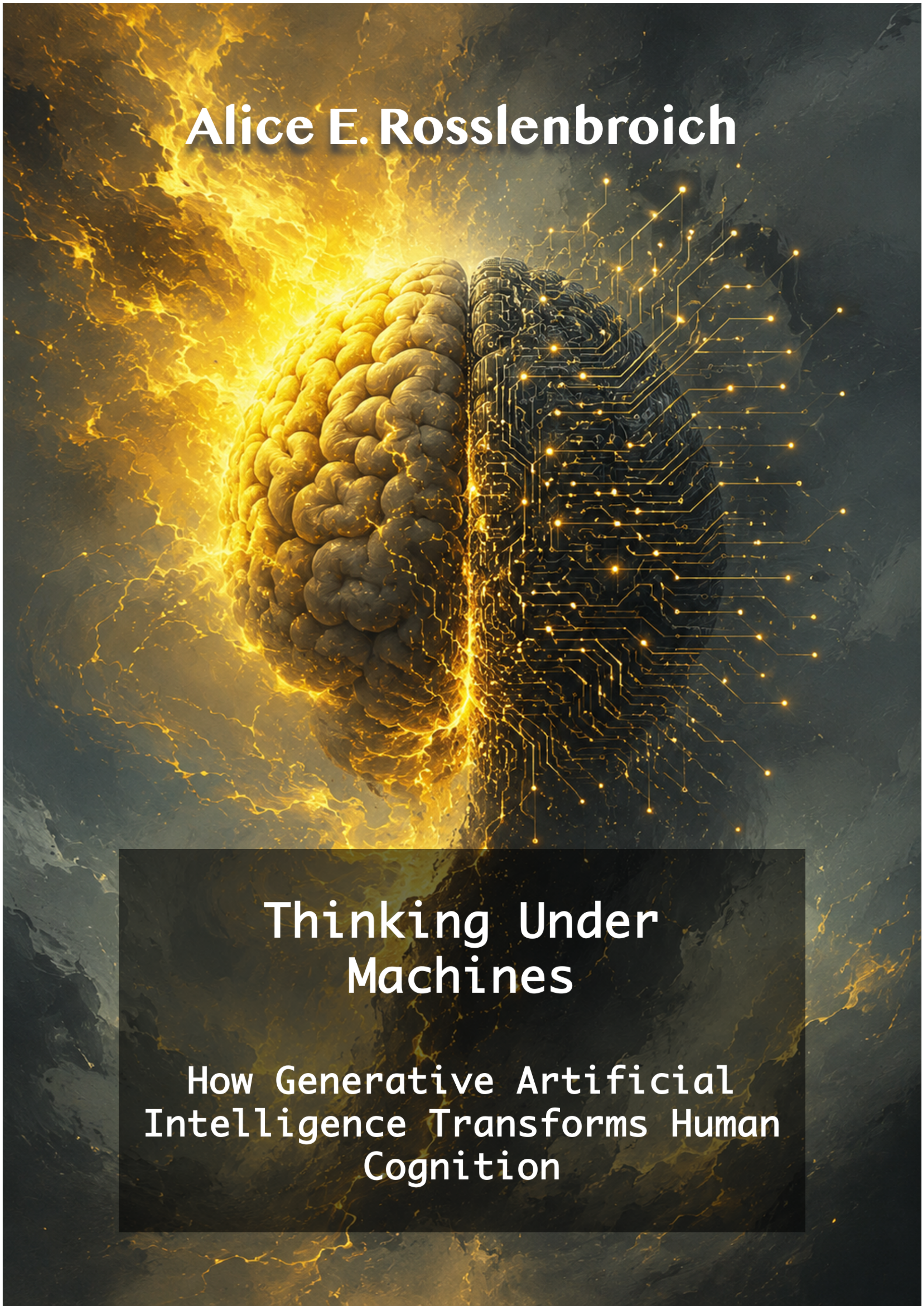
Alice E. Rosslenbroich
Thinking Under Machines
How Generative Artificial Intelligence Transforms Human Cognition

Introduction

The rapid spread of generative artificial intelligence does not mark an ordinary technological leap. For the first time, a tool is available that not only automates physical or administrative work but intervenes in areas of human cognition long considered genuinely intellectual achievements: formulating, structuring, researching, summarizing, arguing, and problem-solving.

Systems such as ChatGPT, Claude, or Gemini generate texts, analyses, and responses within seconds that convey the impression of intellectual and emotional competence in numerous everyday situations. This fundamentally changes the relationship between humans and machines. Digital systems no longer serve exclusively as tools for information processing but increasingly as linguistically addressable assistants for cognitive work.

The social relevance of this development extends far beyond technical questions. Educational institutions are debating the loss of independent writing and learning processes. Companies are integrating AI systems into knowledge work and decision-making structures. At the same time, concern is growing that the permanent outsourcing of mental processes could, over time, affect concentration, critical reflection, and independent judgment.

The history of technological innovation is simultaneously a history of cognitive relief. Writing reduced the necessity of oral memory. The printing press externalized knowledge on an unprecedented scale. Calculators automated mathematical operations; search engines reduced information retrieval to a matter of seconds. Each of these technologies changed the way humans store, process, and retrieve information.

