Journal of Trends and Challenges in Artificial Intelligence

ASPUR PUBLISHING

Vol. 03, No. 03 (2026) 127-140, DOI: 10.61552/JAI.2026.03.002 - http://jai.aspur.rs

TRANSFORMATIONS DRIVEN BY ARTIFICIAL INTELLIGENCE IN EDUCATION, LABOR PRODUCTIVITY, AND NEUROCOGNITIVE PROCESSES

Ana-Gabriela Banquez-Maturana [1]

Received 04.07.2025.
Revised 17.08.2025.
Accepted 12.09.2025.

Keywords:

Generative Artificial Intelligence; Large Language Models (LLM); Higher Education; Labor Productivity and Creativity; Neurocognitive Processes.

Original research

A B S T R A C T

The purpose of the research was to analyze the transformations that artificial intelligence, and in particular large-scale language models, generate in three dimensions: education, labor productivity, and neurocognitive processes. A systematic review of scientific literature was conducted under the PRISMA 2020 protocol, covering the period 2021–2025. Initially, more than 58,000 records were identified in specialized databases and repositories, which, after a rigorous screening process, were reduced to a final sample of 98 articles. The findings show that the benefits of AI depend less on its technical capacity and more on the pedagogical, organizational, and cognitive conditions that guide its implementation. The novel contribution of the study lies in integrating these three dimensions into a common framework.

1. INTRODUCTION

Artificial intelligence, and particularly large language models, have consolidated as one of the most influential technological innovations today, with visible repercussions in education, labor productivity, and neurocognitive processes. In the educational field, various studies have demonstrated that their integration fosters personalized learning, intelligent tutoring, and immediate feedback, which positively influence students' motivation and performance when applied under structured pedagogical frameworks (Bettayeb et al., 2024; Kestin et al., 2025; Wang & Fan, 2025). In parallel, in the labor domain, these tools have increased productivity and reduced performance inequalities among profiles with different levels of experience, while also stimulating individual creativity, albeit with risks of content homogenization and loss of diversity in collective production (Noy & Zhang, 2023; Doshi & Hauser, 2024; Sun et al., 2025). In the neurocognitive domain, recent literature warns that constant interaction with these systems transforms mechanisms of memory, attention, and executive control, generating immediate benefits in efficiency but also threats such as "cognitive debt" associated with the excessive outsourcing of processes (Grinschgl et al., 2021; Gkintoni et al., 2025a; Mischler et al., 2024).

The relevance of studying this phenomenon lies in the need to articulate dimensions that, until now, have been examined in a fragmented manner. Understanding the

[1] Corresponding author: Ana-Gabriela Banquez-Maturana
Email: abanquezm@unicartagena.edu.co

simultaneous impact of language models on education, work, and cognition makes it possible to transcend a merely instrumental vision and emphasize the ethical, pedagogical, and social challenges accompanying their expansion. This approach is essential, since the decisions that guide their incorporation affect not only immediate learning or organizational productivity but also condition the configuration of future competencies and the preservation of individual cognitive autonomy (Munaye et al., 2025; Guizani et al., 2025; García-López & Nájera-Ochoa, 2025). In this sense, the research gains practical relevance by providing evidence that can guide institutional policies and models of technological governance oriented toward responsible and human-centered use (Krakowski et al., 2025; Brewer et al., 2024).

Nevertheless, the reviewed literature reveals significant gaps. In the educational field, longitudinal studies evaluating the impact of these tools on the consolidation of higher-order competencies remain scarce (Deng et al., 2025; Lo et al., 2024). In the workplace, there is still a lack of evidence on how to balance productivity gains with the long-term preservation of creative diversity (Doshi & Hauser, 2024; Sun et al., 2025). In the neurocognitive sphere, research presents methodological and representational limitations that hinder the generalization of findings regarding memory, control, and learning in heterogeneous populations (Baydili et al., 2025; Brook et al., 2025). These gaps raise questions such as: how do language models transform teaching processes without compromising students' cognitive autonomy, how do they affect productivity and creativity in the tension between individual efficiency and collective diversity, and what neurocognitive implications emerge from intensive use of these technologies?

This study was designed as a systematic literature review under the PRISMA 2020 protocol, with a temporal scope of 2021–2025. Initially, more than 58,000 records were identified, which, after a screening and exclusion process, were refined to a final sample of 98 articles classified into three areas: education and learning, labor productivity and creativity, and neurocognition and health. The novelty of the study lies in offering an integrative framework that connects three dimensions usually analyzed in isolation, allowing the construction of a broad vision of the effects of artificial intelligence on human experience (Franceschelli & Musolesi, 2025; García-López & Nájera-Ochoa, 2025). The central hypothesis proposes that the benefits of language models depend less on their technical capabilities than on the pedagogical, organizational, and cognitive conditions in which they are implemented; therefore, their potential only materializes in intentionally designed, ethically regulated, and human-centered contexts capable of maximizing their contributions and mitigating risks of dependency, homogenization, and loss of cognitive autonomy (Krakowski et al., 2025; Zhai et al., 2024; Abbas et al., 2024).

2. METHODOLOGY

This research was structured as a systematic review of scientific literature following the guidelines of the PRISMA 2020 protocol (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), with the purpose of rigorously analyzing the impact of generative artificial intelligence and large language models (LLM) in three fundamental areas: higher education and learning, labor productivity and creativity, and neurocognitive processes and health (Figure 1).

The temporal scope of analysis covered the period 2021–2025, during which 31,395 records were initially identified in education, 9,728 in productivity and creativity, and 99 in neurocognition, complemented with 535 documents in governmental databases (Ciencia.Science.gov), 17,300 in Google Scholar, and 168 in specialized preprint repositories (arXiv, TechRxiv, bioRxiv). The search strings combined English and Spanish terms with Boolean operators, including expressions such as ChatGPT OR Generative AI AND Teaching AND Learning, Large Language Models AND Productivity AND Creativity, and Artificial Intelligence AND Neurocognition AND Cognitive Processes.

The inclusion criteria considered peer-reviewed articles and preprints from recognized repositories, published in English (n=96) or Spanish (n=2), with direct thematic relevance and preference for open access (74 open-access and 24 in hybrid journals). As exclusion criteria, duplicates, grey literature, non-academic documents, and studies without explicit methodology were removed. The screening process was conducted in stages through title review, abstract evaluation, and full-text reading, which allowed the final sample to be refined to 98 articles.

The selected studies were classified into three thematic areas: education and learning with AI/LLM (54 articles), labor productivity and creativity (24 articles), and neurocognitive processes and health (20 articles). They were also distributed chronologically across five years (Figure 2): 2021 (4 articles), 2022 (1), 2023 (9), 2024 (38), and 2025 (37). For each study, the following elements were extracted: author(s), year, title, journal or repository, methodology, limitations, and findings, all organized in an analytical matrix. Finally, quality assessment was performed by evaluating the coherence among objectives, methods, and results, as well as the transparency in presenting the analysis. Conceptual articles were integrated as theoretical reference frameworks, while empirical studies and meta-analyses were prioritized in the final synthesis, which allowed the consolidation of a robust and relevant body of evidence (Table 1).

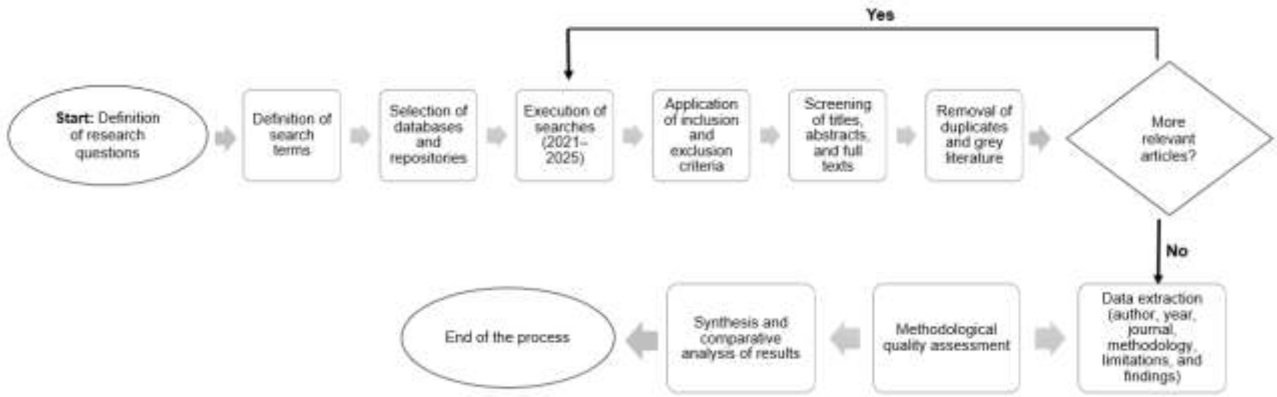

Figure 1. Methodological Flowchart of the Systematic Review
Source: Prepared by the authors, 2025.

