



Measuring cognitive debt in ai supported education

Evaluando la deuda cognitiva en la educación asistida con IA

A. Martínez Gordon^{1*}, M. I. Prieto Barrio², A. Leal Matilla³

¹ Escuela Técnica Superior de Edificación, Universidad Politécnica de Madrid, Email: alejandro.mgordon@upm.es

² Escuela Técnica Superior de Edificación, Universidad Politécnica de Madrid, Email: mariaisabel.prieto@upm.es

³ Escuela Técnica Superior de Edificación, Universidad Politécnica de Madrid, Email: alberto.leal.matilla@upm.es

Received: 05/01/2026 | Accepted: 19/01/2026 | Publication date: 20/01/2026
DOI: 10.20868/abe.2025.3.5649

TITULARES

- Cuantifica la deuda cognitiva y el sesgo de automatización en titulados de ingeniería de la edificación, proponiendo un marco metodológico y índices reproducibles.
- Confirma incrementos significativos de productividad derivados del uso de LLMs genéricos, mientras que la fiabilidad se degrada a medida que aumenta la complejidad de la tarea, lo que justifica un uso basado en la gestión del riesgo.
- Evidencia una fuerte demanda de supervisión humana y de herramientas de dominio validadas, así como disposición a pagar cuando los mecanismos de gobernanza son explícitos.
- Propone la integración de un Asistente Virtual de Curso (VCA/AVA) como herramienta controlada y acotada al currículo, orientada a reducir las alucinaciones y preservar las competencias críticas.

HIGHLIGHTS

- Quantifies cognitive debt and automation bias in building-engineering graduates suggesting a framework and reproducible indices.
 - Confirms large productivity gains from generic LLMs, while reliability collapses as task complexity increases, motivating risk usage.
 - Shows strong demand for human oversight and validated domain tools, with willingness to pay when governance is explicit.
 - Proposes integrating a Virtual Course Assistant (VCA/AVA) as a controlled, curriculum-bounded tool to reduce hallucinations and preserve critical skills.
-

RESUMEN

Este estudio analiza el impacto de la IA generativa en el comportamiento formativo en ingeniería de edificación mediante la estrategia FOCUS. Se administró una encuesta tipo “Likert” de 23 ítems (1–5) a 35 estudiantes. Las respuestas se normalizaron y se agruparon en índices que incluyen dependencia de la automatización (R), compromiso analítico (AE), atrofia del conocimiento (KA), percepción de alucinaciones (H), beneficios percibidos (B), preferencia por herramientas especializadas (SvG) y respaldo a la supervisión humana (HE). Los resultados muestran una elevada percepción de ahorro de tiempo, pero una confianza limitada en los resúmenes generados por IA y una percepción frecuente de alucinaciones. Los estudiantes declararon una alta dependencia de la automatización ($R = 0,711$) junto con un comportamiento significativo de verificación ($AE = 0,693$), lo que se traduce en una deuda cognitiva moderada ($CDI = 0,570$) y un sesgo de automatización bajo-moderado ($ABI = 0,219$). Los resultados evidencian un equilibrio delicado entre ganancias de productividad y riesgos de erosión cognitiva en una disciplina crítica para la seguridad, y respaldan la adopción de herramientas educativas gobernadas, como un Asistente Virtual de Asignatura (VCA), para preservar las capacidades analíticas sin renunciar a la eficiencia que aporta la IA.

Palabras clave: *inteligencia artificial generativa, deuda cognitiva, sesgo de automatización, educación en edificación, asistente virtual de asignatura.*

ABSTRACT

This study examines how generative AI affects research behavior and cognitive engagement in graduate building engineering education using the FOCUS measurement framework. A 23-item “Likert” survey (1–5) was administered to 35 graduate students after completing core-engineering tasks without AI assistance. Responses were normalized and grouped into constructs including automation reliance (R), analytical engagement (AE), knowledge atrophy (KA), hallucination perception (H), perceived benefits (B), preference for specialized tools (SvG), and human oversight endorsement (HE). Results indicate strong perceived timesavings for literature review but limited trust in AI-generated summaries and frequent hallucination perception. Students reported high automation reliance ($R = 0.711$) alongside substantial verification behavior ($AE = 0.693$), resulting in moderate cognitive debt ($CDI = 0.570$) and low-to-moderate automation bias ($ABI = 0.219$). Human oversight endorsement was high ($HE = 0.846$), with a clear preference for specialized over generic tools ($SvG = 0.658$). The findings highlight a trade-off between productivity gains and risks of cognitive erosion in a safety-critical discipline and motivate the adoption of governed educational tools such as a Virtual Course Assistant (VCA) to preserve analytical skills while retaining AI-enabled efficiency.

