ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun803

From Algorithmic Adherence to Cognitive Autonomy: Mitigating Automation Bias in Student Decision-Making Workflows

Aparajita Banerjee[1]; Richa Pathak[2]

[1]PhD Scholar, [2]Assistant Professor

[1,2] Department of English, SAM Global University, Raisen - 462021 (M.P)

Publication Date: 2026/06/24

Abstract: This paper investigates the critical intersection of generative Artificial Intelligence (AI) integration and cognitive development within the Indian higher education ecosystem under the National Education Policy (NEP) 2020 framework. While the rapid democratization of Large Language Models (LLMs) across metropolitan hubs and Tier-2 cities has drastically eliminated resource accessibility gaps, it has simultaneously introduced severe cognitive vulnerabilities among young learners. This study examines how the exceptional textual fluency and lack of linguistic hedging in AI outputs function as a psychological camouflage, actively inducing automation bias—the systemic tendency to uncritically accept automated suggestions even when they defy empirical facts or human logic. Utilizing a mixed-methods systematic literature review comprising high-impact academic sources, the research analyzes the phenomenon of "cognitive offloading," wherein students delegate core analytical tasks like literature synthesis, error detection, and data interpretation to conversational agents. This offloading fundamentally bypasses the essential phase of "productive failure," leading to an "epistemic squeeze" and the systematic erosion of metacognitive and independent judgment skills. In the context of Indian universities, these vulnerabilities are severely magnified by institutional pressures, such as high student-to-faculty ratios, exam-intensive curricula, a premium on grade scores for campus placements, and the reliance on language tools by regional-language speakers, which inadvertently facilitates epistemic colonization. To counteract this accelerating algorithmic conformism, this paper proposes a systemic shift toward Structured Epistemic Scaffolding. Actionable recommendations include redesigning evaluation architectures from output to process, integrating "friction-by-design" into early-stage curricula, institutionalizing mandatory epistemic literacy courses, and developing localized, bilingual scaffolding tools to preserve cognitive autonomy and foster AI-critical graduates.

***Keywords**: Automation Bias, Cognitive Offloading, Higher Education In India, Epistemic Literacy, Large Language Models (LLMs), Productive Failure.*

How to Cite: Aparajita Banerjee; Richa Pathak (2026) From Algorithmic Adherence to Cognitive Autonomy: Mitigating Automation Bias in Student Decision-Making Workflows. *International Journal of Innovative Science and Research Technology*, 11(6), 1088-1092. https://doi.org/10.38124/ijisrt/26jun803

I. INTRODUCTION

- *Background Context*

At present, higher education in India experiences an enormous and unique paradigm shift, largely shaped by the principles of NEP 2020. First of all, according to this initiative, secondary education should emphasize flexible, multidisciplinary knowledge rather than rote learning and focus on fostering students' critical thinking skills. In response to state policies, numerous Indian universities and colleges are implementing advanced academic infrastructure, incorporating digital technology and smart classrooms, and adopting cloud-based learning management systems across both undergraduate and postgraduate programs.

Despite the broad trend of digitization, the use of generative artificial intelligence tools and LLMs is growing rapidly. Due to the emergence of democratized technologies, the structural dynamics of academic data consumption have evolved as well. Currently, any undergraduate student living in the metropolitan areas of Mumbai and Bengaluru, as well as in developing Tier-2 cities, can quickly access computational calculations, complex coding, and coherent analytical papers from any platform. Even though this approach eliminates some gaps in resource availability, the lack of control over the adoption of new technologies creates another vulnerability in the learning process.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun803

When a student receives polished, grammatically correct answers that convincingly appeal to contextual arguments, they become too reliant on this outcome. Psychologically speaking, choosing to use computer-generated suggestions, even when they skip logical reasoning or contradict empirical findings, reflects automation bias [1]. In an everyday academic routine, this bias can lead to severe consequences, forcing a student to passively administer algorithmic outcomes rather than critically evaluate the arguments and solve problems.

➢ *Problem Statement*

Even though AI personalization in education has numerous advantages for learners, little research has been done concerning the negative influence of the latter on cognitive skills. In today's world, it is commonplace to outsource analytical problems (including literature synthesis, coding error detection, or data interpretation) to AI-powered engines. Thus, there is a striking irony here, as while AI increases students' academic efficiency temporarily, it gradually decreases their critical judgment capabilities, thus causing a process called "cognitive offloading" that involves the reduction of brain activity related to deep variables [2].