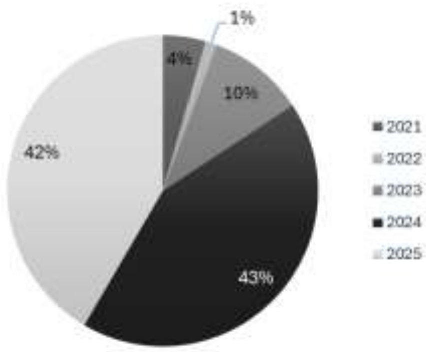

Figure 2. Distribution of articles by year
Source: Prepared by the authors, 2025.

Table 1. Database directory and search results on AI, LLM and education (2021-2025)

Database / Directory	Approach	Search nomenclature	Period (year)	Number of items found
ScienceDirect	Transforming the educational paradigm with AI and LLM	ChatGPT OR Generative AI AND Teaching AND Learning	2021-2025	31.395
Ciencia.Science.gov		Artificial Intelligence AND Large Language Models AND Education		535
Google Scholar		Artificial Intelligence AND Higher Education AND Pedagogy		17.300
arXiv		Artificial Intelligence AND Personalized Learning AND Feedback		168
ScienceDirect	AI/LLM in productivity and workplace creativity	Artificial Intelligence AND Large Language Models AND Productivity		9.728
Ciencia.Science.gov		Generative AI AND Workforce OR Labor Market		224
Google Scholar		ChatGPT OR LLM AND Creativity AND Innovation		17.200
arXiv		Artificial Intelligence AND Human-Centered Design AND Workplace		2
ScienceDirect	Neurocognitive Foundations and LLM	Artificial Intelligence AND Neurocognition AND Cognitive Processes		99
Ciencia.Science.gov		Large Language Models AND Cognitive Offloading AND Memory		49
Google Scholar		Artificial Intelligence AND Cognitive Load AND Learning		17.000
arXiv		Artificial Intelligence AND Cognitive Control		280

Source: own elaboration, 2025.

3. RESULTS

3.1 The transformation of the educational paradigm through artificial intelligence and language models

Large-scale language models have begun to shape a new paradigm in education, transforming both teaching practices and learning processes through personalization, intelligent tutoring, and automated feedback mechanisms that directly impact student motivation and performance (Meyer et al., 2024; Liu et al., 2025; Liu et al., 2024a; Lo et al., 2024). Their integration into formal educational contexts has been shown to have a widespread positive impact when used under appropriate pedagogical conditions, as they facilitate the generation of individualized plans, autonomous practice, and the reinforcement of specific skills, elements that contribute to a more favorable perception of learning (Bettayeb et al., 2024; Wang & Fan, 2025; Kestin et al., 2025; Deng et al., 2025). Consequently, it has been identified that the use of these systems improves self-efficacy, reduces anxiety when faced with complex tasks and promotes the construction of a supporting environment that increases confidence in problem-solving processes (Zhai et al., 2024; Duy et al., 2024; Munaye et al., 2025; Guizani et al., 2025).

Despite these benefits, various analyses have shown that an overreliance on language models carries significant risks for the consolidation of cognitive and practical skills, particularly in areas such as programming and logical reasoning, where excessive use of these tools is associated with a decrease in the ability to internalize essential skills (Yan et al., 2025; Jošt et al., 2024; Welskop, 2023; Filippi & Motyl, 2024). In this sense, it has been documented that students who turn to these applications as simple response machines tend to show a lower capacity for abstraction, a deficit in conceptual retention, and lower performance in academic assessments (Mutanga et al., 2025; Luo et al., 2025; Younas et al., 2025; Jin et al., 2023). Therefore, it is essential to emphasize that the negative effects do not derive from the use of technology itself, but from the way in which it is articulated within a structured pedagogical design oriented towards the development of higher-order skills (Prasad & Sane, 2024; Wecks et al., 2024; Liffiton et al., 2023; Leinonen et al., 2023).

Pedagogical design is therefore a determining factor in understanding the differences in the observed impacts, since the integration of language models as intelligent tutors promotes the scaffolding of learning, increases active student participation, and favors the co-construction of knowledge, while their passive use limits the development of cognitive autonomy (Yang et al., 2025; Heung & Chiu, 2025; Peláez-Sánchez et al., 2024; Abbas et al., 2024). Likewise, the inclusion of these systems in training practices focused on problem-solving contributes to generating more dynamic learning scenarios, where error becomes an opportunity for exploration and adjustment, thereby raising the quality of the educational experience (Silva et al., 2024; Yang et al., 2025; Güner & Er, 2025; Rahe & Maalej, 2025). Similarly, the role of immediate feedback has proven to be fundamental in consolidating learning, by allowing students to correct their productions in real time and achieve a deeper understanding of the content covered (Dinucu-Jianu et al., 2025; Chowdhury et al., 2024; Shen, 2024; Wang et al., 2024).

It is worth noting that the differential effects on students become relevant when it is found that those with lower initial performance experience greater benefits from interacting with language models, given that these systems contribute to calibrating self-assessment and building more accurate representations of their own knowledge (Huber et al., 2024; Lopez-Gazpio, 2025; Sharma et al., 2025; Nie et al., 2024). This compensatory effect reinforces the importance of its use as a mechanism for educational equity, insofar as it offers lagging students opportunities for academic remediation that could hardly be achieved through traditional methodologies (Hellas et al., 2024; Padiyath et al., 2024; Lehmann et al., 2024; Chu et al., 2025). In parallel, it has also been reported that the implementation of pedagogical feedback systems mediated by artificial intelligence allows reducing performance gaps and promoting more inclusive learning, particularly in highly technical areas such as programming (Scholz et al., 2025; Wang et al., 2025;Jyothy et al., 2024; Ganjavi et al., 2024).

The consolidation of these findings allows us to affirm that the transformation of the educational paradigm through the incorporation of language models does not only constitute a technological innovation, but also structurally redefines the dynamics of teaching and learning, providing new forms of personalization, access, and support (Fan et al., 2025; Meyer et al., 2024; Liu et al., 2025; Wang & Fan, 2025). However, it is recognized that the benefits are closely linked to an intentional pedagogical design and active teacher supervision that avoids instrumental dependence and promotes meaningful and sustainable learning (Abbas et al., 2024; Guizani et al., 2025; Munaye et al., 2025; Filippi & Motyl, 2024). In short, language models are configured as a tool with an undeniable potential to strengthen the quality of training processes, provided that their implementation responds to clear pedagogical principles, defined training objectives and constant support that guarantees the consolidation of lasting and transferable skills (Chu et al., 2025; Scholz et al., 2025; Lopez-Gazpio, 2025; Sharma et al., 2025).

3.2 Productivity and creativity at work with the assistance of language models

The integration of language model-based systems has been quantified with average execution time reductions of around 40% and output quality increases of nearly 18% in professional writing tasks with university participants randomly assigned to use ChatGPT; moreover, the likelihood of subsequent actual use doubled after two weeks and remained 1.6 times higher at two months, indicating persistent adoption (Noy & Zhang, 2023). In a field experiment with sales personnel, the "non-adapted" treatment resulted in 0.46 fewer daily meetings than the control and around a 20% decrease in performance, along with approximately 547 fewer units sold per day; in contrast, the "human-centered" treatment added 0.84 daily meetings, more than a 40% relative increase, and approximately 450 additional units sold, with effects that consolidated over time (Krakowski et al., 2025).

In customer service, a mixed study showed that the perception of “informational support” and “emotional support” derived from the use of chatbots is associated with standardized coefficients $\beta = 0.40$ and $\beta = 0.60$, respectively, while the perceived ability of the chatbot to “improve service”, “support the decision” and “streamline business” was related to $\beta = 0.35$, $\beta = 0.33$ and $\beta = 0.20$; Additionally, contextual surveys reported that more than 80% of consumers have interacted with chatbots, 75% doubt their capacity for complex requests and up to 30% would abandon a purchase after a negative experience, in parallel with the fact that 79% of service personnel consider that technology affects their work (Lin et al., 2024).