Generative artificial intelligence, however, differs from earlier technologies in one decisive respect. Whereas classical digital tools primarily accelerated access to information, large language models generate responses, argumentative sketches, summaries, and formulations. They do not merely deliver data but present linguistically ordered results that frequently convey the impression of independent understanding.

It is precisely this linguistic form of problem-solving that makes the technology socially relevant. Interaction no longer takes place primarily through technical interface logic but through natural language. The human describes a problem; the machine generates a proposed solution. This substantially lowers the threshold for outsourcing intellectual work.

Previous research on the use of digital media has already suggested that technological systems can influence human memory and attention processes. The so-called Google Effect study showed in 2011 that people remember information less well when they expect to be able to retrieve it digitally at a later point. What was decisive was not so much a general loss of knowledge as a change in memory strategy: what was remembered was primarily the storage location, no longer the content itself.

Source: https://www.science.org/doi/10.1126/science.1207745

With generative AI, this form of cognitive outsourcing reaches a new quality. For the first time, not only information retrieval or memory performance but also linguistic, analytical, and structuring thought processes can be delegated to external systems. This brings a central question to the fore of educational and cognitive research: what are the consequences when humans are increasingly rarely compelled to actively engage in certain mental processes themselves? Learning, problem-solving, and critical reflection arise not exclusively through the finished result but substantially through the mental processing that takes place during the thinking and working process itself.

The neurosciences have for decades regarded the brain not as a static organ but as a dynamic system of continuous adaptation. Neuroplasticity describes the capacity of neural networks to change structurally and functionally through repeated use. Put simply, frequent activation can strengthen certain neural connections, while rarely used networks may be functionally weakened. Eric Kandel received the Nobel Prize in part for his work on signal transmission in the nervous system and on the biological foundations of learning and memory.

Source: https://www.nobelprize.org/prizes/medicine/2000/kandel/facts/

Against this background, the present study examines not primarily the technical performance of artificial intelligence but its influence on human cognition and on the neurobiological preconditions of learning. At its center stands the question of whether generative AI merely makes workflows more efficient or whether it changes the conditions under which the human brain remembers, learns, and thinks.

A differentiated scientific assessment is therefore necessary. For the assumption that AI use causes immediate neurological damage, no reliable long-term data currently exist. At the same time, current research indicates that intensive AI use can intensify processes of cognitive offloading — that is, the outsourcing of intellectual work. The decisive question is therefore not whether artificial intelligence is fundamentally harmful or beneficial but under what conditions it supports or replaces active mental processing.

1: Neuroplasticity and Cognitive Adaptation

The human brain's capacity for change forms the neurobiological foundation of every learning process. Modern neuroscience understands the brain not as a hard-wired structure but as a highly dynamic system whose neural networks continuously adapt to experiences, environmental stimuli, and patterns of use. This process is referred to as neuroplasticity.

Neuroplastic changes result from the repeated activation of specific neural connections. In 1949, in *The Organization of Behavior*, Donald Hebb formulated a principle of associative learning that remains fundamental to this day: When a neuron is repeatedly involved in the activation of another, the efficiency of that connection can increase. The popular shorthand "neurons that fire together wire together" is a later condensation of this idea, not Hebb's own wording. Repetition can stabilize neural networks; absence of use can, conversely, contribute to functional weakening.

Source: https://pure.mpg.de/pubman/item/item_2346268_3/component/file_2346267/Hebb_1949_The_Organization_of_Behavior.pdf

Learning is therefore not an abstract mental process but a biological process of adaptation. Attention, memory formation, linguistic processing, and analytical thinking are all grounded in repeated neural activation. Particularly relevant in this context is active cognitive effort. Studies in learning research have shown for decades that lasting knowledge consolidation is closely linked to depth of mental processing. Information is stored in long-term memory especially when people independently structure, formulate, retrieve, and critically process content.

The Cognitive Load Theory of psychologist John Sweller describes, in this connection, the limited processing capacity of working memory. Learning arises not from information intake alone but from active processing within this limited capacity. If a problem is fully solved externally, the cognitive load that would, under certain conditions, be productive for deeper learning may be reduced. At the same time, it is important to note that excessive stress can also hinder learning processes; the key issue, therefore, is not stress itself, but rather finding the right balance.

Source: https://doi.org/10.1207/s15516709cog1202_4

These connections are particularly relevant to the use of generative AI. Systems like ChatGPT not only streamline routine work processes but can also handle linguistic and analytical tasks. This creates a fundamental difference from earlier digital tools. While search engines primarily facilitate access to information, language models generate pre-structured responses, arguments, and formulations. Active mental processing can thereby be partially bypassed.

From a neuroscientific perspective, this is relevant because neural stabilization remains coupled to use. Brain research has shown for years that demanding cognitive activity can contribute to the formation of stable functional neural networks. Neuroscientist Stanislas Dehaene emphasizes, in his work on learning and reading research, that concentrated, active processing shapes the functional organization of the brain. Learning does not arise passively but through processes of focused attention, active engagement, feedback on errors, and the subsequent consolidation of new content. Reading, writing, and problem-solving are therefore not purely cultural practices but intervene in neural learning processes.