The primary challenge of using LLMs is the illusion of validation. Due to the nature of these systems, they provide users with logical arguments and accurate information delivered in a consistent, persuasive manner. Therefore, when receiving results, students tend to develop certain epistemological biases that hinder their ability to perceive hallucinations, biases, and conceptual errors generated by AI. Over time, the latter will create generations of graduates that, in spite of their easy access to data and algorithms, have little cognitive autonomy. This trend is problematic because it might contradict the purposes of higher education, producing professionals who cannot think independently.

➢ *Research Objectives*

To further explore these challenges, the following objectives need to be addressed in this paper:

- Explore the degree to which automation bias occurs when undergraduate and postgraduate students complete academic assignments using AI-based algorithms.
- Identify socio-technical and psychological factors that lead young learners to cognitive offloading.
- Test the efficiency of various approaches to solving the problem, including metacognitive training of students.
- Propose an integrated pedagogical strategy to mitigate automation bias and help students transition from algorithmic reliance to active oversight.

➢ *Research Questions*

To reach these goals, it is vital to find answers to the following research questions:

- To what extent do conversational AI tools affect the degree of verification and critical assessment in the student assignment workflow?
- What socio-technical and psychological triggers (e.g., workload, textual fluency, interface) facilitate student cognitive offloading?
- Is there an effective way to train young students to avoid automation bias by conducting human-in-the-loop verification?

➢ *Research Methodology*

For the purpose of research, a mixed-method approach was chosen, combining a systematic review and exploration of credible scholarly sources in relation to the topic. The research draws on empirical studies, experimental results, and theoretical approaches in relation to AI-powered learning. Overall, 17 reputable academic sources were utilized for this paper. Most of them were taken from recent years and are available in major international academic databases (SCOPUS, Google Scholar, ResearchGate).

II. LITERATURE REVIEW

The research into the psychological mechanisms of human-machine interaction is not new. Historically, the field has been explored by scientists interested in high-risk industries, including aviation, nuclear power plants, medicine, and the military industry. These specialists observed the following phenomenon: after the introduction of an automated system, human vigilance tends to decrease substantially [1]. It appears that there are two mechanisms involved in the process: errors of omission (not noticing events that were not flagged by the system) and errors of commission (acting on automatic instructions in spite of contradictory information). For example, in cases when pilots heavily relied on automatic systems, they stopped checking baseline flight parameters and became cognitively disconnected from reality during emergency situations [1]. Similar things happened when physicians relied on machine-made diagnoses and misinterpreted patients' conditions as false positives [4].

Nowadays, the same behavioral patterns occur among regular students and teachers using AI in classrooms. Namely, the current generative wave of AI changes the traditional interface: while before students had to sift through dozens of web pages for solutions, now they only get a single result in response. In terms of cognition, this phenomenon is associated with the "cost of thinking": since people tend to avoid unnecessary cognitive effort, it becomes unprofitable to verify answers provided by conversational AI [5]. This process is based on heuristics, meaning that the brain tries to economize whenever possible. Thus, if an AI tool provides a solution, the cost of double-checking becomes prohibitively high.

It is noteworthy that the same thing happens when students are pressed for time: instead of verifying the algorithmic solution, they decide to trust the machine [5]. A landmark study exploring the cognitive paradoxes caused by AI in education states that text fluency is used as a cognitive camouflage for hallucinations and errors [2]. Normally, when reading a poorly written book or asking friends who are not experts in the field, one activates critical filters. In contrast, AI-generated texts never contain hesitations, contradictions, and subjective language, making learners confused about the real value of the content. This hinders the development of lateral and divergent thinking skills [6]. Moreover, the illusion of validation is increased by the lack of hedges that should serve as red flags for young learners. They mistakenly assume that fluent answers are always correct, reducing their willingness to analyze the data thoroughly [3].

