical conditions of human action, social structures, and the governance or use of AI rather than the individual process of thinking with AI itself. Even Floridi's principle of autonomy, including his "decide-to-delegate model" (Floridi 2023, p. 62), primarily addresses which decisions humans should retain or delegate. This paper starts one step earlier: what happens to intellectual autonomy when AI becomes part of the process through which a judgement is formed in the first place? A *Practical Ethics of Thinking* therefore shifts the focus from responsible action with AI toward the normative conditions of responsible thinking with AI.

Therefore, this article's argument rests on three observations. First, genAI changes the epistemic structure of academic work by providing continuous intellectual resonance. By using LLMs, arguments can immediately be questioned, reformulated, or confronted with alternatives. This resonates with Dewey's pragmatist understanding of inquiry as an iterative process of interaction, testing, and revision. A Practical Ethics of Thinking builds on this understanding but adds a normative dimension necessitated by genAI: once suggestions, arguments, and apparent evidence can be introduced continuously by a machine, researchers must consciously assume and actively define responsibility for the process through which propositions become their own judgements. In this sense, responsible thinking with AI requires not merely reflection, but the deliberate regulation of reflection itself - a connection already implicit in Dewey's insistence that reflective thinking must be regulated by its purpose (Dewey 1910, pp. iv-vi). This way, AI can become a valuable resonance and structuring instrument rather than merely a tool for producing information and texts blindly adopted.

Second, AI-supported inquiry involves epistemic risks. Algorithmically generated results may acquire an aura of objectivity and credibility despite being context-dependent, opaque, or biased. In extreme cases, seemingly objective algorithmic classifications may even reproduce and legitimize discriminatory assumptions (Horwath 2022, pp. 71–72). In the case of genAI, information may also simply be inaccurate or fabricated - a phenomenon commonly described as AI hallucination. Importantly, such hallucinations can nevertheless appear credible because of the fluency, coherence, and authoritative tone of AI-generated responses (Shao 2025). This can produce what this article calls an *Illusion of Evidence*: linguistic plausibility and argumentative structure begin to function as substitutes for evidence. Moreover, continuous interaction with AI may influence the language, argumentative patterns, and conceptual framing of the researcher's own thinking.

Third, Mollick's concept of Co-Intelligence illustrates both the potential and the limits of existing approaches. His insistence on keeping humans in the loop points toward the problem addressed here (Mollick 2024, pp. 46–62). A Practical Ethics of Thinking, however, asks more specifically how humans can preserve authorship, judgement, and intellectual autonomy once AI becomes part of thinking itself.

The article answers this question by specifying and refocusing the concept of Human-in-the-Loop as Human-in-the-Loop-Thinking. Responsible-AI-use requires more than supervision of machine outputs; it requires continuous human engagement for contextual framing, epistemic judgement, and meaning. The proposed Practical Ethics of Thinking therefore emphasizes intellectual autonomy, epistemic vigilance, intellectual self-assertion, transparency, and non-delegable responsibility. Its central claim is straightforward: AI may participate in inquiry, but responsibility, knowledge and judgement cannot be delegated. The challenge is therefore not to

prevent AI from influencing human thought, but to ensure that this influence itself becomes an object of critical reflection.

2 GenAI and the Epistemic Shift

For quite some time, we have known that machines can work faster, harder, and often better than humans (Brynjolfsson and McAfee 2015, p. 15). However, academic work, research, science, and liberal arts have always been fields dominated by humans. With the rapid development and improvement of genAI, it is doubtful that this dominance can be sustained (Dessureault et al. 2025, pp. 6241–6243). Today, within academic and non-academic contexts, we use genAI for many different tasks like researching information, structuring our output, revising our texts, generating presentations and learning material, and summarizing literature. Some of these assignments work out perfectly, particularly as long as the authors have enough expertise about the subject at hand to check and verify its output. Every bad thesis submitted by a student who did not sufficiently revise and verify AI-generated material strengthens the following argument: although AI can support research and thinking in ways that academia has not experienced before, it still needs a human being to make sense of the machine's output.