Source: https://www.penguin.co.uk/books/310169/how-we-learn-by-dehaene-stanislas/9780141989303

Cognitive scientist Maryanne Wolf advances similar arguments in the context of digital media use. In Reader, Come Home, she describes the danger of an increasingly fragmented mode of information processing in which deep reading could be displaced by faster and more superficial forms of reception. Wolf argues not from a technophobic position but from a neurocognitive one: the brain adapts to dominant patterns of use. If slow analytical processes are practiced less frequently, the corresponding skills could become less stable over time.

Source: https://www.harpercollins.com/products/reader-come-home-maryanne-wolf

A further perspective, increasingly attended to in cognitive science, supplements these considerations: the concept of embodied cognition. Proponents of this approach — among them George Lakoff, Mark Johnson, and Francisco Varela — argue that thinking is not exclusively a neural process but is substantially shaped by bodily experience, motor activity, and sensory feedback. Cognition arises, in their view, not in the brain alone but in the interplay of body, brain, and environment.

This is relevant to the question of the impact of generative AI for a specific reason: Writing is not only a mental process but also a physical one. The motor skills involved in writing—especially handwriting—are closely linked, at the neural level, to learning and memory processes. A widely cited study by Pam Mueller and Daniel Oppenheimer from 2014 showed that students who took notes by hand processed content more deeply and retained it better than those who typed — presumably because handwriting compels independent selection and reformulation, while typing favors faster but more superficial transcription. When linguistic production is increasingly delegated to generative systems, what is lost is not only the mental but also the bodily-motor dimension of formulation — an aspect that has received little attention in the current AI debate.

These reflections on the embodied dimension of cognition underscore that the question of the effects of generative AI cannot be reduced to neural processes in the narrow sense alone.

At the same time, it would be scientifically untenable to derive from these mechanisms any claim of immediate damage to the brain by artificial intelligence. Neuroplasticity is not an exclusively negative process. Technological tools can expand cognitive capacities as readily as they reduce them. Numerous studies on computer-assisted learning support show that interactive digital systems can improve learning outcomes, provided users remain actively engaged in problem-solving processes.

The decisive variable is therefore not the existence of artificial intelligence itself but the degree of independent cognitive engagement. From a neuroscientific perspective, what matters is not whether people use tools but which mental processes they continue to carry out themselves in the process.

2: Cognitive Offloading and the Outsourcing of Human Intellectual Performance

The use of external aids to relieve mental processes is not a new phenomenon. Writing, libraries, and later digital storage systems have long enabled people to store information outside biological memory. Modern cognitive psychology refers to this process as "cognitive offloading." This refers to the transfer of cognitive tasks to external systems in order to reduce the load on memory, attention, or working memory.

Cognitive offloading is not in principle a pathological process. On the contrary, the capacity for intelligent information outsourcing is regarded as a central component of human cultural development. The process becomes problematic where not merely storage or routine functions are outsourced but active thought processes are taken over by external systems and thereby trained less frequently.

An early empirical indication of changed memory strategies was provided by the already-mentioned Google Effect study by Betsy Sparrow, Jenny Liu, and Daniel Wegner from 2011. The researchers demonstrated that test subjects remembered information less well when they believed they would be able to retrieve it digitally at a later point. Memory focused more on the storage location than on the content itself.

This study is relevant to the current debate on AI because it shows that digital tools not only change access to information but can also influence memory strategies. However, it does not prove that digital systems generally erode knowledge; rather, it describes a specific change in memory under certain experimental conditions.

In 2016, Evan Risko and Sam Gilbert described cognitive offloading as a fundamental form of distributed cognition. People use their bodies, the environment, and technical artifacts to reduce internal cognitive load. Perceived difficulty and metacognitive assessment play a central role here: the more demanding a task appears, the more readily people turn to external aids.

Source: https://doi.org/10.1016/j.tics.2016.07.002

In everyday digital life, this principle has gained new reach. Smartphones, search engines, and navigation systems take over tasks that previously required active recall or orientation. A 2017 study by Adrian Ward, Kristen Duke, Ayelet Gneezy, and Maarten Bos found indications that the mere presence of one's own smartphone can reduce available cognitive capacity — a "brain drain". This finding, too, should be read with caution: what was examined was a specific experimental setting, not a general or irreversible change in attention.

Source: https://www.journals.uchicago.edu/doi/10.1086/691462

Generative artificial intelligence extends cognitive offloading to complex linguistic and analytical processes. While earlier digital systems primarily provided information, large language models generate independently formulated responses, argumentative structures, and problem solutions. This produces a qualitatively new form of cognitive relief: outsourced are not only memory or search but parts of the process in which a thought is linguistically ordered and argumentatively examined.

This is relevant for learning and knowledge processes because lasting understanding depends not only on the outcome. The independent formulation of a thought, the structuring of an argument, or the independent solution of a problem generate different cognitive demands than the mere reception of finished answers. When a language model produces complete essays, summaries, or analyses, precisely the mental effort that is significant for deeper understanding may be reduced.

Current research suggests that this effect becomes particularly relevant when high levels of trust in AI systems develop. A 2025 study by Michael Gerlich, published in Societies, examined the relationship between AI use, cognitive offloading, and critical thinking. The study found a negative relationship between intensive AI use and critical reflective capacity, as well as a possible connection with increased cognitive offloading.