Keywords: *generative artificial intelligence, cognitive debt, automation bias, building education, virtual course assistant.*

1. INTRODUCTION

Generative artificial intelligence systems—most notably large language models (LLMs) and large reasoning models (LRMs)—are rapidly permeating the daily practices of building engineering education and research. Their appeal is straightforward: they dramatically reduce the time required for literature review, contextualization of unfamiliar topics, preliminary feasibility checks, drafting of technical narratives, and even the interpretation of experimental or project-based data. For graduate students and early-career engineers, these tools operate as powerful cognitive accelerators, lowering the entry barrier to complex domains and enabling faster resolution to enduring problems. In an academic environment increasingly shaped by productivity pressures, AI appears to offer an efficient shortcut through tasks that were traditionally time-intensive and cognitively demanding [1-3].

However, building engineering occupies a distinctly safety-critical position within the broader engineering landscape. Decisions made at the design, specification, or assessment stage can have direct and irreversible consequences on structural integrity, fire safety, durability, regulatory compliance, and long-term service performance. Within such a context, the integration of generative AI introduces not only opportunities but also systemic risks that differ fundamentally from those associated with conventional digital tools. Two of these risks are particularly worrisome. The first is automation bias, defined as the human tendency to over-rely on automated systems and to accept their outputs even when they conflict with domain knowledge, physical intuition, or empirical

evidence. The second is cognitive debt, a longer-term and less visible phenomenon referring to the gradual erosion of analytical reasoning, conceptual understanding, and problem-solving autonomy resulting from the repeated delegation of cognitively intensive tasks to AI systems [4-5].

It is clear that generative AI delivers undeniable short-term productivity gains, however, its unregulated use risks introducing subtle but cumulative failure modes that are difficult to detect in educational settings. Unlike traditional calculation software or rule-based tools, LLMs produce fluent, persuasive, and contextually rich outputs. This rhetorical competence can mask logical inconsistencies, incorrect assumptions, or fabricated content, thereby increasing the likelihood that errors may pass unnoticed—particularly among less experienced users. Importantly, empirical evidence suggests that as task complexity increases, especially in multi-constraint engineering problems involving interacting physical, regulatory, and environmental variables, where models deteriorate progressively [6-8].

Within higher education, these dynamics raise fundamental questions about the certifications and backgrounds of future engineers as competent professionals. If AI systems consistently perform the most cognitively demanding, steps of analysis, synthesis, and interpretation, students may complete their degrees with reduced exposure to the very reasoning processes that underpin safe engineering judgment. Over time, this may result in cohorts of graduates who are highly efficient users of AI tools but insufficiently equipped to detect errors, challenge

assumptions, or operate independently when automated assistance fails or is unavailable [3-4].

Against this backdrop, the present work seeks to move beyond anecdotal concern by providing a quantitative framework for examining AI-related cognitive risks in building engineering education. Using the FOCUS measurement framework, the study translates abstract concepts such as reliance, verification, hallucination awareness, and skill atrophy into observable constructs and indices, enabling the empirical assessment of cognitive debt and automation bias in graduate students. This diagnostic approach is essential: without measurable indicators, institutions lack the ability to monitor trends, compare cohorts, or intervene proactively.

Crucially, the contribution of this work is not limited to risk identification. Building on the observed patterns, it proposes a pragmatic mitigation pathway through the integration of a Virtual Course Assistant (VCA/AVA). Unlike generic conversational AI, a VCA is conceived as a curriculum-bounded, governance-aligned educational tool. Its role is not to replace human reasoning, but to scaffold it: constraining outputs to validated course materials, enforcing explicit verification steps, and embedding human oversight as a structural requirement rather than an optional behavior. By design, such a system aims to preserve analytical competence, reduce hallucination exposure, and limit the accumulation of cognitive debt while retaining the efficiency benefits that make AI attractive in the first place.

In this sense, the study positions generative AI not as an external disruption to engineering education, but as a technology whose impact depends critically on how it is framed, constrained, and institutionalized. When left

ungoverned, AI risks accelerating productivity at the expense of professional judgment; when embedded within structured pedagogical tools such as a VCA, it can instead function as a controlled cognitive aid that supports learning without undermining it. This work therefore contributes to the emerging discourse on responsible AI adoption in higher education by offering both a measurement lens and an actionable design strategy tailored to the specific demands of building engineering.