Even though research is often collaborative, thinking retains an irreducibly individual dimension. Suitable interlocutors are not always available, nor does communication necessarily provide the kind of feedback needed at a particular moment of inquiry. GenAI potentially changes this condition by providing a continuously, immediately, and almost always available counterpart against which thoughts can be articulated, challenged, and reorganized. Here, my ideas resonate: they find a mirror, become condensed, and are questioned, which can help me to develop them further. In some sense, genAI makes my thinking - to myself - more audible, more visible, more revisable, and, in the best case, even more testable.

Knowledge is not simply a finished product; it emerges through a process of inquiry (Dewey 1910, p. v). Usually, it does not pop out of a person's head. The famous story of Archimedes (according to the Roman architect Vitruv in his work *De architectura libri decem*) suggests the opposite: before the "Eureka moment" came the problem, engagement with it, and the search for a solution. The "Eureka moment" was not the beginning of knowledge. It was the provisional conclusion of a process of experiencing, questioning, learning, and thinking.

This understanding of knowledge appears very helpful in order to understand how genAI can enrich - but also endanger - research processes and the generation of academic knowledge. John Dewey's explanation that knowledge results from an active, cyclical process where real-world experiencing triggers a natural sense of questioning when faced with a problem suits an approach that characterizes AI primarily as a tool to reflect thought processes critically and resonate ideas and conclusions. By actively thinking through challenges and testing solutions, individuals undergo meaningful learning that transforms raw experience into lasting understanding. Osawu Tunde relates this idea directly to AI: "[T]he pragmatist philosophy of John Dewey emphasises how technology can improve human inquiry when it is employed as a tool for experiential learning." (Tunde 2025, p. 88)

Experience, reflection, and correction must remain at the core of a researcher's thinking and use of AI. Tunde argues that Dewey's pragmatic epistemology can help to incorporate AI into reflective and critical processes of inquiry (Ibid., p. 91). However, AI does not thereby generate knowledge or truth on its own. Rather than merely assuming routine and supportive tasks, it can play a resonating and dynamic role in the development of knowledge by making trains of thought explicit and opening them to dialogical reflection.

From this point of view, the implementation of genAI as a tool for academic work can transform researching and writing academic papers from a purely end-product-oriented task into a continuous process of development, testing, and correction. AI can help to mirror ideas, question conclusions, and condense findings. It offers interaction and resonance. Human researchers, however, have to assemble, contextualize, and judge the outcome. This is what we commonly call the Human-in-the-Loop. But is this enough?

3 Human-in-the-Loop and Epistemic Responsibility

Researchers must always be aware that genAI is no authority of insight. Particularly when using AI to produce text, caution is essential. AI-output can sound alluringly convincing. It creates appealing plausibility and structures thoughts in a persuasive manner. Therein lies the danger of what this paper has called the Illusion of Evidence: well-written words can quickly appear correct. However, a well-expressed text is not yet a good thought, let alone a breakthrough in research.

New insights must always go through a process of testing. This may involve other researchers, methods, and AI systems; epistemic responsibility for judging whether a claim is sufficiently justified, however, cannot simply be delegated to the machine. Again, Tunde finds a suitable summary for this approach: "People can preserve their epistemic agency by viewing AI results as testable hypotheses rather than unquestionable facts. But this necessitates a degree of digital literacy and philosophical understanding that is not available to everyone, underscoring the necessity for educational systems to give critical thinking abilities top priority in the AI era." (Tunde 2025, p. 91). While this paper agrees with the basic argument, it questions whether all AI users need a deep philosophical understanding. A Practical Ethics of Thinking instead aims to provide five simple and applicable principles that can guide researchers in dealing with the epistemic risks.

So, let us take a pragmatic look at this phenomenon. Imagine the following Gedankenexperiment which I have discussed with my students:

> *You develop your own thesis for a paper. You ask an AI system to provide three counterarguments to challenge your thinking. One of them sounds particularly convincing. You incorporate it into your reasoning and revise your position.*
>
> *Later, you realize that the argument was logically coherent but based on a questionable assumption.*
>
> *Now ask yourself: Did the AI help you think better—or did its linguistic persuasiveness influence your judgement?*

None of these epistemic risks are entirely new. Human interlocutors can also provide convincing but flawed arguments. GenAI, however, combines immediate and continuous availability with linguistic fluency, opaque epistemic origins, and the ability to generate persuasive responses at a scale and speed that can make such influence a permanent feature of the research process.