The evidential value of this study is, however, limited. The results are predominantly correlational. It cannot be directly inferred from them that artificial intelligence directly reduces critical thinking. It is equally conceivable that persons with a lower propensity for analytical reflection turn to AI systems more frequently, or that further factors such as level of education, media literacy, workload, or purpose of use play a role.

Source: https://www.mdpi.com/2075-4698/15/1/6

The study by Microsoft Research and Carnegie Mellon University likewise describes no general cognitive deterioration through generative AI. The authors analyzed the use of generative AI in knowledge work and found that human cognitive engagement can shift. Users tended, particularly when placing high trust in AI-generated content, to carry out fewer independent verification and control processes. The study examines primarily self-reports and examples from knowledge work; it does not permit conclusions about long-term neurological changes.

Source: https://www.microsoft.com/en-us/research/wp-content/uploads/2025/01/lee_2025_ai_critical_thinking_survey.pdf

Cognitive offloading should therefore not be understood as an inherently harmful process. External relief can free cognitive resources and enable more complex problem-solving. Historically, many scientific and social advances have rested on forms of outsourced cognition — from writing to computer technology.

The key question, therefore, is which cognitive processes are being outsourced, and which continue to be actively exercised. If routine tasks are automated, this can support higher-level mental performance. If, on the other hand, fundamental skills such as analysis, structuring, or critical evaluation are permanently replaced, this could lead to a reduction in active cognitive practice over the long term.

Generative AI thereby shifts the historical dimension of technological relief. It not only stores information and accelerates access to knowledge but simulates linguistic and analytical sub-steps of human cognition. The question is therefore no longer merely how quickly people find information but to what extent they continue to perform the intellectual work of structuring and examination themselves.

3: Generative Artificial Intelligence and the Transformation of Cognitive Processes

Generative language models extend the principle of cognitive offloading to a new level. While earlier digital systems primarily supported information retrieval, data storage, or repetitive workflows, large language models intervene directly in linguistic and analytical thought processes. Texts, lines of argument, summaries, and problem solutions can today be produced in partly automated fashion — precisely those forms of mental processing that have hitherto been closely linked to independent thinking and learning.

Language is, in this context, not merely a medium of communication. In many learning and work processes, it is itself part of the cognitive performance. To formulate is to select, weigh, connect, justify, and examine. When this process is partly delegated to a language model, what changes is not only the mode of producing a text but the distribution of cognitive work between human and system.

This is particularly visible in writing. An original text normally arises through several cognitive sub-steps: understanding the problem, research, selection, structuring, formulation, revision, and self-examination. Generative AI can take over individual or multiple steps of this process. The human remains the decision-maker but is not necessarily required to complete all intermediate steps personally.

This shift is not necessarily a bad thing. In professional contexts, it can speed up research processes, generate drafts, or reveal alternative lines of argumentation. It becomes problematic, however, when the generated text is used not as a working document but as a substitute for one's own analysis. In this case, it not only reduces writing time but may also diminish the metacognitive control that arises during the process of drafting and revising.

A widely discussed study from the Massachusetts Institute of Technology in 2025 attempted to address this question experimentally. The work, published under the title Your Brain on ChatGPT, used electroencephalographic measurements (EEG) to analyze the brain activity of individuals writing essays with and without the support of generative AI. Participants were assigned to, among others, an LLM group (Large Language Model), a search engine group, and a group without external aids.

The researchers reported indications of reduced functional connectivity during the use of ChatGPT. In addition, participants later partially recalled self-authored content less well and described a reduced sense of personal authorship. These findings are consistent with the assumption that strongly automated text production can be accompanied by reduced active processing.

The evidential value of the study is, however, limited. It was initially available as a preprint without a completed peer review process, the sample was comparatively small, and the duration of the investigation was brief. A science-journalistic assessment on the platform Nature also warned against over-interpreting the results. Neither long-term neurological changes nor lasting cognitive damage can be derived from the data.

Source: https://arxiv.org/abs/2506.08872

Additional scientific assessment: https://www.nature.com/articles/d41586-025-02005-y

Despite these limitations, the study is relevant to the debate. It provides no proof of damage to the brain but an early indication that different writing conditions can be accompanied by different degrees of cognitive engagement. It is precisely this question — not a blanket claim about "AI and brain damage" — that is scientifically tractable.

Evidence of changed working and thinking processes also emerges outside neuroscientific laboratory studies. The already-mentioned study by Microsoft Research and Carnegie Mellon University analyzed how knowledge workers use generative AI in real tasks. The authors found indications that critical thinking can partly shift from actual problem-solving toward the evaluation of machine-generated results. At the same time, at high levels of trust in AI-generated outputs, reported self-verification activity decreased.

This shift is of particular significance for educational processes. Learning requires not only access to correct solutions but the active construction of understanding. Someone who traces a line of calculation, formulates a thesis, or orders a text argumentatively trains different capacities than someone who merely adopts a finished result.

Empirical educational research presents a mixed picture in this regard. A 2025 study published in Scientific Reports, examining an AI-assisted tutoring system, found that students working with the AI tutor in a specific setting learned more and required less time than in a comparison condition with active in-person teaching. The authors emphasized, however, that the tutor had been designed according to pedagogical best practices. The study therefore does not speak for arbitrary AI use but for carefully designed, interactive learning environments.