2. METHODOLOGY

A structured 23-item Likert survey, listed on table 1, (1 = strongly disagree, 5 = strongly agree) was administered to $N = 35$ graduate students in building engineering after they completed required engineering tasks with and without AI assistance. Items covered literature retrieval and summarization, feasibility prediction, data interpretation, dependence, willingness to pay for specialized tools, and oversight/ethics.

For each item i the responses were recorded as percentages shares $X_{i,1}, X_{i,2}, X_{i,3}, X_{i,4}, X_{i,5}$ that sum to 100. To convert raw pie charts or percentages responses into interpretable metrics, each question was assigned to one or more categories or constructs representing cognitive and behavioral dimensions relevant to AI interaction and engagement in building engineering. For every item i , the arithmetic mean across participants was first computed as expressed below in equation (1).

$$\bar{x}_i = \frac{1}{N} \sum_{j=1}^N x_{i,j} \quad \text{equation (1)}$$

2.1. Normalization and constructs

Item means were computed and linearly normalized to [0,1]. Items were assigned to constructs:

- R: automation reliance
- AE: analytical engagement (verification and preference for human judgment)
- KA: knowledge atrophy (self-reported erosion/dependence-linked risk)
- H: hallucination perception

$x_{i,j}$ represents the score given by the respondent to item i , while the sample size or total number of participants $N = 35$. Afterwards, these mean values [1-5] were normalized to the interval [0,1] by implementing linear scaling as seen on equation (2).

$$n_i = \frac{\bar{x}_i - 1}{4} \quad \text{equation (2)}$$

This normalization ensures comparability across all items, regardless of their raw 1-5 scale. Construct scores were then defined as the average value of all normalized items assigned to each construct.

2.2 Derived indices

These overlaps are deliberate modeling decisions, strictly based on the recognition that AI use in building engineering often yields simultaneous short-term benefits and long-term liabilities. For instance, reliance on AI for drawing conclusions accelerates workflow (a benefit) while risking superficial reasoning (a liability). Explicitly modeling this overlap allowed the FOCUS framework to capture the dual nature of cognitive debt in construction practices.

The constructs were further synthesized into higher-order indices. The Cognitive Debt Index

- B: perceived benefits/productivity
- SvG: specialized vs generic preference
- HE: human oversight endorsement

Overlapping assignments were intentionally permitted, reflecting real-world coupling between short-term gains and long-term liabilities.

(CDI) was defined as the mean of disengagement (1-AE), reliance (R), knowledge atrophy (KA), and hallucination perception (H), as expressed in equation (3).

$$CDI = \frac{(1 - AE) + R + KA + H}{4} \quad \text{equation (3)}$$

The Automation Bias Index (ABI) was defined as the product of reliance and disengagement, as shown in equation (4).

$$ABI = R * (1 - AE) \quad \text{equation (4)}$$

Equation (5) was implemented to determine the agreement percentage considering only the top scores to evaluate the proportion of participants that favoured each item of the survey. While the net agreement percentage points considered both top and bottom answers, as shown in equation (6)

$$P_{agree} = x_{i,4} + x_{i,5} \quad \text{equation (5)}$$

$$Net = (x_{i,4} + x_{i,5}) - (x_{i,1} + x_{i,2}) \quad \text{equation (6)}$$

Lastly, a dependence index (D) was extracted directly from Item 16, and normalized to [0,1]. These composite indices quantify the trade-off between efficiency and cognitive erosion, directly informing implications for engineering education and professional practices.

Table 1: Implemented survey

Measuring Cognitive Debt in AI Supported Education
Alejandro Martínez Gordon, María Isabel Prieto Barrio, Alberto Leal Matilla