In this regard, awareness is as imperative as knowledge and expertise when using AI for academic work. We must be aware that genAI is trained on vast amounts of data, perspectives, and interactions - even interactions with ourselves. The quality and character of the data is a major factor, but to the user, this process is neither transparent nor comprehensible (Horwath 2022, pp. 73-74). At the same time, LLMs have a documented tendency toward sycophancy, prioritizing agreement with users and validating their assumptions or beliefs even where critical disagreement would be warranted (Liu et al. 2026). Therefore, users must actively demand honesty, critique, and transparency from the tool. Praise and agreement from the AI shall not be

confused with applause. When used correctly, the machine can help us to develop new perspectives and make our thinking more precise. Simultaneously, blind adoption of the AI's resonance can lead to misunderstanding by adding misfitting perceptions or introducing wrong assumptions. Such 'imitation of knowledge' can lead to the aforementioned Illusion of Evidence. The added value of AI in this regard is not blind validation of our thinking or agreement with our conclusions. It rather results from the machine's examination of our process of thinking, its structure, and conclusion. But only as long as the researcher critically reflects the reflections, compares it to and contextualizes it with his/her findings. Just under this condition AI can help to generate knowledge. Yet it can neither substitute for judgement nor provide human justification, and it cannot free researchers from their responsibility to justify their claims.

Luciano Floridi frames such a conclusion. He declares information- and communication technologies (ICTs) like AI to be more than a mere tool. His approach resonates with this paper's assumption that AI influences how we perceive information, what appears plausible as well as relevant. According to Floridi, ICTs influence our perception of reality. In this sense, AI and similar technologies can be understood as an epistemic infrastructure (Infosphere) shaping a space where thinking and knowledge can develop and arise (Floridi 2015, pp. 63-64). Concerning the train of thought of this paper, following Alan Turing's concept, Ruth Hagengruber specifies how knowledge is created and gains its character as knowledge and truth. For this to happen, it needs human and cultural recognition (Hagengruber 2017, pp. 334-336).

Does this mean that using AI for research makes knowledge creation a co-production between humans and machines? Ethan Mollick introduced the concept of Co-Intelligence in order to describe the collaboration of human and AI. He conceptualizes co-intelligence as a collaborative relationship in which AI becomes a thinking partner or co-worker rather than merely a tool for automation. In this dynamic, AI enhances human capabilities and shares cognitive tasks, while humans supply the necessary context, strategic direction, and ethical oversight (Mollick 2024, pp. XIX-XX and 46-62).

However, this paper argues that the connection of Floridi's, Hagengruber's, and Mollick's approaches, while valid and important, is not sufficient to fully frame the epistemic dimension of human-machine collaboration. Mollick already acknowledges the processual character of this collaboration and emphasizes the need to keep humans in the loop. Floridi does so as well, but frames human responsibility within a broader social and ethical context. Yet responsibility does not merely mean remaining involved throughout the process. It begins when AI-generated suggestions enter the process of inquiry and start shaping what the researcher considers plausible, relevant, or worth pursuing. The deeper AI enters the process of thinking, the earlier human responsibility must begin.

4 Towards a Pracitcal Ethics of Thinking

Realizing a thought, generating knowledge, and judging whether a claim can be considered true are ultimately human tasks. Even if we assume that AI can contribute to these processes, the machine does so only on the foundation of human-fed information, human-led training, and human-bred contexts. Ultimately, humans decide what counts as knowledge, while societies and scientific communities determine what can be recognized as justified knowledge or accepted as truth. Hagengruber emphasizes precisely this social and cultural dimension of knowledge: knowledge is embedded in a "cultural network of knowledge confirmation" and emerges through context-dependent processes of acceptance and rejection (Hagengruber 2017, p. 336).