Source: https://www.nature.com/articles/s41598-025-97652-6

A 2025 study published in the Proceedings of the National Academy of Sciences (PNAS) on the use of GPT-4 in mathematics instruction presents the other side. Pupils achieved better performance in the short term during the use of GPT-4. When access was subsequently withdrawn, those who had previously been most reliant on AI-generated solution paths performed worse. The authors interpret this as an indication that generative AI can improve performance in the short term, but that sustainable skill acquisition depends on whether learners remain actively engaged in the problem-solving process.

Source: https://www.pnas.org/doi/10.1073/pnas.2422633122

These studies yield no uniform verdict on generative AI. Rather, a distinction becomes necessary: AI can promote learning processes when it guides questioning, examining, reasoning, and step-by-step understanding. It can impede learning processes when it replaces central thinking operations and learners merely adopt results.

The long-term neurocognitive consequences of this development have not yet been conclusively researched. The current state of research permits neither alarmist claims about a general intellectual decline through AI nor technological euphoria about unlimited cognitive expansion. The empirical evidence points to a differentiated picture: generative artificial intelligence changes the conditions of human thought processes. Whether the outcome is cognitive weakening or expansion depends substantially on the form of human participation.

4: Attention, Reward Systems, and Digital Efficiency

The effect of generative artificial intelligence on human cognition cannot be explained through learning and memory processes alone. Equally relevant is the question of how AI systems influence fundamental mechanisms of attention, motivation, and cognitive effort. Modern neuroscience understands thinking not as an unlimited resource but as a biologically costly process. Concentration, problem-solving, and analytical reflection are accompanied by increased neural energy consumption.

The human brain evolved under conditions of scarce resources. An evolutionary tendency toward efficiency optimization therefore emerged. In many situations, people prefer those behaviors that yield usable results with the least possible cognitive effort. It is precisely at this point that generative AI systems meet fundamental mechanisms of human information processing.

Psychologist Daniel Kahneman describes two distinct forms of human information processing in his research. One operates quickly, intuitively, and with comparatively low cognitive effort, while the other encompasses slower, analytical, and consciously controlled thought processes. Complex problem-solving, critical examination, and deep understanding rest substantially on this second, reflective form of processing.

Source: https://us.macmillan.com/books/9780374533557/thinkingfastandslow

Generative AI reinforces this dynamic because it makes analytical work seemingly effortlessly available. The user receives summaries, arguments, or problem solutions within seconds without having to pass through extended phases of concentrated processing. This lowers the threshold for externalizing complex mental processes.

Neuroscientific research shows that dopaminergic systems play a central role in motivation, expectation processing, and reward learning. Rapid problem solutions and reduced uncertainty can therefore be experienced as cognitively relieving and immediately rewarding.

Generative artificial intelligence transfers the logic of digital efficiency for the first time directly to human thought processes. It is no longer only entertainment or social feedback that is accelerated but writing, structuring, and problem-solving. The technology reduces what may be called cognitive friction — the mental resistance traditionally associated with deep analysis, uncertainty, and intellectual effort.

It is precisely this friction that is significant for learning and knowledge processes. Learning psychology describes, under the concept of *desirable difficulties*, productive challenges that can promote lasting knowledge consolidation. Robert Bjork and colleagues argue that demanding mental processing frequently leads to more stable memory traces than frictionless information intake. The strength of this effect, however, depends strongly on context, difficulty, prior knowledge, and individual learning preconditions.

Source: https://bjorklab.psych.ucla.edu/research/

A related aspect, hitherto little discussed, concerns the psychological dimension of cognitive dependency. Psychologist Albert Bandura described, with the concept of self-efficacy, the subjective conviction of a person that they can accomplish specific tasks through their own efforts. This conviction arises substantially through one's own experiences of success — that is, through the repeated experience that demanding problems can be solved through one's own exertion.

Generative AI can influence this mechanism in both directions. On the one hand, it can enable experiences of success by lowering barriers and guiding learners to results they could not have achieved alone. On the other hand, there is the risk that people who repeatedly experience AI systems solving tasks more quickly, more comprehensively, or with more linguistic precision than they themselves may, over the long term, lose confidence in their own cognitive capacities. A weakened sense of self-efficacy can, in consequence, reduce motivation, willingness to learn, and readiness to engage independently with difficult problems — regardless of whether actual cognitive capacity has in fact diminished. The psychological effect of perceived machine superiority can thereby produce real behavioral consequences that extend beyond purely neurocognitive processes.

Against this backdrop—and beyond purely neurocognitive processes—a key tension in the modern use of AI emerges: technologies that make thinking more efficient can simultaneously reduce the cognitive effort required for deep learning under certain conditions.

This development is particularly evident in how modern digital media engage human attention. Even before the rise of generative AI, neuroscientists and media researchers were debating the effects of permanent digital stimulus systems on concentration and information processing. Nicholas Carr argued, in The Shallows, that intensive internet use could promote fragmented patterns of attention. This thesis is influential but should not be treated as a conclusive neuroscientific consensus; it constitutes rather an important contribution to the debate on digital reading and attention practices.

Source: https://wwnorton.com/books/the-shallows

Maryanne Wolf also describes, in her research on reading cognition, that deep reading activates complex neural and cognitive processes that support analysis, perspective-taking, and abstract thinking. Shortened digital forms of reception could, over the long term, weaken these processes if intensive reading practice is increasingly replaced by fragmented information intake.