Item	Question	Construct
1	AI (e.g., ChatGPT, Gemini) reduces the time required to perform literature reviews.	B
2	AI-generated summaries for scientific articles are as reliable as those produced by humans.	B
3	Specialized academic tools (e.g., Edison scientific) are more accurate than generic AI (Grok, Gemini) at identifying relevant literature.	SvG
4	AI frequently generates false claims or data (“hallucinations”) in academic searches.	H
5	I would use AI to simulate project feasibility (e.g., material properties) before laboratory experiments.	R
6	AI predictions of feasibility are accurate enough to make decisions.	AE
7	Domain-specialized AIs (e.g., Edison scientific) provide more reliable predictions than generic models (Grok, ChatGPT).	SvG
8	AI reduces costs by replacing preliminary studies.	B
9	AI improves my ability to detect patterns in complex data.	B
10	I trust expert human interpretation of data more than AI.	AE, HE
11	Domain-specific AI tools interpret experimental data better than generic models.	SvG
12	When AI and my analysis disagree, I prioritize my human judgment.	AE, B
13	AI accelerates extracting conclusions from experiments.	R, B
14	Use of AI decreases my analytical depth when solving problems.	AE
15	I always verify original sources cited by AI.	AE, HE
16	I feel dependent on AI to advance my research.	KA
17	AI inhibits the development of critical skills in students.	KA
18	I would pay for a specialized academic AI if it guarantees greater accuracy than free tools.	R
19	Generic AIs (ChatGPT, Gemini, Grok) are sufficient for most academic tasks.	R, KA
20	Tools designed by academics offer more useful features.	SvG
21	Human supervision is indispensable even when using specialized AI.	HE
22	The university should include ethical AI training in curricula.	HE
23	AI increases my research productivity without compromising quality.	R, B

3. RESULTS

Analysis of individual survey items reveals a clear and internally consistent pattern in how graduate students perceive and use generative AI during building engineering tasks. At the item level, responses reflect a strong recognition of instrumental efficiency gains, coupled with persistent skepticism regarding epistemic reliability and decision-level trust.

3.1. Survey and behavioral patterns

Participants overwhelmingly agreed that generative AI significantly reduces the time required for literature review and contextual familiarization, as exhibited in table 2. These values confirm that AI is primarily perceived as a productivity accelerator at the information acquisition stage. In practical terms, students appear to rely on AI to compress the early exploratory phases of academic and technical work, including keyword expansion, topic scoping, and rapid synthesis of large bodies of text.

Table 2: Constructed indices

R	AE	KA	H	B
0.711	0.693	0.609	0.653	0.677
SvG	HE	CDI	ABI	D
0.658	0.846	0.570	0.219	0.448

In contrast, trust sharply declines when AI outputs transition from supportive to substitutive roles. Confidence in AI-generated summaries being as reliable as those produced by humans was notably low. This divergence between perceived speed and perceived reliability is critical: it indicates that students distinguish clearly between efficiency and epistemic authority. AI is valued as a fast assistant, but not as an autonomous knowledge arbiter.

Perception of hallucinations further reinforces this cautious stance. Respondents reported frequent encounters with fabricated, inconsistent, or unverifiable information ($H=0.653$). Importantly, this value exceeds the midpoint of the normalized scale, suggesting that hallucinations are not viewed as rare anomalies but as a recurrent characteristic of current AI systems. This perception likely contributes to the observed reluctance to fully trust AI outputs without verification.

The same pattern emerges in feasibility-related tasks, which are particularly relevant in building engineering. While students expressed moderate willingness to use AI for early-stage feasibility assessments their confidence in the accuracy of AI predictions being sufficient to directly inform decisions was low. This discrepancy highlights a functional boundary: AI is currently considered as a screening or exploratory tool, but not as a determinant in decision-making processes that may carry structural, safety, or regulatory implications.

Taken together, these results indicate that students do not reject AI outright, nor do they uncritically embrace it. Instead, they adopt a selective reliance strategy, exploiting AI where perceived risk is low and reverting to human judgment when stakes increase.

3.2. Constructed indices

The construction of indices has provided a more comprehensive picture regarding cognitive behaviour. For starters, automation reliance construct shows a relatively high value ($R = 0.711$), indicating that AI tools are deeply embedded in students' workflows. This reliance does not necessarily imply blind trust; rather, it reflects frequent and habitual use, particularly in the early and intermediate stages of task execution.

Simultaneously, analytical engagement remains high ($AE = 0.693$). Students report consistent behaviors related to verification, cross-checking, and human judgment override. This coexistence of high reliance and high engagement is a key finding: it suggests that, at least at the graduate level, AI use has not yet displaced critical thinking entirely. Instead, AI appears to be layered on top of existing analytical practices.

However, this apparent equilibrium is destabilized by the knowledge atrophy construct ($KA = 0.609$). Despite active verification behaviors, students acknowledge a gradual erosion of independent analytical effort when AI is readily available. This self-reported atrophy does not imply immediate incompetence, but rather signals a long-term risk trajectory, where repeated delegation of reasoning tasks may weaken cognitive endurance, problem decomposition skills, and domain intuition over time.