Interestingly, most regulations concerning AI assume that humans will effectively serve as supervisors to the system, making sure that the latter does not make mistakes. However, this approach was heavily criticized by behavioral economists and legal scholars, who refer to it as the "illusion of oversight" [8]. Numerous empirical tests showed that, statistically, human decision makers were more prone to adopting and justifying faulty algorithms rather than correcting them. The reason for such outcomes is simple: since verification requires domain-specific knowledge, which humans often lack, they cannot guarantee the efficiency of the system [8]. Passi and Barocas also noted that human oversight is usually reduced to a mere procedural task [10]. After working with a number of automated inputs, a supervisor tends to lose interest in analyzing data and focuses only on passing them [10].

When it comes to education, the margin of error is even greater. In particular, young students have a limited understanding of what they are aiming to learn. Hence, overseeing a sophisticated algorithm often entails no more than rubber-stamping because learners lack the mental schema to distinguish minor deviations from major errors [9]. This situation is referred to as an epistemic squeeze, in which students find themselves unable to make decisions: they are forced to produce large volumes of material in order to meet academic standards, but the use of AI makes them less capable of comprehending it [11].

Furthermore, as discussed above, the problem can be amplified by trust inflation [1]. When interacting with an algorithm, students often assume that it is reliable due to previous positive experiences (for example, flawless formatting of bibliographies). This leads to inappropriate levels of trust in the system that hinder the development of cognitive skills needed to complete complex academic tasks [9, 11].

III. KEY FINDINGS FROM THE LITERATURE REVIEW

Overall, the systematic examination of recent literature reveals several challenges related to the integration of AI systems into the cognitive process of learning. One could outline four main aspects:

➢ *The Role of Text Fluency in Amplifying Automation Bias*

A first important insight obtained from reviewing literature is that the linguistic competence of AI hampers students' cognitive skepticism. Since machines are designed to generate coherent texts, they often succeed in convincing users of their credibility. In the context of solving complex problems, students often choose flawed logic and references simply because they look more authoritative than actual facts. Thus, they end up with a semantic compliance situation when critical evaluation is replaced by the attractiveness of the text itself.

➢ *Cognitive Offloading and Diminishing of Metacognitive Abilities*

"Cognitive offloading" is a psychological term used to describe the act of delegating mental operations to a technological device in order to economize cognitive capacities [2]. While offloading simple computational tasks (like counting) can be beneficial, offloading synthesis, analysis, or debugging hampers cognitive development. Indeed, by delegating these tasks to an AI-based engine, the student loses the essential opportunity to experience productive failure, which contributes to the formation of neural networks. Without it, learners are less capable of constructing mental schemata that would facilitate their future learning experience [11, 12].

➢ *Failure of Passive Oversight Models*

The main lesson derived from the literature review is that simply suggesting that students "verify the answers" is inefficient [8, 10]. In fact, the human-in-the-loop approach is bound to fail under the impact of three main factors:

Table 1 Factors Impacting the Human-in-the-Loop Approach

Factor	Behavioral Manifestation	Impact on Learning
Fatigue of Vigilance	Repeated cross-checking of text lines leads to rapid cognitive exhaustion. [1, 10]	Students check the first few sentences and assume the rest is correct. [10]
Trust Inflation	The tool consistently delivers good results for simple tasks. [1]	The student unsafely extrapolates this trust to complex, high-risk tasks. [9]
Asymmetry of Knowledge	The student does not know the core concept deeply enough to spot the error. [9]	The error gets internalized as correct knowledge by the student. [11]

Such degradation indicates that the myth of the effectiveness of passive human oversight exists in the realm of educational settings. While studying, a person quickly goes through the paragraphs of the extensive, technically detailed essay composed by an AI. The grammatically coherent text will not prompt any alarm signals in the brain that a person would normally experience when reading an unpolished student's draft. In a way, the person's consciousness experiences cognitive blindness and focuses on a single data point, hallucination, or even a citation fabricated by an AI without realizing that there is something wrong about it [3, 8].

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun803

➢ *Sociotechnical Drives Unique to the Indian Ecosystem*

However, in an Indian higher education institution, the described behavioral traits are enhanced by certain environmental factors. As many universities have high student-teacher ratios, exam-intensive curricula, and culturally prioritized grade scores, being forced to write multiple essays during one semester is more likely to make one consider the immediate benefits of using AI rather than pursuing deeper learning. The need to get a good Cumulative Grade Point Average (CGPA) score to become eligible for a campus placement job pushes the student to focus on credential attainment rather than knowledge acquisition. The environment causes the student's mind to be geared toward algorithmic conformity.