In persuasive messaging, the probability of request fulfillment increased from 63% (control) to 77% with the foot-in-the-door technique, to 84% with verbal anthropomorphism, and to 95% with the combination of both; the estimated odds ratios were 2.50 and 3.98 with respect to the baseline (N = 153) (Adam et al., 2021). In the infrastructural layer that determines latencies and operating costs, disaggregating inference phases reduced the time to first token by up to 97%, the job completion time by 47–50%, and resource consumption by 38%, with additional improvements of 44% in TTFT, 17% in JCT, and 1.4× in performance-to-cost in multi-server deployments; Furthermore, prefill latency decreased by 86.4% under shard partitioning, while dynamic reservation policies cut JCT by 10–12% under composite loads (Hu et al., 2024; Hu et al., 2025).

In relational analytics with LLM, field reordering, prompt reuse, and prefix caching produced speedups of 1.5–3.4× in end-to-end latency and cost savings of up to 32% over proprietary models, with increases in prefix-hit rate and near-optimal solutions running in seconds (vs. hourly solves) with losses ≤ 2% relative to the optimum (Liu et al., 2024b). In financial marketing roles, the multi-component PU-learning methodology outperformed baselines in AUC and adjusted AUC in simulations with 100 replicates (e.g., AUC 0.699 and adjusted AUC 0.767 for n = 800; AUC 0.737 and adjusted AUC 0.782 for n = 1600), with consistent improvements in mean square error of coefficients and preference matrices (Chen et al., 2023). In job recommendation, a generative approach increased AUC by 1.5% to 5.2% compared to baseline predictors and reduced LogLoss by up to 29.8%, with robust improvements across multiple datasets (Zheng et al., 2023).

The leveling effect between less and more skilled profiles is manifested in the narrowing of performance inequality when LLM assistance is enabled, given that participants with weaker skills captured greater marginal benefits in time and quality during the experiment with professional tasks; moreover, sustained adoption suggests that the gain is not restricted to the laboratory (Noy & Zhang, 2023). In the field, “human-centric” orchestration was a condition for the technology to enhance productivity, while the “non-adaptive” configuration was associated with losses of around 20%, implying that leveling depends on a design that addresses cognitive styles and roles (Krakowski et al., 2025).

In short creativity, a small model scored short synopses 14% higher than humans overall and produced surprising associations in 15% of cases versus 2–3% for large models, but showed lower internal consistency (e.g., 68.3–95–100% internal consistency depending on condition), with attribute-specific improvements in readability (22%), comprehensibility (17%), relevance (23%), informativeness (11%), and appeal (18%) (Marco et al., 2024). In creative writing tasks supported by AI-generated ideas, granting one idea increased “usefulness” by 5.4% and five ideas increased it by 9%; “Novelty” increased by 6.3–10.7%, “good writing” by 26.6%, and “enjoyment” by 22.6%, with 88.4% of participants requesting ideas and an average of 2.55 calls (24.5% asking for five ideas), although greater similarity was observed between stories within the treated groups, which reduces the collective diversity of content (N = 293 writers; N = 600 evaluators) (Doshi & Hauser, 2024). Convergently, in a recent synthesis, 9.4% of individuals outperformed GPT-4 on creativity metrics, while in problem solving, models contributed about 33% of the best answers on social tasks and 40–57% on scientific tasks; overall, about one-third of the top ten answers came from LLMs (Sun et al., 2025). AI literacy and interaction quality shape individual and collective performance; Structural models with K-12 teachers showed that behavioral intention to learn AI was explained by 75% and perception of social good applications by 70%, with direct effects $\beta = 0.62$ (perception → intention) and $\beta = 0.29$ (self-efficacy → intention), in addition to $\beta = 0.77$ (literacy → self-efficacy), $\beta = 0.18$ (literacy → perception), and $\beta = 0.62$ (literacy → ethical awareness); the fit indices were CFI = 0.927–0.965 and RMSEA = 0.071–0.089 in different submodels (Du et al., 2024).

In higher education, a systematic review identified that 18 works reported as "disseminations" and a large volume of experiences with structured prompting, task decomposition, and guided self-reflection are associated with improvements in assessed products, which guides the design of competency-based curricula (Lee & Palmer, 2025). In the mapping of teaching research, 44 educational AI studies were analyzed from 751 initial records; 70% of the works came from the Web of Science, 46% were conducted in university settings, and the collection techniques included video (22%), interviews (22%), questionnaires (18%), and interaction logs (7%), among others (Celik et al., 2022).

On divergent thinking measures, GPT-4 outperformed the human average on originality with main effects

$F(1,600) = 622.10$ and $\eta^2 = 0.51$, and stimulus effects $F(1,600) = 584.50$ and $\eta^2 = 0.49$; however, the human maximum outperformed the model maximum, with inter-rater ICC consistencies of 0.88–0.93 and moderate correlations between semantic distance and judgments ($r \approx 0.52$–0.55) (Hubert et al., 2024; Koivisto & Grassini, 2023). In story authorship detection, accuracy reached 94% for expert judges, 81% for non-experts, and 71% for an evaluative LLM, which shows detectable stylistic patterns in part of the generated production (Ismayilzada et al., 2024). Automatic evaluation with LLM-as-judge showed correlations with humans that can be considered high; in particular, for recent families such as Qwen 2.5 and Llama 3.3, coefficients higher than those obtained with EM/F1 metrics are reported, with references to thresholds of $r > 0.80$ for strong correspondence (Ho et al., 2025).

In pre-writing co-creation, a study with 15 participants documented that LLM support is perceived as "second-minded" in the pre-writing phase, with an easing of the burden on ideation and planning and an increase in post-writing curation and editing work (Wan et al., 2024). In a conceptual synthesis and comparative evidence, conditions under which human-AI collaboration enhances the problem-finding phase more than the problem-solving phase have been discussed, widening the formulation space before converging, with implications for the governance of creative processes in organizations (Sabbah & Li, 2025; Chen & Chan, 2024; Bellemare-Pepin et al., 2025; Franceschelli & Musolesi, 2025; Laverghetta et al., 2025; Lockhart, 2025).

In on-demand work platforms and microtasks, the proportion of LLM-generated texts was estimated at 33–46% using a combination of typing dynamics and detection classifiers; a specialized detector flagged 46% (95% CI: 31–61%) of summaries and another 33% (95% CI: 20–45%) as synthetic, with a macro-F1 of 97–99% in validation, while a general classifier only recognized 26% of LLM texts, introducing measurement bias if domain-adapted detectors are not used (Veselovsky et al., 2023). In retail-related sector settings, almost all participants (93%) purchased a product during a shopping experience; 65% did so in physical stores and 35% online, with a distribution by categories that included 47.7% fashion and 19.6% consumer electronics, among others, which contextualizes customer inspiration patterns in the face of hybrid channels (Frasquet et al., 2024).

In applied creativity with synopsis, already cited, the percentages of internal coherence ranged between 68.3% and 100% depending on the condition, and "surprising" associations were observed in 15% of the SLM material compared to 1.6–3.3% in alternatives, with data sets of 42,049 examples for training-validation-testing (80/10/10) (Marco et al., 2024). In competency extraction from textual descriptions, a fine-tuned 8B LLaMA-3 achieved a total F1 of 64.8% (64.7% in validation), with F1 of 54.3% for "skills" and 74.2% for "knowledge", outperforming BERT/SpanBERT baselines (≈ 57.7–58.9%) and recent methods (≈ 62.6–64.2%), enabling partial automation and profile traceability (Herandi et al., 2024). In human-AI collaboration on creative tasks, the effect of modality (sequential co-authoring, task splitting, assisted ideation) and user expertise modulate the quality and perceived usefulness of products; This pattern is articulated with the evidence that gains are greater in less expert profiles and with the warning that stylistic homogeneity can grow when the use of AI ideas becomes widespread (Chen & Chan, 2024; Doshi & Hauser, 2024).