However, we must take a closer look at the individual researcher. Researchers themselves must add judgement, interpretation, and contextualization to the process. Doing so requires expertise in the respective field and knowledge of the subject under investigation. GenAI can support processes of insight, but whatever plausibility the machine may create, it cannot substitute rational reasoning and human judgement. Epistemic tasks can be delegated; epistemic responsibility for accepting, interpreting, and defending their results cannot. In this sense, researchers must accept this obligation - not only for the final outcome, but right from the beginning of the research process and throughout their use of AI.

Such vigilance does not only concern the process of thinking and drawing conclusions, but also the process of writing. LLMs may appear to be marvelous authors, but they have their own style, rhythm, and recurring patterns of expression. If researchers want their own style to remain recognizable, they must make sure that the narrative is still told in their own voice. Blindly adopting AI-generated text also carries the risk of adopting alien patterns of thinking and expression. One important question which authors therefore have to ask themselves when using AI and adopting its output is: Does this still sound like me, and are these still my own thoughts translated into written form? We need to maintain a critical distance from the machine.

Authorship means that researchers are able - and willing - to reason and substantiate their own thoughts. They must be able to explain them, develop them further, and defend them. Intellectual hygiene in this regard means that we reflect frequently, that, as we express our ideas ourselves, and that the arguments we ultimately make are genuinely our own. In order to retain authorship, researchers should remain capable of producing the initial intellectual formulation of their argument themselves - at least its core.

This paper offers some simple guidelines for how researchers can maintain authorship and intellectual self-assertion. Within a Practical Ethics of Thinking, these can be summarized in five principles. This number may change as the framework develops further. This paper also does not claim that the individual principles are entirely new or previously unknown. Its contribution lies rather in bringing them together within the specific context of genAI-supported research processes.

The first principle is *intellectual autonomy*. GenAI can function as a space of resonance and hypothesis testing. It may even suggest a hypothesis itself, but it must not become the origin of an argument that the researcher adopts without intellectual appropriation and independent judgement. There is a difference between letting AI help you to test your ideas, sharpen them, and relate certain findings to each other, or to let it do all of these things unchecked. Just ask yourself: Is using the machine strengthening or weakening my own thinking? When doing research, you may “outsource” certain routine assignments but you cannot delegate thinking and judgement.

Drawing on Sperber et al. (2010), the second principle is *epistemic vigilance*. This paper adapts the notion of epistemic vigilance to human–AI interaction. Whereas the concept originally concerns vigilance toward communicated information and its sources, here it denotes the researcher's deliberate scrutiny of AI-generated claims, arguments, and apparent evidence. As we learned, plausibility is not truth or secured knowledge. Do not confuse these concepts and ask yourself whether you would come to the same conclusion and express it the same way as the AI if you based your work exclusively on literature and data you read yourself! The Illusion of Evidence can be a real challenge for researchers. In order to preserve your own voice, treat AI-output as a proposition, not as evidence.

As discussed above, the third principle has already been established: *intellectual self-assertion*. Due to the allure of AI generated arguments in plausible text form, it is easy to ignore your own

expertise. There is a real danger that we underestimate and forget the importance of the contexts and relations between our own thoughts. We blindly follow the machine's narrative, arguments, and conclusions. However, as long as we can trace the AI's suggestions back to their origins and are able to explain and defend every adopted argument, we maintain contextual control and, ultimately, authorship.

Transparency, as the fourth principle, should be an implicit commitment rather than an external requirement. Researchers should be open about the use and the extent of AI use in their output - hence, I will gladly admit to have used AI in support of writing this paper. But transparency is also something that can serve the process itself and not only a final output. Bind the AI-system you are using to this commitment. If the machine is feeding you information, demand transparency about its sources and the degree of certainty it assigns to the information provided. Such indications of uncertainty must themselves be treated critically; they are signals for further verification, not reliable measures of epistemic certainty. Thus, you still have to check - and sometimes even double-check – those information, but making uncertainty explicit supports the process by generating cause for further examination.