Generative AI could reinforce this development, though in a different way from social media. While social media can primarily fragment attention, language models present the possibility of increasing cognitive passivity: users receive not merely fragments of information but already synthesized thought products. The mental distance between question and finished answer shrinks considerably.

This poses new challenges, particularly in the field of education. Learning does not depend solely on correct answers, but on actively engaging with uncertainty, mistakes, and cognitive load. If this process is too strongly automated, the willingness to engage independently with complex problems and to sustain intellectual effort could, over the long term, diminish. Some educational debates describe this as a possible weakening of the capacity for critical and independent problem-processing ("epistemic resilience").

At the same time, a purely negative interpretation would not be scientifically tenable. Numerous neuroscientific and cognitive-psychological models proceed from the assumption that the brain tends toward energy-efficient information processing. From this it may not, however, be concluded that every form of technological relief is automatically cognitively harmful. Historically, cultural and scientific advances rest substantially on the use of external tools to extend human capacities.

The scientifically relevant question is therefore not whether people outsource cognitive processes, but which cognitive processes continue to be actively exercised. At present, there is only limited long-term data on how sustained use of generative AI affects attention, memory, or problem-solving over the long term. Several recent studies therefore focus more on potential risks and theoretical mechanisms than on neurological consequences that have already been conclusively demonstrated.

Herein lies the particular neurobiological relevance of generative AI. For the first time, a system is emerging that not only complements human cognition but simulates linguistic and analytical sub-steps. The balance between one's own mental activity and external problem-solving thereby shifts. Whether this produces, over the long term, cognitive expansion or intellectual dependency depends substantially on whether people use AI as a tool for active reflection — or as a permanent substitute for independent thought.

5: Education, Knowledge Work, and the Social Consequences of Cognitive Automation

The effects of generative artificial intelligence are not confined to individual learning or attention processes. As language models become increasingly integrated into education, the world of work, and information systems, the social organization of knowledge itself is changing. Particularly affected are those domains that have traditionally rested on independent analysis, linguistic processing, and critical reflection: schools, universities, the sciences, journalism, and knowledge-based professions.

In the educational domain, the effect of generative AI is especially visible. For decades, writing was regarded not merely as an examination performance but as a tool of intellectual development. The process of independent formulation compels learners to structure information, establish connections, and produce argumentative coherence. It is precisely these processes that can be partly automated by AI systems.

This creates a pedagogical problem. When pupils, students, or learners adopt finished answers, what is often lost is not only the writing work itself but also part of the underlying thought process. Educational research and education policy are therefore increasingly debating whether generative AI undermines traditional assignment formats and what new forms of assessment and learning are required.

UNESCO published guidelines in 2023 on the use of generative AI in education. The organization emphasizes that AI systems can support learning processes but that risks simultaneously arise for autonomy, data protection, equality of opportunity, critical thinking, and human responsibility in educational processes. What is decisive is a human-centered integration, not an unregulated transfer of learning and assessment processes to machines.

Source: https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research

Particular attention in this context is warranted by the question of the effects of generative AI on children and young people. In developmental-psychological and neurobiological terms, this group differs fundamentally from adult users. The prefrontal cortex, which is centrally responsible for planning, impulse control, critical judgment, and complex problem-solving, continues to mature until well into the third decade of life. Cognitive capacities that are not regularly and actively trained during this phase may be structurally less stably developed than in already-mature neural networks.

This does not imply a blanket warning against the use of AI by children and adolescents. It does, however, give rise to the necessity of a developmentally sensitive assessment: what can count as productive cognitive relief for adult knowledge workers may present itself differently for learners in critical phases of maturation. Precisely when fundamental capacities such as argumentation, structured writing, or analytical reading are to be acquired for the first time, the early outsourcing of these processes to generative systems can reduce the necessary practice of those capacities. Education policy and pedagogy therefore face the task of developing age-differentiated frameworks for the use of generative AI — beyond blanket prohibitions as well as beyond unreflective integration.

The Organization for Economic Co-operation and Development (OECD) is also examining, in its project Artificial Intelligence and the Future of Skills, how AI capabilities can be compared with human competencies in education and work. The focus here is less on immediate changes to the brain than on the question of which capacities will remain particularly relevant in the future and how educational systems should respond to the growing performance of AI.

Source: https://www.oecd.org/en/publications/ai-and-the-future-of-skills-volume-2_a9fe53cb-en.html

Current empirical studies partly support this concern. The already-mentioned PNAS study on the use of GPT-4 in mathematics instruction indicates that lasting competence acquisition depends substantially on whether learners continue to be actively engaged in the problem-solving process despite AI support.

Source: https://www.pnas.org/doi/10.1073/pnas.2422633122

This trend affects not only formal education but also, increasingly, knowledge work. Professions that have traditionally relied on research, analysis, and linguistic precision are currently undergoing a profound transformation.

The effects of generative AI on cognition are, in this context, not to be understood exclusively as an individual phenomenon. Thinking takes place in science, organizations, and democratic societies substantially as a collective process as well. Teams develop shared problem solutions; scientific communities generate knowledge through the contest of different perspectives; democratic publics form judgments through exchange and deliberation. Generative AI changes these collective processes too — in a manner that has to date scarcely been systematically investigated.