Regarding the organizational and labor policy context, various analyses have considered that the diffusion of LLMs will coexist with job creation and loss, which shifts the emphasis toward continuous reskilling and role redesign strategies. Although this is a conceptual document, the argument is supported by comparisons of scenarios and sectoral trajectories (Wang, 2023). In parallel, technological integration frameworks have been proposed combining generative AI and blockchain to address traceability, governance, and risk, without providing primary quantitative metrics but offering design guidelines for information-intensive processes (Brewer et al., 2024). Finally, in biomedical research unrelated to job performance or organizational AI, experimental parameters of mechanical control and dosage (e.g., 0.1 MPa for 48 h and 20 mg/kg/day of quercetin) have been reported that illustrate practices of methodological rigor in applied studies, although their content is not extrapolated to the impact of AI on productivity and creativity at work (Ren et al., 2024).

3.3 Neurocognitive Foundations of the Human Brain in the Age of AI with LLM

The expansion of LLMs has reconfigured the neurocognitive ecosystem by accelerating the search, synthesis, and control of information, while introducing tensions between instrumental efficiency and the maintenance of autonomous reasoning and memory processes, with implications for individual and organizational epistemic agency (García-López & Nájera-Ochoa, 2025; Mischler et al., 2024). Computational neuroscience evidence shows alignments between contextual extraction hierarchies in LLMs and cortical trajectories during language comprehension, where higher model performance improves the prediction of neural responses and reduces the depth needed to encode features, suggesting shared principles of predictive computing (Mischler et al., 2024).

In a complementary manner, brain network dynamics reveal that effective connectivity includes substantial competitive interactions, with 25% ± 8% negative edges in humans, 38% ± 7% in macaques, and 28% ± 4% in mice, implying a balance between cooperation and resource competition that conditions control allocation under intelligent assistance (Luppi et al., 2024). In cognitive control, frontal midline theta has been shown to index the need for control and discriminate strategies by decoding control states in a generalizable way in N=176 adults who faced inhibition, proactive/reactive control, and conflict, establishing a regulatory mechanism applicable to algorithmic support contexts (Eisma et al., 2021).

Likewise, the literature oncognitive offloadingdocuments immediate efficiency and multitasking gains when offloading costs are low, with N=172 participants showing greater accuracy on a concurrent N-back task by offloading more information under no-time-cost conditions (Grinschgl et al., 2023). However, cognitive offloading compromises subsequent memory by optimizing performance in the execution phase and decreasing recall and consolidation when deliberate practice is subtracted, an effect modulated by the anticipation of memory tests and by the balance between convenience and cost of offloading (Grinschgl et al., 2021).

In the child population, it has been observed thatoffloadingIt depends on metacognitive judgments and the incentive structure of the task, with increases in recall when there is an offloading option and with external recording behaviors in around 67% of the items, which shows savings of immediate work resources and sensitivity to the perception of effort (Iley & Medimorec, 2024). These findings support the conceptual axis of “cognitive load 2.0”, understood as the articulation between Cognitive Load Theory, neuroeducation and adaptive AI technologies that rebalance intrinsic, extrinsic and germinal loads through neurophysiological instrumentation and real-time adjustment (Gkintoni et al., 2025a).

A systematic review using the PRISMA protocol recorded 523 initial and 103 included studies on adaptive systems guided by neurophysiological markers, highlighting the use of EEG and fNIRS to adapt difficulty and feedback, with implications for personalization, data security, and equity of access in educational settings (Gkintoni et al., 2025a). Evidence indicates that neuroadaptive systems allow modulating extrinsic load by suppressing irrelevant elements and amplifying relevant scaffolds, while boosting germinal load by optimizing schema encoding under continuous EEG/fNIRS effort measurements, favoring efficiency without degrading deep processing (Gkintoni et al., 2025b).

In turn, multisensory and second language learning in adults shows robust neuroplasticity, synthesized in 80 studies using fMRI, MEG, and DTI, with associations between multimodal immersion, increases in frontotemporal connectivity, and improvements in working memory, attention, and cognitive flexibility, relevant to designing AI supports that strengthen executive functions (Gkintoni et al., 2025b). It is evident that “cognitive debt” describes the progressive erosion of higher-level functions due to the chronic delegation of control and memory to artifacts, a phenomenon supported by industrial automation data where, after six weeks, the average cycle time increased and the “right-first-time%” decreased in groups that relied on automation versus refresher training (Rashid & Rötting, 2021).

In organizations, a case study exposes vicious cycles of expertise erosion under cognitive automation; technological dependence reduces deliberate practice, consolidates rigidities, and feeds back into outsourcing, calling for knowledge governance and task redesign to safeguard cognitive autonomy (Rinta-Kahila et al., 2023). Similarly, multitasking scenarios with unreliable automation illustrate performance declines explained by the "out-of-the-loop" phenomenon, which degrades situational awareness and manual skills; therefore, the reliability and adaptability of assistance are crucial to avoid losses of stability-flexibility in control (Stasch & Mack, 2025).

In parallel, AI applied to mental health shows high metrics in multimodal classification tasks, with 98.2% accuracy, 99.15% sensitivity, and 99.53% specificity for multimodal emotional detection, and accuracies of 97.86% in stress with ECG/GSR, although limitations in sample diversity and training times persist (Baydili et al., 2025). It is worth noting that AUC≈0.94 are reported when distinguishing people with Alzheimer's from controls using speech extraction with transformer-based models, reinforcing the potential of digital biomarkers in psychiatry and neurology when physiological signals and language are combined (Baydili et al., 2025).

A review in Alzheimer's disease that integrated 431 studies reported an average AUC of 0.887 in AD and 0.821 in MCI, with only 2 studies with external validation and 3 with calibration, calling for methodological transparency and out-of-sample evaluation for clinical translation (Qi et al., 2025). Due to the estimated global burden of dementia of >55 million people, with 60%–70% attributable to Alzheimer's disease, biomarker standardization and calibration reporting are priorities for reliable deployments in digital care and longitudinal monitoring (Qi et al., 2025). To enable open science and integrative biomarkers of affective state and sleep in development, a resource with high-quality neuroimaging (multi-echo fMRI and compressed dMRI), ecological momentary assessment, and behavioral data has been released, with an initial release of n=10 and a target of 100 participants (Brook et al., 2025).

As a result, the integration of brain signals and behavior with standardized protocols allows for the study of covariations among mood variability, sleep architecture, and network organization, facilitating the validation of translational multimodal models in youth populations (Brook et al., 2025). On the neuroethical level, the use of BCIs for cognitive enhancement demands safeguards regarding identity, autonomy, and equity, given the sensitivity of neural data and the dependence on ongoing technical support. Therefore, frameworks for consent, fair distribution of access, and accountability for failures are required (Gordon & Seth, 2024). Therefore, the expansion of BCI technologies and their convergence with language analytics must be accompanied by robust privacy policies, interoperability standards, and audit mechanisms to maintain the integrity of personal agency during enhancement processes (Gordon & Seth, 2024).

In terms of control fundamentals, medial frontal theta oscillation serves as both a “common language” of control and a differential signature of strategies, suggesting that adaptive systems must calibrate support based on markers of control demand and signals of switching between proactive and reactive modes (Eisma et al., 2021). Indeed, the competitive interactions detected in effective connectivity limit the resources available for such adjustments, such that excessive assistance can shift the cooperation-competition balance and consolidate patterns of dependence that reduce the practice of endogenous control (Luppi et al., 2024).

At the learning level, the combination of CLT, neuroeducation, and AI allows for the reduction of extrinsic load by filtering out distractors and the optimization of germinal load by promoting schematization, according to a systematic synthesis of 103 studies that recommend multimodal designs sensitive to physiological markers (Gkintoni et al., 2025b). Likewise, evidence of neuroplasticity in adults undergoing multisensory and L2 learning supports positive effects on executive functions and metacognition, proposing that “cognitive debt” can be counteracted with rich experiences that require control and updating of representations (Gkintoni et al., 2025a).

Regarding the risk of atrophy, experimental and field results indicate that chronic delegations without refresher training increase cycle times and decrease first-pass quality, reinforcing the need for skill preservation strategies to mitigate erosion (Rashid & Rötting, 2021). The organizational articulation of knowledge with LLM offers improvements in productivity, access, and traceability of information assets, but requires bias governance, explainability, and privacy to prevent structural dependencies and preserve decision-making autonomy (García-López & Nájera-Ochoa, 2025).