We already know the fifth and final principle: *responsibility is not delegable*. This is not only true for the outcome of a research process which is being supported by genAI. This is true right from the beginning. Researchers must be aware that they remain responsible throughout the entire process. Always ask yourself if the ideas and thoughts are still your own and not watered down or completely changed by the machine without your deliberate and conscious approval. Are you positive that every step along the road represents your train of thought? Do they fit into the framework of normative values that defines your research process? As long as your arguments reflect your interpretation, and as long as you can defend those, you do not merely assume responsibility; you exercise it.

In practice, these principles translate into five simple actions, operating on different levels and at different stages of the research process. To preserve *intellectual autonomy*, produce the core of a self-authored draft. *Epistemic vigilance* requires treating AI output critically and verifying uncertain claims. *Intellectual self-assertion* means tracing adopted arguments back to their origins and judging them through one's own expertise. *Transparency* requires both disclosing AI use and demanding clarity about sources and uncertainty during the process. Finally, *non-delegable responsibility* means keeping human judgement as the reference point throughout the whole process: researchers remain accountable for ensuring that their work is consistent with their own reasoning, standards, and values.

Thus, let us return to the Gedankenexperiment introduced above. The researcher asked the AI for three counterarguments, found one of them convincing, and changed the original position - only to realize later that the argument rested on a questionable assumption. Applying a Practical Ethics of Thinking would not prohibit such an interaction with AI. Quite the opposite: asking the machine to challenge an argument can be a valuable form of epistemic resonance. However, the five principles change how the researcher deals with possible answers. *Intellectual autonomy* requires that the original argument remains the researcher's own point of departure. The researcher cannot unreflectingly adopt the AI-generated counterargument. *Epistemic vigilance* means treating the counterargument as a proposition to be tested rather than as evidence. Testing it by other means could uncover its flaw before it becomes a problem. *Intellectual self-assertion* requires the researcher to understand, contextualize, and independently judge the argument before incorporating it into the reasoning. That would also serve as a true test on the basis of his or her own expertise.

Transparency would, on the one hand, mean openly disclosing the use of AI and acknowledging it as a possible source of error. On the other hand, it would require inspecting the AI's sources as well as examining which sources, assumptions, or evidence can independently support the counterargument, and how certain the underlying information actually is. Finally, *non-delegable responsibility* means that the researcher must make the decision whether the argument is convincing enough to change the original position. It is her/his judgement that matters in the end. If he or she is wrong, it is his or her responsibility. The crucial question of the Gedankenexperiment can therefore be answered more precisely: AI may indeed help us to think better - but only if its linguistic persuasiveness and other factors do not replace our own judgement. The purpose of a Practical Ethics of Thinking is not to prevent AI from influencing our thinking. It is to ensure that such influence becomes an object of reflection itself and that the human researcher remains the author of the resulting thought, knowledge, and judgement.

5 Conclusion

This paper has argued that the integration of genAI into academic work is more than a technological change. Once AI becomes a space of resonance in which ideas are challenged, reorganized, and developed, it also becomes part of the process through which researchers form judgements. This creates opportunities for reflection and inquiry, but also epistemic risks such as the Illusion of Evidence. A Practical Ethics of Thinking therefore begins before the final output: researchers must preserve intellectual autonomy, remain epistemically vigilant, assert their own thinking, demand transparency, and accept that responsibility for knowledge and judgement cannot be delegated. Human-in-the-Loop must, in this sense, become Human-in-the-Loop-Thinking.

This aspect may become even more important as AI capabilities continue to develop. Dessureault et al. identify the alignment of increasingly capable AI with human values as a central ethical challenge and warn that the emergence of artificial superintelligence could make maintaining alignment with human intentions increasingly difficult (Dessureault et al. 2025, p. 6260). A Practical Ethics of Thinking approaches a structurally similar problem from the opposite direction and on a much smaller, but already relevant scale: it asks not only whether AI remains aligned with human values and intentions, but whether humans remain sufficiently aware of their own values, intentions, and judgements while interacting with AI. Before we can ask how machines should remain aligned with us, we should therefore also ask how we can remain intellectually self-assertive when thinking with them.