The hallucination perception construct ($H = 0.653$) reinforces this concern. Awareness of hallucinations does not eliminate their impact; instead, it introduces an additional cognitive burden, requiring constant vigilance. Over extended periods, this vigilance may paradoxically contribute to fatigue and selective trust, especially under time pressure.

On the benefit side, perceived benefits remain substantial ($B = 0.677$), confirming that students experience real value from AI-assisted workflows. Importantly, these benefits are not purely abstract; they translate into faster task completion, smoother drafting processes, and reduced friction in navigating unfamiliar domains.

Preferences regarding tool design are also unambiguous. The specialized-versus-generic construct ($SvG = 0.658$) indicates a clear

inclination toward domain-specific AI tools over general-purpose conversational models. This suggests that students recognize the limitations of generic systems in safety-critical disciplines and implicitly favor bounded, validated, and context-aware solutions.

Finally, human oversight endorsement reaches one of the highest construct values ($HE = 0.846$). This result is decisive: students overwhelmingly agree that AI outputs should remain subject to human review, particularly in engineering contexts. Oversight is not perceived as an optional safeguard, but as a structural necessity.

From these constructs, the derived indices provide a compact synthesis of risk and behavior. The Cognitive Debt Index ($CDI = 0.570$) indicates a moderate but non-negligible accumulation of latent cognitive risk. This value reflects the tension between sustained analytical engagement and creeping atrophy. The Automation Bias Index ($ABI = 0.219$) remains relatively low, suggesting that overt over-trust is currently limited. However, the presence of dependence ($D = 0.448$) signals that reliance is already substantial and may amplify bias effects under stress, time constraints, or increased task complexity.

3.3. Productivity gains

Collectively, the results depict a transitional cognitive regime. Productivity gains from generative AI are unequivocal, and students are neither naïve nor complacent regarding its limitations. Verification practices are widespread, and awareness of hallucinations is high. Nevertheless, the data reveal that cognitive debt can accumulate even in the presence of caution. The combination of high reliance, moderate atrophy, and persistent hallucination exposure suggests that erosion of deep reasoning is not a

function of misuse alone, but can arise from normal, well-intentioned use.

This finding is particularly consequential for building engineering education. The discipline relies heavily on layered reasoning, physical intuition, and the ability to detect implausible results before they propagate into design or construction decisions. The observed pattern implies that without deliberate pedagogical and technological interventions, students may graduate with strong AI-assisted efficiency but weakened resilience in independent reasoning.

In this sense, the results do not argue against the use of generative AI. Instead, they underscore the necessity of structured integration, where AI is embedded within educational frameworks that actively preserve analytical competence rather than passively assuming it will persist.

4. DISCUSSION

Productivity is real, but precision collapses with complexity as it is revealed that generic models are effective at low-to-medium complexity tasks (context gathering, drafting, idea generation), but reliability collapses when tasks demand multi-constraint reasoning (e.g., interacting load combinations, fire strategy, moisture/thermal coupling, interpolation outside measured ranges). This aligns with broader critiques that reasoning performance degrades sharply as complexity rises, producing confident but fragile outputs [2, 5-6, 8-9].

Furthermore, specialized tools plus governance outperformed “general chat” or LLMs as respondents preferred domain-specialized tools (SvG = 0.658) and strongly endorsed human oversight (HE = 0.846). This behaviour is consistent with AEC-oriented recommendations: use AI to accelerate workflows, but constrain it

through validated tooling, traceability, and verification protocols [1-2,5-6,10].

As such, it is clear that to ensure an analytical competence in future professionals the emergence of tools such as virtual course assistants (VCA or AVA in Spanish) would provide a mechanism to introduce “specialized and governed” AI to future building engineers by embedding control, accountability, and pedagogical intent directly into the AI interaction itself. Rather than functioning as an unrestricted conversational agent, the VCA constrains its outputs to course-specific content, relevant standards excerpts, verified datasets, and instructor-approved methodologies, thereby significantly reducing exposure to hallucinated or out-of-scope information [2, 5-6].