Furthermore, reviews of AI in psychiatry show contributions of EEG, ECG, blood biomarkers, speech, and social data for diagnosis and follow-up, with high performances and generalization challenges that require expanding population diversity and calibration reports (Baydili et al., 2025). Consequently, prudent performance evaluation must go beyond accuracy and incorporate external validation, sensitivity analysis, and fairness metrics, especially in digital Alzheimer's biomarkers where clinical adoption requires reproducibility (Qi et al., 2025). Likewise, the availability of multimodal open resources with common protocols facilitates replicability and cross-sectional comparison, enabling more precise estimates of effects and learning curves of adaptive algorithms in development (Brook et al., 2025).

The conceptual synthesis suggests that the immediate efficiency gains brought by LLM and automation must be balanced with designs that maintain deliberate practice and endogenous control, avoiding the accumulation of cognitive debt and stabilizing the cooperation-competition trade-off in cortical networks (Grinschgl et al., 2021; Luppi et al., 2024). Furthermore, the horizon of consciousness and its interface with smart technologies requires conceptual and regulatory caution, given that the technological extension of higher-level functions raises questions about the continuity of the self, criteria of responsibility, and the distribution of benefits (Gusev et al., 2025).

Consequently, a neurocognitive framework for the AI era must integrate brain-model alignments in language processing, theta regulation of control strategies,offloadingwith temporal benefits and costs, “cognitive load 2.0” supported by EEG/fNIRS, digital biomarkers with rigorous validations, adult neuroplasticity, BCI ethics and knowledge governance with LLM (Mischler et al., 2024; Eisma et al., 2021; Grinschgl et al., 2023; Gkintoni et al., 2025a; Qi et al., 2025; Gkintoni et al., 2025b; Gordon & Seth, 2024; García-López & Nájera-Ochoa, 2025).

4. DISCUSSION

Evidence confirms that the impact of artificial intelligence (AI) in educational and social contexts depends less on its technical capabilities and more on the pedagogical, organizational, and cognitive conditions that frame its use. In line with this approach, Abbas et al. (2024) show that students' perceived usefulness is modulated by factors of trust, technological anxiety, and social norms, demonstrating that the psychosocial context is as or more decisive than the power of the underlying algorithms. Similarly, Celik et al. (2022) highlight that the promises of AI for teaching face obstacles linked to cognitive overload and lack of teacher training, placing teachers as key agents in realizing the benefits of these technologies.

In line with the above, Kestin et al. (2025) demonstrated that an AI tutor designed under well-established pedagogical principles produced greater learning gains than active classroom teaching, provided that practices such as cognitive load management, personalized feedback, and the promotion of a growth mindset were incorporated. Coinciding with this result, Meyer et al. (2024) showed in high school students that feedback generated by LLM not only improved text revision but also motivation and positive emotions. As noted in previous studies, Lo et al. (2024) emphasize that student engagement depends on how activities are structured, and can encourage self-regulation and critical thinking; however, in some cases it also generates dependency and discourages reflection.

In contrast, Krakowski et al. (2025) found that the benefits of AI are maximized when institutions adopt a human-centered approach, aligning algorithms with support structures and coordination processes. In contrast, Welskop (2023) warns that the absence of clear policies on the use of ChatGPT in universities can lead to the dehumanization of the teaching process. In line with this warning, Brewer et al. (2024) suggest the integration of blockchain as a traceability mechanism, emphasizing that organizational governance is an essential element for risk mitigation.

Although some authors highlight the immediate value of cognitive offloading, Grinschgl et al. (2021, 2023) documented that this strategy can compromise memory and deep learning. However, Zhai et al. (2024) confirm that overreliance on dialogues with AI impairs critical skills, indicating that the real benefit only appears when AI acts as a support for human reasoning. In line with this idea, Dinucu-Jianu et al. (2025) show that aligning LLM with problem-solving pedagogies through reinforcement learning turns the interaction into a genuine learning experience, reinforcing the need for explicit pedagogical frameworks.

In the creative field, Doshi and Hauser (2024) show that AI enhances individual creativity; however, it reduces the collective diversity of content. In contrast, Hubert et al. (2024) argue that in divergent tasks, generative models outperform humans in the number of ideas, although not in conceptual depth. In line with what Franceschelli and Musolesi (2025) proposed, this tension confirms that the benefits or risks depend on how its use is organized: AI can stimulate new forms of thinking if it is oriented towards critical reflection, or encourage superficial productions if it is used as an uncritical substitute for human creativity.

In the clinical setting, Liu et al. (2024b) show that ChatGPT passes medical exams in English-speaking contexts, in contrast to its poor performance in non-English-speaking environments. While this finding demonstrates the power of the models, it also confirms their dependence on cultural and contextual conditions. Similarly, Baydili et al. (2025) emphasize that AI in psychiatry can complement behavioral diagnoses and analyses; however, only under the supervision of human clinical judgment. In line with these observations, the educational field needs to conceive AI as contextualized support within clear professional frameworks.

At the organizational and workplace levels, Noy and Zhang (2023) documented an increase in individual efficiency with generative AI; however, they warn of risks of homogenization and skill loss. In contrast, Wang (2023) adds that LLMs generate a paradox of simultaneous job creation and destruction, which poses structural challenges for university training oriented to a changing market. Although this efficiency can translate into immediate benefits, Lehmann et al. (2024) found that the unstructured use of ChatGPT harms learning, reinforcing the idea that the risks arise not from the technique, but from the absence of adequate pedagogical and organizational frameworks.

On the motivational level, Heung and Chiu (2025) summarize in their meta-analysis that ChatGPT can increase student engagement, although this effect is heterogeneous and depends on the way it is integrated into the classroom dynamics. In agreement, Huber et al. (2024) show that playful learning with LLM increases motivation, while Guizani et al. (2025) point out that the lack of ethical and methodological guidelines can neutralize these benefits. In agreement with these results, the literature converges in that the decisive factor is not the capacity of the AI, but rather the way in which its use is regulated and intended.

In this sense, the reviewed studies allow us to affirm that AI is neither intrinsically beneficial nor harmful in education, health, or creativity. Although it can act as an effective tutor (Kestin et al., 2025), motivational support (Meyer et al., 2024), or creative catalyst (Hubert et al., 2024), it can also lead to dependency and superficiality (Zhai et al., 2024; Grinschgl et al., 2023). Hence, the theoretical implications point to the need for an interdisciplinary approach that integrates pedagogy, neuroscience, ethics, and organizational management, while the practical implications show that only a human-centered and regulated design can ensure that AI amplifies human capabilities rather than replacing or eroding them.

5. CONCLUSION

The research developed allowed for an integrative analysis of the transThe insights that artificial intelligence, and language models in particular, generate in education, labor productivity, and neurocognitive processes. The main finding is that the impact of AI depends less on its technical capacity and more on the pedagogical, organizational, and cognitive conditions in which it is implemented, making intentional, human-centered design the true determinant of its benefits or risks.

The novel contribution of this work lies in articulating three dimensions: educational, occupational, and neurocognitive which while analyzed in a fragmented manner in the literature, have rarely been integrated into a comprehensive framework that allows for understanding the breadth of the impact of artificial intelligence on human experience. This integrative approach broadens the scope of discussion by going beyond a purely technical perspective and highlighting the pedagogical, organizational, and cognitive challenges that emerge with the widespread use of these tools.

From a methodological perspective, the study was based on a documentary and comparative analysis of recent academic literature, which allowed for the systematization of disparate findings and the construction of a coherent interpretative framework. This procedure not only strengthened the validity of the work but also provided a replicable methodology for future research focused on studying complex, multidimensional phenomena.

At the theoretical level, the research contributes to enriching existing conceptual frameworks by connecting learning theories, productivity studies, and cognitive neuroscience approaches with the impact of language models, thus offering a unifying framework that jointly explains the benefits, risks, and conditions for the effectiveness of AI.

Likewise, the practical contribution translates into recommendations applicable in educational and work contexts, proposing intentional pedagogical designs that enhance student autonomy and motivation, as well as organizational models that ensure human-centered technological integration. In the neurocognitive field, useful guidelines are proposed to avoid the accumulation of "cognitive debt" and promote responsible uses of artificial intelligence that strengthen individual agency.

While this study is limited by its exclusive reliance on documentary sources, which prevents the findings from being compared with direct empirical evidence, this restriction does not weaken the validity of its conclusions, but rather opens an opportunity for further research that combines theoretical analysis with field studies and experimental measurements.