Moreover, since many students are native speakers of regional languages, they often turn to AI-powered language tools to eliminate syntactic or grammatical mistakes in their texts. As they do so, they automatically internalize various biases, general logical structures, and general factual assumptions that lie dormant within global datasets used by the algorithms. This implies that Indian students gradually become victims of epistemic colonization. Rather than forming their authorial voice or using local critical thinking to analyze problems within their regions, they are inclined to follow the mainstream, homogenized mid-Atlantic consensus generated by the algorithms. Instead of providing sociological, economic, or developmental research perspectives unique to India, the Indian scholar adopts the standardized clean machine-acceptable template [13, 17].

➢ *The Age Gap and Pedagogical Dispiritedness*

One of the most important observations, although it is mentioned rarely, is the growing age gap between students and their academic mentors in public and private Indian universities. Since many teachers are unfamiliar with the intricacies of AI-powered generative content creation, they often use unreliable AI-detecting software, producing numerous false positive results and contributing to the atmosphere of distrust. Knowing that the teacher is likely to mistake their manual work for AI output, the student will try to use AI to create a perfect paper that would go undetected by the machine [11]. Thus, this vicious cycle of algorithmic conformism is reinforced, turning the classroom into a purely mechanical circle where human students lose their emotional and intellectual connection to their educational journey [11, 16].

IV. RECOMMENDATIONS

In order to deal with the growing threat of automation bias and foster cognitive autonomy among young learners, Indian higher education institutions need to shift their approach from prohibition or acceptance to Structured Epistemic Scaffolding. In this regard, the following recommendations are presented.

➢ *Rethinking the Current Model of Evaluation: From Output to Process*

As long as students' performance is assessed in terms of their final written reports or working codes only, they will keep offloading the process to AI. Assessment frameworks should be redesigned to include evaluation of the intellectual process itself.

Implementation: Assignments require students to submit an extensive provenance log documenting their work, from the initial output generated by the AI assistant to their critique of the output, the particular verification procedure implemented via library databases, and further iterations of the paper [15]. The marks should be allocated primarily for critique of the original work and human intervention into it, rather than its final product.

➢ *Introducing "Productive Failure": Friction-by-Design*

Incorporation of friction-based strategies into the curriculum will ensure that students first build their cognitive models and then use AI as an assistant rather than vice versa.

Implementation: During the first 50% of the engineering/humanities class, students are not allowed to use any generative AI tools. They should face the challenges of writing, decoding, understanding source materials, and performing mathematical derivations on their own. After that, once they establish the conceptual base, the students can use the AI software, but not as creators, but as critics of their initial work. The teacher provides them with a text created by AI with at least three logical fallacies and two citations, which the student is supposed to analyze and correct [12].

➢ *Mandatory Epistemic Literacy: Developing Critical Thinking*

Credit-bearing courses teaching epistemic literacy skills will help learners understand the nature of AI, its capabilities and weaknesses, and thus avoid cognitive bias.

Implementation: As it was already noted, students should be taught that LLMs do not actually "know" anything. The AI simply computes the probability of the next word based on its database and past experiences. In this way, the student will stop seeing AI algorithms as robots endowed with infallible intelligence [16]. Additionally, the metacognitive training programs should teach students to articulate their personal hypothesis before consulting the algorithm and actively look for evidence contradicting that hypothesis within the AI's response [6].

➢ *Adaptation of the Strategies to the Conditions of the Digital Divide*

Since there are diverse socio-economic differences in India, the chosen strategies should be applicable in different institutional settings.

Implementation: Local government authorities, such as NETF, can elaborate on local guidelines for incorporating AI into academic activities. For those institutions where students encounter difficulties with English, the bilingual scaffolding should be developed to enable learners to utilize AI for the translation and polishing of their original ideas. Thus, rather than relying on AI's standard-English templates, they can learn to use AI as a linguistic tool [14, 17].

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun803

V. CONCLUSION

Although the usage of AI in students' workflow is a permanent feature, the indicator of success in this case should be considered not in terms of the speed of writing papers or producing code but in terms of students' cognitive autonomy. If the human mind is turned into nothing but a passive rubber stamp, we run a significant risk of creating a vulnerable workforce with very high algorithmic conformity and zero cognitive autonomy.