A first risk concerns the homogenization of thinking and argumentation. When large population groups use the same language models for analysis, research, and formulation, the diversity of perspectives produced could diminish. Language models are trained on similar data bases and tend toward statistically probable, hence often consensual, responses. This creates the danger of an algorithmically produced narrowing of the space of opinion and argument — a form of collective groupthink promoted not through social dynamics but through technical systems.

A second risk concerns democratic judgment formation. Democracy requires that citizens be able to independently verify information, weigh arguments, and form political judgments based on their own reflection. If this process is increasingly replaced or shaped by AI-generated frameworks of opinion, a new form of epistemic dependence emerges—not at the individual level, but at the societal level. The capacity for critical cognition thus becomes a question of relevance to democratic theory.

Journalism is not merely about producing written content. It involves evaluating sources, providing context, identifying contradictions, examining conflicting interests, and analyzing social power dynamics. Generative AI can simulate many of these processes linguistically, but it does not automatically replace their epistemological value.

That is precisely where the main danger lies. Language models generate plausible outputs without possessing genuine understanding or any sense of truth. The systems calculate linguistic probabilities and can in the process produce factually incorrect or invented statements. OpenAI describes such hallucinations as plausible but false claims that can occur even with apparently simple questions.

Source: https://openai.com/index/why-language-models-hallucinate/

This has significant implications for journalism and academia. When AI-generated content is adopted without verification, the responsibility for verifying and contextualizing that content increasingly shifts from the active researcher to the machine. At the same time, generative systems make it much easier to mass-produce superficial content.

Several media researchers therefore warn of a possible industrialization of synthetic information: a digital environment in which large quantities of linguistically convincing but substantively uncertain content are produced. In such systems, critical source examination becomes not less but more important.

A further aspect should be noted: generative AI changes not only the production of knowledge but also the human relationship to authority and the credibility of information. Historically, knowledge and expertise were associated with human experience, training, and institutional credibility. Language models, by contrast, generate the impression of competent authority independently of actual verification.

International educational organizations are increasingly viewing AI literacy as a fundamental skill for the future. What is meant is not merely technical proficiency in using AI systems but the capacity to critically examine machine-generated content, recognize uncertainties, and assess the limitations of algorithmic systems.

This demand for AI literacy gains additional urgency when the global distribution of AI access and media literacy is brought into view. Generative AI is currently being developed in ways that are neither technically nor educationally equitably accessible. High-quality language models, powerful infrastructure, and critically reflective competencies in use are at present concentrated predominantly in economically strong countries and privileged population groups. This creates a new dimension of digital inequality — no longer merely in the sense of unequal access to information but in the sense of unequal cognitive tools and the capacity to use them critically.

Two countervailing risks should be distinguished. On the one hand, countries and population groups with good access to powerful AI systems could develop knowledge and productivity advantages that exacerbate existing global inequalities. On the other hand, in contexts with poorly developed media and AI literacy, there is a danger that generative systems are used less critically and reflectively — with correspondingly greater risks to epistemic autonomy and educational quality. UNESCO and the OECD emphasize, in this connection, that a human-centered AI integration must not ignore questions of global justice. The cognitive dimension of generative AI is thereby simultaneously a question of distributive politics.

In this context, the OECD *Digital Education Outlook 2026* highlights a potential "AI learning paradox": While generative AI can lead to better performance and more efficient results in the short term, it can also undermine the actual acquisition of skills if students no longer engage in key thinking and problem-solving processes on their own. The discussion also addresses the risk of "false mastery"—that is, the impression of high competence coupled with limited independent understanding.

Source: https://www.oecd.org/content/dam/oecd/en/publications/reports/2026/01/oecd-digital-education-outlook-2026_940e0dd8/062a7394-en.pdf

The European Parliament addressed this problem in 2026 in a scientific briefing on the role of artificial intelligence in the classroom. Highlighted were cognitive dimensions such as reasoning, memory consolidation, metacognitive regulation, attention, and epistemic judgment. The briefing warns against equating digital operational competence with unimpaired cognitive development.

Source: https://www.europarl.europa.eu/thinktank/en/document/IUST_BRI%282026%29784575

The current state of research therefore speaks neither for a general cognitive impoverishment nor for unlimited technological optimization. Rather, the evidence suggests that the effects of generative AI depend substantially on the educational environment, intensity of use, format of tasks, and degree of active engagement.

The social relevance of this development extends far beyond individual learning processes. Democracies depend on citizens who can examine information, identify contradictions, and independently assess complex relationships. When central mechanisms of critical cognition are increasingly automated, the challenge that arises is not only technological but also social.

At the same time, it would be scientifically inappropriate to view generative AI solely as a threat. AI systems can facilitate access to knowledge, reduce barriers to learning, and support creative processes. This opens up significant opportunities, particularly for people with linguistic, organizational, or cognitive limitations.

The crucial question, therefore, is not whether artificial intelligence should be used in education and knowledge work, but under what conditions it enhances human judgment rather than replacing it. It is precisely at this juncture that it will be determined whether generative AI becomes a tool for intellectual enhancement—or a system of creeping cognitive dependence.