Consequently, we encourage continued research in this field, exploring future lines such as the longitudinal evaluation of the impact of AI on the development of cognitive skills, the design of institutional policies that balance innovation with governance, and neuroscientific experimentation applied to educational and work contexts. Exploring these directions will not only deepen academic understanding of the phenomenon but also guide concrete strategies for the ethical, inclusive, and sustainable integration of artificial intelligence into contemporary society.

Acknowledgement: The author wishes to thank Universidad de La Salle and Universidad de Cartagena for their support in the development of this article.

References

Abbas, M., Jam, F., & Khan, T. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education*, 21(1). DOI: 10.1186/s41239-024-00444-7

Adam, M., Wessel, M., & Benlian, A. (2021). AI-based chatbots in customer service and their effects on user compliance. *Electronic Markets*, 31(2), pp. 427–445. DOI: 10.1007/s12525-020-00414-7

Baydili, İ., Tasci, B., & Tasci, G. (2025). Artificial intelligence in psychiatry: A review of biological and behavioral data analyses. *Diagnostics (Basel, Switzerland)*, 15(4), pp. 434. DOI: 10.3390/diagnostics15040434

Bellemare-Pepin, A., Et al. (2025). Divergent creativity in humans and Large Language Models. En *arXiv [cs.CL]*. http://arxiv.org/abs/2405.13012

Bettayeb, A. M., Abu Talib, M., Sobhe Altayasinah, A. Z., & Dakalbab, F. (2024). Exploring the impact of ChatGPT: conversational AI in education. *Frontiers in education*, 9. DOI: 10.3389/feduc.2024.1379796

Brewer, J., Patel, D., Kim, D., & Murray, A. (2024). Navigating the challenges of generative technologies: Proposing the integration of artificial intelligence and blockchain. *Business Horizons*. DOI: 10.1016/j.bushor.2024.04.011

Brook, J. B., Salo, T., Luo, A. C., Bagautdinova, J., Rush, S., Alexander-Bloch, A. F., ... & Satterthwaite, T. D. (2025). An open, fully-processed data resource for studying mood and sleep variability in the developing brain. En *bioRxivorg*. DOI: 10.1101/2025.05.01.651544

Celik, I., Dindar, M., Muukkonen, H., & Järvelä, S. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research. *TechTrends : For Leaders in Education & Training*, 66(4), pp. 616–630. DOI: 10.1007/s11528-022-00715-y

Chen, S., Qiu, Y., Li, J., Fang, K., & Fang, K. (2023). Precision marketing for financial industry using a PU-learning recommendation method. *Journal of Business Research*, 160(113771), pp. 113771. DOI: 10.1016/j.jbusres.2023.113771

Chen, Z., & Chan, J. (2024). Large language model in creative work: The role of collaboration modality and user expertise. *Management Science*, 70(12), pp. 9101–9117. DOI: 10.1287/mnsc.2023.03014

Chowdhury, S., Zouhar, V., & Sachan, M. (2024). AutoTutor meets Large Language Models: A language model tutor with rich pedagogy and guardrails. En *arXiv [cs.CL]*. DOI: 10.48550/ARXIV.2402.09216

Chu, Z., Wang, S., Xie, J., Zhu, T., Yan, Y., Ye, J., ... & Wen, Q. (2025). LLM Agents for Education: Advances and Applications. En *arXiv [cs.CY]*. http://arxiv.org/abs/2503.11733

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, 227(105224), pp. 105224. DOI: 10.1016/j.compedu.2024.105224

Dinucu-Jianu, D., Macina, J., Daheim, N., Hakimi, I., Gurevych, I., & Sachan, M. (2025). From problem-solving to teaching problem-solving: Aligning LLMs with pedagogy using reinforcement learning. En *arXiv [cs.CL]*. DOI: 10.48550/ARXIV.2505.15607

Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*, 10(28), pp. 1-9. DOI: 10.1126/sciadv.adn5290

Du, H., Sun, Y., Jiang, H., Islam, A. Y. M., & Gu, X. (2024). Exploring the effects of AI literacy in teacher learning: an empirical study. *Humanities & Social Sciences Communications*, 11(1), pp. 1–10. DOI: 10.1057/s41599-024-03101-6

Duy, N. B. P., Phuong, T. N. M., Chau, V. N. M., Nhi, N. V. H., Khuyen, V. T. M., & Giang, N. T. P. (2024). AI-assisted learning: an empirical study on student application behavior. *Multidisciplinary Science Journal*, 7(6), pp. 2025275. DOI: 10.31893/multiscience.2025275

Eisma, J., Rawls, E., Long, S., Mach, R., & Lamm, C. (2021). Frontal midline theta differentiates separate cognitive control strategies while still generalizing the need for cognitive control. *Scientific Reports*, 11(1), pp. 14641. DOI: 10.1038/s41598-021-94162-z

Fan, G., Liu, D., Zhang, R., & Pan, L. (2025). The impact of AI-assisted pair programming on student motivation, programming anxiety, collaborative learning, and programming performance: a comparative study with traditional pair programming and individual approaches. *International Journal of STEM Education*, 12(1). DOI: 10.1186/s40594-025-00537-3

Filippi, S., & Motyl, B. (2024). Large language models (LLMs) in engineering education: A systematic review and suggestions for practical adoption. *Information (Basel)*, 15(6), pp. 345. DOI: 10.3390/info15060345

Franceschelli, G., & Musolesi, M. (2025). On the creativity of large language models. *AI & Society*, 40(5), pp. 3785–3795. DOI: 10.1007/s00146-024-02127-3

Frasquet, M., Ieva, M., & Mollá-Descals, A. (2024). Customer inspiration in retailing: The role of perceived novelty and customer loyalty across offline and online channels. *Journal of Retailing and Consumer Services*, 76(103592), pp. 103592. DOI: 10.1016/j.jretconser.2023.103592

Ganjavi, C., Eppler, M., O'Brien, D., Ramacciotti, L. S., Ghauri, M. S., Anderson, I., ... & Cacciamani, G. E. (2024). ChatGPT and large language models (LLMs) awareness and use. A prospective cross-sectional survey of U.S. medical students. *PLOS Digital Health*, 3(9), pp. e0000596. DOI: 10.1371/journal.pdig.0000596

García-López, I. M., & Nájera-Ochoa, J. (2025). Mapeo de Literatura sobre el Impacto de los LLM en la Gestión del Conocimiento: Tendencias y Desafíos. *Vinculatégica EFAN*, 11(4), pp. 61–77. DOI: 10.29105/vtga11.4-1149

Gkintoni, E., Antonopoulou, H., Sortwell, A., & Halkiopoulos, C. (2025a). Challenging Cognitive Load Theory: The role of Educational Neuroscience and Artificial Intelligence in redefining Learning Efficacy. *Brain Sciences*, 15(2). DOI: 10.3390/brainsci15020203

Gkintoni, E., Vassilopoulos, S. P., & Nikolaou, G. (2025b). Brain-inspired multisensory learning: A systematic review of neuroplasticity and cognitive outcomes in adult multicultural and second language acquisition. *Biomimetics (Basel, Switzerland)*, 10(6). DOI: 10.3390/biomimetics10060397

Gordon, E., & Seth, A. (2024). Ethical considerations for the use of brain-computer interfaces for cognitive enhancement. *PLoS Biology*, 22(10), pp. e3002899. DOI: 10.1371/journal.pbio.3002899

Grinschgl, S., Papenmeier, F., & Meyerhoff, H. (2021). Consequences of cognitive offloading: Boosting performance but diminishing memory. *Quarterly Journal of Experimental Psychology (2006)*, 74(9), pp. 1477–1496. DOI: 10.1177/17470218211008060

Grinschgl, S., Papenmeier, F., & Meyerhoff, H. (2023). Mutual interplay between cognitive offloading and secondary task performance. *Psychonomic Bulletin & Review*, 30(6), pp. 2250–2261. DOI: 10.3758/s13423-023-02312-3

Guizani, S., Et al. (2025). A systematic literature review to implement large language model in higher education: issues and solutions. *Discover Education*, 4(1). DOI: 10.1007/s44217-025-00424-7

Güner, H., & Er, E. (2025). AI in the classroom: Exploring students' interaction with ChatGPT in programming learning. *Education and Information Technologies*, 30(9), pp. 12681–12707. DOI: 10.1007/s10639-025-13337-7