At the same time, it enforces structured verification behaviors by requiring explicit citations to course materials or normative documents and by integrating mandatory intermediate checking steps before any final answer is delivered, reinforcing analytical engagement and limiting automation bias. Crucially, the VCA is designed not only to support learning but to preserve competence over time, deliberately incorporating periodic AI-free exercises and requiring manual derivations or reasoning steps before allowing AI-assisted extensions, directly counteracting the gradual knowledge atrophy observed in unrestricted AI use [9-10].

Finally, by maintaining auditable logs of prompts, sources, model outputs, and human overrides, the VCA enables program-level monitoring of cognitive debt and automation bias trends, allowing early detection of unhealthy dependency patterns as a self-correcting procedure. In this way, the VCA will directly address the observed risk profile, by preserving

the benefits of AI while structurally limiting cognitive erosion.

5. CONCLUSIONS

From a practical standpoint, the integration of generative AI into building engineering programs must be explicitly structured to reflect the safety-critical nature of the discipline. Generic large language models can be tolerated for low-risk activities such as drafting, contextualization, or exploratory literature navigation, but their use should be progressively constrained as task criticality or complexity increases, requiring domain-specific tools and deterministic crosschecks for calculations, feasibility assessments, and design-related reasoning.

Moreover, AI has shown productivity gains in building-engineering graduate workflows ($B = 0.677$) students also perceive frequent hallucinations ($H = 0.653$) and exhibit meaningful reliance ($R = 0.711$) with moderate cognitive debt ($CDI = 0.570$). While verification behavior remains strong ($AE = 0.693$) and automation bias is comparatively low ($ABI = 0.219$), the coexistence of benefits and atrophy risk ($KA = 0.609$; $D = 0.448$) warrants immediate governance in a safety-critical discipline. A VCA/AVA implementation provides a concrete pathway to constrain scope, enforce verification, preserve competence, and retain efficiency—aligning educational practice with responsible AI guidance while reflecting AEC-specific adoption realities.

6. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support and contributions of their colleagues.

7. REFERENCES

[1] N. Rane, “Role of ChatGPT in Construction Industry,” Social Science Research

Network, Oct. 2023, doi: <http://dx.doi.org/10.2139/ssrn.4598258>.

[2] R. Samsami, “Optimizing the Utilization of Generative Artificial Intelligence (AI) in the AEC Industry: ChatGPT Prompt Engineering and Design,” *CivilEng*, vol. 5, no. 4, pp. 971–1010, Oct. 2024, doi: <https://doi.org/10.3390/civileng5040049>.

[3] P. Mahajan, “The Silent Erosion: Global Generational Cognitive Decline in the Age of AI and the Future of Human Intellectual Agency,” *Social science research network*, Aug. 2025, doi: <https://doi.org/10.2139/ssrn.5386814>.

[4] R. Armitage, “Your brain on ChatGPT,” *British Journal of General Practice*, vol. 75, no. 758, pp. 410–410, Aug. 2025, doi: <https://doi.org/10.3399/bjgp25x743181>.

[5] N. Rane, S. Choudhary, and J. Rane, “Enhancing Sustainable Construction Materials Through the Integration of Generative Artificial Intelligence, such as ChatGPT or Bard,” *Social Science Research Network*, Jan. 2024, doi: <https://doi.org/10.2139/ssrn.4681678>.

[6] D. Onatayo, A. Onososen, and A. O. Oyediran, “Generative AI Applications in Architecture, Engineering, and Construction: Trends, Implications for Practice, Education & Imperatives for Upskilling—A Review,” *Architecture*, vol. 4, no. 4, pp. 877–902, Oct. 2024, doi: <https://doi.org/10.3390/architecture4040046>.

[7] P. Shojaee et al., “The Illusion of Thinking: Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity,” *Apple Inc*, Jun. 2025. Retrieved from URL: <https://ml-site.cdn-apple.com/papers/the-illusion-of-thinking.pdf>

[8] K. Wach, “The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT,” Jun. 2023. Retrieved from URL: <https://eber.uek.krakow.pl/index.php/eber/article/view/2113>.

[9] M. Bobula, “Generative artificial intelligence (AI) in higher education: a comprehensive review of challenges, opportunities, and implications,” *Journal of learning development in higher education*, vol. 30, no. 30, Mar. 2024, doi: <https://doi.org/10.47408/jldhe.vi30.1137>.

[10] M. Söllner et al., “ChatGPT and Beyond: Exploring the Responsible Use of Generative AI in the Workplace,” *Business & Information Systems Engineering*, vol. 67, no. 89–303, Feb. 2025, doi: <https://doi.org/10.1007/s12599-025-00932-8>.