Conclusion: Between Cognitive Expansion and Intellectual Dependency

The current debate on artificial intelligence is marked by sharp contrasts. On one side stand technological narratives of progress that regard generative AI as the beginning of a new era of human productivity. On the other, warnings predominate about intellectual impoverishment, declining concentration, and the loss of independent thought. The scientific evidence of recent years, however, speaks neither for unqualified optimism nor for simplified scenarios of alarm. Rather, a considerably more complex picture is emerging.

For the assumption that generative artificial intelligence causes immediate neurological damage to the human brain, no reliable long-term scientific data currently exist. Neither neuroscience nor cognitive psychology has so far provided indications of a blanket degeneration through AI use. Numerous public claims about an alleged intellectual decline of society therefore exceed the empirical evidence.

At the same time, it would be equally untenable scientifically to underestimate the effects of generative AI on human cognition. The available studies indicate that language models can change the conditions of mental activity. Processes such as writing, researching, structuring, and problem-solving can today be externalized to a degree that is historically unprecedented. This shifts not only the efficiency of human labor, but potentially also the way knowledge is processed and thought is organized.

Particularly relevant in this context is the principle of *cognitive offloading*. The brain tends fundamentally to reduce cognitive load and to use external systems for relief. Generative AI extends this principle to complex linguistic and analytical processes. The technology delivers not only information but generates already pre-structured thought products. It is precisely through this that the necessity of active mental processing can diminish.

From a neuroscientific perspective, this is significant because learning and cognitive consolidation remain tied to use. Attention, memory formation, and analytical thinking rely on the repeated activation of neural networks. If certain mental processes are permanently outsourced, the opportunity to practice them may also be reduced. This particularly affects skills that require intensive cognitive engagement: critical reflection, deep reading, independent reasoning, and complex problem-solving.

However, current research suggests that it is not the existence of artificial intelligence itself that is decisive, but rather the way it is used. Systems that actively engage people in thinking and learning processes can support and expand cognitive performance. AI-assisted tutoring systems, adaptive learning environments, or assistive technologies demonstrate that generative models possess, under certain conditions, pedagogical and creative potentials.

AI becomes problematic when it replaces rather than complements human effort. If language models permanently take over the role of independent analysis, formulation, or source verification, this could lead to a shift in cognitive habits over the long term. This is particularly relevant in education, academia, and journalism—that is, in fields whose social function relies heavily on critical scrutiny and the responsible handling of knowledge.

The ambivalence of generative AI is particularly evident in journalism. Language models can speed up research processes, organize data, and streamline workflows. At the same time, they generate linguistically plausible content without any sense of truth of their own. This increases the risk of an information environment in which linguistic persuasiveness becomes increasingly disconnected from factual reliability. This does not render the ability to critically evaluate sources obsolete; rather, it makes it more important than ever.

Moreover, it is currently often difficult to distinguish which changes arise specifically through generative artificial intelligence and which are already part of general digital media use. Numerous observed changes — such as fragmented information processing or increased cognitive outsourcing — were already discussed long before the rise of generative language models in connection with search engines, social media, and permanent smartphone use. The isolated scientific investigation of specific AI effects is therefore still at its beginning.

From a regulatory perspective, the first signs of a response are already emerging. With the EU AI Act, which took effect in 2024, the European Union adopted the world's first comprehensive set of rules for classifying and regulating artificial intelligence according to risk categories. AI systems used in educational contexts are subject to stricter requirements regarding transparency, traceability, and human oversight. However, the legal framework alone does not answer educational questions: How AI should be integrated into learning processes remains a task for educational researchers and educators.

Several countries have already developed concrete curricular responses to this challenge. Singapore has for several years been systematically integrating AI literacy into national educational programs, combining technical competence in use with critical reflection on algorithmic systems. Finland pursues, with the Elements of AI program, a broadly based approach to foundational education in artificial intelligence directed at the entire population. These examples demonstrate that a reflective integration of generative AI into educational systems is possible — it requires, however, targeted political decisions and pedagogical development work that go beyond the mere deployment of technology.

At the same time, the empirical state of research on central questions remains considerably incomplete. Reliable long-term studies on the neurocognitive effects of sustained AI use are largely absent. Differential effects by age, prior knowledge, and context of use are also barely investigated, as is the question of how collective knowledge production in science and the public sphere changes under conditions of mass AI use. The current debate therefore necessarily moves in a space between plausible theoretical mechanisms and still-outstanding empirical clarification — a circumstance that both public discussion and education policy decisions should take more firmly into account.

Against this background — of regulatory approaches, international examples, and still-open research questions — the social challenge of the coming years can be more precisely determined.

The social challenge of the coming years will therefore not be primarily technical in nature. What is decisive is rather the question of which forms of human cognition continue to be actively cultivated. Democracies, scientific systems, and educational institutions rest on people who can independently examine information, tolerate uncertainty, and critically reflect on complex relationships. It is precisely these capacities that do not arise through passive reception but through active intellectual engagement.

The available scientific findings indicate that generative artificial intelligence changes less the brain itself than the conditions under which human thinking is trained. AI can expand intellectual capacity when it supports analytical processes. It can, however, equally promote cognitive passivity when it permanently replaces independent intellectual work.

Whether artificial intelligence will ultimately lead to an expansion of human cognition or to growing mental dependence is therefore not solely a technical question. What remains crucial is the role that humans themselves will continue to play within these systems—as active thinkers or as mere recipients of machine-generated answers.