Gusev, E., Sarapultsev, A., & Komelkova, M. (2025). Evolutionary trajectories of consciousness: From biological foundations to technological horizons. *Brain Sciences*, 15(7), pp. 734. DOI: 10.3390/brainsci15070734

Hellas, A., Leinonen, J., & Leppänen, L. (2024). Experiences from integrating large language model chatbots into the classroom. En *arXiv [cs.CY]*. DOI: 10.48550/ARXIV.2406.04817

Herandi, A., Li, Y., Liu, Z., Hu, X., & Cai, X. (2024). Skill-LLM: Repurposing general-purpose LLMs for skill extraction. En *arXiv [cs.CL]*. http://arxiv.org/abs/2410.12052

Heung, Y., & Chiu, T. (2025). How ChatGPT impacts student engagement from a systematic review and meta-analysis study. *Computers and Education: Artificial Intelligence*, 8(100361), pp. 100361. DOI: 10.1016/j.caeai.2025.100361

Ho, X., Huang, J., Boudin, F., & Aizawa, A. (2025). LLM-as-a-Judge: Reassessing the Performance of LLMs in Extractive QA. En *arXiv [cs.CL]*. http://arxiv.org/abs/2504.11972

Hu, C., Huang, H., Xu, L., Chen, X., Wang, C., Xu, J., ... & Shan, Y. (2025). ShuffleInfer: Disaggregate LLM inference for mixed downstream workloads. *ACM Transactions on Architecture and Code Optimization*, 22(2), pp. 1–24. DOI: 10.1145/3732941

Hu, C., Huang, H., Xu, L., Chen, X., Xu, J., Chen, S., ... & Shan, Y. (2024). Inference without interference: Disaggregate LLM inference for mixed downstream workloads. En *arXiv [cs.DC]*. http://arxiv.org/abs/2401.11181

Hubert, K., Awa, K., & Zabelina, D. (2024). The current state of artificial intelligence generative language models is more creative than humans on divergent thinking tasks. *Scientific Reports*, 14(1), pp. 3440. DOI: 10.1038/s41598-024-53303-w

Huber, S. E., Kiili, K., Nebel, S., Ryan, R. M., Sailer, M., & Ninaus, M. (2024). Leveraging the potential of large language models in education through playful and game-based learning. *Educational Psychology Review*, 36(1). DOI: 10.1007/s10648-024-09868-z

Iley, C., & Medimorec, S. (2024). Children's metacognition and cognitive offloading in an immediate memory task. *Psychological Research*, 88(5), pp. 1448–1455. DOI: 10.1007/s00426-024-01978-1

Ismayilzada, M., Stevenson, C., & van der Plas, L. (2024). Evaluating creative short story generation in humans and large language models. En *arXiv [cs.CL]*. http://arxiv.org/abs/2411.02316

Jin, S. H., Im, K., Yoo, M., Roll, I., & Seo, K. (2023). Supporting students' self-regulated learning in online learning using artificial intelligence applications. *International Journal of Educational Technology in Higher Education*, 20(1). DOI: 10.1186/s41239-023-00406-5

Jošt, G., Taneski, V., & Karakatič, S. (2024). The impact of Large Language Models on programming education and student learning outcomes. *Applied Sciences (Basel, Switzerland)*, 14(10), pp. 4115. DOI: 10.3390/app14104115

Jyothy, S. N., Kolil, V. K., Raman, R., & Achuthan, K. (2024). Exploring large language models as an integrated tool for learning, teaching, and research through the Fogg Behavior Model: a comprehensive mixed-methods analysis. *Cogent Engineering*, 11(1). DOI: 10.1080/23311916.2024.2353494

Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: an RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports*, 15(1), pp. 17458. DOI: 10.1038/s41598-025-97652-6

Koivisto, M., & Grassini, S. (2023). Best humans still outperform artificial intelligence in a creative divergent thinking task. *Scientific Reports*, 13(1), pp. 13601. DOI: 10.1038/s41598-023-40858-3

Krakowski, S., Haftor, D., Luger, J., Pashkevich, N., & Raisch, S. (2025). Human-centered artificial intelligence: A field experiment. *Management Science*, *mnsc.2022.03849*. DOI: 10.1287/mnsc.2022.03849

Laverghetta Jr, A., Chakrabarty, T., Hope, T., Pronchick, J., Bhawsar, K., & Beaty, R. E. (2025). How do humans and language models reason about creativity? A comparative analysis. En *arXiv [cs.CL]*. http://arxiv.org/abs/2502.03253

Lee, D., & Palmer, E. (2025). Prompt engineering in higher education: a systematic review to help inform curricula. *International Journal of Educational Technology in Higher Education*, 22(1). DOI: 10.1186/s41239-025-00503-7

Lehmann, M., Cornelius, P., & Sting, F. (2024). AI meets the classroom: When does ChatGPT harm learning? En *arXiv [cs.CY]*. DOI: 10.2139/ssrn.4941259

Leinonen, J., Denny, P., MacNeil, S., Sarsa, S., Bernstein, S., Kim, J., ... & Hellas, A. (2023). Comparing code explanations created by students and large language models. En *arXiv [cs.CY]*. DOI: 10.48550/ARXIV.2304.03938

Liffiton, M., Sheese, B. E., Savelka, J., & Denny, P. (2023). CodeHelp: Using large language models with guardrails for scalable support in programming clases. En *arXiv [cs.CY]*. DOI: 10.48550/ARXIV.2308.06921

Lin, X., Wang, X., Shao, B., & Taylor, J. (2024). How chatbots augment human intelligence in customer services: A mixed-methods study. *Journal of Management Information Systems : JMIS*, 41(4), pp. 1016–1041. DOI: 10.1080/07421222.2024.2415773

Liu, S., Biswal, A., Kamsetty, A., Cheng, A., Schroeder, L. G., Patel, L., ... & Zaharia, M. (2024a). Optimizing LLM queries in relational data analytics workloads. En *arXiv [cs.LG]*. http://arxiv.org/abs/2403.05821

Liu, M., Okuhara, T., Chang, X., Okada, H., & Kiuchi, T. (2024b). Performance of ChatGPT in medical licensing examinations in countries worldwide: A systematic review and meta-analysis protocol. *PloS One*, 19(10), pp. e0312771. DOI: 10.1371/journal.pone.0312771

Liu, Z., Zuo, H., & Lu, Y. (2025). The impact of ChatGPT on students' academic achievement: A meta-analysis. *Journal of Computer Assisted Learning*, 41(4). DOI: 10.1111/jcal.70096

Lo, C., Hew, K., & Jong, M. (2024). The influence of ChatGPT on student engagement: A systematic review and future research agenda. *Computers & Education*, 219(105100), pp. 105100. DOI: 10.1016/j.compedu.2024.105100

Lockhart, E. (2025). Creativity in the age of AI: the human condition and the limits of machine generation. *Journal of Cultural Cognitive Science*, 9(1), pp. 83–88. DOI: 10.1007/s41809-024-00158-2

Lopez-Gazpio, I. (2025). Integrating large language models into accessible and inclusive education: Access democratization and individualized learning enhancement supported by generative artificial intelligence. *Information (Basel)*, 16(6), pp. 473. DOI: 10.3390/info16060473

Luo, X., Rechardt, A., Sun, G., Nejad, K. K., Yáñez, F., Yilmaz, B., ... & Love, B. C. (2025). Large language models surpass human experts in predicting neuroscience results. *Nature Human Behaviour*, 9(2), pp. 305–315. DOI: 10.1038/s41562-024-02046-9

Luppi, A. I., Perl, Y. S., Vohryzek, J., Mediano, P. A., Rosas, F. E., Milisav, F., ... & Kringelbach, M. L. (2024). Competitive interactions shape brain dynamics and computation across species. En *bioRxivorg*. DOI: 10.1101/2024.10.19.619194

Marco, G., Rello, L., & Gonzalo, J. (2024). Small language models can outperform humans in short creative writing: A study comparing SLMs with humans and LLMs. En *arXiv [cs.CL]*. http://arxiv.org/abs/2409.11547

Meyer, J., Jansen, T., Schiller, R., Liebenow, L. W., Steinbach, M., Horbach, A., & Fleckenstein, J. (2024). Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. *Computers and Education: Artificial Intelligence*, 6(100199), pp. 100199. DOI: 10.1016/j.caeai.2023.100199