Beyond the individual level addressed here, one structural condition deserves at least a brief outlook: the future of academic work. Metcalf warns - at least for the humanities - that the ability of LLMs to mass-produce technically sound but incremental research may intensify existing academic incentives toward quantity and thereby crowd out more radical innovation (Metcalf 2026, p. 16/23). This risk is real for every field of study. If AI merely enables academics to produce more papers in less time, the old logic of publish or perish may become even more problematic. The technological transformation of academic work should therefore be accompanied by a cultural one: from "publish more" to "think better".

Yet this also points toward a more promising scenario. If genAI assumes routine and preparatory tasks, provides structure, and serves as a critical space of resonance, it may give researchers something increasingly scarce in academia: time to think. AI could then contribute not to replacing human intellectual work, but to creating better conditions for it. Whether this potential is realized ultimately depends on researchers and academic institutions themselves. The goal

should not be to produce as much knowledge as possible, but to preserve what academic inquiry is ultimately about: understanding, knowledge, judgement, and the courage to develop genuinely new ideas.

AI-Use-Disclosure: *Generative AI was used in the preparation of this paper as a tool for language refinement, structural feedback, and critical discussion of the argument; all substantive claims, interpretations, and final editorial decisions remain the author's own.*

References

1. Brynjolfsson, E./McAfee, A.: The Second Machine Age. Wie die nächste digitale Revolution unser aller Leben verändern wird (Plassen Verlag: 2015).
2. Dessureault, J.-S., Lamontagne, R., Parisé, P.-O.: "The Ethics of Creating Artificial Superintelligence: a Global Risk Perspective" in: *AI Ethics* 5, 6241–6263 (2025)
3. Dewey, J.: How We Think ([original] D. C. Heath & Co.: 1910), [used] Amazon Reprint.
4. Floridi, L.: Die 4. Revolution. Wie die Infosphäre unser Leben verändert (Suhrkamp: 2015).
5. Floridi, L.: The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities (Oxford University Press: 2023).
6. Hagengruber, R.: "Creative Algorithms and the Construction of Meaning" In: Pietsch, W., Wernecke, J., Ott, M. (eds.): Berechenbarkeit der Welt? (Springer VS: 2017).
7. Horwath, I.: "Algorithmen, KI und soziale Diskriminierung" in: Kordula Schnegg et al. (eds.), Inter- und multidisziplinäre Perspektiven der Geschlechterforschung (Innsbruck University Press, 2022).
8. Liu, J. et al.: When Flattery Backfires: How Sycophancy and Interaction Context Shape Perceived Authenticity and Trust in Large Language AI Models [Extended Abstract] (2026; https://dl.acm.org/doi/10.1145/3772363.3798575; last access 28.09.2026).
9. Metcalf, T.: "AI-Produced Humanities Research: On the Dangers of Technical Incrementalism" in: Philosophy & Technology, Volume 39, April 2026.
10. Mollick, E.: Co-Intelligence: The Definitive, Bestselling Guide to Living and Working with AI (WH Allen: 2024).
11. Mühlhoff, R.: Automatisierte Ungleichheit: Ethik der Künstlichen Intelligenz in der biopolitische Wende des Digitalen Kapitalismus (2020; https://rainermuehlhoff.de/media/publications/m%C3%BChlhoff_2020_automatisierte-ungleichheit_DZPhil2.pdf; last access 28.09.2026).
12. Shao, A.: "New sources of inaccuracy? A conceptual framework for studying AI hallucinations" In: Harvard Kennedy School Misinformation Review, August 2025, Volume 6, Issue 4.
13. Sperber, D. et al.: Epistemic Vigilance (2010; https://dan.sperber.fr/wp-content/uploads/EpistemicVigilance.pdf; last access 28.09.2026).
14. Tunde, O.: "Ethical Reflections on the Impact of Artificial Intelligence on Human Epistemic Employment of Critical Thinking" in: Jos Journal of Humanities, Vol. 7, No. 1, June 2025.
15. Vitruv: De architectura libri decem [Zehn Bücher über Architektur] (Reclam: 1981).