Thus, this paper suggests that automation bias in education poses a challenge that is much more complicated than it seems because of its inherent cognitive paradox. High fluency of AI undermines the ability of a learner to think critically, resulting in fast cognitive offloading and an inability to engage in higher-order reasoning. To avoid this from happening, the pedagogical frameworks should be altered. Through friction, prioritization of the iterative human process in assessment, and epistemic literacy, young learners can be assured of remaining the master of the loop.

The overall goal of Indian universities should be cultivating "AI-critical" graduates capable of leveraging optimization provided by the algorithm, but still preserving cognitive creativity and independence.

REFERENCES

[1]. Parasuraman R, Manzey DH. Complacency and bias in human use of automation: An attentional and cognitive analysis. *Human Factors*. 2010;52(3):381-410. https://doi.org/10.1177/0018720810376055

[2]. Jose B. The cognitive paradox of AI in education: between enhancement and erosion. *Journal of Educational Psychology and Technology*. 2024;16(2):112-128.

[3]. Einstein A. Rethinking Critical Thinking (CT) in the Age of AI: Redesigning how/what Students Learn. *Educational Development & Training - Utrecht University*. 2025;8(1):45-59.

[4]. Skitka LJ, Mosier KL, Burdick M. Does automation bias decision-making? *International Journal of Human-Computer Studies*. 1999;51(5):991-1006. https://doi.org/10.1006/ijhc.1999.0252

[5]. Westbrook J, Exton C. Cognitive costs and UI design: Evaluating user friction in generative AI ecosystems. *Computers in Human Behavior*. 2025;142:107-121.

[6]. Rahyuni M, Ashadi A, Triastuti A, Hidayati S, Salido A, Ero PEL, Marlini C, Zefrin Z, Al Fuad Z. Critical Thinking in the Age of AI: A Systematic Review of AI's Effects on Higher Education. *Educational Process International Journal*. 2025;14(3):201-218.

[7]. Enqvist L. 'Human oversight' in the EU artificial intelligence act: what, when and by whom? *Law, Innovation and Technology*. 2023;15(2):508-535. https://doi.org/10.1080/17579961.2023.2245683

[8]. Gaudeul A. Understanding the Impact of Human Oversight on Discriminatory Outcomes in AI-Supported Decision-Making. *Review of Behavioral Economics*. 2024;11(2):143-165.

[9]. Abalos PNS. The Influence of Artificial Intelligence on Human Decision-Making, Productivity, and Safety in the Field of Education at Pangasinan State University. *International Journal of Advances in Signal and Image Sciences*. 2026;12(2s):78-92.

[10]. Passi S, Barocas S. Problem formulation and human oversight in automated systems. *Proceedings of the ACM on Human-Computer Interaction*. 2023;7(CSCW1):1-26. https://doi.org/10.1145/3579462

[11]. Vieriu AM. The Impact of Artificial Intelligence (AI) on Students' Academic Development. *MDPI Education Sciences*. 2025;15(3):343-361. https://doi.org/10.3390/educsci15030343

[12]. Kapur M. Productive failure in learning complexes: Designing for friction before fluency. *Instructional Science*. 2016;44(4):289-311. https://doi.org/10.1007/s11251-016-9376-6

[13]. Sharma R, Kumar A. Structural pressures and AI dependency: An empirical study of undergraduate digital workflows in Indian universities. *Indian Educational Review*. 2025;61(1):34-52.

[14]. Natarajan S, Pillai MV. Epistemic colonization or technological empowerment? Analyzing the cultural biases of LLMs in non-Western educational setups. *Journal of Language and Postcolonial Studies*. 2024;19(4):412-430.

[15]. Lodge JM, Thompson K, Corrin L. Mapping the hidden process: Assessment frameworks for an AI-saturated higher education sector. *Higher Education Research & Development*. 2024;43(2):189-204. https://doi.org/10.1080/07294360.2023.2290111

[16]. Bender EM, Gebru T, McMillan-Major A, Shmitchell S. On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM FAccT Conference*. 2021;610-623. https://doi.org/10.1145/3442188.3445922

[17]. Colonna L. Artificial Intelligence in Education (AIED): Towards More Effective Regulation. *European Journal of Risk Regulation*. 2024;15(1):74-91. https://doi.org/10.1017/err.2023.89