Mischler, G., Li, Y. A., Bickel, S., Mehta, A. D., & Mesgarani, N. (2024). Contextual feature extraction hierarchies converge in large language models and the brain. *Nature Machine Intelligence*, 6(12), pp. 1467–1477. DOI: 10.1038/s42256-024-00925-4

Munaye, Y. Y., Admass, W., Belayneh, Y., Molla, A., & Asmare, M. (2025). ChatGPT in education: A systematic review on opportunities, challenges, and future directions. *Algorithms*, 18(6), pp. 352. DOI: 10.3390/a18060352

Mutanga, M. B., Msane, J., Mndaweni, T. N., Hlongwane, B. B., & Ngcobo, N. Z. (2025). Exploring the impact of LLM prompting on students' learning. *Trends in Higher Education*, 4(3), pp. 31. DOI: 10.3390/higheredu4030031

Nie, A., Chandak, Y., Suzara, M., Malik, A., Woodrow, J., Peng, M., ... & Piech, C. (2024). The GPT surprise: Offering large language model chat in a massive coding class reduced engagement but increased adopters exam performances. En *arXiv [cs.CY]*. DOI: 10.31219/osf.io/qy8zd

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science (New York, N.Y.)*, 381(6654), pp. 187–192. DOI: 10.1126/science.adh2586

Padiyath, A., Hou, X., Pang, A., Viramontes Vargas, D., Gu, X., Nelson-Fromm, T., ... & Ericson, B. (2024). Insights from social shaping theory: The appropriation of large language models in an undergraduate programming course. En *arXiv [cs.HC]*. http://arxiv.org/abs/2406.06451

Peláez-Sánchez, I., Velarde-Camaqui, D., & Glasserman-Morales, L. (2024). The impact of large language models on higher education: exploring the connection between AI and Education 4.0. *Frontiers in education*, 9. DOI: 10.3389/feduc.2024.1392091

Prasad, P., & Sane, A. (2024). A self-regulated learning framework using generative AI and its application in CS educational intervention design. *Proceedings of the 55th ACM Technical Symposium on Computer Science Education*, 1, pp. 1070–1076. https://dl.acm.org/doi/10.1145/3626252.3630828

Qi, W., Zhu, X., Wang, B., Shi, Y., Dong, C., Shen, S., ... & Cao, S. (2025). Alzheimer's disease digital biomarkers multidimensional landscape and AI model scoping review. *Npj Digital Medicine*, 8(1), pp. 366. DOI: 10.1038/s41746-025-01640-z

Rahe, C., & Maalej, W. (2025). How do programming students use generative AI? En *arXiv [cs.HC]*. DOI: 10.48550/ARXIV.2501.10091

Rashid, Z., & Rötting, M. (2021). Evaluation of manual skill degradation due to automation in apparel manufacturing. *Applied Sciences (Basel, Switzerland)*, 11(23), pp. 11098. DOI: 10.3390/app112311098

Ren, J., Xin, R., Cui, X., Xu, Y., & Li, C. (2024). Quercetin relieves compression-induced cell death and lumbar disc degeneration by stabilizing HIF1A protein. *Heliyon*, 10(17), pp. e37349. DOI: 10.1016/j.heliyon.2024.e37349

Rinta-Kahila, T., Penttinen, E., Salovaara, A., Soliman, W., & Ruissalo, J. (2023). The vicious circles of skill erosion: A case study of cognitive automation. *Journal of the Association for Information Systems*, 24(5), pp. 1378–1412. DOI: 10.17705/1jais.00829

Sabbah, J., & Li, F. (2025). When humans and large language models collaborate, problem-finding illuminates. *Innovation (North Sydney, N.S.W.)*, pp. 1–34. DOI: 10.1080/14479338.2025.2504428

Scholz, N., Nguyen, M. H., Singla, A., & Nagashima, T. (2025). Partnering with AI: A pedagogical feedback system for LLM integration into programming education. En *arXiv [cs.CY]*. http://arxiv.org/abs/2507.00406

Sharma, S., Mittal, P., Kumar, M., & Bhardwaj, V. (2025). The role of large language models in personalized learning: a systematic review of educational impact. *Discover Sustainability*, 6(1). DOI: 10.1007/s43621-025-01094-z

Shen, S. (2024). Application of large language models in the field of education. *Theoretical and natural science - the open access proceedings series for conferences*, 34(1), pp. 147–154. DOI: 10.54254/2753-8818/34/20241163

Silva, C. A. G. D., Ramos, F. N., De Moraes, R. V., & Santos, E. L. D. (2024). ChatGPT: Challenges and benefits in software programming for higher education. *Sustainability*, 16(3), pp. 1245. DOI: 10.3390/su16031245

Stasch, S., & Mack, W. (2025). When automation fails - Investigating cognitive stability and flexibility in a multitasking scenario. *Applied Ergonomics*, 125(104461), pp. 104461. DOI: 10.1016/j.apergo.2024.104461

Sun, L., Yuan, Y., Yao, Y., Li, Y., Zhang, H., Xie, X., ... & Stillwell, D. (2025). Large language models show both individual and collective creativity comparable to humans. *Thinking Skills and Creativity*, 57(101870), pp. 101870. DOI: 10.1016/j.tsc.2025.101870

Veselovsky, V., Ribeiro, M., & West, R. (2023). Artificial artificial artificial intelligence: Crowd workers widely use large language models for text production tasks. En *arXiv [cs.CL]*. http://arxiv.org/abs/2306.07899

Wan, Q., Hu, S., Zhang, Y., Wang, P., Wen, B., & Lu, Z. (2024). It felt like having a second mind: Investigating human-AI co-creativity in prewriting with large language models. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW1), pp. 1–26. DOI: 10.1145/3637361

Wang, Y. (2023). *The large language model (LLM) paradox: Job creation and loss in the age of advanced AI*. DOI: 10.36227/techrxiv.24085824

Wang, J., & Fan, W. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: insights from a meta-analysis. *Humanities & Social Sciences Communications*, 12(1). DOI: 10.1057/s41599-025-04787-y

Wang, S., Xu, T., Li, H., Zhang, C., Liang, J., Tang, J., ... & Wen, Q. (2024). Large Language Models for education: A survey and Outlook. *ArXiv*, *abs/2403.18105*. DOI: 10.48550/arXiv.2403.18105

Wang, X., Lee, C., & Mutlu, B. (2025). LearnMate: Enhancing online education with LLM-powered personalized learning plans and support. En *arXiv [cs.HC]*. http://arxiv.org/abs/2503.13340

Wecks, J. O., Voshaar, J., Plate, B. J., & Zimmermann, J. (2024). Generative AI Usage and Exam Performance. En *arXiv [econ.GN]*. DOI: 10.48550/ARXIV.2404.19699

Welskop, W. (2023). Chatgpt in higher education. *International Journal of New Economics and Social Sciences*, 17(1), pp. 9–18. DOI: 10.5604/01.3001.0053.9601

Yan, Y. M., Chen, C. Q., Hu, Y. B., & Ye, X. D. (2025). LLM-based collaborative programming: impact on students' computational thinking and self-efficacy. *Humanities & Social Sciences Communications*, 12(1). DOI: 10.1057/s41599-025-04471-1

Yang, T., Hsu, Y., & Wu, J. (2025). The effectiveness of ChatGPT in assisting high school students in programming learning: evidence from a quasi-experimental research. *Interactive Learning Environments*, pp. 1–18. DOI: 10.1080/10494820.2025.2450659

Yang, X., Pujara, H., & Li, J. (2025). Learning by teaching: Engaging students as instructors of Large Language Models in computer science education. En *arXiv [cs.CY]*. DOI: 10.48550/ARXIV.2508.05979

Younas, M., Ismayil, I., El-Dakhs, D. A. S., & Anwar, B. (2025). Exploring the impact of artificial intelligence in advancing smart learning in education: A meta-analysis with statistical evidence. *Open Praxis*, 17(3), pp. 594–610. DOI: 10.55982/openpraxis.17.3.842

Zhai, C., Wibowo, S., & Li, L. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learning Environments*, 11(1). DOI: 10.1186/s40561-024-00316-7

Zheng, Z., Qiu, Z., Hu, X., Wu, L., Zhu, H., & Xiong, H. (2023). Generative job recommendations with large language model. En *arXiv [cs.IR]*. http://arxiv.org/abs/2307.02157

Ana Gabriela Banquez Maturana
University of the Cartagena
Colombia
abanquezm@unicartagena.edu.co
ORCID: 0000-0002-8354-6396